Potential of neural network triggers for the Water-Cherenkov detector array of the Pierre Auger Observatory

Master's thesis by

Paul Filip

at the Institute for Astroparticle Physics

First Reviewer: Prof. Dr. Ralph Engel
Second Reviewer: Prof. Dr. Guido Drexlin
Supervisors: Dr. David Schmidt

Processing time: 01.04.2022 – 02.05.2023

Review and Declaration

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

Karlsruhe, 02.05.2023

<div style="text-align: right;">

_____

Prof. Dr. Ralph Engel

</div>

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, 02.05.2023

<div style="text-align: right;">

_____

Paul Filip

</div>

Abstract

This thesis investigates the potential of station-level triggers for the surface detector of the Pierre Auger Observatory. The development of an extensive software framework to generate realistic WCD time traces from measured background and simulated air showers enables the development of novel classification algorithms based on concepts of machine learning. An in-depth evaluation of several candidate algorithms demonstrates the particular promise of LSTM networks, whose performance exceeds those of classical triggers. A multi-layer instance of this LSTM type neural network increases the efficiency of station-level triggers from $\sim 8\%$ to $\sim 18\%$ for triggering on stations with incident secondaries for proton primaries ranging from $10\,\mathrm{PeV}$ to $32\,\mathrm{EeV}$. This results in gains in event detection efficiency of up to 15.7 percentage points at the highest examined zenith angles of $65°$. The gains in efficiency with LSTMs were achieved without increasing the rate of background triggers, and only using 44 parameters. As such, LSTMs appear to be viable candidates for improving the trigger efficiencies of Water-Cherenkov detectors to extensive air showers without increasing demands on bandwidth between individual stations and central data acquisition, and without requiring more powerful FPGAs than those used in modern air shower arrays.

Contents

1. Introduction

From bit flips in computer hardware giving one candidate an impossible number of votes in a federal election [1], to full-scale extinction events considerably reducing the biodiversity on earth during prehistoric times [2, 3], the influence that cosmic rays have on our day-to-day life ranges from the smallest to biggest imaginable scales.

These incredibly energetic particles are of singular interest to physicists. Not only is this because of to the information they carry about the environment near distant stars and galaxies, but also due to their unmatched energy distribution, which is unachievable with earth-based particle accelerators. The relativistic messengers from outer space help us test interaction models, push the frontier of ultra-high-energy physics and teach us of our galactic and extragalactic neighbourhood.

The Pierre Auger Observatory is a world-leading experiment designed to detect cosmic rays of the highest energies. It achieves this by observing extensive air showers in the atmosphere and drawing conclusions based on measured data. A hybrid detection approach refines the accuracy with which arrival direction, energy, and the type of a primary particle are reconstructed. The larger of the two available detectors is the surface detector, which is sensitive to the shower footprint on earth. It consists of 1600 individually operating stations in an isometric triangular grid with unit spacing $1.5\,\mathrm{km}$.

With this compartmentalized structure, each station determines autonomously which information to forward to a central data acquisition system. This is realised via a hierarchy of trigger algorithms that scan the measured data locally for each detector. While these triggers are fully efficient at the tail end of the energy spectrum, shower cascades of lower energies $E \lesssim 10^{18}$ are seldomly detected.

In an attempt to extend the sensitivity of the surface detector to lower energies, the following thesis revisits the current trigger algorithms. Possible designs of new triggers based on deep neural networks are discussed, and the overall detection sensitivity is examined with a focus on implementability in the local station software.

With more sensitive triggers, more candidate shower events will be recorded. This paves the way for newfound analysis of the cosmic ray energy spectrum, and might enable detection of exotic primaries like photon- or neutrino cosmic rays. Such discoveries will be certain to send waves through the community of astroparticle physics.

The structure of this thesis offers a summary of cosmic ray physics and atmospheric particle cascades in chapter 2 and chapter 3. The Pierre Auger Observatory is introduced in chapter 4. Supplementary information regarding the implementation of neural networks and discussion of training data can be found in chapter 5 and chapter 6. chapter 7 and chapter 8 are dedicated to an in-depth description of the current trigger

implementations and newly developed triggers based on machine learning. The results of the analysis are summarized in chapter 9

2. Physics of cosmic rays

The term cosmic ray is used to describe particles that travel through space at velocities close to the speed of light. Their kinetic energy, which far exceeds their rest energy, ranges from several (GeV) for particles of solar origin to energies exceeding 10^{20}eV for nuclei of extragalactic origin. The latter are typically referred to as **U**ltra **H**igh **E**nergy CRs (UHECRs) in literature.

This chapter aims to introduce the general physical principles needed to describe cosmic rays. Immediately following is a summary of the discovery of cosmic rays in section 2.1. An overview of the origin of CRs is presented in section 2.2. Discussions regarding CR propagation in space is given in section 2.3. The resulting primary and energy composition are given in section 2.4 and section 2.5 respectively.

2.1. History

A first hint at the existance of high-energy particles in the upper atmosphere was given by Hess in 1912, who found that the discharge rate of an electroscope is increasing at higher altitudes. Millikan coined the term cosmic "rays" for these particles, as he argued the ionizing radiation must be part of the electromagnetic spectrum [4]. This was later, at least partially, falsified with the discovery of the east-west effect [5]. Hess' observation however withstood the tests of time and was ultimately recognized with the Nobel prize in physics in 1936 [6]. Two years later, in 1938, Pierre Auger showed via coincidence measurements that cosmic rays in fact originate from outer space, and gave a first description of extensive air showers [7]. Another 60 years later, the Pierre Auger collaboration would adopt his experimental setup and name in their search for cosmic rays of the highest energies.

In the meantime, numerous results from different cosmic ray detectors all over the globe have helped propel the related fields of particle physics, astrophysics and cosmology to new insights. Observations from cosmic ray physics serve as a valuable cross-check to the hadronic interaction models developed e.g. at CERN [8]. New theories modeling the final moments in the life of stars have arisen thanks to results from e.g. Kamiokande [9]. Publications by the Pierre Auger collaboration regarding the CR energy spectrum and flux help refine knowledge of our cosmic neighbourhood [10, 11].

2.2. Acceleration

Following the detection of a cosmic ray on earth, the logically following questions are "where did it originate from?", and "how was it accelerated?". It is clear that particles with extreme energies must be created under extreme conditions. In particular regions with large **e**lectro**m**agnetic (EM) fields, either in field strength or spatial extent, play a big role in the acceleration of CRs for reasons listed in the following pages. Several acceleration mechanisms have been identified.

2.2.1. Diffusive shock acceleration (Fermi I)

Super **N**ova **R**emnants (SNR) typically feature a plasma sphere propagating outwards from the former stars core into the **I**nter **S**tellar **M**edium (ISM). In this region of plasma any magnetic field lines will be comoving, according to Alfvén's theorem [12]. First realised by Fermi, such SNR shock fronts serve as source of high-energy CRs [13].

If a low-energy particle is injected into the SNR shock front, it will eventually be reflected by the local $\vec{B}$-field. If the diffusion length within the plasma is much smaller than the spatial extent of the SNR, the shock front can be modelled as a plane, and the process is analogous to an elastic reflection against a wall. Consequently, if $\frac{d\vec{B}}{dt} = 0$, this does not cause the particle to gain any energy, espically because $W = \vec{F}_L \cdot \vec{r} \propto (\vec{v} \times \vec{B}) \cdot \vec{r} = 0$. However, because the $\vec{B}$-field is moving radially outward alongside the plasma, a net energy gain of

$$\Delta E = +\beta_{SNR} \cdot E_0 \tag{2.1}$$

arises, where $\beta_{SNR} = |\vec{v}_{SNR}| / c$ and E_0 are the velocity of the shock-front and the initial energy of the particle. From chapter 7 in [13] it follows that ionization losses within the shock front are not completely negligible. Hence a particle must have a sufficient energy such that ΔE in Equation 2.1 exceeds possible ionization losses. The corresponding threshold for the primary energy above which acceleration occurs is dubbed the injection energy, and is of the order of 200 MeV for protons.

Furthermore, because typically $\beta_{SNR} \leq 0.10$, a single acceleration cycle is not enough to explain the CR energies observed on earth. Instead, multiple cycles are needed. This requires additional, focusing $\vec{B}$-fields, provided for example by the ISM, which alter the trajectory of injected particles such that they can be reflected off the shock-front again.

With each cycle, the particles rigidity $R = |\vec{p}|c / q$ increases, until its gyroradius $\rho = R/|\vec{B}|$ exceeds the spatial extent of the focusing $\vec{B}$-field and the particle escapes into space. With an effective ejection probability p per cycle, the energy after n cycles and the expected flux w.r.t energy, $\Phi(E)$, becomes roughly

$$E(n) = E_0 \ (1 + \beta_{SNR})^n \ . \tag{2.2}$$

$$N(n) = N_0 \ (1 - p)^n$$

$$\Leftrightarrow \quad \log\left(\frac{N(n)}{N_0}\right) = n \cdot \log\left(1 - p\right)$$

$$\Leftrightarrow \quad \overset{(2.2)}{=} \log\left(\frac{E(n)}{E_0}\right) \frac{\log(1 - p)}{\log(1 + \beta_{\mathrm{SNR}})}$$

$$\Leftrightarrow \quad N(E) = N_0 \cdot \left(\frac{E(n)}{E_0}\right)^{\log(1-p) \ / \ \log(1+\beta_{\mathrm{SNR}})}$$

$$\Rightarrow \quad \Phi(E) = \frac{\mathrm{d}N}{\mathrm{d}E} \propto E(n)^{\alpha - 1}, \tag{2.3}$$

where $\alpha = \frac{\log(1-p)}{\log(1+\beta_{\mathrm{SNR}})}$ in Equation 2.3 is a spectral coefficient whose exact value will depend on the age of the SNR (β_{SNR} decreases with age), the injected particle (different primaries have different injection energies and ejection probabilities), as well as many other factors that are often not known a priori. It can however be observed that the expected spectrum is a power law in the ranges from injection energy to a cutoff at the highest energies, which arises due to the finite lifetime of SNRs.

Results from several studies (e.g. [11, 14, 15]) hint that the presented first order Fermi acceleration mechanism is the main source of galactic CRs, extrasolar particles that originate from within the milky way, with energies ranging up to orders $O(\mathrm{TeV})$.

2.2.2. Stochastic scattering acceleration (Fermi II)

Second order (or stochastic) Fermi acceleration is the more general case of subsection 2.2.1 and represents the original idea developed by Fermi in [13]. The underlaying principle of scattering particles off plasma clouds remains unchanged. However, if the diffusion length within the cloud exceeds its radius of curvature, the energy gain per collision instead becomes

$$\Delta E \propto (\beta_{\mathrm{SNR}})^2 \cdot E_0. \tag{2.4}$$

Logically, this represents a much more inefficient acceleration mechanism, but is nevertheless observed in nature under certain circumstances (c.f. [16]).

2.2.3. Centrifugal acceleration in rotating $\vec{B}$-fields

Some astrophysical objects such as pulsars or **A**ctive **G**alactic **N**uclei (AGNs) possess strong magnetic fields ranging from 1 T for some AGNs [17] to $\approx 10\,\mathrm{GT}$ for magnetars, a subset of pulsars with extremely high magnetic flux densities [18].

If such objects rotate at an angular velocity Ω, which is in general nonzero, charged particles at a radial distance r from the rotation axis will undergo centrifugal acceleration. In particular, their Lorentz factor γ behaves like Equation 2.5 [19].

$$\gamma := \frac{E}{m_0 c^2} = \frac{\gamma_0}{1 - \left(\frac{\Omega r}{c}\right)^2},$$
(2.5)

where m_0 is the rest mass of the particle and γ_0 the prior Lorentz factor before acceleration. It follows that a test particle can in theory gain an arbitrarily high energy from this process by outspiraling towards the light cylinder surface, where $\Omega \cdot r = c$. In reality however, these processes are stopped by e.g. inverse Compton scattering at some point [20]. In any case, [19] and [20] conclude that values of $\gamma \approx 10^7 - 10^8$ are possible, corresponding to protons at $\approx 10\,\text{PeV} - 10\,\text{PeV}$ or iron nuclei at $\approx 500\,\text{PeV} - 5\,\text{EeV}$ energy.

2.2.4. Direct electrostatic acceleration

The presence of non-static $\vec{B}$-fields implies the existance of (in vacuum) comparably strong $\vec{E}$-fields and a corresponding electrical potential difference Φ across different regions within the magnetosphere. A back-of-the-envelope calculation reveals the proportionalities

$$|\vec{E}| \propto \frac{\Omega\, r_0}{c} \cdot |\vec{B}|,$$
(2.6)

$$\Phi \propto r_0 \cdot |\vec{E}|,$$
(2.7)

where r_0 is the radius of the central object rotating at an angular frequency Ω. Consequently, an ion with atomic number Z can be accelerated to energies $E = Z \cdot e \cdot \Phi$, which can in some cases easily exceed 10^{20}eV [21].

Some caveats to this consideration need to be mentioned. Screening effects from plasma clouds surrounding the central body are expected to limit the electrical field strength, and maximum acceleration energy by extension. Additionally, losses via e.g. Bremsstrahlung have been neglected in the above calculation, limiting the maximum attainable energy in theory even further.

2.2.5. Other types & general classification

Several acceleration mechanisms have been discussed. A plethora of other interactions that are able to accelerate elementary particles to fantastic energies remain unmentioned, or even undiscovered, as CR physics is an active area of research. In general though, the driving force behind all considered (and non-considered) acceleration mechanisms are thought to be (electro-) magnetic fields, given their infinite range and large coupling strengths compared to other fundamental forces. Consequently, the maximum energy
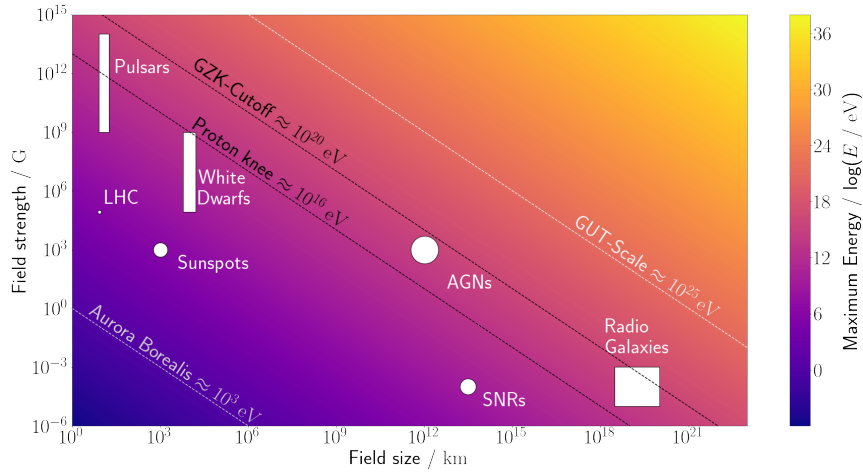
Figure 2.1.: Rough estimate of field strength and size of different CR sources as well as the corresponding maximum energy estimated with Equation 2.8 ($\beta = Z = 1$). Isoenergetic lines mark notable points in the energy spectrum discussed in section 2.5.

a specific CR accelerator with magnetic field $\vec{B}$ and size L moving at velocity βc can in theory provide for a particle with charge Ze is given by the Hillas formula [22]:

$$E_{\max} \, [\text{PeV}] = |\vec{B}| \, [\mu\text{G}] \cdot L \, [\text{pc}] \cdot Z \cdot \beta \tag{2.8}$$

This allows for an elegant classification of different cosmic ray sources, in part discussed on the previous pages, according to the Hillas plot shown in Figure 2.1.

2.2.6. Acceleration of uncharged particles

Particles like neutrons, neutrinos or photons possess a electromagnetic charge $q = 0$. Assuming Equation 2.8 holds in these cases, they should thus not appear in the CR spectrum. This is in disagreement with reality, where energies from $100\,\text{GeV}$ up to $\approx 100\,\text{EeV}$ have been observed [23, 24]. Consequently, additional interaction channels are required to explain the existance of such uncharged CRs.

High energy γ-rays in particular can be created by accelerated, charged particles via Bremsstrahlung. This occurs for example during centrifugal acceleration near pulsars or AGNs (compare subsection 2.2.3). Furthermore, inverse Compton scattering with a (charged) high-energy cosmic ray can imply a significant energy gain for a photon [25].

Uncharged CRs are also frequently produced in nuclear interactions, such as e.g. deeply inelastic scattering of charged CRs with the ISM. This espically contributes to the CR photon spectrum, as high energy π^0 are often byproducts of these scattering processes. The uncharged pions then decay into two photons. Finally, neutrons or neutrinos can originate from interactions involving the weak force. When a proton converts

to $p \rightarrow n + e^+ + \bar{\nu}_e$ during the decay of a UHECR ion, the resulting decay products, positron, neutron and electron-antineutrino inherit the parents' energy, and must thus be of sufficiently high energy as well.

2.3. Propagation

Once a cosmic ray has been observed coming from some arrival direction (θ, ϕ), backtracking its' trajectory to an eventual origin is, ignoring external factors, in the literal sense, straight forward. It has been shown however that several effects need to be considered for an accurate treatment of CR propagation through space.

For uncharged CRs (γ, n, ν), this task is simple. Such particles are not deflected by cosmic $\vec{E}$ and $\vec{B}$-fields. Possible interactions either demand the destruction of the particle (pair production, weak decay), or occur close to the source (e.g. Compton scattering), in which case the observed arrival direction will still be coincident with the actual source [26]. Gravitational lensing effects in some cases alter the trajectory of extragalactic photons. Such phenomena (if present in the first place) are however well understood in the scope of general relativity, and can be corrected for [27, 28].

2.3.1. Intergalactic propagation & transport equation

Contrary, charged particles ($e^\pm$, p, ions) propagate along non-trivial paths within galaxies due to deflections from solar- and galactic EM-fields. While the galactic field is coherent over large scales, numerous irregular magnetic domains, seeded in part by individual stars complicate CR propagation to essentially a three-dimensional random walk [29]. It is thus challenging to pinpoint the origin of a charged cosmic ray.

Nevertheless, related queries, such as for example the question whether or not a particle of given energy is likely to be of extragalactic origin can be answered by examining the distribution of cosmic rays within a region of spacetime. The behaviour of a population of n_i particles of type i can be approximately recreated via the below transport equation:

$$\frac{\partial n_i}{\partial t} = \underbrace{Q_i}_{\text{Source}} + \underbrace{\nabla D_i \left(\nabla n_i \right)}_{\text{Diffusion}} - \underbrace{\frac{\partial k_i(E)}{\partial E}}_{\text{Energy}} - \underbrace{\left(\frac{n_i}{\tau_{\text{spal.}, i}} - \sum_{j>i} \frac{n_j \, p_{ij}}{\tau_{\text{spal.}, j}} \right)}_{\text{Spallation}} - \underbrace{\left(\frac{n_i}{\tau_{\text{rad.}, i}} - \sum_{j>i} \frac{n_j \, d_{ij}}{\tau_{\text{rad.}, j}} \right)}_{\text{Weak decay}}$$

- **Source Q_i :**

  The source term is responsible for the creation of CR particles (of type i). The exact form of Q_i will depend on the considered creation process. For example, the near instantaneous creation of n_γ photons in a **G**amma-**R**ay **B**urst (GRB) at time t_0 and location $\vec{r}_0$ can be modelled like $Q_\gamma = n_\gamma \, \delta(\vec{r} - \vec{r}_0) \, \delta(t - t_0)$.

- **Diffusion** $\nabla D_i \, (\nabla n_i)$:

  The random walk mentioned above is accounted for in the diffusion term, which takes a similar form to the Stokes-Einstein equation. The diffusion coefficient(s) D_i in the most general case take a tensor form due to anisotropic diffusion in different directions. Furthermore, D_i is different for each particle type, as the deflecting EM-fields couple to the respective charges q_i, which need not be equal in principle.

- **Energy loss** $\partial k_i(E) \, / \, \partial E$:

  During propagation, a cosmic ray can interact with the ISM, and lose energy in the process. If this happens often enough, the CR is eventually thermalized and does not contribute to the population i any longer. Different interaction channels for different CR types i require different loss models $k_i(E)$ for each type.

- **Spallation** $\left(n_i \, / \, \tau_{\mathrm{spal.},\, i} - \sum n_j \, p_{ij} \, / \, \tau_{\mathrm{spal.},\, j} \right)$:

  Nuclear spallation describes the process of violent disintegration of a target nucleus upon being struck by an energetic projectile. The resulting fragments can retain energies up to the projectiles energy. The spallation term in the transport equation considers both the destruction (first term), as well as creation (second term) of CRs i from heavier types j. It is assumed that spallation from type $j \rightarrow i$ occurs at a constant probability of p_{ij} in a characteristic time frame $\tau_{\mathrm{spal.},\, j}$.

- **Weak decay** $\left(n_i \, / \, \tau_{\mathrm{rad.},\, i} - \sum n_j \, d_{ij} \, / \, \tau_{\mathrm{rad.},\, j} \right)$:

  If a particle j is weakly unstable ($\tau_{\mathrm{rad.},\, i} < \infty$) there is a nonzero chance d_{ij} it decays into a daughter nuclei of some type i during propagation. The decay term reflects this and describes both decay from heavier and decay into lighter nuclei.

Insights to the physical implications of this parametrization can be gathered from a simplified example. Consider the case of a galaxy with height $2H$ and width W. It is $H \ll W$, and thus only diffusion along the $\pm z$-direction will be examined. Considering n_0 protons located at $z = 0$ initially, and ignoring interactions with the ISM, the transport equation reduces to the first two terms, with $D = (0, 0, D_z)^{\mathrm{T}}$ and $Q = n_0 \, \delta(z) \, \delta(t_0)$.

It can quickly be verified that a solution to the transport equation in this case is given by a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = \sqrt{2 D_z t}$. The diffusion coefficient D_z is a measure of how quickly the population spreads out (along the $\pm z$-direction). According to [31], D_z can be parametrized via the particles energy E_p, and characteristics of the present $\vec{B}$-fields.

$$D_z = \frac{1}{3} \frac{E_p}{m_p \, c^2} \cdot \frac{|\vec{B}| \cdot \langle L_{\vec{B}} \rangle}{\sqrt{\mu_0 \, \rho_{\mathrm{ISM}}}}, \tag{2.9}$$

where $\langle L_{\vec{B}} \rangle$ is the characteristic length scale of deflecting $\vec{B}$-fields, μ_0 and ρ_{ISM} are the magnetic vacuum permeability and density of the interstellar medium respectively. After some time t, a fraction $r_{\mathrm{esc.}}(t)$ of particles will have a z-coordinate $|z| > H$, and

(a) Escape fraction

(b) AD anisotropy
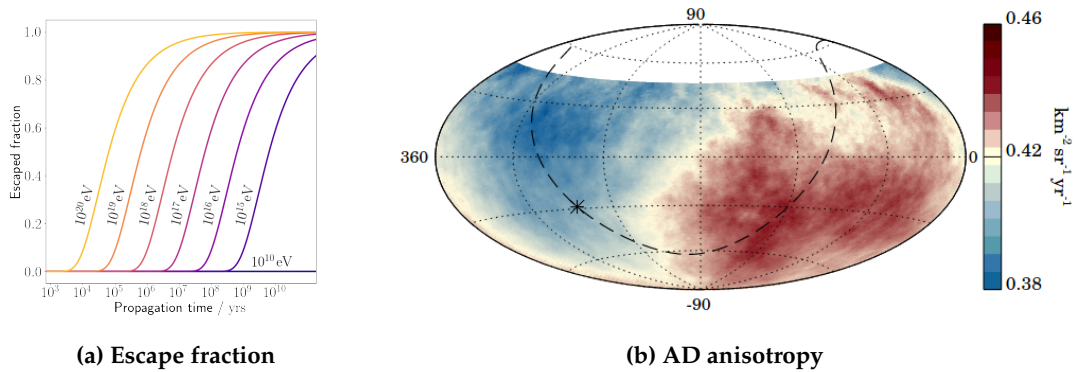
Figure 2.2.: (a) $r_{\text{esc.}}(t)$ according to Equation 2.10 for protons of different energies. **(b)** Dipole in the arrival direction of CRs with $E > 8\,\text{EeV}$. Image copied from [30].

exit the galactic disc accordingly. In reality, this is not equivalent to the particle leaving the galaxy, as large-scale halo structures extend above and below the visible disc [32]. These halos are ignored here. $r_{\text{esc.}}(t)$ can thus be calculated according to Equation 2.10. For some selected energies, a plot of the escaping ratio over time is offered in Figure 2.2a.

$$r_{\text{esc.}}(t) := 1 - \int_{-H}^{H} \frac{n(z,t)}{n_0}\,\mathrm{d}z \tag{2.10}$$

It can be concluded that low energy CRs do not travel outside their host galaxy within reasonable timeframes. Meanwhile UHECRs with energies exceeding $E > 10^{18}\,\text{eV}$ escape swiftly on ballistic trajectories and are thus likely have an extragalactic origin, not last also due to the limited energies that CR sources in the milky way can provide.

This observation is consistent with a dipole in the **A**rrival **D**irection (AD) of UHECRs observed by the Pierre Auger observatory. The dipole points roughly in the opposite direction of the galactic core, marked with an asterisk in Figure 2.2b.

2.3.2. Extragalactic propagation & GZK-Cutoff

In the last paragraph the (likely) extragalactic origin of UHECRs was discussed. Such particles must traverse millions of lightyears of extragalactic space before inducing a large air shower on earth. As the energy of these primaries increases above the **G**reisen-**Z**atsepin-**K**usmin threshold (GZK), their propagation through space is thought to be severely impeded. At energies above $\sim 10^{20}\,\text{eV}$ the **C**osmic **M**icrowave **B**ackground (CMB) consisting of photons in the microwave range are blueshifted to energies $E_{\gamma,\text{CMB}} > 300\,\text{MeV}$. A proton with the corresponding energy can thus absorb such CMB photons and convert to its' excited spin state, the Δ^+-baryon. The Δ^+-baryon decays nearly instantanouely to (for example) the ground state again, by radiating away a π^0 and losing energy in the process [33].

The mean free path of this interaction, also labelled GZK horizon is both energy- and primary-dependant. For 75 EeV protons, it is $\approx 100\,\text{Mpc}$ [34]. Cosmic rays exceeding the GZK threshold should ergo not be observed from faraway sources, and an overall reduction in flux at these energies should be recorded [35].

Indeed, results published by the Auger collaboration (see section 2.5) are consistent with this assumption. Whether or not the GZK-supression is the main cause for this soft spectrum at the highest energies remains unclear for now, but might be answered with the ongoing AugerPrime upgrade of the Pierre Auger observatory.
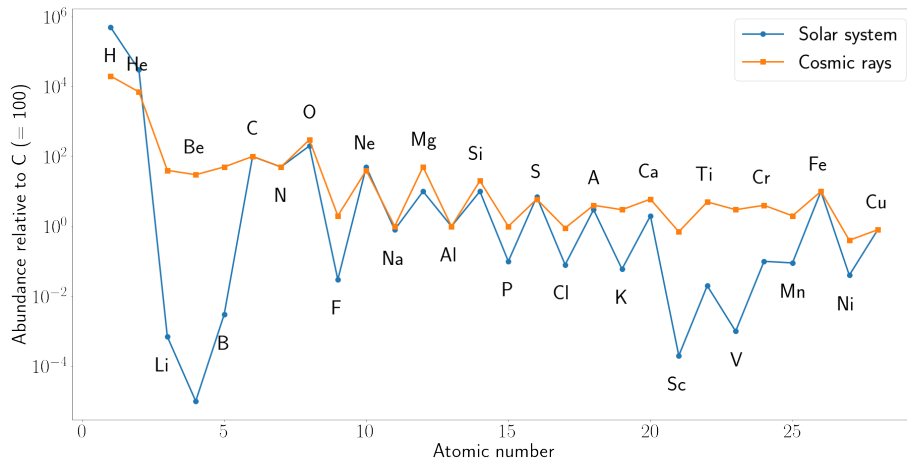
2.4. Composition

The composition of CRs largely mirror the relative abundancies of elements in the universe with some noteable exceptions shown in Figure 2.3. Elements like beryllium (Be) or vanadium (V) are atypical products of supernovae and thus not as common as e.g. oxygen (O) [36, 37] in the solar system. This leads to a dip in the corresponding abundancy spectrum. The same dip is not observed in CR primary abundancies. While it is a priori existant upon creation of cosmic rays, it gradually gets "filled up" via e.g. spallation processes during their propagation until an equilibrium state is reached.

This equilibrium state depends sensitively on the characteristic age of cosmic rays, i.e. the mean travel time until a particle escapes the galaxy. Measuring CR composition hence enables the estimation of this parameter. Such an analysis is conducted in [38], where it is found that the observed abundancies are consistent with a characteristic age of 1.7×10^7 yr for such high energy particles.

Contrary, hydrogen (H) and helium (He) are underrepresented (w.r.t their natural abundancy in the solar system) in cosmic ray particles. This is likely due to the comparably high ionization energy of both elements, which leads to less readily available hydrogen/helium ions. Since the acceleration mechanisms discussed in section 2.2 all couple to the net charge q of a particle, unionized hydrogen and helium are not accelerated [39].

2.5. Energy spectrum

It has been discussed in section 2.2 that the expected CR flux w.r.t energy for supernova remnants is a powerlaw in the rough range of $200\,\text{MeV} < E \lesssim 100\,\text{TeV}$. Observations by various experiments extend this result to even higher energies. Their combined results are shown in Figure 2.4. However, while the general assumption of a powerlaw $\Phi(E) \propto E^\alpha$ holds over a large range of energies, kinks and other feature in the spectrum indicate that the spectral index α is not uniform, and instead a function of energy. The approximate form of $\alpha(E)$ will be discussed in the following by examining several key regions of the energy spectrum.

Figure 2.3.: Composition expressed as abundancy relative to carbon for different sources. The ragged, alternating structure stems from an increased stability of nuclei with an even amount of protons (c.f. for example [40]). Data from [41]
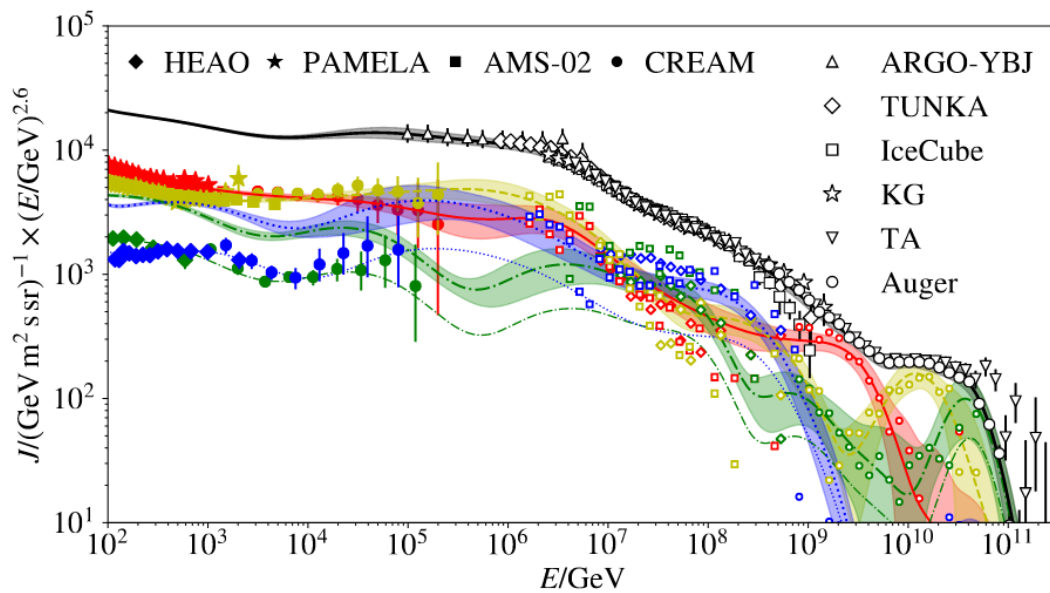
Figure 2.4.: Measurements of the cosmic ray flux, multiplied by a factor $E^{2.6}$ for all types (black), and broken down by primary. Shown are protons (red), helium (yellow), oxygen (green) and iron (blue). Plot adopted from [42].

- **Proton knee:**

Below an energy of $\approx 1\,\mathrm{PeV} = 10^6\,\mathrm{GeV}$, it is $\alpha(E) \approx 2.7$, while above a spectral index of $\alpha(E) \approx 3.0$ is found [41]. This softening of the spectrum (i.e. fewer particles of higher energy) could be attributed to several effects.

  – **Dark matter channel** (partially falsified)

  Weakly **I**nteractive **M**assive **P**articles (WIMPs) are a popular candidate for **D**ark **M**atter (DM), as they correctly estimate the cosmological evolution of the universe [43]. A WIMP with a sufficiently high mass could explain the kink in the energy spectrum. CRs with $E > m_{\mathrm{WIMP}}c^2$ could in theory produce DM in deeply-inelastic scattering processes. The degree of steepening in the spectrum is a measure for how readily the process $X \rightarrow \mathrm{WIMP} + Y$ occurs. The steeper the spectrum gets, the more particles are converted to DM, and the higher the corresponding cross sections are. Most theories involving WIMP creation can be excluded, since detectors at earthbound particle accelerators should have observed DM production at the observed steepening from 2.7 to 3.0 [44].

  – **Escape during propagation**

  While the analysis in subsection 2.3.1 concludes that the escape time for $10^{15}\,\mathrm{eV}$ protons to leave their host galaxy is at least of the order $10^8\,\mathrm{yr}$ (compare Figure 2.2a), the calculations leading to this result are extremely simplified. If a more accurate treatment finds that particles with rigidities corresponding to the relevant energies are no longer confined by galactic magnetic fields, the kink could originate from particles leaving the milky way and not contributing to the flux observed on earth any longer.

  – **Limited source energy**

  A last possible explanation might lie in Equation 2.8. A prevalent acceleration mechanism for CRs below $10^{15}\mathrm{eV}$, such as shock acceleration in SNRs for example, might not be able to provide energies exceeding this threshold due to physical constraints. The spectrum above the proton knee is thus populated by CRs that originate from different acceleration mechanisms with differen relations $\alpha(E)$.

- **Iron knee:**

The spectrum exhibits various similar kinks at slightly higher energies than the one of the proton knee. While it is assumed that they are ultimately caused by the same physical principles, each one corresponds to a different primary particle. Representatively, the iron knee at $E \approx 10^{17}\,\mathrm{eV}$ is discussed here.

Because iron has both a higher mass and charge ($Z = 26$, $A = 56$) compared to the proton ($Z = A = 1$), different processes couple to the different nuclei with disparate strength. In particular, an iron core has a higher magnetic rigidity $R \propto \frac{A}{Z}$ than a proton of equivalent energy. This is explained by the fact that, while the iron core experiences a larger Lorentz force ($\propto Z$), the resulting acceleration ($\propto \frac{Z}{A}$)

is not as strong due to a disproportionally larger mass ($\propto A$). It is thus logical to expect differences in the creation, propagation, and shower characteristics (see section 3.3 for details) of different CR primary particles. By extension, the spectral index $\alpha(E)$ should be contrasting for each distinct particle type i, giving rise to different fluxes $\Phi(E)_i$, which is confirmed by Figure 2.4.

The ultimate cause of the different knees remains unknown currently. In the context of the ongoing AugerPrime upgrade, the Pierre Auger observatory will scan the energy spectrum at higher precision [45]. This will allow to test whether the location of the various kinks scale with A or with Z, and consequently shed light on the processes giving rise to these features.

- **Ankle:**

  At an energy of roughly $\sim 1\,\mathrm{EeV} = 10^9\,\mathrm{GeV}$ an inflection point is found, where the spectrum hardens again from an index $\alpha(E) = 3.0$ to ≈ 2.7. This might mark the final transition from predominantely galactic to intergalactic cosmic rays. If intergalactic CR sources such as AGNs have a harder spectrum (but lower luminosity) than galactic sources, the ankle could be well explained by a smooth transition from the latter to the former spectrum [46]. Other explanations focus on a change of the primary composition, which does seem to be apparent in Figure 2.4 at the correct energy [47]. In any case, more data needs to be gathered to come to an informed conclusion on the ultimate cause.

- **Suppression:**

  At energies beyond $10^{20}\,\mathrm{eV}$ a sharp drop in the flux can be noted. This represents the tail end of the spectrum, beyond which events are so rare that cosmic ray observatories can mostly just identify upper limits with their available statistics. The cause for the drop is actively debated, and might lie in the GZK cutoff discussed in subsection 2.3.2. Another explanation might be that the tail end represents the point at which even the largest and most powerful (in terms of energy) CR sources in the universe are not able to accelerate particles any further (c.f. Figure 2.1).

3. Extensive air showers

Extensive air showers describe the particle cascades that are the result of a high-energy cosmic ray interacting with the atmosphere of the earth. While the microstate of a given air shower is inherently chaotic, macroscopic variables such as the number of particles in the cascade, its' multiplicity, allow conclusions on the primary cosmic ray.

In this chapter the various processes which give raise to the jets of energetic particles are discussed. Because hadronic primaries that carry intrinsic SU(3)-color charges are fundamentally different from leptonic ones which do not, this is done in a two-fold way. The funamental principles kicking off electromagnetic cascades are explained in section 3.1. Supplementary information regarding hadronic showers is listed in section 3.2. Finally, the effect of differing hadronic primaries is considered in section 3.3.

3.1. Electromagnetic showers

The dominating interaction of $E > 10\,\mathrm{MeV}$ photons in matter is e^+e^- pair production, whereas for electrons/positrons the creation of a γ via Bremsstrahlung prevails at high energies. This is shown in Figure 3.1. Consequently, an entire cascade of $e^\pm$ and photons can emerge from a single primary particle, as realised by Heitler in [50].

Of particular interest in these showers are, apart from the primary particles energy E_0 and arrival direction (θ, ϕ), the atmospheric depth X_{max} at which it reaches its' maximum multiplicity, as well as the **L**ateral **D**istribution **F**unction (LDF), that parametrizes the distribution of particles along the shower axis. An important variable that influences both values is the radiation length X_0. It represents the characteristic length at which an $e^\pm$ loses $1 - \frac{1}{e} \approx 63\%$ of its energy. It also corresponds to the mean free path of a photon in matter up to a factor 7/9 [51]. Neglecting said factor and assuming that new particles on average inherit half of the parent energy, describing the multitude of particles contained in an electromagnetic shower becomes a counting exercise in the context of the Heitler-model.

With each radiation length, the number of particles N in the shower double, while the energy per particle, E_{pp}, halves. After traversing an atmospheric depth of $n \cdot X_{\mathrm{max}}$, typically measured in units of $\frac{\mathrm{g}}{\mathrm{cm}^2}$, they consequently read

$$N(n) = 2^n, \qquad E(n) = \frac{E_0}{2^n}. \tag{3.1}$$

After sufficient interactions, the energy of each individual particle E_{PP} will have degraded to such an extent that other processes are no longer negligible compared to

(a) γ interactions

(b) $e^{\pm}$ interactions
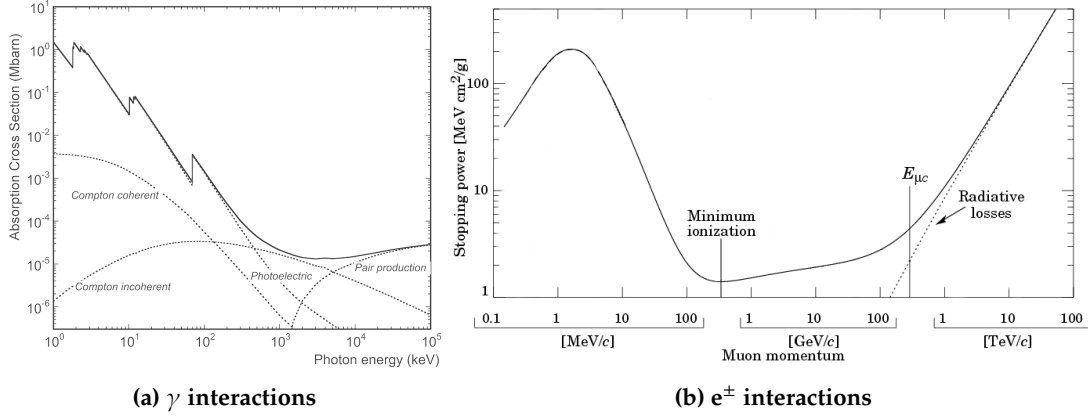
Figure 3.1.: (a) Cross section for different energy loss processes of a photon in tungsten. The sudden spikes correspond to the transition energy of increasingly higher-energy electron shells. From [48]. **(b)** Stopping power of copper, representatively on an antimuon μ^{+}, with respect to its' momentum. Plot adopted with changes from [49].

Bremsstrahlung and pair production. This occurs at the critical energy $E_{c,\text{EM}}$ below which the shower rapidly stops creating new particles and dies out as a result. It follows via Equation 3.2 and 3.3 that both X_{max} as well as N_{max} increase with E_0. The multiplicity arising from these assumptions alongside a stylized propagation of the thus created shower is represented in Figure 3.2.

$$E_{\text{PP}}(n_{\text{max}}) \overset{!}{=} E_c \overset{(3.1)}{=} \frac{E_0}{2^{n_{\text{max}}}}$$

$$\Leftrightarrow \qquad n_{\text{max}} = \left\lfloor \log_2\left(\frac{E_0}{E_{c,\text{EM}}}\right) \right\rfloor$$

$$\Rightarrow \qquad X_{\text{max}} = n_{\text{max}} \cdot X_0 = \left\lfloor \log_2\left(\frac{E_0}{E_{c,\text{EM}}}\right) \right\rfloor . \tag{3.2}$$

$$\Rightarrow \qquad N_{\text{max}} = 2^{n_{\text{max}}} = \left\lfloor \frac{E_0}{E_{c,\text{EM}}} \right\rfloor . \tag{3.3}$$

The number of particles at a given distance from the shower axis (y-axis in Figure 3.2) is essentially random, but follows a statistical basis, the lateral distribution function. The LDF can either be derived approximately from first principles [52] or empirically, as is done in [53]. The latter arrives at a closed form approximation for the local density ρ of particles given a shower with multiplicity N at a distance r from the shower axis as

$$\rho_{\text{EM}}(N, r) = \frac{0.4\, N}{r_{\text{M}}^2} \left(\frac{r_{\text{M}}}{r}\right)^{0.75} \left(\frac{r_{\text{M}}}{r + r_{\text{M}}}\right)^{3.25} \left(1 + \frac{r}{11.4\, r_{\text{M}}}\right) . \tag{3.4}$$

In Equation 3.4, the Molière radius r_{M} characterizes the lateral spread in multiple scattering processes. It is of order $r_{\text{M}} \approx 100\,\text{m}$ for interactions that are relevant here, and in general depends on the density of the considered material [54].
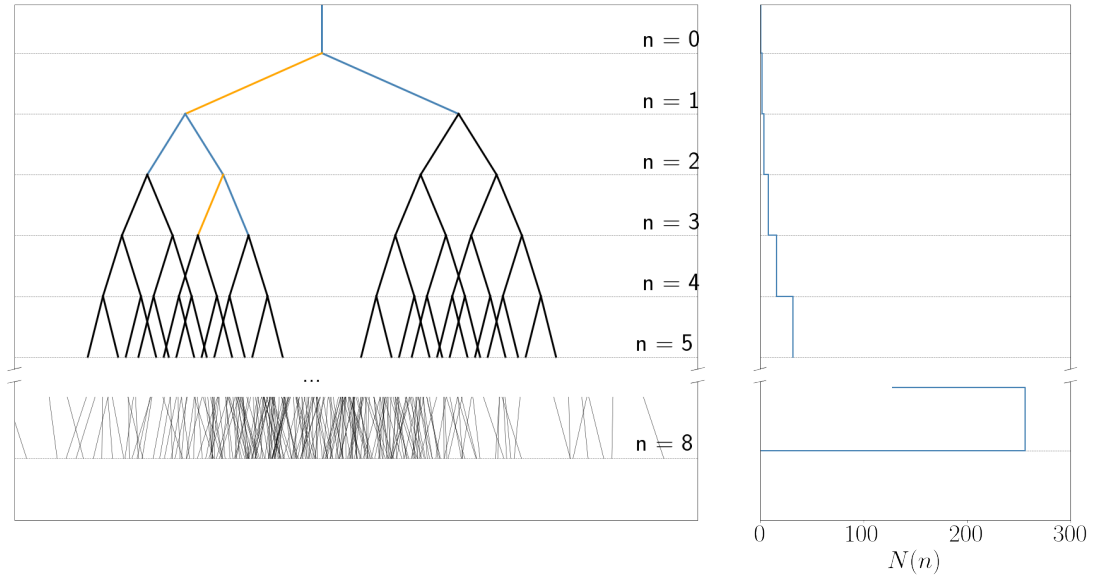
Figure 3.2.: Shown on the left is the stylized propagation of an extensive air shower through the atmosphere according to the Heitler-model, quantized in units of X_0. The energy of the primary particle is of order $2^8 \cdot E_{c,\text{EM}}$, which allows for 8 bifurcation steps, and $N_{\max} = 256$ shower particles. The multiplicity of the shower after each step is shown in the right subplot.

3.2. Hadronic showers

Hadronic primaries will readly produce color-charged secondaries, as has been shown many times in particle accelerators. In order to model the development of hadronic showers, the model discussed in Equation 3.1 thus needs to be adjusted. An example theory has been developed by Matthews in 2005. Following the reasoning in [55], after traversing an atmospheric depth corresponding to the hadronic interaction length, a proton creates on average $N_\pi \approx 15$ pions, of which two thirds are charged, and one third is uncharged. The corresponding decay channels of the light π-mesons with the largest **B**ranching **R**atios (BR) are

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \qquad (\text{BR} \approx 0.9999, \ \tau = 2.6033 \times 10^{-8}\,\text{s} \ [33]),$$
$$\pi^- \rightarrow \mu + \bar{\nu}_\mu \qquad (\text{BR} \approx 0.9999, \ \tau = 2.6033 \times 10^{-8}\,\text{s} \ [33]),$$
$$\pi^0 \rightarrow 2\gamma \qquad (\text{BR} \approx 0.9882, \ \tau = 8.5 \times 10^{-17}\,\text{s} \ [33]).$$

With a mean lifetime of just attoseconds, the π^0 decay instantly before being able to continue the cascade process. In this fashion, the uncharged particles initiate a Heitler shower such as the one in section 3.1, by providing high-energy photons. It follows that every hadronic shower has an electromagnetic component. Moreover, assuming that the inherited energy from the parent particle is roughly uniformly distributed among its' children, one third of the remaining energy in the hadronic component is lost to the electromagnetic component per hadronic interaction length.

Meanwhile, the charged pions repeat the procedure of creating secondary mesons, kicking off the hadronic component of the shower in the process.

Similar to the reasoning in section 3.1, a primary of given energy initiates a shower of a specific multiplicity $N_{\max}$. This is reached after $n_{\max}$ steps, where the energy per particle $E_{\mathrm{PP}}(n_{\max})$ is below the critical energy $E_{c,\,\mathrm{had.}}$ at which the mesons ionize rather than continue the cascade. After this last step, the charged pions eventually decay into muons and neutrinos. The shower characteristics are thus given by

$$N_{\mathrm{had}}(n) = \left(\frac{2\,N_\pi}{3}\right)^n, \qquad E_{\mathrm{PP}}(n) = \frac{E_0}{N_\pi^n}, \qquad n_{\max} = \left\lfloor \log_N \left(\frac{E_0}{E_{c,\,\mathrm{had}}}\right)\right\rfloor, \qquad (3.5)$$

wheras the maximum multiplicity (ignoring neutrinos) in the shower is calculated as

$$N_{\max,1} = \underbrace{\frac{3}{2}\left(\frac{2}{3}N_\pi\right)^{n_{\max}}}_{\text{Muon component}} + \underbrace{\sum_{k=1}^{n_{\max}-1}\frac{N(k)}{3}\cdot\left\lfloor\frac{E_{\mathrm{PP}}(k)}{E_{c,\,\mathrm{EM}}}\right\rfloor}_{\text{EM component}}. \qquad (3.6)$$

The muons stemming from pion decay follow a different LDF than the electromagnetic component. Again following the analysis in [53], the muonic LDF can be recovered as

$$\rho_\mu(N,t) = 18\left(\frac{N}{10^6\cdot r}\right)^{\frac{3}{4}}\cdot\left(1+\frac{r}{320}\right)^{-\frac{5}{2}}. \qquad (3.7)$$

While the above Equation 3.7 drops off slower $O(r^{-\frac{3}{2}})$ compared to the electromagnetic component ($O(r^{-3})$), the immediate vicinity of the shower axis contains mostly photons and leptons from the EM subshower. Further out, the muonic component takes over. This is visualized in Figure 3.3. Due to this reason, and the fact that muons can carry considerable amounts of energy faraway from the shower axis, the muonic footprint of a shower often appears much more "patchy" compared to the EM portion. This knowledge is espically useful when distinguishing between hadron- and photon-induced air showers (compare [56]).

3.3. Composite primaries

As is evident from the discussion in chapter 2, not only single protons (which are strictly speaking also composite) or elementary particles like photons, electrons, etc. appear in the cosmic ray spectrum. In theory, any somewhat stable particle is a possible primary. The consequence of different primaries on resulting shower characteristics is subtle, but large enough such that it can be used for identification purposes.

Assuming the constituents in a CR nucleus all coherently interact with an air molecule, one arrives at the superposition principle for extensive air showers. It states that for a
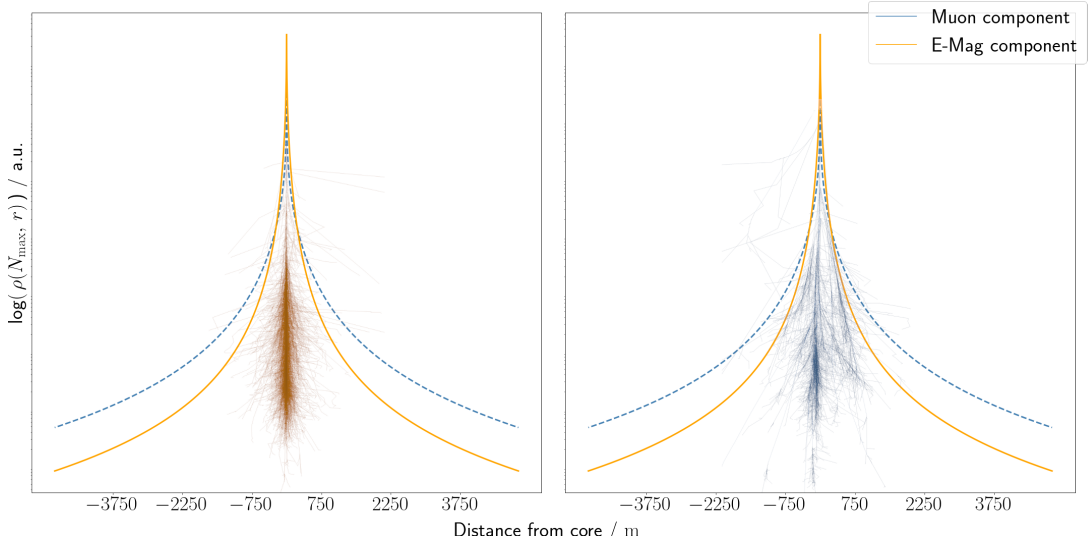
Figure 3.3.: The lateral distribution function for the muonic (steelblue) and electromagnetic (orange) component of a vertical, 100 GeV proton shower at roughly sea level ($r_\mathrm{M} = 100\,\mathrm{m}$). The inset plots represent the xz-projection of the color-coded shower component. Both component images adopted with changes from [57].

composite primary with $A = N + Z$ neutrons and protons, each constituent particle will initiate a subshower with initial energy of $E'_0 = E_0 / A$, where E_0 is the initial energy of the composite particle. For this scenario, N_max and n_max are known from the preceding sections. It follows that air showers from heaver primaries occur at higher altitudes (lower atmospheric depth X) and with higher particle counts.

$$N_{\mathrm{max}, A} = A \cdot N_\mathrm{max}, \qquad n_{\mathrm{max}, A} = \left\lfloor \log_N \left(\frac{E_0}{E_{c, \mathrm{had}}} \right) - \log_N A \right\rfloor, \qquad (3.8)$$

where $N_{\mathrm{max}, A}$, $n_{\mathrm{max}, A}$ refer to the resulting extensive air shower characteristics that are induced by a particle with mass number A. On top of this, because more massive particles initiate more subshowers of comparably lower primary energy, and thus have lower n_max, less energy is transferred from the hadronic to the electromagnetic shower component. This results in differing fractions of muonic to electromagnetic signal in the shower footprint.

3.4. Validity of shower simulations

The Heitler model and Heitler-Matthews model discussed in section 3.1 and section 3.2 respectively make only very rudimentary assumptions on the underlaying physics of particle cascades. Nevertheless, the equations recovered from these assumptions are already a close approximation of real world processes up to X_max.
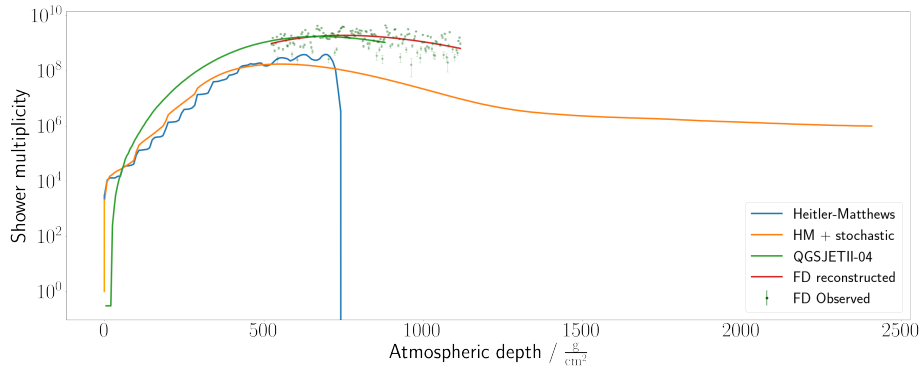
Figure 3.4.: Comparison of the number of charged particles for different physics models to reconstructed observations from a 3 EeV proton shower. Shown in steelblue and orange is the Heitler-Matthews model as described in section 3.2, and extended with a stochastic component. An example of a more in-depth simulation in red, and observed shower multiplicities in green.

Of course, adding a stochastic component to the above assumptions (c.f. [58]) improves predictions. But even full-fledged Monte-Carlo simulation software frameworks like GEANT4 [59] or CORSIKA [60] show discrepancies between observed and predicted shower development when analysed in depth. This is shown in Figure 3.4.

While shower-to-shower fluctuations can explain discrepancies to a degree, there also exist systematic differences between the simulated and observed extensive air showers. These are largely owed to imprecise knowledge of the underlaying physical processes. For example, hadronic interaction models (e.g. QGSJETII-04) rely on extrapolation of measured cross sections in the GeV-TeV scales to the relevant CR energies [61]. While this is not an unfair assumption given the scale invariance of deep inelastic scattering [62], it is clear, that the approach cannot accurately encompass all effects that may take place at the high energies present in atmospheric particle cascades.

In conclusion, the particle cascades evolving from relativistic CRs impinging on earth are still not fully understood. Several simulation frameworks have been developed, which each have their own shortcomings. It is therefore important to compare not only results from simulations using one framework to observations, but also different simulations with each other.

3.5. Detection methods

3.5.1. Cherenkov light

When a charged particle exceeds the phase velocity of light in a medium with refractive index n, the optical equivalent of a sonic boom occurs. Photons travel in a shockwave along an angle $\theta = \arccos\left(n^{-1}\beta^{-1}\right)$ relative to the trajectory of the particle while $\beta \geq \frac{c}{n}$. This process is well understood and can be used to detect high energy cosmic rays.

Imaging Air Cherenkov Telescope

The refractive index of air ranges from 1 in the near vacuum of the upper atmosphere to $\approx 1 + 2.9 \cdot 10^{-4}$ at sea level [63]. This implies that particles with a Lorentz factor $\gamma \gtrsim 41.5$ emit Cherenkov light, which is satisfied for e.g. muons and protons in the low GeV ranges and above.

It can therefore be expected that extensive air showers, which contain high energy charged particles, produce considerable amounts of Cherenkov light when propagating throught the atmosphere. This light, and by extension the air shower can be detected with **I**maging **A**ir **C**herenkov **T**elescopes (IACTs). Ground based experiments such as VERITAS [64] or H.E.S.S [65] detect gamma-rays using this technique.

Water Cherenkov Detector

The lightspeed in water is roughly 33% slower than in vacuum. Cherenkov radiation in water therefore occurs more easily than in air. Using this reasoning, a water tank equipped with means of detecting the emitted Cherenkov light, via e.g. **P**hoto**M**ultiplier **T**ubes (PMTs), should be able to measure traces of an air shower.

Indeed, this exact measurement principle of a **W**ater **C**herenkov **D**etector (WCD) was and is adopted in a variety of cosmic ray observatories such as the Pierre Auger observatory [66], HAWC [67], or Kamiokande [68], for example.

3.5.2. Fluoresence

Ionization losses have been ignored in the discussion of the formation of extensive air showers. This is of course not completely accurate. During the development of an extensive air showers, particles excite, or even ionize the permeated medium. Consequently, spontaneous emission of photons due to recombination, or transition back to a ground state can be observed. The amount of fluorescence light produced in this way is a gauge for the number of particles present in the shower at a given moment.

In Air

The predominant element in air is nitrogen (78%), whose transitions lay in the UV-band [69]. After a nitrogen molecule is excited by a passing shower particle, a photon with wavelength $300\,\text{nm} \leq \lambda \leq 430\,\text{nm}$ is emitted isotropically due to relaxation. The low attenuation of ultraviolet radiation in air allows the fluorescence light to travel large distances before being absorbed [70]. This enables cameras like EUSO-TA [71] or the **F**lourescence **D**etector (FD) of the Pierre Auger observatory (see [69] and section 4.1) to observe traces of extensive air showers from faraway during their development.

However, because of the low light yield of just $5\,\text{photons/MeV}$ [72], the detectors must operate in low UV-noise conditions, which places an upper limit on their duty cycle.

In Scintillators

Conventional plastic scintillators have a light yield, that is $1000-10000$ times higher than that of air [73]. Such scintillators therefore pose an effective way of measuring a shower footprint on the ground. The Pierre Auger **S**urface **D**etector (SD) is equipped with scintillators during the ongoing AugerPrime upgrade. Predecessors like KASCADE [74] have been using scintillation light to detect cosmic rays as well.

3.5.3. Radio Emission

Relativistic charged particles in the cascade are subject to deflections due to the geomagnetic field. This deflection is largest for $e^{\pm}$ due to their comparably tiny masses. Albeit the deflection of a typical electron in the shower is miniscule and the subsequent emission of Bremsstrahlung tiny, coherence effects along the entire shower front can greatly amplify the electric field strengths obtained from this effect [75]. This gives rise to radio signals emitted by the extensive air shower, which can in principle be detected via antennas.

One challenge in constructing an efficient CR radio detector is the requirement for a radio-quiet environment, where a large enough **S**ignal-to-**N**oise-**R**atio (SNR) permits analysis of measured data. This is the case for the AERA component of the Pierre Auger observatory, which has been in operation since 2010 [76]. Its results have in part lead the proposal of a vast radio array, GRAND [77], which will have an enormous exposure to CR showers, if built.

4. 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 [78].

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 [79]. Some further 20 years later, the Pierre Auger collaboration has published over 110 papers [80] 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 4.1. 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 4.1. The SD is described in section 4.2. A more in depth read on detector specifications and design choices is represented by the Pierre Auger observatory design report [78], where a lot of information stated in this chapter is conglomerated from. Notes on the event reconstruction are listed in section 4.3 and summarized from [81] and [69].

4.1. 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 $\approx 0.5\,\frac{\text{px}}{\text{deg}^2}$. 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 chapter 2). A schematic of the setup of each eye is given in Figure 4.2a.

The individual telescopes consist of $3.6\,\text{m}$ by $3.6\,\text{m}$, 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
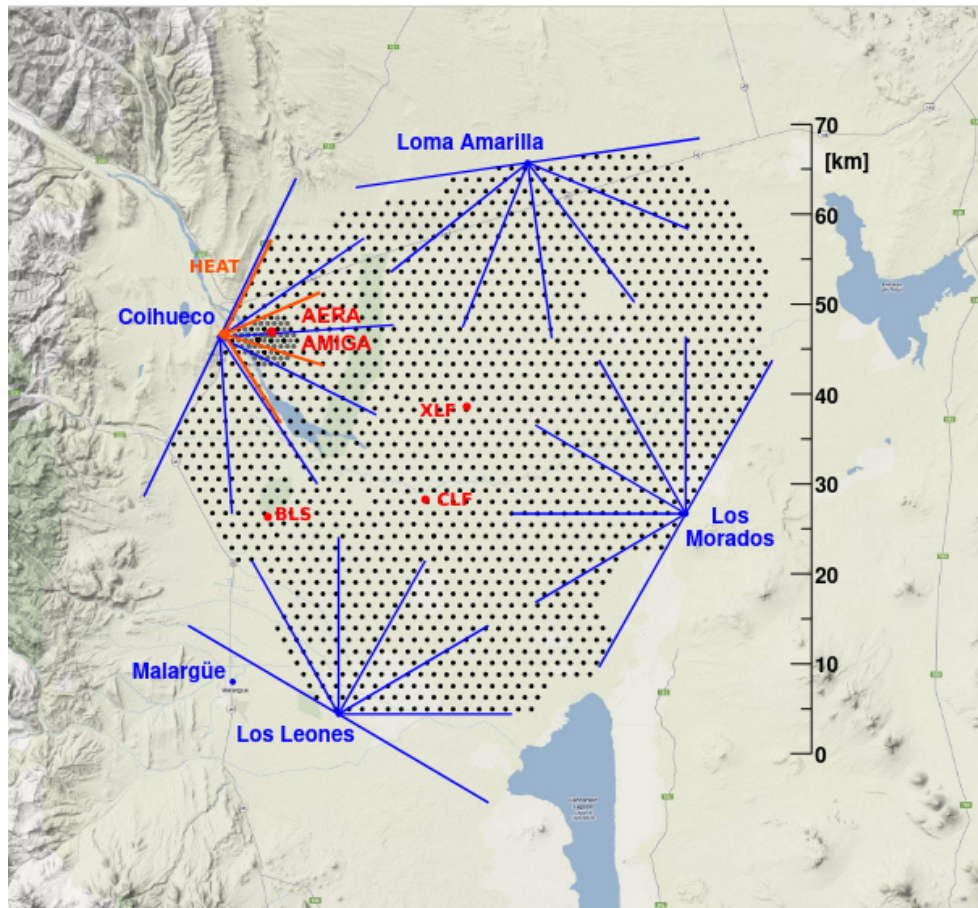
Figure 4.1.: Overview of the Pierre Auger observatory. The four different FD sites (respective FOV shown with blue lines) sit at the edge of the detector area and monitor the night sky above the SD array consisting of 1600 water tanks (black dots). A denser spacing of stations near Coihueco is equipped with additional electronics such as e.g. radio antennas (AERA) and muon detectors (AMIGA). Image taken from [82]

.

(a) FD eye setup   **(b) SD station setup**
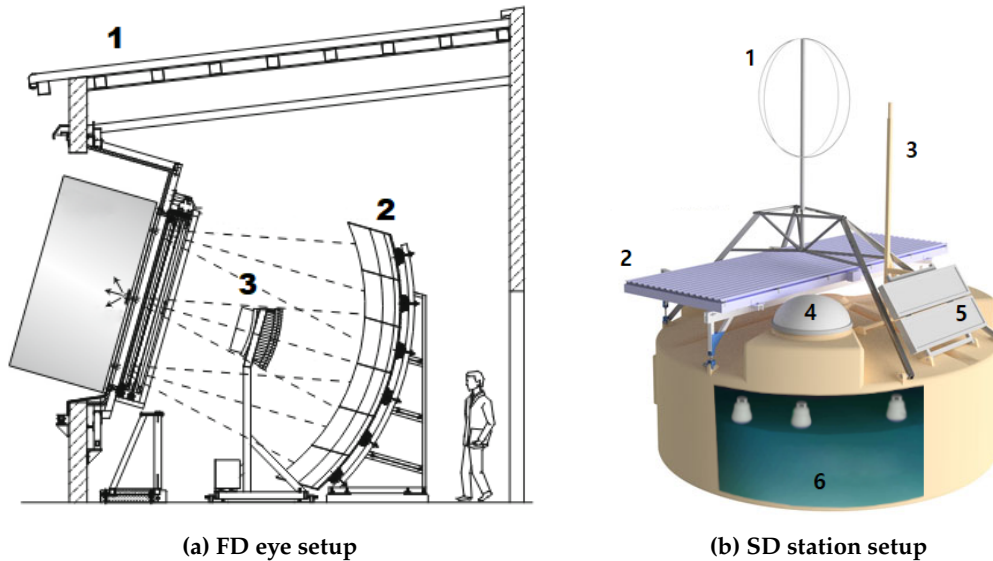
Figure 4.2.: (a) Schematic view of an FD eye with housing (1), main mirror (2) and camera (3). Image taken from [69] **(b)** Setup of an SD WCD with radio antenna (1), SSD (2), communication and GPS antenna (3), electronics box (4), solar panesl (5) and the WCD (6). Image adopted with changes from [83] and [84]

$\angle$ 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).

4.2. 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 [85, 86]. Figure 4.2b shows a schematic blueprint of each SD station.

4.2.1. 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 (≈ 8.33 ns binning) [87]. 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 ($\geq O(10^3)$ particles hitting the tank) as well as small shower signals ($O(1)$ particle hitting the tank) with sufficient accuracy [81]. Once an FADC bin has been recorded and checked for possible triggers (c.f. chapter 7) via **F**ield **P**rogrammable **G**ate **A**rrays (FPGAs) 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 4.2.3, subsection 4.2.2) or processed by a higher-level CPU for event reconstruction purposes (see section 4.3).

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 ($1200\,\frac{\mathrm{bit}}{\mathrm{s}}$, [88]), 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 via the official Auger reconstruction and simulation framework, $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ [89]. In any case, both algorithms are listed in the following subsections and discussed in more detail in the referenced literature.

4.2.2. $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ 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 μs) 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 $\sigma_{B_{front}}$ ($\sigma_{B_{end}}$) is given by the standard deviation of the remaining bins [90]. The baseline between the flat front and end estimate is then interpolated based on the difference

$$\Delta B = B_{end} - B_{front}. \tag{4.1}$$

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

  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 $+5 > \frac{\Delta B}{\sigma_{\Delta B}} \geq 0$:**

  Small fluctuations of the baseline are expected and the norm. If these fluctuations are positive ($B_{end} > B_{front}$) 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 ($B_{end} < B_{front}$) 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 4.3a.

- **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_{front} for the bins 1-300, the remaining baseline is first linearly interpolated according to Equation 4.2,

$$b_i = B_{front} - \Delta B \cdot \frac{i - 300}{1448}, \qquad 300 \geq i \geq 2048, \tag{4.2}$$

  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 4.3.

(a) Step-function approximation   (b) Charge-linear interpolation
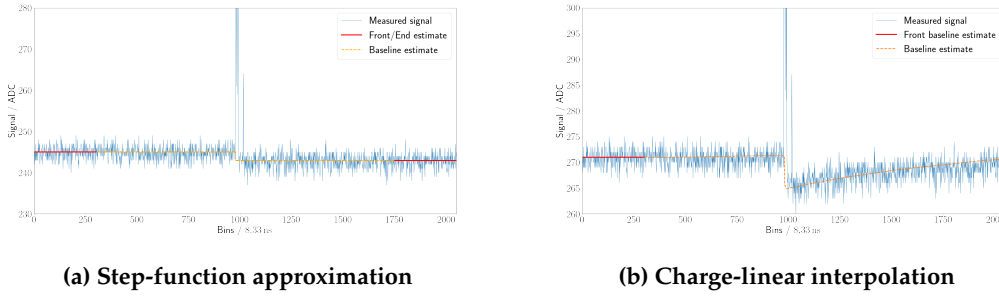
Figure 4.3.: (a) A simple step function is sufficient to accurately model a PMTs' noise level at small downward fluctuations. **(b)** For larger discrepancies the more involved charge-linear interpolation is used. Note that the signal undershoot is exaggerated for visualization purposes in both examples.

$$q_i = \sum_{k=0}^{i} (T_k - b_k) \exp\left(-\frac{8.33\,\text{ns}}{\tau} \cdot (i - k)\right) \tag{4.3}$$

In Equation 4.3, 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 [91]. Assuming the magnitude of the signal undershoot is directly proportional to the deposited charge q, a correction of the baseline thus becomes

$$b_i = B_{\text{front}} + \frac{q_i}{q_{1898}} \cdot \Delta B. \tag{4.4}$$

The parametrization in Equation 4.4 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 4.3b shows an example baseline estimate after three such iterations. In general, it converges to a robust estimate within five repetitions [91].

Estimation of I_{VEM} and Q_{VEM}

The conversion factor between ADC counts and VEM$_{\text{Peak}}$, 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 [88].

(a) VEM$_{\text{Peak}}$ Histogram

(b) VEM$_{\text{Ch.}}$ Histogram
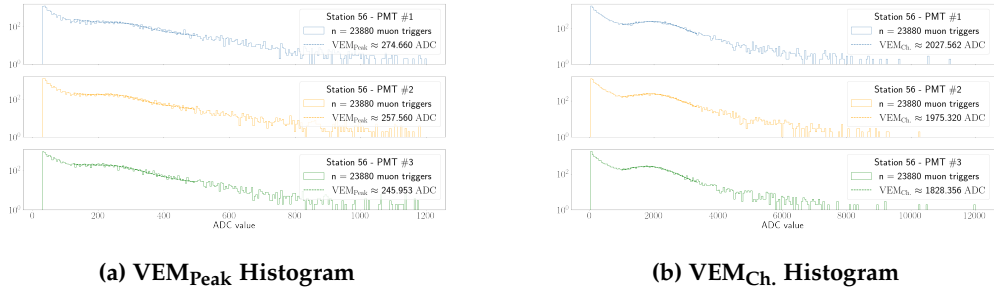
Figure 4.4.: (a) The maximum value of each muon trace is histogrammed in order to gain information about the current value of I_{VEM} of a station. **(b)** The conversion factor from recorded ADC values to Q_{VEM} is given from the histogrammed sum of each muon trace.

By histogramming the maximum value (sum) of each trace, the plot shown in Figure 4.4a (Figure 4.4b) 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 chapter 2. Noteable are characteristic deviations from this powerlaw, as these contain information about I_{VEM} and Q_{VEM}:

- 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 4.6) In similar plots from related works (c.f. [88, 92]) 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 [93]. Finally, one arrives at an estimate for the conversion factor between ADC counts and VEM$_{\text{Peak}}$, VEM$_{\text{Ch.}}$.

4.2.3. 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 4.2.3). This baseline estimate is used to set the thresholds of the hardware triggers discussed in chapter 7.

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 [78].

In the beginning, the to-be-estimated value $I_{\text{Peak}}^{\text{est.}}$ ($Q_{\text{Peak}}^{\text{est.}}$) 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 $t_{70} = 1.75\,I_{\text{Peak}}^{\text{est.}}$ above baseline plus a given PMT exceeding $2.5\,I_{\text{Peak}}^{\text{est.}}$ 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 4.1. 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 $I_{\text{Peak}}^{\text{est.}}$ ($Q_{\text{Peak}}^{\text{est.}}$) 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 $I_{\text{Peak}}^{\text{est.}}$ ($Q_{\text{Peak}}^{\text{est.}}$).

Table 4.1.

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

4.2.4. Trigger procedure

The flux of cosmic rays espically at the highest energies is barely of the order of $1\,\text{km}^{-2}\,\text{yr}^{-1}$ [22]. Consequently, most signals observed by the Auger observatory stem 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
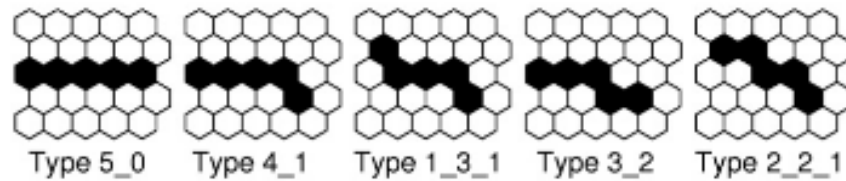
Figure 4.5.: Fundamental shape of tracks considered straight. Image from [69].

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 [69, 81].

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 4.5) 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 section 7.2. 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 forwarded to CDAS if the station observed a T1/T2 event within an appropriate timespan of order $O(\mu s)$ 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.

4.3. Event Reconstruction

If an event has been detected (subsection 4.2.4) 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 4.3.2), energy (subsection 4.3.3) and primary particle (subsection 4.3.4). 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.

4.3.1. 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 [81].

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 [94].

4.3.2. Arrival direction

The shower footprint measured by the SD (example given in Figure 4.6a) 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 [95]) 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

(a) SD shower footprint

(b) Angular resolution
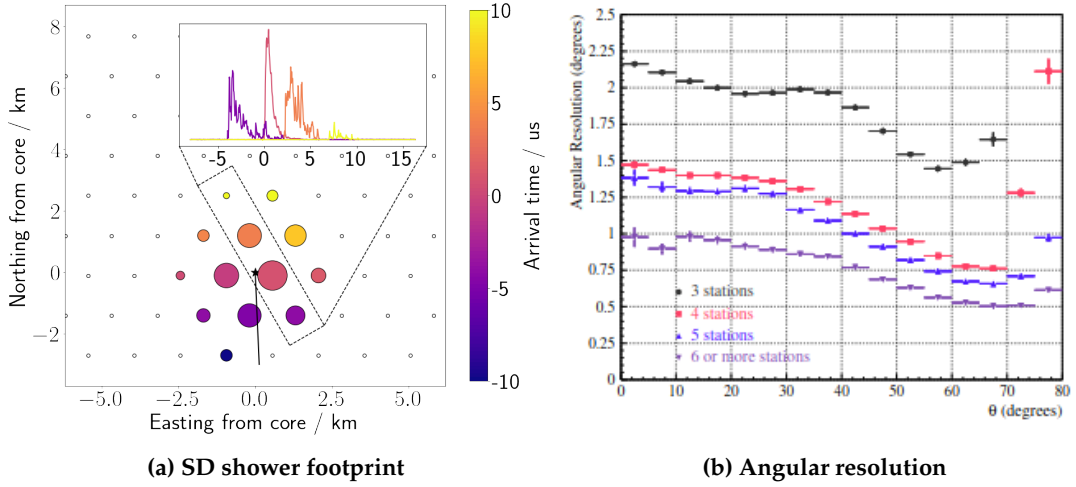
Figure 4.6.: (a) An example shower footprint recorded by the individual SD tanks (circles). The measured signal strength and arrival time is encoded in the size and color for each station. Tanks that haven't recorded any signal are shown colorless. For a subset of stations the respective VEM trace and 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 $3\,\text{EeV}$. Image from [95].

larger energies of the primary particle. This can be seen in Figure 4.6b. 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 $0.6°$ [95].

4.3.3. Energy estimator

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 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 $\text{SPD} = 1000\,\text{m}$ is obtained. This standardizes the comparison of results across many different events, even if a shower has only triggered few stations [96].

Due to attenuation effects in the atmosphere S_{1000} is a function of θ. It has been shown in [97] 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 4.5 and Equation 4.6.

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

(4.5)

The reference angle is chosen as $\theta_{\mathrm{ref}} = 38°$, as this is the median inclination of detected events [98]. It follows

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

(4.6)

Further corrections are applied to S_{38} in order to counteract influence of the local weather or geomagnetic effects [98]. 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 chapter 3.

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 [99].

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

(4.7)

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

4.3.4. 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 section 3.3. 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 [100], Figure 2.4)

5. Neural networks

The idea of a **N**eural **N**etwork (NN) is to attempt to capture the human thinking process in machine code. For this purpose, a network architecture connects some input (e.g. a picture) to an output (e.g. digits 0-9). Much like in a human brain, the architecture consists of multiple smaller chunks, neurons and layers, which connect in some way to form an emergent intelligent system.

As described, neural networks do not yet hold the abilities to achieve their designated tasks, and can hardly be called intelligent. They need to be trained. This is done by presenting an example input (called training data) to the network. The network output is compared to the desired output for the given input via some loss function. During training, the network attempts to minimize this loss function. How it is minimized is often a design choice, and in general will depend on the network architecture, which in turn is influenced by the type of data and kind of task the NN should accomplish.

In the following several, network architectures which are relevant for this work are detailed. The most simple option of a **D**ense NN (DNN) is given in section 5.1 in order to introduce several key concepts. **C**onvolutional NNs (CNNs) used for example in image recognition are explained in section 5.2. Lastly **R**ecurrent NNs (RNNs) that find an application in time series analysis are shown in section 5.3

5.1. Dense neural network

Dense neural networks are subdivided into layers, which themselve consist of individual neurons. Each neuron conglomerates information from a previous layer according to some weights w_{jk} and a bias b_j and propagates it through some nonlinear activation function $\sigma^{(i)}$. That is, the propagation of an input to the output layer through intermediate, hidden layers $\mathcal{L}^{(i)}$ can be described with the below matrix form:

$$\mathcal{L}_j^{(i)} = \sum_{k=0}^{n^{(i-1)}} \sigma^{(i)}\left(w_{jk}\mathcal{L}_j^{(i-1)} + b_j\right). \tag{5.1}$$

In Equation 5.1 $\mathcal{L}_j^{(i)}$ is the value of neuron j in layer i, and $n^{(i)}$ is a reference to the number of neurons in layer i. The activation function achieves two important goals. First, it limits the numerical value neurons can have. This ensures numerical stability during training, and is typically achieved by choosing a sigmoidal activation function. Secondly, the nonlinearity of the activation function ensures that the propagation
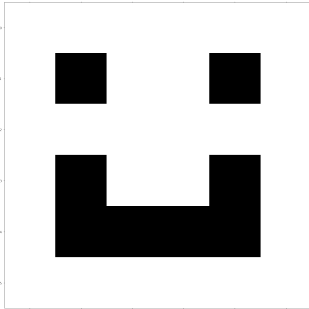
function of the entire network cannot mathematically be reduced to a single layer, as this disallows the network to learn nonlinearly separable patterns [101].

Important to note is the fact that the usage of one densely connected layer does not restrict the network architecture to consist solely of such layers. In fact, the network architectures discussed here and in the following section can all be used interchangably. This is a common practice in model building [102, 103].

5.2. Convolutional neural network

Convolutional neural networks introduce convolutional layers, which aggregate information from nearby input values. Nearby in this case referring for example to proximate pixels in a 2D image, or neighbouring voxels in a 3D scan. Even succesive inputs in a 1-dimensional time series can be convoluted. In general, the working principle of a convolutional layer can be extended to an arbitrary input shape and size, but will be representatively explained here for a two-dimensional, image-like input.

The convolution in a single layer is done by one or several filters, matrices, that are scalar-multiplied to subchunks of the input data. For example a mock filter like

$$\mathcal{F} = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 & +1 & -1 \\ -1 & +1 & +1 & +1 & +1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

could be swept across an entire image area. During the iteration, areas of the image that resemble the filter will result in a large (relative to the input data) positive scalar value.

In this fashion, an image can be efficiently scanned for specific patterns (lines, edges, etc.) with just few parameters. Namely, these are the numerical values in the filter matrices, which can be optimized during training. The resulting output of a single layer containing different filters can interatively be propagated into subsequent convolutional layers (searching for curves, corners, etc.) until full-scale image detection of complex structures becomes possible.

5.3. Recurrent neural network

Recurrent neural networks not only propagate an input value forward through their architecture, but also introduce cyclic connections between distinct layers. For example,

(a) simple RNN   **(b) LSTM**
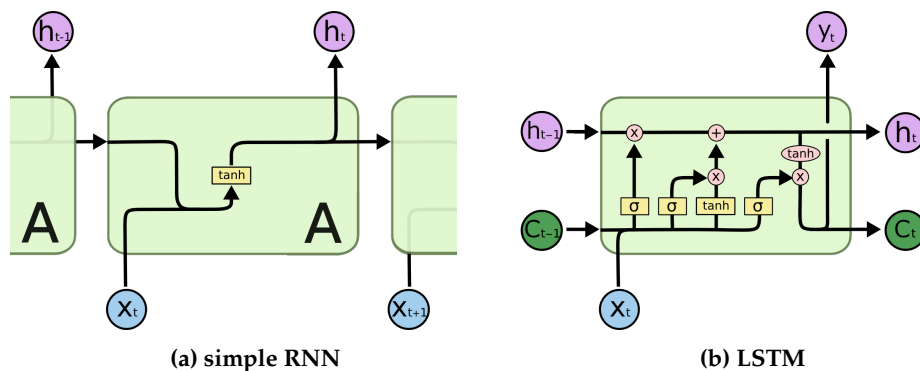
a) An example RNN architecture, with one layer A, where the network output h_t at step t is used as an additional input for step $t + 1$. Image altered from [105]. **b)** The architecture of a single LSTM layer relies on multiple gates, that update the network configuration, or memory of the layer. From [105] with changes. .

in the simplest conceptional case there exists a feedback loop that connects the output of a single network layer to its' input(c.f. Figure 5.2a). Other feedback configurations are of course possible, and generally preferred, as they can address the problem of vanishing/exploding gradients [104].

Moreover, due to the cyclic connections in the network architecture, RNNs are espically qualified for time-series analysis. i.e. where temporally sucessive inputs are highly correlated. A popular example is the **L**ong **S**hort-**T**erm **M**emory (LSTM) architecture, which is visualized in Figure 5.2b.

The output of a single LSTM layer at timestep $t \neq 0$ is calculated from two variables, the hidden state h_t, as well as the cell state C_t. The former exists in some form in every RNN, while the cell state C_t corresponds to a long-term memory that is unique to the LSTM architecture [106]. Unlike the name implies, the hidden state corresponds to the output of the network, and is thus in no way hidden. Due to vanishing/exploding gradients, the hidden state does not preserve information for more than few iterations. For this reason, it is often referred to as the short-term memory of the network. Data propagates through an LSTM layer according to the below steps (c.f. as well Figure 5.2b):

- A sigmoid layer outputs a number between 0 and 1 depending on the input x_t. The cell state from the previous iteration, C_{t-1}, is weighted by this number. In this fashion, the network can choose to discard or keep information from previous iterations. The first sigmoid layer is thus aptly named the forget gate.

- Next, the new cell state is calculated by extracting important information from the input. This is achieved via a feature extraction layer, which is weighted by another instance of a sigmoidal layer. The thus recovered new long term memory is added to the weighted cell state C_{t-1} and represents the new cell state C_t. This is the input gate of the LSTM.

- Lastly, the hidden state h_t is computed at the output gate of the LSTM. A feature extraction layer processes C_t and provides an output, that is weighted by one last sigmoid layer dependant on the input x_t.

5.4. Other architectures

With the current rise of deep learning applications in every facet of data processing, a plethora of other architectures have been established to achieve specific tasks. Some of these are purely generative (e.g. **G**enerative **A**dversarial **N**etworks (GANs) [107]) or not out of the box suitable for classification due to other reasons. Of particular interest, specifically for this work, are architectures that have been shown to outperform others in time-series analysis. Apart from the LSTM discussed in section 5.3 the **G**ated **R**ecurrent **U**nit (GRU) [108] is a popular implementation of a recurrent neural network. Last but not least, transformers [109] have consistently been among the most promising architectures for time-series analysis. However, because they require a lot of computational resources they are not compatible with the use-case that this work is pertaining to.

6. Neural network training data

Over their relatively brief existance, neural networks have been shown to perform increasingly impressive tasks (e.g. [110], [111], 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 6.1. Next, the process to extract signal originating from CRs is detailed in section 6.2. 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 6.3.

6.1. 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 4.2) with occasional spikes upwards due to low-energy particles impinging on the detector. Coupled with electronic noise from the many digital components in the station electronics, the **U**pgraded **U**nified **B**oard (UUB), this constitutes the data that is collected inbetween air shower events.

6.1.1. Accidental muons

Most low-energy background particles present in the detector are muons. These are produced in the upper atmosphere during cascading processes analog to chapter 2. Due to the low primary energy the electromagnetic component of the shower is thermalized before it reaches surface level. The muonic component by itself does not contain enough information to enable an accurate reconstruction of primary energy and origin. This fact, coupled with the high flux of events at lower energies ($\Phi|_{E=100\,\text{GeV}} \approx 1\,\text{m}^{-1}\,\text{s}^{-1}$
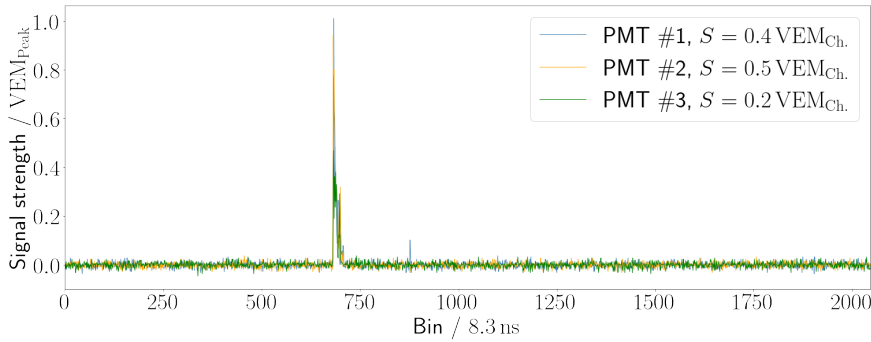
Figure 6.1.: The simulated time trace from a single muon. The maximum peak of the time trace is equal to $1\,\mathrm{VEM_{Peak}}$. The integral over each PMT, S, sums to $\approx 1\,\mathrm{VEM_{Ch.}}$.

[112]) make these events unsuitable for analysis. Stray muons, even though they originate from extensive air showers, must consequently be considered background events.

The rate at which such particles traverse a WCD tank is $f_{\mathrm{Acc.}} \approx 4.665\,\mathrm{kHz}$ [113]. Their arrival time is Poisson-distributed. This implies that generally, one in 14 time traces contains signal from a low-energy background event. Coincidences of two accidental muons occur on a sub-percent level. Any higher order of coincidences is likely originating from a single air shower process. The typical signal recorded by the surface detector from a single muon is presented in Figure 6.1.

A library of background traces of this type was provided by David Schmidt [113]. However, only the largest response of the three WCD PMTs is available for this library. Due to the lack of information one is either forced to assume the response to a low-energy background particle is the same across all PMTs, or neglect the response of the two remaining PMTs altogether upon injectiong a background muon into a signal trace (c.f. section 6.3). In both cases, neural networks are provided an easily detectable pattern to discern such particles from "real" shower signal. As a result, it should be refrained from training AI triggers on this dataset.

6.1.2. Electronic noise

Electronic noise is the umbrella term given to the distortions that some signal is subject to during digital readout. Such noise can have many different origins. An illustrative example is given by the **L**aser **I**nterferometer **G**ravitational wave **O**bservatory, which excludes the 60 Hz band and its' harmonics from analysis. This is owed to the fact that the DC frequency standard in the United States introduces systematic uncertainties in the detector [114]. In the electronics of Pierre Augers' SD array, electronic noise is assumed to be Gaussian. That is to say that the ADC values of a time trace that was measured while no particle produced signal in the tank are normally distributed around the baseline. The standard deviation can be estimated from monitoring data, as is shown in Figure 6.2.

6.1.3. Random traces

Both above mentioned phenomena can be simulated, and the simulation results used as background training data for the neural networks discussed in the next chapter. A more accurate method however, and the approach elected for this work is to utilize directly measured data from the field. Thanks to the work of David Nitz, there exist collections of so called random-traces[1] that were gathered by forcing DAQ readout via a manually set trigger.

In particular, two datasets of UUB random-traces have been created until now. They were taken from 13th-18th March 2022, and 14th-18th November 2022 respectively. The first set contains data a total of sixteen million time traces distributed over four different SD stations. For reasons explained in subsubsection 6.1.3, only data from the first set is used in the analysis presented in this work.

Characteristics

Contrary to the naming of the random trigger, it occurs at a deterministic time. More accurately, the process of measuring random-traces is as follows; A single time trace $(2048 \cdot 8.333\,\text{ns} = 17.07\,\text{µs})$ is written to the local station buffer every 10 ms. Once the buffer has accumulated enough data, it is written to a USB storage device. Because of a bottleneck in the last step, collecting four million traces takes roughly 22 h per station [115].

It is thus not the trigger that is unpredictable, but the data measured by each trigger. Due to the read/write process being independent of the measured data (as opposed to the algorithms in chapter 7) the latter must be considered to be essentially random. For the most part, random-traces are assumed to consist solely of electronic noise. However, signal of cosmic origin - be it accidental muons or extensive air showers - will appear in the data at a rate at least consistent with section 7.1.

A crude analysis of the type of noise in the random-traces can be made by examining the spectral density of the dataset, shown in Figure 6.2. Harmonic modulations visible in both spectra might originate from an offset between last and first bin of the random-traces. If this offset is nonzero, the periodic extension of $f(x)$ exerts a step-function-like behaviour. The Fourier transform consequently reflects this [116]. Still, several features of $|\hat{f}(\xi)|^2$, espically present at 10 MHz, imply the presence of systematic noise in the UUB. Nevertheless, the large scale form of the spectral density is compatible with at least two noise models, that cannot be distinguished based on the data at hand:

- $|\mathbf{f}(\xi)|^2 \propto \exp\left(\frac{(\xi-\mu)^2}{2\sigma^2}\right)$. The spectral density is Gaussian. This implies the noise is Gaussian distributed as well, confirming the assumption in subsection 6.1.2.

---

[1]to avoid possible confusion between this dataset and a *random* trace in the statistical sense, the traces recorded by David Nitz are referred to as random-trace, with a hypen.
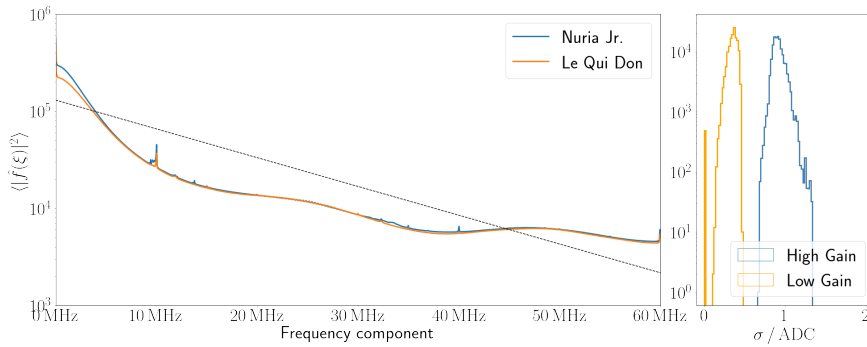
Figure 6.2.: (*Left*) The random-trace spectral density for two stations. Plotted with a dashed black line is reference attenuation curve falling at $-6\,\mathrm{dB/Oct}$. The spike at $10\,\mathrm{MHz}$ is of unknown origin and represents systematic noise in the UUB electronics. (*Right*) Example variance of all UUB stations in the surface detector array. The data shown in this plot was recorded on November 15th 2022.

- $|\mathbf{f}(\xi)|^2 \propto \exp{(-m\xi + b)}$. The spectral density is proportional to ξ^{-n} for some power n. The case $n = 2$ ($-6\,\mathrm{dB/Oct}$ attenuation) seems to describe the observations well, hinting that the generating function could be Brownian.

Calibration

The random-trace files contain raw measurement data in units of ADC for the HG and LG channel of the three WCD PMTs. In a first step to standardize this information, the baseline is substracted from each FADC bin. This is done via the baseline finding algorithm described in subsection 4.2.2 and [90, 91]. Note that this approach differs from the baseline finding algorithm that runs on each station (c.f. subsection 4.2.3). However, the difference is negligible ($<<\ 1\,\mathrm{ADC}$) for traces that do not contain any signal, which is the case for the vast majority of the dataset.

Next, the baseline-substracted time traces are converted from units of ADC to $\mathrm{VEM_{Peak}}$. This conversion is not straight forward, as it requires knowledge of I_{VEM} at the time of data taking. Each station estimates this value in periodic time intervals in the context of monitoring diagnostics.

For the second dataset of random-traces (taken from 14th-18th November 2022) a UNIX timestamp packaged with each time trace may be related to monitoring data. This reveals that no information regarding I_{VEM} was forwarded to CDAS for any station while it recorded random-traces. As a result, and because the available stations show large fluctuations in I_{VEM} before and after data taking (shown in Figure 6.3) the entire dataset is unfortunately rendered useless for this work.

For the first collection of random-traces, monitoring data is available, but there exists no timing information for the individual traces. Only the date of the measurement is known. The elected procedure to evaluate data as accurately as possible is thus to calculate the day average of I_{VEM} and Q_{VEM} and take this as the best (first) estimate for each trace. As can be seen in Figure 6.5, this eliminates half of the remaining dataset,
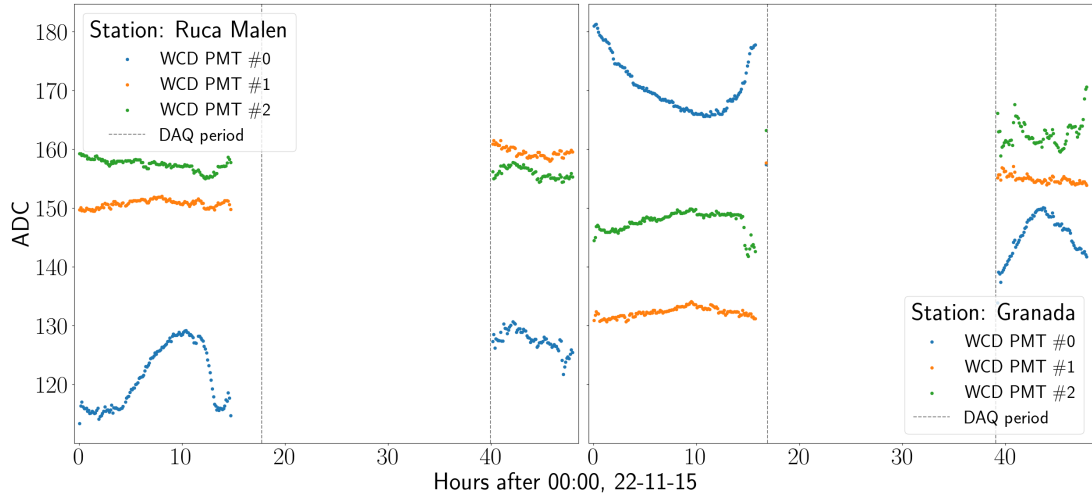
Figure 6.3.: The stations for which the second set of random-traces was recorded show large fluctuations in I_{VEM} before and after data acquisition. The assumption of a constant I_{VEM} for the collected datasets will consequently impose large errors in the subsequent analysis.

as two of the four stations show a large variance in I_{VEM}. The day average in these particular cases is not a good estimator for calibration purposes. For this reason, only random traces from the stations Nuria and Le Qui Don, measured in the March dataset are used for analysis purposes, as their $I_{\text{VEM, Rand.}}$ is stable.

As a crosscheck to verify the goodness of the approximation, the T2 trigger rate as reported by the calibration process is related to the trigger rate obtained by direct calculation over all (calibrated) random-traces. By extension, this also serves as a unit test for the classical triggers as they will be implemented in chapter 7. The results of this analysis are shown in Figure 6.4. As is clear from the plot, the rates calculated from the two different approaches are not in accordance. This indicates systematic errors in the calibration (a wrong implementation of trigger algorithms is disfavoured from the discussion in chapter 7). The errors are consistent across both considered stations. It is found that calculated trigger frequencies are $\approx 25\%$ lower than what is taken from monitoring. It is unclear why this discrepancy occurs, as it implies that the stations do not use the same threshold values for triggering as they report.

In any case, $I_{\text{VEM, Rand.}}$ must be adjusted to reflect this. A 25% increase in trigger rate is relatable to a $\approx 10\%$ decrease in I_{VEM}. The calibration scaling factors for random-traces thus become the values listed in Table 6.1.

Table 6.1.: Calibration constant I_{VEM} for random-traces.

PMT #	Nuria	Le Qui Don
1	159.34 ADC	145.79 ADC
2	161.37 ADC	144.63 ADC
3	149.91 ADC	154.62 ADC
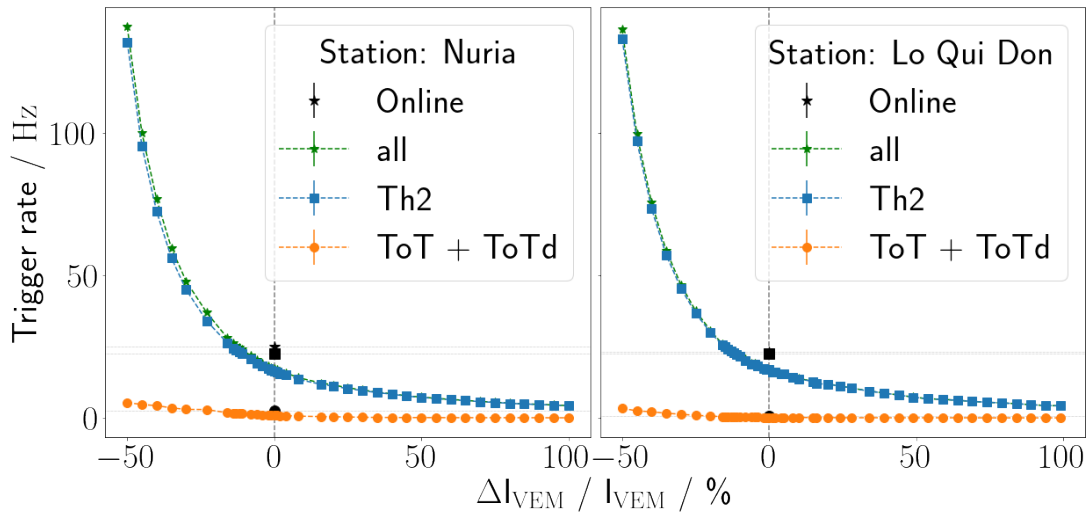
Figure 6.4.: The online reported T2 trigger rate (black) does not match the calculated trigger rate. Only a decrease ΔI_{VEM} by 1/10th of the original I_{VEM} gives a close approximation of the observed rate when manually calculating trigger frequencies.
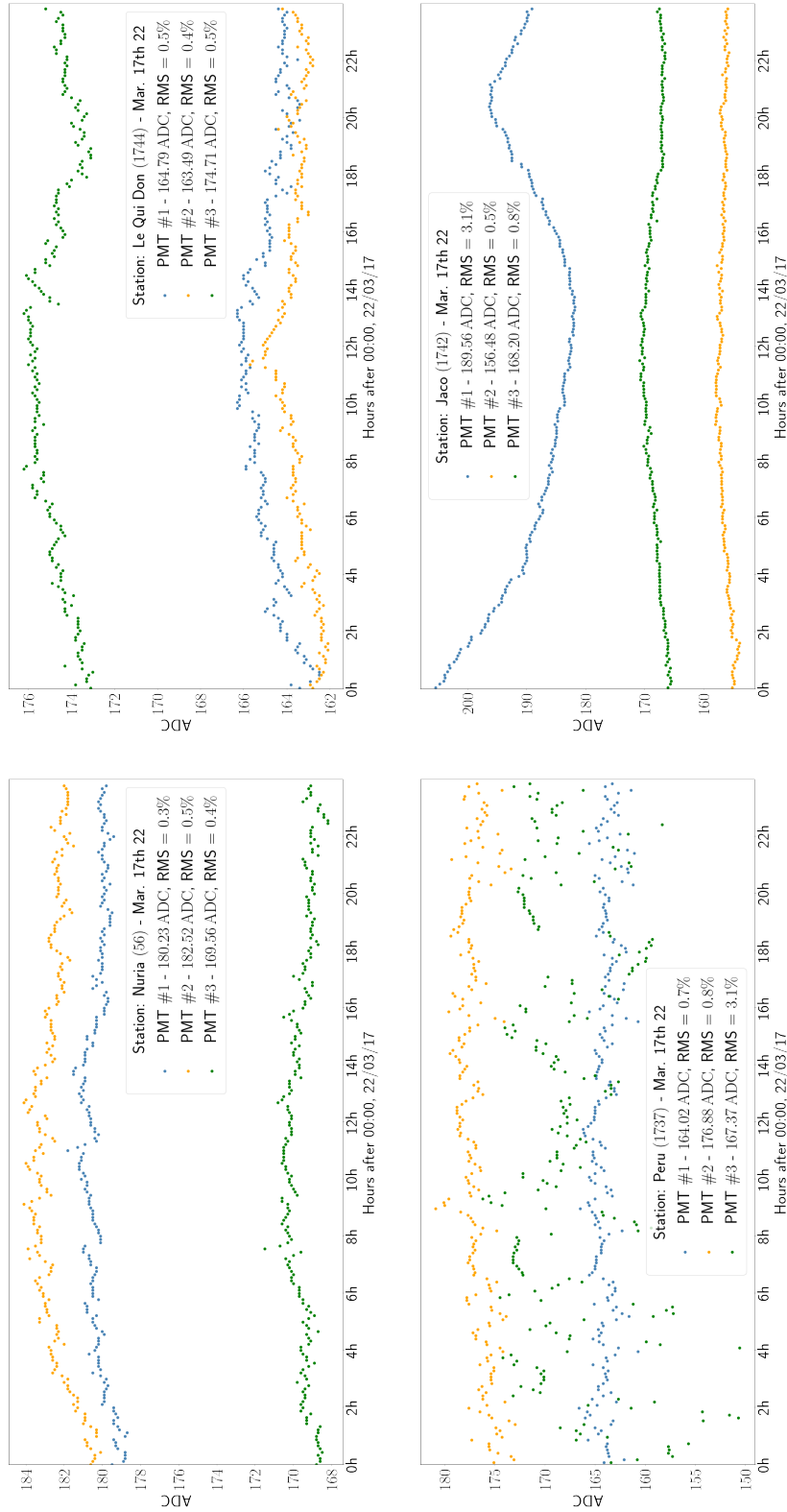
Figure 6.5.: Monitoring values for the four stations available in the random-trace dataset measured in March '22. The bottom two stations show a large variance in I_{VEM} for at least one PMT. The two stations in the first row (Nuria, Le Qui Don) are more stable ($\sigma / \mu < 1\%$).

6.2. Signal dataset

In the context of this work, a "signal" label (as opposed to background label) is assigned to *any* detector response caused by extensive air showers. Admittedly, this choice of classification is not ideal, as the particle density far from the shower axis grows sparse. Time traces recorded at those locations look very similar to ones raised by accidental muons. In any case, ramifications and possible solutions to this problem are further discussed in the following chapters.

In order to isolate the signal stemming from shower particles only, an $\overline{\text{Off}\underline{\text{line}}}$ simulation using Geant4 is executed on CORSIKA source files [60]. These are a total of 40557 simulated showers with a proton primary of energy $16 < \log(E/\text{eV}) < 19.5$. All showers are simulated using the hadronic interaction model QGSJET-II.04 [8].

In the process, the electronic feedback of the WCD PMTs is evaluated without any disturbance factor (see section 6.1). That is to say that the time trace obtained from such simulations is identically zero (in units of ADC) at any point in time where no ionizing particles are present in the WCD. An example trace that visualizes this is shown in Figure 6.7. Next, the trigger conditions both for individual stations and on the event-level are altered to trigger on everything. This step is needed in order to save all traces to the simulation output, an **A**dvanced **D**ata **S**ummary **T**ree (ADST). If this was not the case, only traces that already satisfy current trigger conditions would be written to disk. A neural network training on such data could therefore at best be as efficient as the current triggers.

The choice of this approach forces some detours in the ADST readout. Instead of extracting the VEM calibrated traces directly, individual component traces, i.e. the PMT signal caused by muons, electrons and photons individually are summed to yield a total ADC trace. Signal stemming from hadrons or other components in the cascade is neglected. This does not impose any errors in the analysis, as the hadronic component especially is located close to the shower core, where the EM- and muonic component of the shower alone should already enable easy detection. Finally, the total trace as calculated above is extracted to a more easily accessible data format alongside shower metadata like primary energy, zenith, but also SPD, and particle count in the station the trace was recorded from. The distribution of shower metadata for all available simulations is shown in Figure 6.6.

6.3. Trace building

With the componentwise traces at hand, the total trace as would be recorded in the WCD PMTs for a given event, can be constructed. First, a trace container with default UUB trace length (2048 bins $\cdot$ 8.3 ns / bin = 17.07 µs) and three components per bin (the three WCD PMTs) is initialized with all values equal to zero. Next, an arbitrary random-trace is selected as baseline. Since the FPGAs fundamentally count in the integer domain, the ADC data in the random-trace contains only whole numbers. As is, this wouldn't correctly model rollover when adding integer random-traces to
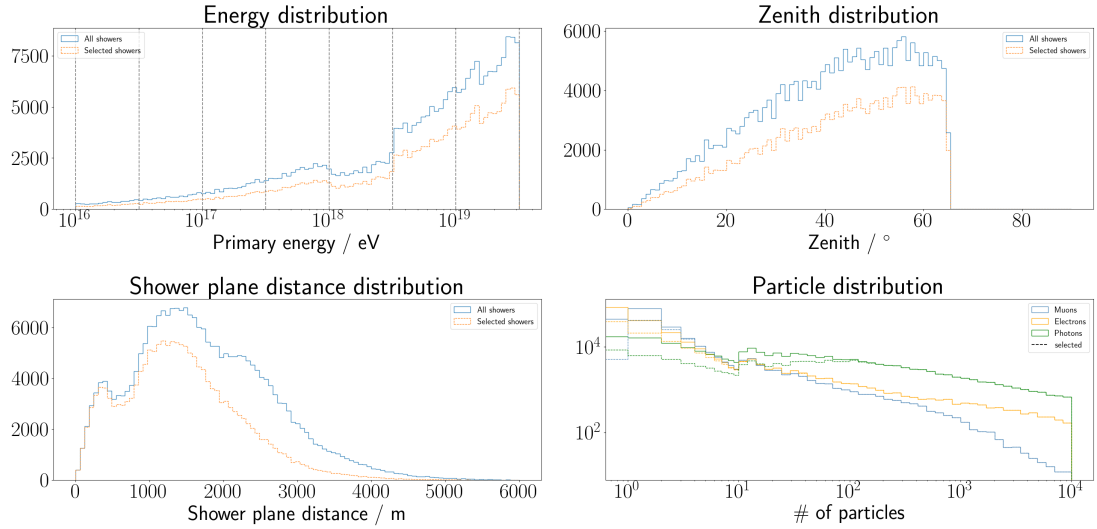
Figure 6.6.: The distribution of primary energy (*top left*), shower inclination (*top right*), separation from shower axis (*bottom left*), and number of particles in a tank (*bottom right*) is histogrammed for each trace available in the dataset. Since showers with higher primary energy reach a larger number of stations, traces from showers with increasing energy are overrepresented in this dataset. The dashed lines in each subplot represent the population of traces that deposit a signal $S > \text{VEM}_{\text{Ch.}}$ in the tank.

floating point simulation information. While two ADC signals by themself might not exceed the threshold to cross to the next higher value, their sum would might; $\lfloor 0.7\,\text{ADC} \rfloor + \lfloor 0.4\,\text{ADC} \rfloor = 0\,\text{ADC} \neq 1\,\text{ADC} = \lfloor 0.7\,\text{ADC} + 0.4\,\text{ADC} \rfloor$. To account for this, uniformly distributed random numbers from 0 (inclusive) to 1 (exclusive) are added to the random-trace.

Furthermore, the $I_{\text{VEM, Rand.}}$ from random-traces (c.f. subsection 6.1.3) will in general be different from I_{VEM} simulated by $\overline{\text{Off}}\underline{\text{line}}$ ($I_{\text{VEM, Off.}} = 215.781\,\text{ADC}$ compare [89]). Thus, the random-trace must be scaled by a factor $\frac{I_{\text{VEM, Off.}}}{I_{\text{VEM, Rand.}}}$ before being added to the container.

If desired, the frequency of accidental muons can be artificially increased by adding background simulations to the trace container as well. This is done either by directly specifying a number of random injections. If the number of accidental muons is nonzero, a sample of random muonic events from [113] is drawn and each sample added to every PMT at a random uniform position somewhere in the trace. Figure 6.7 shows an example where five muons are injected on top of random-trace background.

Last, the actual shower signal is added to the trace container. In principle, it can be added at any random position, similar to the random injections. However, for continuity reasons and ease of comparing to plots generated with other software, the latch bin for signal insertion is hardcoded to be the same as in $\overline{\text{Off}}\underline{\text{line}}$, bin 660. Since otherwise the data is in the correct ADC format, no further manipulation of the data is necessary.
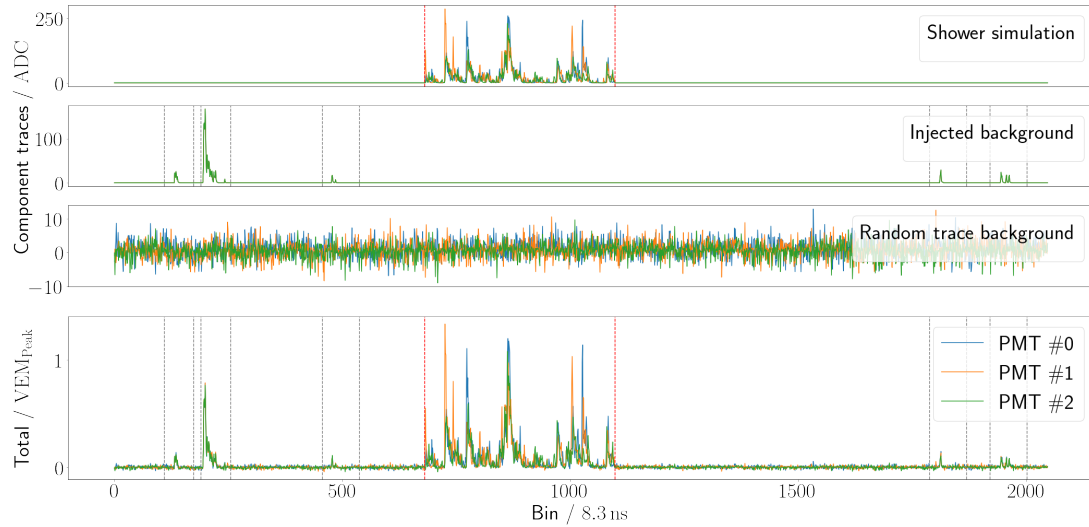
Figure 6.7.: The individual component traces (in units of ADC, top three plots) make up the eventual VEM trace (in $\mathrm{VEM_{Peak}}$, bottom plot). The dashed red (gray) lines signify where in the time trace the shower signal (injected muon signal) is located.

The trace container now holds all necessary components together, but remains in units of hypothetical, floating point ADC. To convert to $\mathrm{VEM_{Peak}}$, each bin in each PMT trace is floored to mimic the FPGA digital counting, and divided by the appropriate scaling factor I_{VEM} times a correction factor that stems from a bias in the online peak estimation algorithm (see [117]). If traces need to be UB compatible, the FADC bins must be filtered and downsampled in an intermediate step. This influences the scaling factor $I_{\mathrm{VEM,\,compat.}}$ in a major way (compare subsection 6.3.1). If the so-called full-bandwidth, UUB time trace is analyzed, the appropriate factor becomes $I_{\mathrm{VEM,\,Off.}}$.

6.3.1. Filtering and downsampling

The triggers discussed in the next chapter 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 ([117], [118], [119]).

These studies were conducted using the predecessor, the **U**nified **B**oard (UB), of the current electronics that is being installed during the AugerPrime upgrade of the observatory. Most importantly, the 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 2048, but it also drastically reduces the efficiency (in particular for ToT-like triggers) of the above discussed algorithms. Whereas a new 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. If the window size of e.g. the ToT trigger
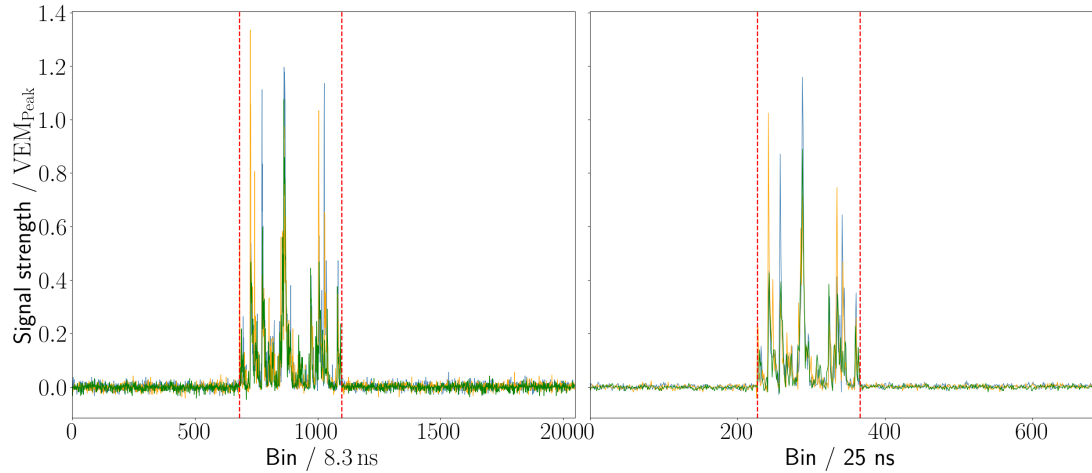
Figure 6.8.: (*Left*) A simulated trace as it would appear to UUB electronics. The ionizing particles originating in the extensive air shower hit the tank around bin 660 ($\approx 5.5\,\mu s$). (*Right*) The same trace but filtered and downsampled to emulate UB electronics.

were to remain constant, only a third of the original signal becomes available for a given window frame.

The modus operandi elected by the Pierre Auger collaboration to circumvent this problem while developing triggers appropriate for the full $120\,\mathrm{MHz}$ bandwidth is to emulate UB electronics using the UUB electronics. This means that measured FADC bins are filtered and downsampled before any trigger runs over them. Software implementations by which this is achieved are listed in Appendix A. The effect the filtering and downsampling has on measured data is visualized in Figure 6.8.

While the features of the time trace largely remain intact, the absolute signal strength decreases due to a smearing effect imposed by filtering. Overall, this amounts to a 30% difference in amplitude between UUB full-bandwidth traces and their filtered and downsampled counterparts. Since both measurements are derived from the same signal in the WCD though, this implies that I_{VEM} must be adjusted by 30% as well if traces are to be filtered and downsampled. This results in a compatibility scaling factor $I_{\mathrm{VEM, compat.}} = 163.235\,\mathrm{ADC}$ [89].

Recent contributions within the Auger collaboration ([120, 121]), and to some extent also this work have shown that issues arise in the comparison of **L**ateral **T**rigger **P**robabilities (LTPs) that are run in this compatibility mode. Namely, the UUB trigger efficiency (where full-bandwidth traces are filtered and downsampled) is lower than that of UB stations. This implies that the filtering and downsampling algorithms in Appendix A either make imprecise assumptions about the station electronics, or $I_{\mathrm{VEM, compat.}}$ or the trigger thresholds themselves need to be adjusted further. This fact has to be kept in mind when discussing results and comparing lateral trigger probabilites from classical triggers and neural networks.

6.3.2. Sliding window analysis & labelling

As will become apparent in chapter 7, three of the four trigger algorithms operate by examining only a window of measurement info, rather than evaluate the whole time trace as it is constructed in section 6.3. In a similar fashion, it is reasonable to assume neural networks do not need to receive the $3\,(\text{PMT}) \cdot 2048\,(\text{UUB bins}) = 6144$ input values from the entire trace to make an informed choice of whether or not a given signal stems from an extensive air shower.

For this reason, samples from the time trace are drawn via a sliding window analysis. A number of n_{bins} are extracted from the trace, to be analyzed by some classification algorithm. In order to not select the same information repeatedly, the window is moved by n_{skip} bins forward and the process can begin anew. Because the signal part is always added at a prespecified point in the total trace, the starting bin of the sliding window analysis is randomized with modulo n_{skip}. Unless explicitly specified otherwise, the hyperparameters in this sliding window analysis are set as

$$n_{\text{bins}} = 120, \qquad n_{\text{skip}} = 10. \tag{6.1}$$

The choice of n_{bin} coincides with the window length of the ToT-like triggers introduced in chapter 7, as they have been shown to work well for classifying the data at hand. Other values of n_{bin} will however be discussed in chapter 8. Whether or not a specific window contains signal from an extensive air shower - which is important for labelling data in the context of neural network training - is a simple exercise. The modular approach in section 6.3 allows to simply check for nonzero bins in the shower signal component of the trace. In practice, upon creating a new combined trace, the first and last positive bin in the shower component are identified. This is e.g. visualized with dashed red lines in Figure 6.7. If any overlap - even just a single bin - exists between the sliding window and this signal region, the extracted window is consequently labelled as signal. If this is not the case, the window is labelled as background.

Of course, quality cuts can be applied, and a decision to count a given trace window can be made based on several statistics. This is further discussed in chapter 8.

7. Classical station triggers

As mentioned in chapter 4, 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 4.2.4) 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, general comments about evaluation of trigger performances are given in section 7.1. Then, the precise implementation of SD station level triggers, as well as their individual performance is given in section 7.2.

7.1. Performance evaluation

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:

<div align="center">

Predicted C

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

True C

</div>

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 correctly raises a T2 trigger. The sensitivity - or **T**rue **P**ositive **R**ate (TPR) - is the ratio of correctly identified signals over all signals.

$$\text{TPR} = \sum_{\{\text{TP}\}} \frac{1}{w}, \qquad \text{where } w = \text{TP} + \text{FN}, \qquad (7.1)$$

$$\Leftrightarrow \qquad = \frac{\text{TP}}{\text{TP} + \text{FN}}. \qquad (7.2)$$

The TPR for a given sample heavily depends on the underlying distribution of data. While the distribution of zenith angles is correctly modeled, the simulated showers at hand do not accurately match the energy flux seen in real data (compare Figure 6.6 and Figure 2.4). For this purpose, the TPR for a given dataset is transformed to a conditional trigger efficiency ϵ by recalculating the weights w_i based on the primary energy E_i of a shower. First, the overrepresentation of higher energy traces is eliminated. Every prediction is weighted by a factor $b_{E_i}/b_{E_{\mathrm{ref}}}$, where b_{E_i} is the number of traces in a bin corresponding to energy E_i (see Figure 6.6). Afterwards, a flux factor is multiplied to the weight. It is of the form $E_i^{-3}/E_{\mathrm{ref}}$. Lastly, the thus calculated sum is normalized to the weight $W = \sum\limits_{\{\mathrm{TP,FN}\}} \frac{1}{w_i}$ of the entire dataset.

The physical interpretation of ϵ for a classifier is as follows: (At station level) It represents the ratio of air showers that *are* detected over all air showers that *can* in principle be detected. It is a conditional probability dependant on the WCD receiving a particle from an air shower. An efficiency of $\epsilon = 1$ thus implies that if at least one particle in the shower cascade hits a tank, the tank raises a T2 trigger.

The false discovery rate is a measure of how readily the station-level 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 raises the amount of T2 triggers, and hence load on the central analysis server of the observatory.

$$ \mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}. \tag{7.3} $$

In this work, the trigger rate on random-traces f_{Random} is used as an alias for the FDR. Of course, this is not completely accurate, as these to contain a small fraction of signals from air showers interesting to the science case of the Pierre Auger Observatory. In any case, the distinction does not matter. Any trigger algorithm must have a sufficiently low trigger frequency on measured data - be it extensive air showers or background events - as to not overload CDAS readout capabilities.

Consequently a pseudo-score can be assigned to each classification algorithm in order to compare them. This score a is given by Equation 7.4.

$$ a = \frac{\epsilon}{f_{\mathrm{Random}} \; [\mathrm{Hz}]} \tag{7.4} $$

The physical interpretation of this variable is not straight forward. If labels were known for random-trace datasets, a would be equivalent to the **S**ignal-to-**N**oise **R**atio (SNR) of the classifier. However, since f_{Random} and the T2 efficiency across all primary energies and zenith angles must be determined using two different sets of data, only a (positive) correlation between the two exists.

In conclusion, all algorithms should maximize a, i.e. boost the T2 trigger efficiency, while keeping the random-trace trigger rate as low as possible.

7.1.1. Lateral Trigger Probability (LTP)

Ultimately, all test statistics constructed by classical station triggers (ignoring MoPS, c.f. section 7.2) are correlated with the deposited charge S in the WCD. Because S is heavily influenced by the primary particle energy, zenith and distance to the shower core, as well as to a lesser extent by shower age and statistical fluctuations, it makes sense to parametrize the trigger efficiency $\epsilon(E, \theta, \mathrm{SPD})$ in terms of these observables.

From a heuristic consideration, it can immediately be concluded that large separations between station and shower axis affect efficiencies negatively, because the lateral particle distribution function monotonically decreases with increasing r (compare Figure 3.3 on page 19). Similarly, inclined showers with a large θ are more attenuated compared to vertical showers, as they have to traverse a larger atmospheric depth ($\propto \sec(\theta)$) before reaching the detector. Lastly, primaries with large E on average deposit higher S in the WCD due to unleashing bigger particle cascades. Consequently, ϵ is positively correlated with E.

The functional form that can be obtained by evaluating trigger efficiencies for a given (slice of) E and θ is labelled the **L**ateral **T**rigger **P**robability (LTP). It will be one of the main comparison metrics, by which different trigger algorithms are compared in this work. For classical triggers, two methods to extract the LTP are presented here. This is to show that both yield comparable results, and the latter method is a fair estimator for the neural network LTPs discussed in chapter 8.

$\overline{\mathrm{Off}}\underline{\mathrm{line}}$ lateral trigger probability

$\overline{\mathrm{Off}}\underline{\mathrm{line}}$ can simulate the SD detector response given a simulated shower footprint as given by e.g. CORSIKA. As such, calculating the LTP for a given event condenses to counting the number of triggered and non-triggered stations at specific distances from the shower axis. If this is done for a large enough sample size of showers, one eliminates noise induced by shower-to-shower fluctuations and arrives at an independent estimator for the probability of a T2 trigger at a given shower plane distance r, for primaries of energy E and zenith θ. As per [122], the closed form approximation of the LTP is given as

$$\mathrm{LTP}\,(r) = \begin{cases} \dfrac{1}{1 + \exp\left(\frac{r - R_0}{\sigma_R}\right)}, & r \leq R_0 \\[2ex] \dfrac{1}{2} \exp\left(C(r - R_0)\right), & r > R_0 \end{cases} \tag{7.5}$$

In Equation 7.5, R_0, σ_R and C are all fit constants that will, in general, depend on E and θ. Most importantly, R_0 marks the shower plane distance where $(R_0) = 0.5$. This is connected to a steepening of the rising flank in the efficiency curve. Whereas an exponential function with decay constant $C < 0$ describes data well for large r, a logistic function with scaling factor $1/\sigma_R$ must account for the asymptotic transition to full efficiency closer to the core.

It must be mentioned that the motivation behind this parametrization is data- and not physics driven. In particular, LTP(r) is not smooth in R_0 if the parameters C and σ_R are not finetuned as indicated in Equation 7.6.

$$\lim_{r \to R_0^+} \frac{\partial \, \text{LTP}(r)}{\partial r} = \frac{C}{2} \stackrel{!}{=} \frac{1}{4\sigma_R} = \lim_{r \to R_0^-} \frac{\partial \, \text{LTP}(r)}{\partial r} \qquad (7.6)$$

In this work however, a different parametrization is used to estimate the T2 response of a station. The functional form of this adjusted trigger probability is a clipped logistic function, and given in Equation 7.7.

$$\text{LTP}^*(r) = \min\left(1, \; \epsilon^* \left(1 - \frac{1}{1 + e^{-\frac{r - R_0}{\sigma_R}}}\right)\right) \qquad (7.7)$$

The reasoning for this choice is as follows:

- The original parametrization, LTP(r), eventually approaches 1. This hints to a problem. It is not a guarantee that some trigger algorithm will detect all extensive air showers. Espically neural network triggers might be sensitive to only a subset of showers. This is reflected in the latter form, LTP*(r), by introducing an additional fit parameter, the pseudo-efficiency $0 \leq \epsilon^* < 2$. In the case of $\epsilon^* \geq 1$, the domain of the function is correctly mapped to $[0, 1]$.

- There exists an imbalance in training data. Due to the geometry of the SD array, more traces at smaller r are available. In an attempt to reduce possible biases resulting from low statistics at small SPD, the form is kept as simple as possible.

- LTP*(r) is guaranteed to be continously differentiable in R_0. For values $\epsilon^* > 1$ this is replaced with a kink at $R^* = R_0 - \frac{\log\left((1 - 1/\epsilon^*)^{-1} - 1\right)}{\sigma_R}$, where LTP*($r$) would exceed 1 if not for clipping. There exists some mathematical motivation for this however. For $\epsilon^* > 1$ a phase transition, namely to full efficiency, exists at this point, discontinuities in the lateral trigger probability are thus allowed/expected. Whether full efficiency is ever reached for a given classifier is however questionable due to shower-to-shower fluctuations.

As stated above, this approach only marginally takes into account shower-to-shower-fluctuations. Such statistic perturbations are in part responsible for a smearing of the (initially) hard transition from sub- to full efficiency. The parametrization used by the Pierre Auger collaboration takes this into consideration by design. The presented LTP*(r) does - at least explicitly - not. As a result, one could expect a bias, where LTP*(R^*) over- or underestimates the actual trigger probability. This is however not the case when examining the residuals of the performance fit that is done in subsection 7.2.5. A plot showcasing this is offered in Figure 7.1. Still, results at low energies ($\log(E \, / \, \text{eV}) < 17.5$) must be taken with a hint of skepticism, as only very little data is available for such showers.
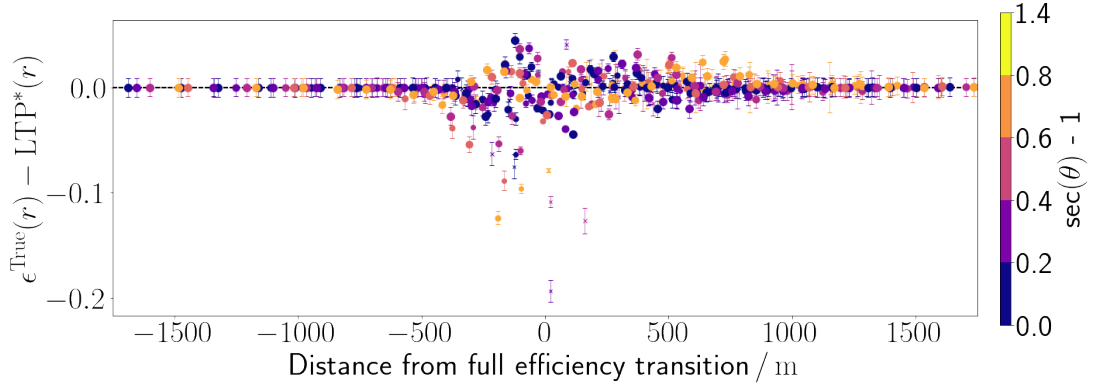
Figure 7.1.: The residuals from comparing ϵ^{True} to $\text{LTP}^*(r)$ vanish at large (small) r. No systematic bias is observed in the transitional region around R^*. Large outliers are caused by the dataset being limited in size at low energies (marked with an x).

Bayesian folding

Given a time trace of the form in chapter 6, it is easy to determine whether or not a given trigger algorithm analyzing this trace would raise a T2. The trigger probability obtained this way must however not be confused with the lateral trigger probability discussed above. It is the conditional probability of a T2 given that a tank receives an air shower signal, $P(\text{T2} \mid C_1)$. The complete LTP can be calculated in the following form:

$$\text{LTP} := P(\text{T2}) = P(\text{T2} \mid C_1) * P(C_1). \tag{7.8}$$

$P(C_1)$ in Equation 7.8 is the **L**ateral **P**article **P**robability (LPP) and quantifies the chance of a station receiving a signal from an extensive air shower. It is the probabilistic interpretation of the lateral distribution function (c.f. Figure 3.3). The LTP of an ideal classifier that is able to identify individual particles would be equal to the LPP.

For this work, the LPP is calculated by comparing the simulated stations which received at least one particle from an air shower in section 6.2 to a catalog of known stations in the vicinity of the shower core. The ratio of simulated stations divided by all stations at a specific SPD is the LPP at a given (slice of) energy E and zenith θ. A function like in Equation 7.7 is fitted to the data to extrapolate values at arbitrary SPD. The result of this analysis for all energies and zenith angles is shown in Figure 7.2. The best fit parameters are listed in Appendix B.

The comparison between the LTP gathered via Bayesian folding is compared to the $\overline{\text{Off}}\underline{\text{line}}$ station counting approach in section 7.2. This is done for every classical trigger individually. It is found that no remarkable difference between both results exists. As a consequence, this method of evaluating the LTP allows for an easy comparison of neural network triggers to the algorithms discussed in the following.
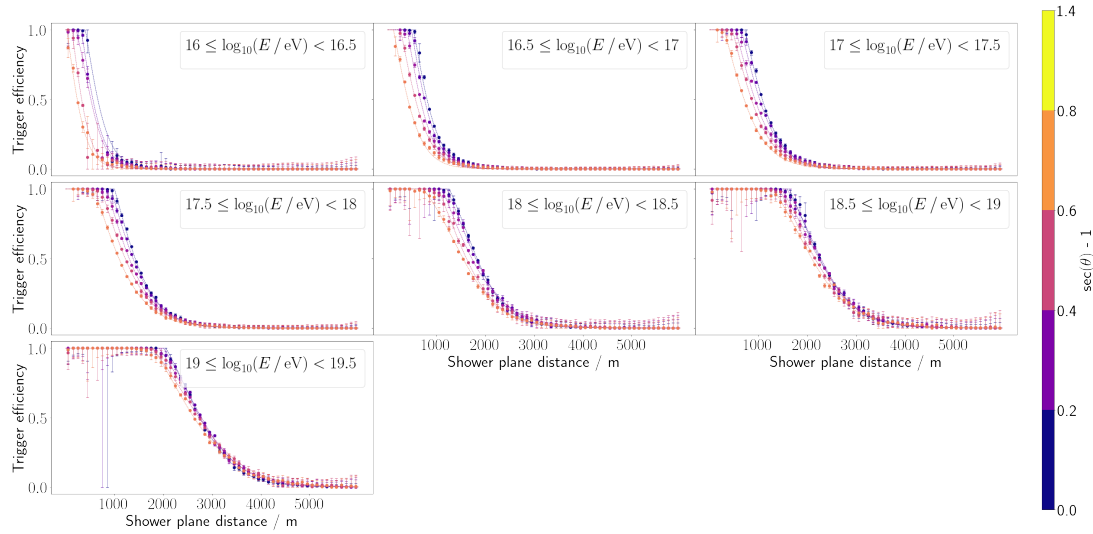
Figure 7.2.: The probability of a station receiving a particle in relation to its separation from the shower axis. The probability at a constant SPD with increasing primary energy, but decreases with increasing zenith angle.

## 7.1.2. Calculation of T3 efficiency

When transitioning from station- to event-level, the important variable becomes the T3 efficiency. It states the probability of a shower being detected in multiple SD stations in such a way that a spatial/temporal pattern is recognized in CDAS and a T3 trigger is raised. In the end, the only measured data that is available for higher level analysis comes from extensive air showers which passed this final hurdle.

With the lateral trigger probability at hand, a simple Monte-Carlo simulation can recover the T3 efficiency. Recall that the probability of a single station i detecting a shower with energy E and arrival direction (θ, ϕ) at distance distance r_i (which is a function of core position, zenith and azimuth) from the shower axis is given by Equation 7.9.

$$\text{LTP}^*(r_i) = \min\left(1,\ \epsilon^*(E, \theta)\left(1 - \frac{1}{1 + e^{-\frac{r_i - R_0(E,\theta)}{\sigma_R(E,\theta)}}}\right)\right). \tag{7.9}$$

In the simplest case, this results in a T3 trigger if the three closest stations raise a T2 within 11 μs of one another. By simulating a random core position and determining the detector response, one can calculate the event detection numerically. The approach presented here does not take into account timing differences resulting from the finite propagation speed of the shower front. Instead, it is assumed that all stations simultaneously receive a signal. This is only accurate if the primary particle initiated a perfectly vertical cascade infinitely far away from the SD. However, due to the permissivity of the T3 triggers such considerations need not be accounted for. Even in the most suboptimal case of a horizontal shower ($\theta = 90°$), the furthest station in the
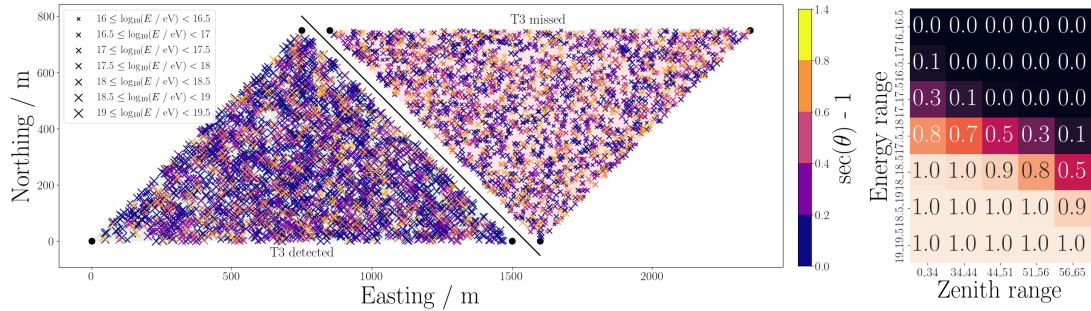
Figure 7.3.: (*Left*) An example of randomized core positions color- and size-coded according to zenith and primary energy. Showers that raised a T3 are shown in the green unit triangle. The ones that do not are displayed in the red part. The stations at the vertices of the SD unit triangle are shown as black circles. (*Right*) The ratio of triggered vs. all showers, with respect to primary energy and shower inclination.

triangle receives a signal latest around $\frac{1.5\,\mathrm{km}}{c} = 5\,\mu$s after the closest one. Exemplary simulations and resulting T3 efficiencies for classical triggers are shown in Figure 7.3 and further discussed in subsection 7.2.5.

7.2. Implementation

7.2.1. Threshold trigger (Th)

The **Th**reshold trigger (Th) is the simplest, as well as longest operating trigger algorithm [123] 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 code block below.

```
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 ThT1_trigger
    if pmt1 <= th1 and pmt2 <= th1 and pmt3 <= th1:
        raise ThT2_trigger
    else:
        continue
```

Logically, with increasing signal strength S in the PMTs, the likelihood of having observed an extensive air shower increases. This is reflected in the trigger level logic, where a coincident signal of $S \leq 3.20\,\text{VEM}_{\text{Peak}}$ is immediately forwarded to CDAS, whereas a signal $1.75\,\text{VEM}_{\text{Peak}} \leq S < 3.20\,\text{VEM}_{\text{Peak}}$ only raises a Th-T1 trigger. The algorithm is insensitive to signals that do not exceed at least $1.75\,\text{VEM}_{\text{Peak}}$ 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 7.1 in order to keep trigger frequencies constant.

Table 7.1.: Numerical values from [118]

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

Performance

The average trigger rate for the Th-T2 trigger per station is *defined* to be $\approx 20\,\text{Hz}$ (c.f. subsection 4.2.3). Comparing this to the nominal T3 trigger rate at CDAS level ($\approx 0.03\,\text{Hz}$ [94]) over the entire array, it becomes obvious that a lot of background events pass this threshold. Consequently, the trigger has a very high false discovery rate on a station to station level FDR $\approx \frac{20\,\text{Hz} - 0.03\,\text{Hz}/1600}{20\,\text{Hz} + 0.03\,\text{Hz}/1600} = 0.999998$.

The efficiency of the threshold trigger is comparably poor. Only every fifth trace ($\epsilon = 0.2$) in the training set (c.f. section 6.2) results in a T2. This number must however be taken with context. In chapter 6, a signal is considered to be any kind of detector response from an extensive air shower. This includes single muons injected into WCDs faraway from the shower core. As such, the dataset that triggers are being tested on contain a lot of information algorithms were designed to ignore. This drops the efficiency considerably. Nevertheless, it serves as a gauge to compare this trigger to the ones discussed on the following pages.

While this may seem like an indigent method of shower detection, the threshold trigger plays a central role in T3 formation and is invaluable in the search for neutrino cosmic rays. The EM component of many such neutrino showers is heavily attenuated due to their inclination ($\theta \geq 65°$). The muonic component is the only signal that reaches the SD detector. The threshold trigger ensures the array is sensitive to such events, at the cost of a high background noise.

The lateral trigger probability for the Th-T2 type trigger is shown in Figure 7.4
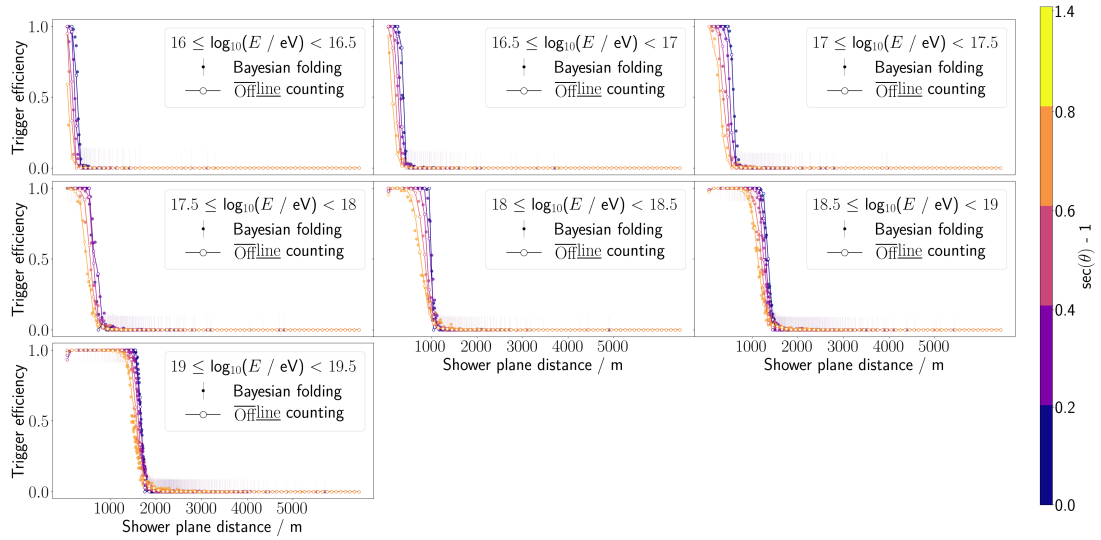
Figure 7.4.: The lateral trigger probability for the threshold T2 trigger. Shown is the comparison between the different procedures discussed in section 7.1. Both results agree across all considered primary energies and zenith angles.
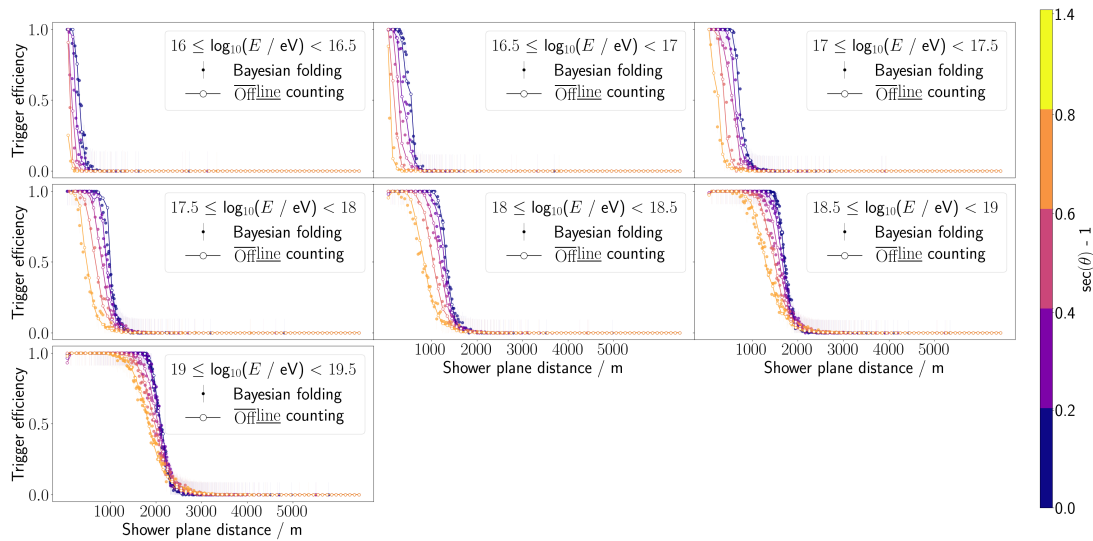
7.2.2. Time over Threshold trigger (ToT)

The **T**ime **o**ver **T**hreshold trigger (ToT) is sensitive to much smaller signals than the Threshold trigger discussed in subsection 7.2.1. For each PMT in the water tank, the past 120 bins ($\approx 1\,\mu$s) are examined for values that exceed $0.2\,\mathrm{VEM_{Peak}}$. If 13 or more bins above the threshold are found in the window - ordering or succession do not matter - the PMT is considered to be active. The ToT trigger requires at least two PMTs to be active in order to issue a T2. As such, the algorithm is theoretically sensitive to events that deposit just $\approx 0.5\,\mathrm{VEM_{Ch.}}$. 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

buffers = [[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:
```

Figure 7.5.: The lateral trigger probability for the time-over-threshold trigger for Bayesian folding and $\overline{\text{Off}\underline{\text{line}}}$ station counting.

```
18              count_active_PMTs += 1
19
20      if count_active_PMTs >= 2:
21          raise ToTT2_trigger
22      else:
23          step_count = buffer_index + 1
24          continue
```

Performance

The nominal operation of the ToT algorithm sees a trigger rate of 2 Hz. While this still corresponds to a relatively large FDR, the signal to noise ratio is at least an order of magnitude better than Th-T2, simply by arguments of trigger frequency. Moreover, coincidences between neighbouring stations are likely to be extensive air showers. Events selected from ToT issued T3s have a purity of 90%, and are the main detection channel for showers with inclination $\theta < 60°$ [94]. The ToT trigger itself has an efficiency of $\epsilon = 0.3969$ when evaluated over the complete neural network training dataset (section 6.2). The trigger probability w.r.t shower plane distance is shown for different primary energies and different arrival directions in Figure 7.5.

7.2.3. 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
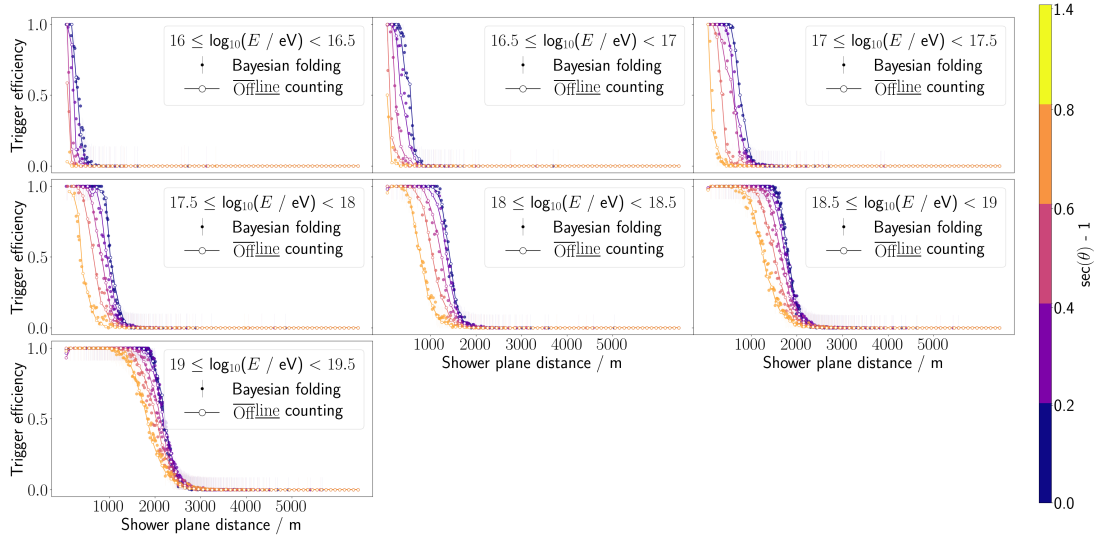
Figure 7.6.: The lateral trigger probability for the time-over-threshold-deconvoluted trigger. Errorbars are plotted transparently in order to not overcrowd the figure.

to the ToT trigger in subsection 7.2.2. Only the FADC input stream from the three PMTs is altered according to Equation 7.10.

$$d_i = \left(a_i - a_{i-1} \cdot e^{-\Delta t/\tau}\right) / \left(1 - e^{\Delta t/\tau}\right) \tag{7.10}$$

In Equation 7.10, 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 4.2.1). The choice of the numerical constants τ and Δt is explained in more detail in [124].

Performance

The performance of the ToTd trigger is very similar to that of the ToT discussed in subsubsection 7.2.2. It posesses a comparable efficiency $\epsilon = 0.4027$ to its' convoluted counterpart. The lateral trigger probability for this algorithm is shown in Figure 7.6.

7.2.4. 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 two bins that are monotonically increasing in value, in a window of 120 bins. Once such a positive step has been identified, a (MoPS) trigger veto is applied to the next

$$n_{\text{skip}} = \lfloor (\log_2(\Delta y) + 1) - 3 \rceil \tag{7.11}$$

bins, where Δy refers to the total vertical increase in the step from first to last bin. Note that in Equation 7.11, the notation $\lfloor x \rceil$ is used as shorthand notation to round x to the nearest integer. If Δy is bigger than $y_{\min} = 3\,\text{ADC}$ (to filter random fluctuations), but does not exceed $y_{\max} = 31\,\text{ADC}$ (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. An in-depth discussion of the different hyperparameters for this trigger is offered e.g. in [125].

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 when comparing performances of classical triggers to those in chapter 8. The implications of this choice are layed out in the following paragraph.

Performance

By examining monitoring data, it follows that $1 - 2$ MoPS triggers are issued per station each minute. This corresponds to a trigger rate of $0.02\,\text{Hz}$ to $0.03\,\text{Hz}$. This is orders of magnitude lower than other discussed trigger mechanisms. While the MoPS trigger consequently can be seen as a relatively noise free trigger, events in which a MoPS is critically required to form a T3 are extremely sparse. From a total of 20000 (simulated) showers at lower energies, none had a three-fold coincidence where at least one station only detected a MoPS T2 trigger. Considering this result, it is expected that T3 efficiencies are largely independant of the fact whether MoPS is considered or not. This is consistent with previous analysis on proton primaries in [126].

7.2.5. Combined performance

In the field, all above discussed algorithms are run simultaneously. That is, a T2 trigger is issued whenever any of the Th-T2, ToT, ToTd, or MoPS trigger become active. This results in an overall trigger rate of roughly $22\,\text{Hz}$ to $23\,\text{Hz}$, with a combined efficiency of $\epsilon = 0.4070$. The combined lateral trigger probability with respect to shower plane distance is shown in Figure 7.7. The overall T3 efficiency as calculated in subsection 7.1.2
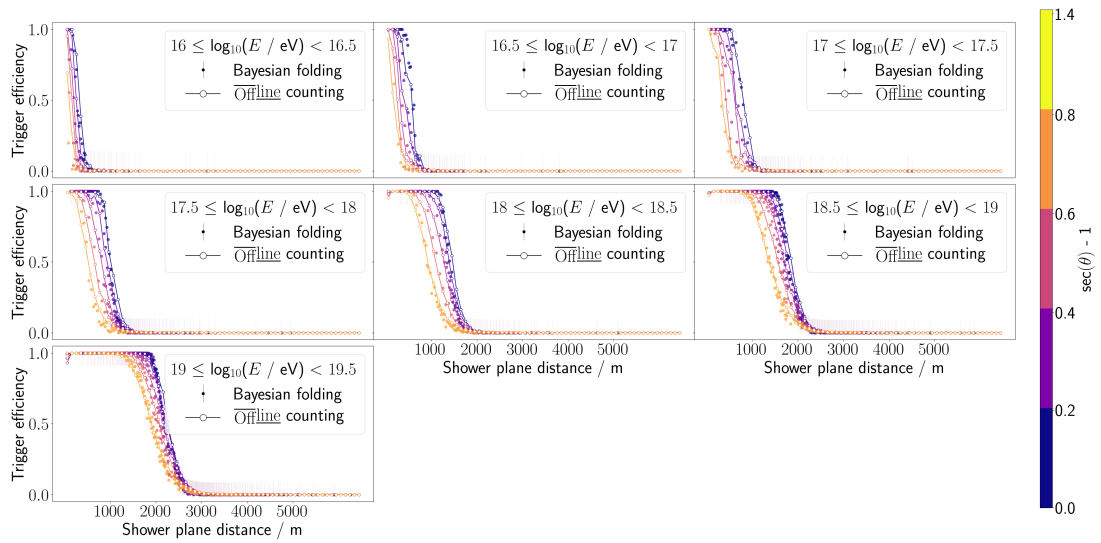
Figure 7.7.: The lateral trigger probability for the combination of all previously discussed trigger algorithms.

is shown in Figure 8.1. As can be seen in the plot, detection of air showers is guaranteed at primary energies of around $10^{18.5}$ eV and upwards.

8. Neural network triggers

8.1. Motivation

The station-level triggers in the previous chapter have been shown to perform well for the science case of the Pierre Auger Observatory. However, it has also been concluded that potential significant discoveries might be made at energies currently outside of the sensitive region of the detector. The data of many low energy events is not read out by intention in order to keep DAQ readout at feasible levels.

Attempts at improving the overal efficiency of the SD triggers can be made. This is only possible to a certain level. At lowest energies, the particle cascade is not big enough to result in coincident triggers in at least 3 WCD stations. As per subsubsection 7.1.1, the lateral trigger probability a given classification algorithm can maximally achieve is given by the LPP (c.f. Figure 7.1). The T3 detection probability of such an ideal trigger, and consequently the maximal efficiency for an array with 1.5 km spacing is compared to the efficiency of classical triggers in Figure 8.1.

Of course, efficiency can be improved simply by adjusting trigger thresholds of the algorithms in section 7.2. However, the more lenient these thresholds are, the more background events will be detected. This quickly results in trigger rates that are unmanagable for the infrastructure at the Pierre Auger observatory. The probability with which time traces correctly raise a T2 is shown alongside the resulting random-trace trigger rate for different thresholds of classical algorithms in Figure 8.1.

Ideally, neural network architectures developed in this chapter should undercut the random-trace trigger rate of classical triggers, while retaining an overall higher accuracy. That is, they lay below and right of the operating point in Figure 8.1. For any algorithm that achieves this, the corresponding LTP will be greater than that of classical triggers, resulting in higher event detection efficiency, while not exceeding the bandwidth limitations of the underlying hardware.

8.2. Implementation

The python library TensorFlow [127] is used as a backend to implement the individual classifiers. All discussed architectures are built and trained with the release version 2.8.4 [128]. Adjustments to the trainable parameters are calculated according to a momentum-based stochastic gradient descent (Adam [129]) on a batch level. In this context, a single batch is made up of all sliding windows of all traces that are recorded from a single air shower event. Since batch size grows quickly with increasing energy,
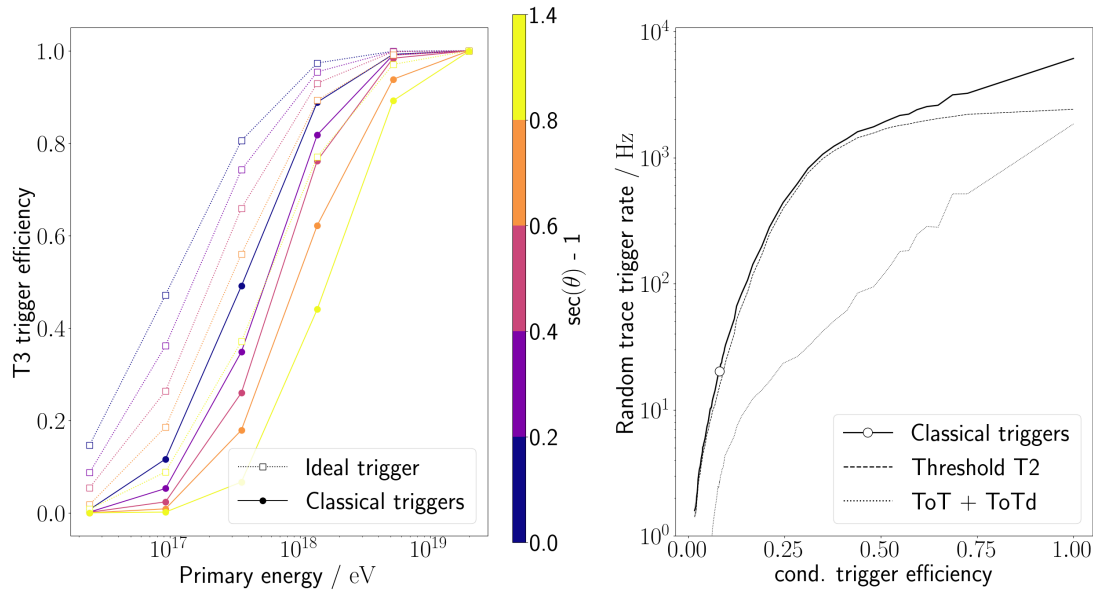
Figure 8.1.: (*Left*) Comparison of an ideal trigger sensitive to any shower signal from primary energies $E \geq 10\,\text{PeV}$ to classical triggers. (*Right*) The noise level over calculated efficiency for classical triggers. The tail ends of the potential curve are calculated by adjusting the trigger thresholds from 0.05 to 2.5 times of the nominal values. The black circle corresponds to the operational setting of a 20 Hz rate.

a generative approach, where traces are created (c.f. section 6.3) at runtime upon requirement, is used in building training data in order to make the process as RAM-efficient as possible. This has important implications. As trace building relies heavily on randomization, the actual training data will not be the same if the random number generators are not seeded beforehand. This has been taken into account. All networks are - unless specifically stated otherwise - trained and validated using the same seeded input data.

8.3. Design consideration

8.3.1. Architecture

The hardware specifications at the FPGA level, where trigger conditions are currently checked, are limited. For example, the Artix-7 chip used in UUB stations is equipped with 5 MB RAM only [130]. For this reason, NN architectures should be kept as simple as possible. Most importantly, weights, biases and other trainable parameters will need to be hardcoded into station firm- or software. Because of the minimal available storage space, this number needs to be low.

This immediately disqualifies powerful candidates like autoencoders or transformers (compare section 5.4) from consideration. Only relatively simple dense-, convolutional-,

or recurrent neural network architectures are viable contenders and might be implementable in the SD electronics.

8.3.2. Choice of input data

As stated in the previous chapters, neural networks learn by example. It is imperative that the input data propagating through the network during training resembles later use-case inputs as much as possible. However, some choices in how data is represented can and need to be made in order to benefit loss minimization.
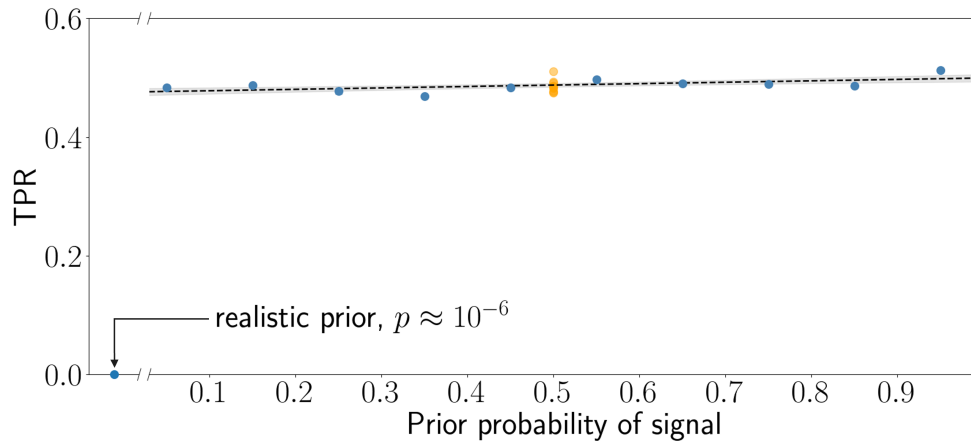
Prior probability

The flux of cosmic rays with energies exceeding the proton knee is tiny ($O(1 \, \mathrm{m}^{-1} \, \mathrm{yr}^{-1})$ [42]). While the size of the SD guarantees decent exposure over the entire array, an individual station will predominantly measure background. In fact, the prior probability p of encountering events beyond the proton knee in a random time trace is roughly one in one million. Of course, if this were reflected in the training dataset, neural networks would show very poor performance. On the notion of a broken clock being correct twice a day, a naive classifier would naively label every input as background and retain almost perfect accuracy. Such behaviour is not desired. The prior probability must be artificially inflated. The influence of prior probability on subsequent trigger sensitivity is shown in Figure 8.2. No strong correlation between prior and network performance is found in the range $0.05 \leq p \leq 0.95$. As long as the fraction of signal over entire training set is statistically relevant, the network learns to discern between signal and background events. For all further analysis a conservative prior of $p = 0.5$ is picked.

8.3.3. Further hyperparameters

The number of ways both input data and the network architecture can be tweaked to minimize loss is extensive. Further optimization methods, like the choice of labelling, are motivated and discussed in the performance reports in section 8.4.

One promising approach of optimizing trigger performance that is not examined in this thesis is the adjustment of the neural network loss function. Several reasons exist that motivate an analysis in this direction:

- **Classwise weighting:** The classes C_1 and C_2 are seen as equally important when calculating efficiencies by default. As noted in subsection 8.3.2 an imbalance in distributions exists. Observing background is much more likely than observing an air shower. The loss function will reflect this to an extent if classification of background is weighted more heavily than classification of signal. Even a more complex loss sensitive to CR flux (compare Equation 7.3) or other parameters is possible.

Figure 8.2.: A very faint positive slope ($m = 0.02 \pm 0.01$) is observed when relating prior probability to trigger sensitivity (blue dots) of an example convolutional neural network with the architecture in subsection 8.4.1. A charge cut of $t_S = 0.5 \, \text{VEM}_{\text{Ch.}}$ is applied for signal labelling. The observed fluctuations could however be attributed to statistics in the training fit. An ensemble of networks with the same architecture trained on the same data and with a single prior shows a comparable spread (orange dots).

- **Random-trace trigger rate pull term:** For the most part, the analysis presented here will introduce a neural network distinguished by some feature, optimize its performance, and check whether the resulting random-trace trigger rate is in an acceptable range. However, it is possible to go the opposite way and specify a target trigger rate first and then train the network. In this fashion, one could hope that the algorithm efficiency is optimized for a given bandwidth by adding a pull term to the loss function. The challenge of this approach is weighting the pull term in a way that it does not dominate the loss function in different fringe cases. An important drawback of this method is that training will be very slow.

8.3.4. Training procedure

In order to make informed statements about the capability of a given network architecture, an ensemble of 10 networks are trained - unless specified otherwise - for 10 epochs. Per epoch, 80% of the available showers are examined to calculate gradients, the remaining 20% are set aside for validation purposes. This is a standard procedure to prevent the network from overfitting training data [131]. Since the training dataset is extensive ($0.8 \cdot 40557 \approx 32445$ showers), convergence to a local minimum is typically observed already within a single epoch. For this purpose, training is halted early if the training accuracy is greater than $\alpha = 0.95$ and no decrease in loss has been observed for at least $75\% \stackrel{\wedge}{=} 24334$ showers of the epoch
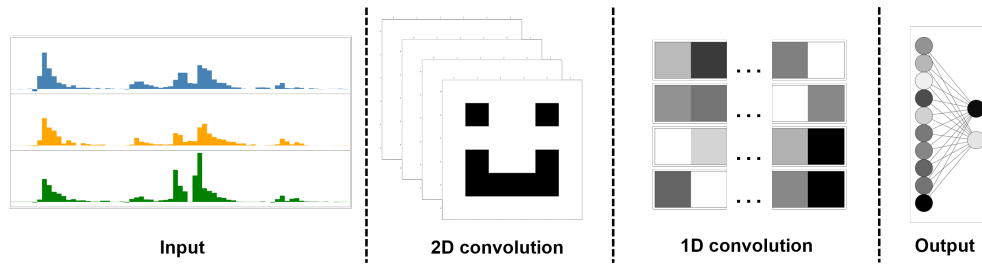
Figure 8.3.: The convolutional neural networks take a two-dimensional input consisting of a time trace excerpt of the three WCD PMTs. On this input data, a two-dimensional convolution is performed with four different filters, resulting in four one-dimensional vectors that are propagated further. A one-dimensional convolutional layer scans the previous outputs and extracts two features per vector, these are finally reduced to a binary output via a dense layer with softmax activation. For extended CNNs, a dense layer with 10 neurons is interposed with a relu activation function.

8.4. CNN Trigger

8.4.1. Architecture

The convolutional neural networks presented in this chapter all have a very similar architecture, which was shown to slightly outperform other candidates during prototyping tests. It is likely that this is due to design ideas that are closely related to the way information is structured in the time traces:

- A **2D convolutional input layer** (c.f. section 5.2) pools together the information of the three different WCD PMTs.

- A **1D convolutional layer** resolves temporal correlation in the output of the previous layer.

- A **Dense layer** reduces the extracted features to a binary choice of signal and background.

In subsequent tests, it is furthermore concluded that extending the architecture by an intermediate dense layer between the time convolution layer and the dense output layer is on average beneficial for signal to noise ratio. However, because both approaches arrive at a similar best score, the smaller, non-extended network is seen as more viable for the use case in the Pierre Auger Observatory. An example performance of both models is given in Figure 8.4. The corresponding number of trainable parameters is given in Table 8.1. The precise implementation designs for both cases are shown in Figure 8.3.

Table 8.1.: Number of training parameters n_{Train} for different network architectures

| Type	Input size	Kernel size	n_{train}	w/ dense extension
CNN	$(3, 120)$	$(3, 3)$	**140**	**834**
CNN	$(3, 120)$	$(3, 10)$	**216**	**534**
CNN	$(3, 120)$	$(3, 30)$	**444**	**714**
CNN	$(3, 40)$	$(3, 3)$	**84**	**210**
CNN	$(3, 60)$	$(3, 3)$	**100**	**290**
CNN	$(3, 90)$	$(3, 3)$	**120**	**390**
CNN	$(3, 240)$	$(3, 3)$	**220**	**890**
LSTM	$(3, 120)$	–	**12**	(single layer)
LSTM	$(3, 120)$	–	(three layers)	**44**

8.4.2. Performance

Training a convolutional neural network with the data from chapter 6 as is results in near perfect accuracy. The lateral trigger probability obtained after training is equal to the LPP (see Figure 8.5 and Figure 7.2), which quantifies the probability of particle existance at station level. This implies that the networks are able to identify even single particles in the WCD with very high confidence. However, they have not learned to distinguish muonic events from actual extensive air showers. This results in a very high random-trace trigger rate of $f_{\text{Random}} > 2\,\text{kHz}$. In order to decrease this value, the training data must be altered.

Muon cut

As discussed in chapter 6, atmospheric muons from very low-energy showers represent the dominant background (that is not just electronic noise) in measured data. The most straight forward approach of lowering the network sensitivity for such signals is requiring muonic coincidences. This means only data from tanks which were hit by at least 2, 3, ..., n_μ muons are assigned a signal label, where n_μ is a tweakable hyperparameter. Traces can be selected based on other particles as well. $\overline{\text{Off}\underline{\text{line}}}$ saves the number of muons, electrons and photons for simulated stations. All three, or even a (weighted) sum are candidate discriminators. In this work, only the muon cut is analyzed with greater detail, as it is thought to be more closely tied to the problem of distinguishing signal from background.

Several ensembles are trained with different n_μ to test the quality of this cut. Their subsequent performance is shown in Figure 8.6. The results are inconclusive. While low cut thresholds requiring a coincidence of $\gtrsim 2$ muons seem to represent a solid way of lowering f_{Random}, the trigger rates are not brought down to acceptable levels by this cut alone. Even worse, the training fits do not converge uniquely at higher cut
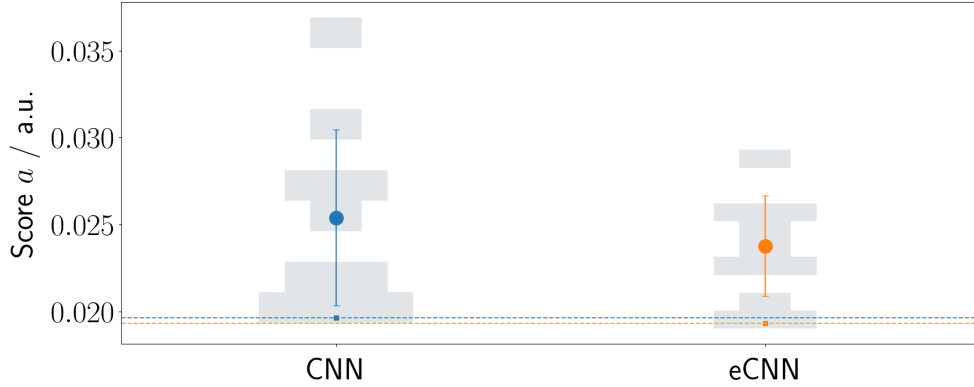
Figure 8.4.: The network score as calculated in Equation 7.4 is compared from original network architecture (CNN, left, in blue) to the extended architecture featuring an additional dense layer (eCNN, right, in orange). Shown are 10 networks of each architecture trained on the same dataset. Both architectures find a similar best solution (squares), while the mean and spread of the population (circles) are smaller for the extended network.
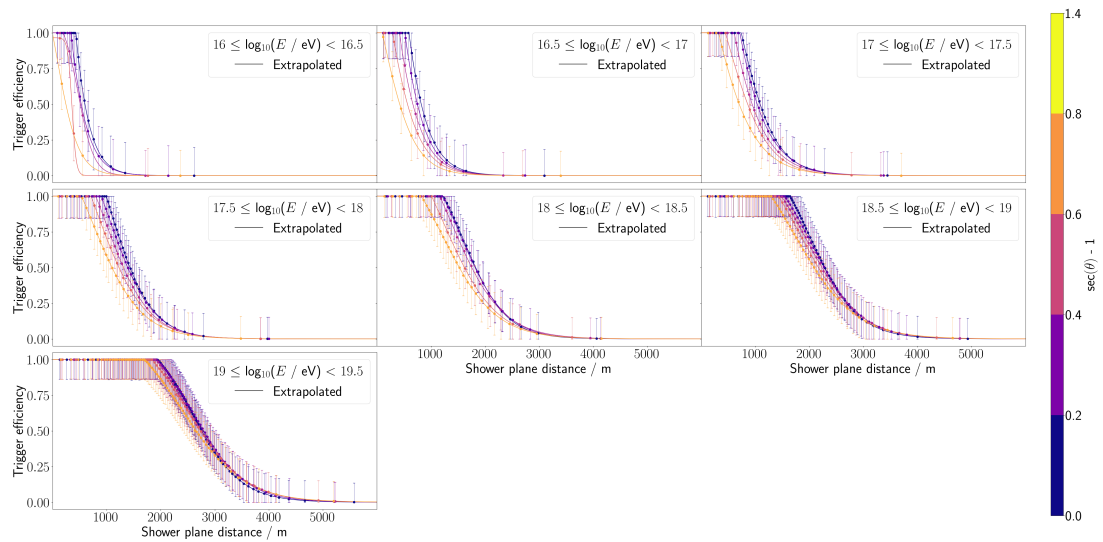
Figure 8.5.: The lateral trigger probability is equal to the lateral particle probability shown in Figure 7.2 for almost all E and θ. Discrepancies occur at low primary energies and high inclinations due to low statistics complicating the LTP fit. In any case, this implies that the chance of raising a T2 provided a station was hit by an air shower is 1.
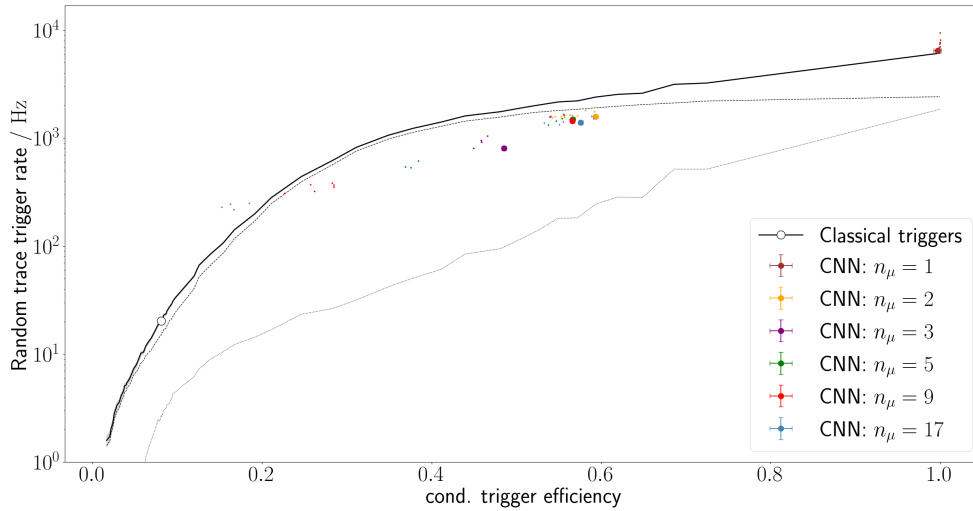
Figure 8.6.: Relation between random-trace trigger rate (lower is better) and conditional trigger efficiency (higher is better). The performance of convolutional neural networks trained on data with a muon cut of different thresholds. The optimizers don't seem to identify a unique local minimum at very high cut thresholds. For all intermediate cuts $1 < n_\mu < 17$ the convolutional neural networks offer a better trigger efficiency compared to the Th-T2 trigger at the same random-trace trigger rate. The performance of the ToT trigger is not reached.

thresholds. This indicates the number of muons is not (at least at higher muon counts) a great discriminator. This may be attributed to several reasons:

- **Lack of strict temporal coincidence**: The imposed cut only requires at least n_μ muons to hit the tank at some point during a shower event. Without a more stringent requirement it is possible that a network sees not a trace stemming from n_μ muons, but rather n_μ traces from individual particles during a sliding window analysis. Since all of these fit the cut criterion and are hence labelled as signal, the network does not learn to ignore such traces.

- **Variety in single muon traces**: The detector response to a single muon varies based on muon inclination, detector specifics and statistics. All of these are parameters that are not available to the CNN. The combination of all fluctuations might be too complex for a convolutional neural network with limited size to learn.

Nevertheless, a hybrid cut requiring at least 2 muons in a tank on top of some other trace characteristic might hold potential in further analysis. A muon cut by itself is shown to not work ideally for the data at hand.

Charge cut

Another method of teaching a network to ignore certain signals is given by imposing a cut t_S on the integral S of the trace. One advantage of this approach is the fact that
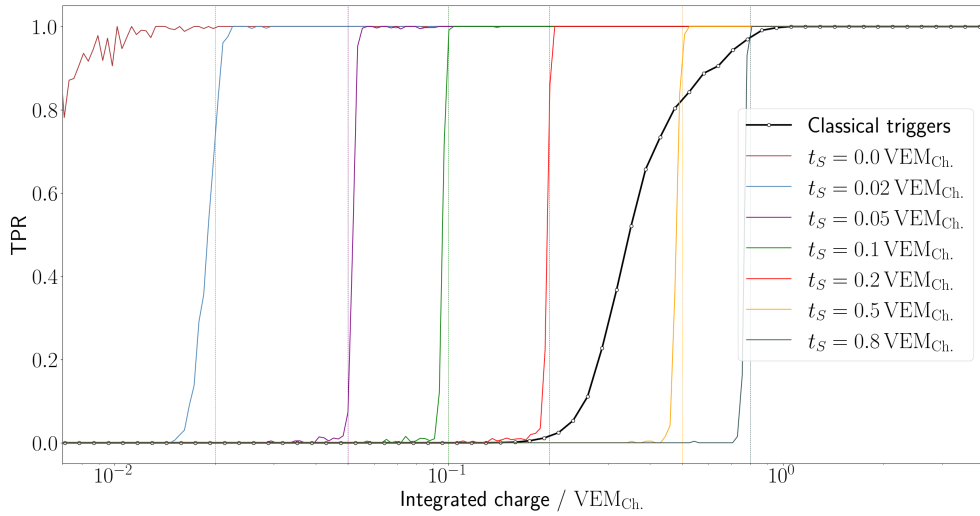
Figure 8.7.: The true positive rate (TPR) in terms of deposited signal for different classifiers. Convolutional neural networks that were trained with a cut threshold of t_S will not trigger on traces that have a lower integral $t_S > S$, dropping the sensitivity in this region. This is precisely the desired effect. Classical triggers become fully efficient at around $1\,\mathrm{VEM_{Ch.}}$.

the choice of labelling in this case is fully determined by the input data. Another edge is the granularity of the cut. Whereas only integer values of n_μ are sensible in the previous paragraph, the integral is discretized in much smaller steps of $1\,\mathrm{VEM_{Ch.}}^{-1}$. This allows for finer tuning of the network performance.

By increasing t_S, the network will learn to ignore increasingly larger signal. This is shown in Figure 8.7. The performance arising from different t_S is shown in Figure 8.8.

Full Bandwidth vs. Filtering & Downsampling

Charge cutting provides a powerful method to control the conservativity of the classifier and force discrimination between muonic background and extensive air showers. Equipped with this tool, other questions can be answered in an attempt to optimize network performance.

For example, classical triggers are run in compatibility mode, where 3 UUB bins (8.3 ns) are filtered and downsampled to resemble a single UB bin (25 ns). Some information about the trace shape is lost in the process. Logically, a neural network should be able to identify original UUB signal traces with more ease than their filtered and downsampled counterparts. Ideal triggers should have a higher efficiency when tested on full bandwidth traces.

This does not render compatibility mode uninteresting for analysis. An effective way of minimizing the trainable parameters in a neural network is reducing its' number of input parameters. Filtering and downsampling encodes - albeit imperfectly - the input data and reduces the dimensionality by a factor of three. This is done in such a
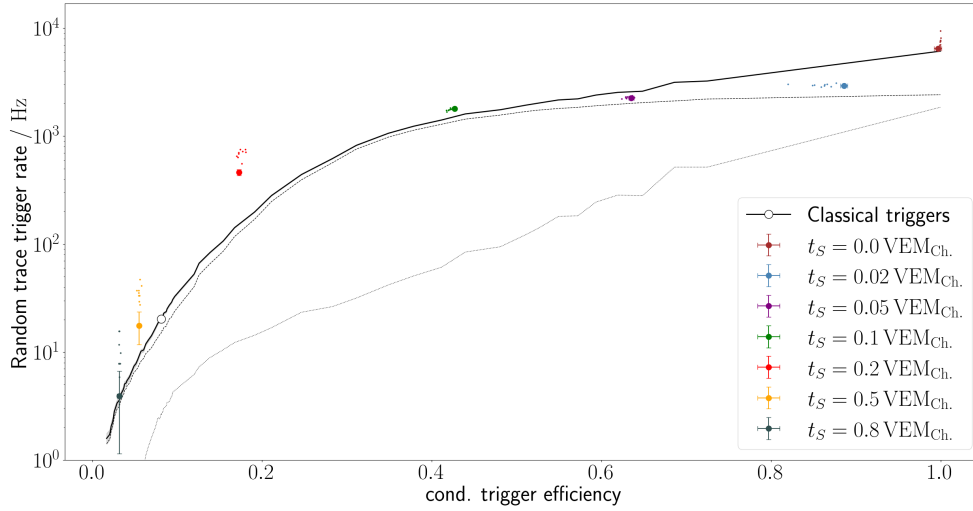
Figure 8.8.: Relation between random-trace trigger rate (lower is better) and conditional trigger efficiency (higher is better). Convolutional neural network performance in relation to different integral thresholds. With an increasingly stringent cut, the networks learn to ignore more and more low-energy events, which quickly drops the trigger efficiency. On the other hand, the classifiers random-trace trigger rate drops to more acceptable levels.

way that the UB-like trace contains trace information from a time window that is three times as long for a given number of bins.

Consequently, by filtering and downsampling input data it becomes possible to hand the classifiers additional information while keeping the network size at a fixed level. The effects this has on predictive power of neural networks is shown in Figure 8.9.

Classifiers trained on compatibility mode traces consistently have a lower random-trace trigger rate while operating at the same efficiency. Whether this is caused by the above discussed difference in (temporal) input size or other trace characteristics is determined in the next paragraph.

Input size

To test whether filtering and downsampling input data gives the classifiers an edge when predicting labels, multiple ensembles are trained on the same dataset with the same charge cut $t_S = 0.5\,\mathrm{VEM_{Ch.}}$. Each ensemble differs in the size of the input trace. If compatibility-trained networks are still outperforming full bandwith ones that have an input size three times the size, filtering & downsampling is the preferred method for classifying air showers with convolutional neural networks.

Indeed, this is found. While the compared networks converge to a similar trigger efficiency, the networks that receive filtered and downsampled traces show a lower random-trace trigger rate. Furthermore, a larger input size slightly positively influences trigger efficiency at the cost of increased network size (c.f. Table 8.1). This is expected.
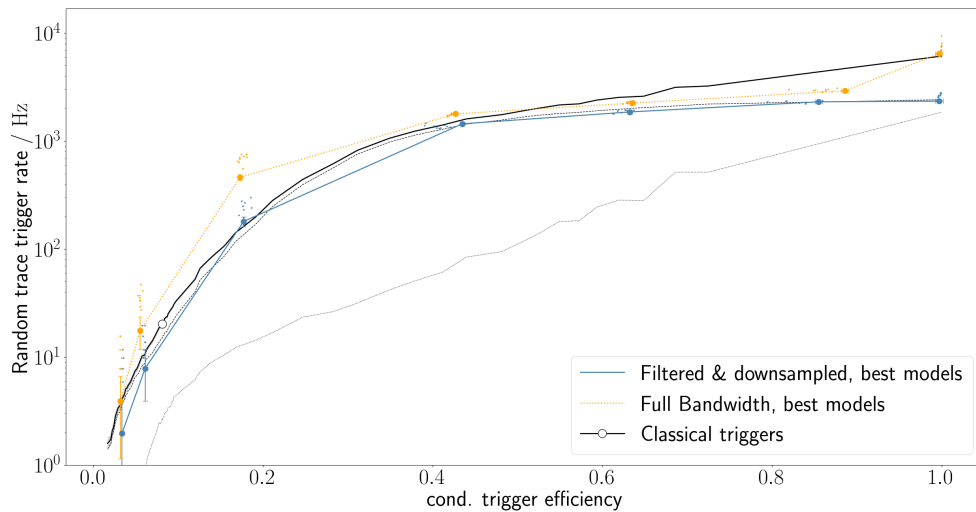
Figure 8.9.: Relation between random-trace trigger rate (lower is better) and conditional trigger efficiency (higher is better). Convolutional neural networks trained on compatibility mode traces compared to networks trained on full bandwidth traces. Compatibility-trained networks consistently have a lower random-trace trigger rate

A larger input size will result in a larger number of trainable parameters. With more trainable parameters, the training loss can be minimized further simply by arguments of fit dimensionality.

In conclusion, for neural networks whose size must be restricted, filtered & downsampled input traces yield better performance compared to full bandwidth input. The shown comparisons are not entirely fair, as the former encompasses a longer span in time, and thus more information, compared to the latter at equal input size. In order to completely clarify whether filtering & downsampling is preferable, the same tests should be run on networks of much, much larger size. This is however out of the relevant scope for this thesis.

Kernel size

The presented models so far show a performance that is similar to the Th-T2 threshold trigger. In an attempt to boost performance, the network architecture is altered. Namely, the kernel size of the 2D convolutional input layer is increased. While this extends the network size, it could be expected that the consideration of a larger window during convolution helps the network extract more meaningful features, and increases predictive power in the process.

Three architectures are examined to check for a trend in prediction efficiency. The original $(3, 3)$-kernel only takes a minimum window into consideration. It has the lowest accuracy out of all models. The two other candidates have a kernel size of $(3, 10)$ and $(3, 30)$ respectively. Both have a very similar performance as can be seen in Figure 8.11.
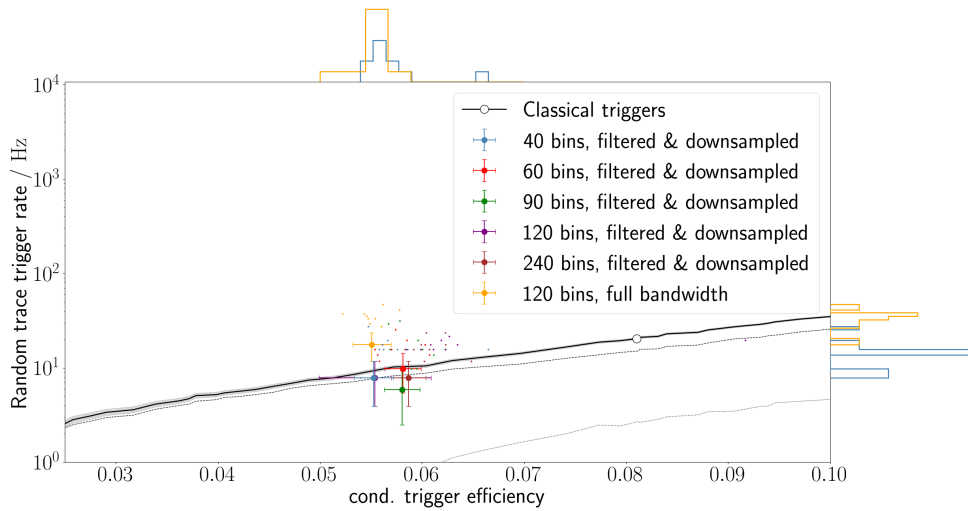
Figure 8.10.: Relation between random-trace trigger rate (lower is better) and conditional trigger efficiency (higher is better). The comparison between a compatibility mode convolutional neural network (blue) that has the same input size in time to a full bandwidth network (orange) reveals the advantage of filtering & downsampling. Both networks have a comparable efficiency, but the compatibility mode network is less sensitive to background triggers from random-traces.

Trigger efficiency is not strongly correlated to kernel size beyond a certain level. An increase in network score is not achieved by increasing it. All architectures have a performance very similar to the threshold trigger.

8.5. LSTM Trigger

8.5.1. Architecture & training procedure

From chapter 5 it follows that LSTMs are especially adept at handling temporally correlated data. The deployment of RNN-based triggers thus seems promising for the problem at hand. The special architecture of LSTMs must however be incorporated in a larger network, as the input of an LSTM is a vector of single dimension, and incompatible with the data structure of a time trace. In the following, the time traces from each PMT will individually be passed through an LSTM layer. The information extracted by the different layers will then be combined via a densely connected layer that propagates features to a binary output representing the signal labels. This philosophy of propagating data is very similar to the ToT-like triggers, which first check trigger conditions per PMT, and then require coincident trigger conditions for at least two PMTs (compare subsection 7.2.1).

While the chosen architectures represent very small networks (< 50 trainable parameters, c.f. Table 8.1), calculating gradients for the special layout is intensive. A training step for LSTM networks takes approximately ten times as long as for CNN networks. For this
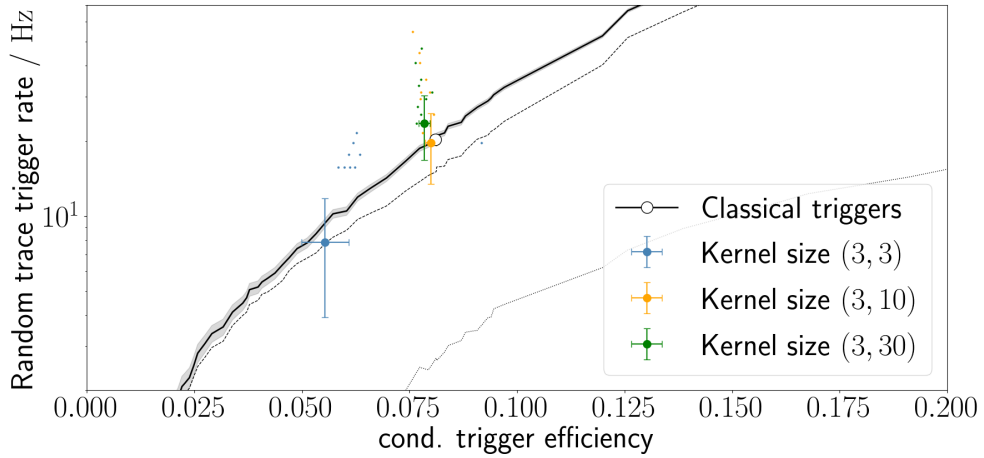
Figure 8.11.: All listed networks are trained with $t_S = 0.5\,\text{VEM}_{\text{Ch.}}$. An increase in kernel size benefits trigger efficiency to some extent. On the other hand, random-trace trigger rate increases. No clear trend in network score can be seen.

purpose, the LSTMs are only trained for a maximum number of 5 epochs. In practice, this restriction was never needed, as the recurrent networks showed convergence after 1 epoch already.

The hyperparameters used in training LSTMs are the ones that show the most promising results in the analysis of CNN trigger performance. Most importantly, this means that the input data of classifiers in this section are filtered & downsampled.

8.5.2. Performance

Single layer LSTM

Since the arrival direction (θ, ϕ) is a priori random, there should be no distinguished PMT in the WCD that takes a precedent role in triggering. This implies that the input time trace should be treated equally whether it is coming from PMT #1, PMT #2, or PMT #3. Logically, a single layer for all three input dimensions should show the same performance as three different layers.

This cannot be verified. An attempt to fit such an architecture to data fails. It is theorized that this is likely due to limitations of the underlaying TensorFlow optimization method. A subsequent analysis in the next paragraph will be conclude that no absolute ordering of layers exist, i.e. any permutation of three distinct layers will on average arrive at the same prediction for signal or background label. With a more elaborate training setup, it is expected that a single-layer LSTM performs as good as a multi-layer LSTM. With a number of just 12 parameters per layer, this will drop the network size considerably.
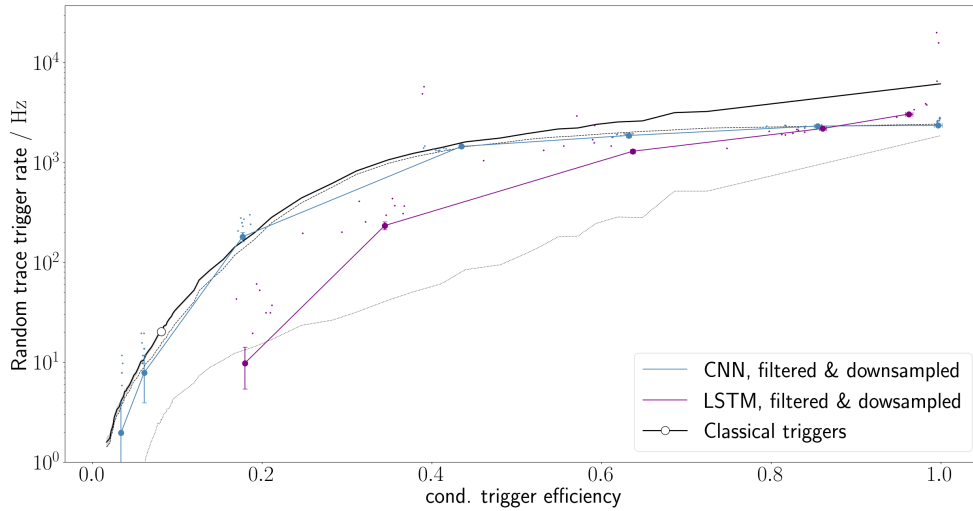
Figure 8.12.: LSTMs are consistently better than CNNs at charge thresholds $t_S >$ 0.05 VEM$_{Ch.}$. The purple line connects the best performing models (as measured by network score a) for the ensembles with t_S (from right to left) equal to 0.02 VEM$_{Ch.}$, 0.05 VEM$_{Ch.}$, 0.1 VEM$_{Ch.}$, 0.2 VEM$_{Ch.}$, and 0.5 VEM$_{Ch.}$. At the highest examined cut $t_S = 0.5$ VEM$_{Ch.}$, LSTMs show a score that is better than the ToT trigger.
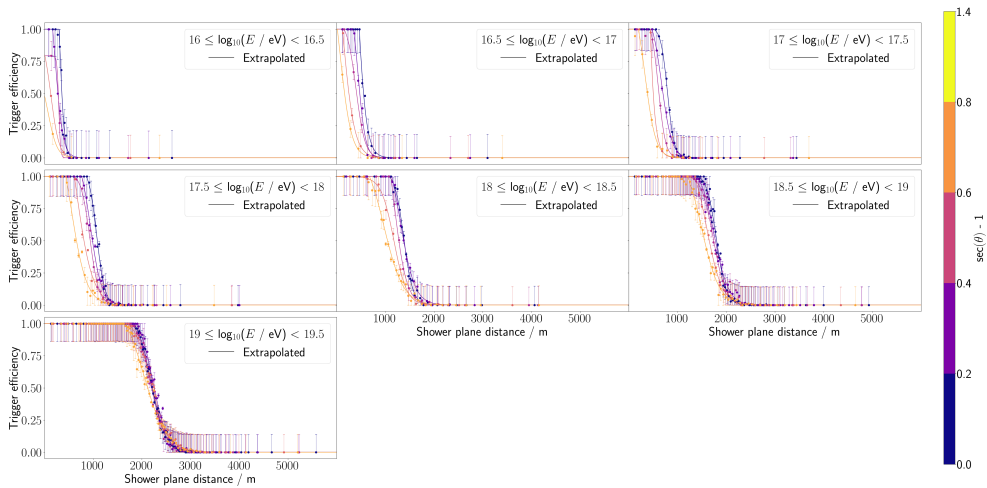
Multi layer LSTM

Using a more involved network with three distinct LSTM layers as well as dense output layer, the performance after training for given charge cuts is presented in Figure 8.12. As can be seen, the performance of LSTMs is - for specific hyperparameters - at least on par with that of the ToT trigger.

The model with the highest ratio of trigger efficiency to random-trace trigger rate increases the sensitivity of station level triggers at full efficiency from ≈ 1 VEM$_{Ch.}$ to 0.5 VEM$_{Ch.}$, as shown in Figure 8.13. This is achieved via a higher lateral trigger probability, especially for inclined showers (compare Figure 8.14). A resulting overall increased T3 efficiency as shown in Figure 8.15 can be observed.

The potential of these multi-layer LSTMs seems promising. For classical triggers, a promise between acceptable background T2 rate and trigger efficiency must be found. The considered networks improve both aspects, and deliver a lower total T2 trigger rate while retaining a higher trigger efficiency. However, the necessity of using three distinct layers over a single LSTM layer is not based on any reasoning. This issue may be indicative of the network architecture deciding labels based on unphysical artifacts in the simulation data, and must be examined. By switching the ordering of the layers from $[1, 2, 3]$ to any possible permutation and recalculating labels for the validation dataset, possible shortcomings of the network should be revealed. The results of this analysis are shown in Figure 8.16, and find no clear trend or preferred ordering of layers. This implies that the LSTM architecture does in fact trigger on the physical information given as input, and not any artifact present in the data.
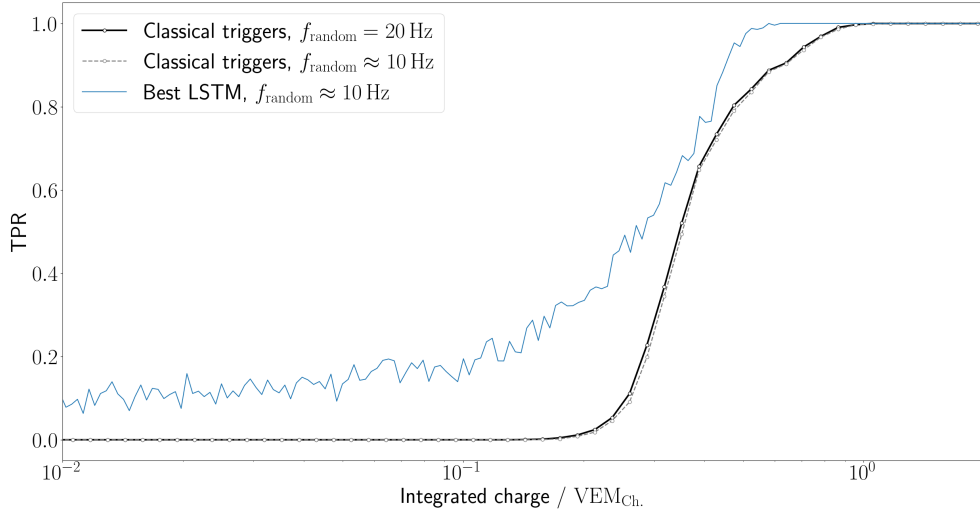
Figure 8.13.: The sensitivity of the best performing LSTM trained with $t_S = 0.5\,\text{VEM}_{\text{Ch.}}$ is superior to that of classical triggers for all deposited signals. This is true both for the currently used classical trigger thresholds, for which $f_{\text{random}} = 20\,\text{Hz}$, as well as for the direct comparison of classical triggers with thresholds that yield the same trigger rate $f_{\text{random}} \approx 10\,\text{Hz}$ as the considered LSTM. Full sensitivity, where all signals raise a T2 is reached at approximately $0.5\,\text{VEM}_{\text{Ch.}}$.

Figure 8.14.: The lateral trigger probability for the best performing LSTM, where the classifier score a as calculated in Equation 7.4 is maximized. This is achieved by using a multi-layer LSTM trained with a charge cut $t_S = 0.5\,\text{VEM}_{\text{Ch.}}$.
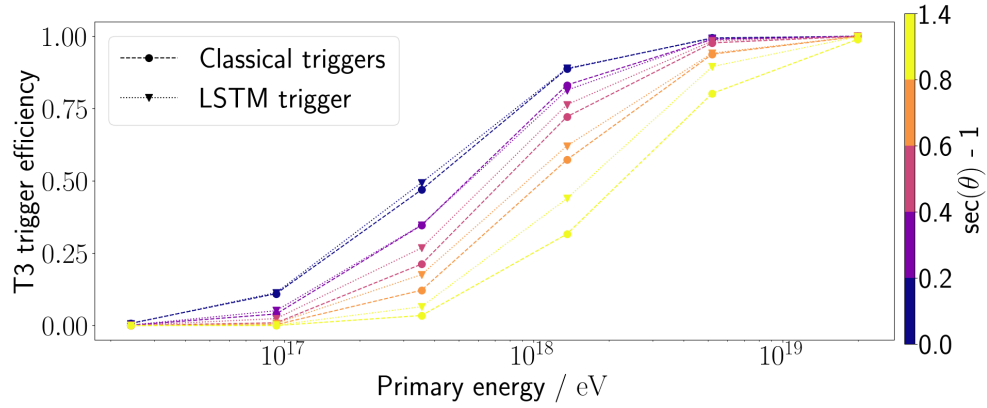
Figure 8.15.: Resulting T3 efficiency of the best performing LSTM with charge cut $t_S = 0.5\,\mathrm{VEM_{Ch.}}$, compared to classical T3 efficiency of the current triggers operating with thresholds that yield the same random-trace trigger frequency. With a higher overall T3 efficiency, the LSTM trigger clearly shows a higher signal to noise ratio.
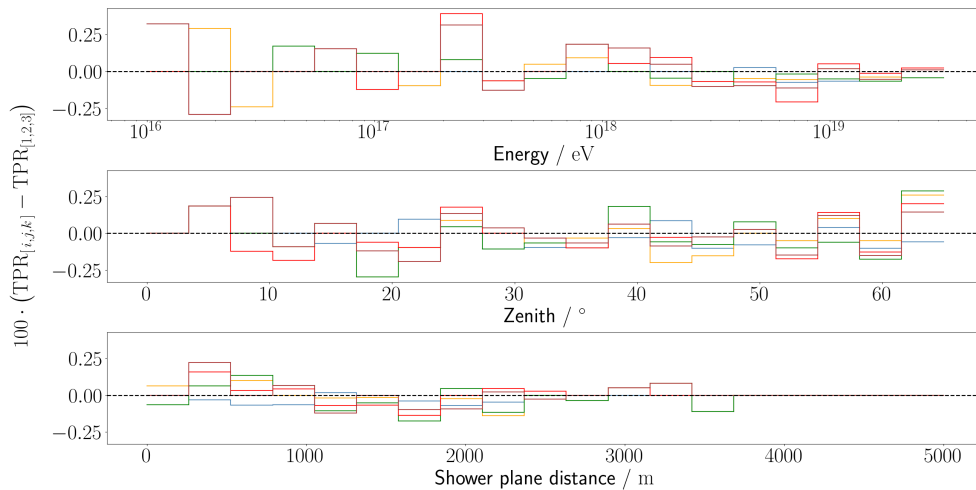
Figure 8.16.: Absolute deviation of true positive rate between original layer ordering ([1, 2, 3]) and the five possible permutated architectures ([1, 3, 2], [3, 1, 2], ...). No clear improvement/regression of performance is noted with respect to any shower variable. Peak deviations are on a sub-percent level.

9. Summary and conclusions

A lot is left to learn about the physics of high-energy cosmic rays. Of special interest in research is the detection of photon or neutrino primaries. This is in part due to different - and possibly unknown - underlying acceleration mechanisms boosting the particles to incredible energies, but also the possibility of source identification for neutral particles. The Pierre Auger Observatory offers a unique hope in answering currently unanswered questions. Due to its enormous exposure, it may be sensitive to very rare events that have not been seen anywhere else in the world.

As of the completion of this thesis, only upper limits for the flux of exotic primaries at ultra-high energies have been identified. With the ongoing AugerPrime detector upgrade, the surface detector of the Pierre Auger Observatory is equipped with new electronics and additional measurement channels. This will push reconstruction and detection boundaries to new levels. In this context, the entire chain of operations from shower detection to event reconstruction is revisited and examined for improvement. This thesis represents one such revision and attempts to revamp the station level triggers of the surface detector, which are required to initialize the process of event readout.

A lot of the presented results are non-exhaustive. The possibility for tweaks to hyperparameters is virtually endless. Especially the performance of neural network triggers on non-proton primaries, and on additional shower simulations using different interaction models remains. Still, several milestones and key insights can be identified from the previous chapters.

Bayesian folding accelerates prototyping studies

The standard way of calculating lateral trigger probabilities and T3 efficiencies is done by running an $\overline{\text{Off}}\underline{\text{line}}$ simulation and examining the appropriate detector response. In this work, it was shown that a reparametrization of the LTP to easily accessible parameters arrives at the same results. This massively speeds up the calculation of critical detector parameters. More importantly, the entire process can be completely detached from $\overline{\text{Off}}\underline{\text{line}}$ once simulation data has been collected. This enables the usage of arbitrary external tools for implementing new triggers (such as TensorFlow) and will streamline analysis especially in prototyping studies, such as this one.

neural-network-based algorithms profit from filtered & downsampled input data

For the considered networks, a trend to lower random-trace trigger rates was observed when training on filtered & downsampled traces, while the trigger efficiency remained the same. This implies that - at least for the considered input sizes - compatibility mode is preferred for neural networks. It is theorized that this is caused due to the filtering part in the conversion algorith, as it smoothes out electronic noise in the time trace., and therefore reduced the complexity a network must interpret.

CNN architecture on par with classical threshold trigger performance

The trigger efficiency of convolutional neural networks coincides with that of the classical threshold Th-T2 trigger within the range of acceptable random-trace trigger rates. This in itself is a promising result. It shows that neural network trigger algorithms can be at least as efficient as the currently employed triggers. It is expected that with a more in-depth analysis of additional hyperparameters the CNN trigger performance can be boosted further.

LSTM architecture outperforms CNN architecture & classical triggers

The LSTM architecture is designed to efficiently recognize patterns in a temporal input. Consequently, the performance boost seen from using such architectures is very promising. With minimal tweaks to hyperparameters, the recurrent networks show a signal to noise ratio that is larger than that of the main SD trigger at shower inclinations below $\theta < 60°$, the time-over-threshold algorithm. Due to very time-intensive training periods, a lot of the hyperparameter optimization done for CNNs remains unexplored for LSTMs. The full potential of LSTM-based triggers is thought to not have been reached yet.

Hints at higher efficiency gains for very inclined showers

The gain in T3 efficiency when using LSTM triggers over classical ones is largest for inclined showers with $\theta \approx 65°$. Since no showers at higher inclinations were simulated, the performance of neural networks on such data remains unknown. Massive improvements to sensitivity for such events would be expected. This has far-reaching implications, as it might enable more accurate detection and reconstruction of highly inclined showers, where a large aperature for neutrino primaries are expected.

Acknowledgements

This thesis would not have been possible without the help of many people.

For this reason, I would like to express my utmost gratitude to Prof. Dr. Ralph Engel and Prof. Dr. Guido Drexlin for inviting me to the institute for astroparticle physics, and for refereeing this work.

Thanks to Dr. Markus Roth who enabled me to travel to Argentina, where I could visit the observatory in person, and experience la vida argentina. More importantly also thanks for his feedback and insights that he shared with me during the last year, and for being an excellent group leader.

Dr. David Schmidt and Dr. Darko Veberič were always available for any question that I had, and played an instrumental role in shaping the analysis with their thoughtful comments, and critical discussions. Dr. Schmidt especially continued to motivate me and helped solve any problems I encountered during my work, even while he was located on a completely different hemisphere with several timezones between us complicating communication.

One person that is criminally underrepresented in these acknowledgements is Sabine Bucher. I cannot thank her enough for all her tireless work and organizational prowess. I am confident in saying that without her skills, the institute for astroparticle physics would not be as productive as it is. Thank you Sabine for all your support in filling out numerous business trip applications, and organizing a large part of my trip to Argentina.

Finally, I want to express a heartfelt appreciation to all my office colleagues, Auger group members and IAP associates. Your questions, recommendations and general interest helped me realize the relevance of this work. All the moments we shared outside of working hours sensitized me to the importance of collaboration

Bibliography

[1] David Glaude. *L'incident de Schaerbeek*. `http://wiki.ael.be/index.php/VoteElectroniqueInversionSpontan%E9eEtAl%E9atoireDunePositionBinaire`. Accessed: 30th Apr. 2023. Nov. 2005.

[2] Adrian L Melott et al. "Did a gamma-ray burst initiate the late Ordovician mass extinction?" In: *International Journal of Astrobiology* 3.1 (2004), pp. 55–61.

[3] Brian D Fields et al. "Supernova triggers for end-Devonian extinctions". In: *Proceedings of the National Academy of Sciences* 117.35 (2020), pp. 21008–21010.

[4] Robert A Millikan and G Harvey Cameron. "The origin of the cosmic rays". In: *Physical review* 32.4 (1928), p. 533.

[5] Thomas H Johnson. "A note on the nature of the primary cosmic radiation". In: *Physical Review* 54.5 (1938), p. 385.

[6] Nobel Media AB. *Press release*. `https://www.nobelprize.org/prizes/physics/1936/summary/`. Accessed: 18th Jul. 2022. 1936.

[7] Pierre Auger et al. "Extensive cosmic-ray showers". In: *Reviews of modern physics* 11.3-4 (1939), p. 288.

[8] Sergey Ostapchenko. "Status of QGSJET". In: *AIP Conference Proceedings*. Vol. 928. 1. American Institute of Physics. 2007, pp. 118–125.

[9] I Goldman et al. "Implications of the supernova SN1987A neutrino signals". In: *Physical review letters* 60.18 (1988), p. 1789.

[10] J Abraham et al. "Measurement of the energy spectrum of cosmic rays above 1018 eV using the Pierre Auger Observatory". In: *Physics Letters B* 685.4-5 (2010), pp. 239–246.

[11] Alexander Aab et al. "Searches for anisotropies in the arrival directions of the highest energy cosmic rays detected by the Pierre Auger Observatory". In: *The Astrophysical Journal* 804.1 (2015), p. 15.

[12] Hannes Alfvén. "Existence of electromagnetic-hydrodynamic waves". In: *Nature* 150.3805 (1942), pp. 405–406.

[13] Enrico Fermi. "On the origin of the cosmic radiation". In: *Physical review* 75.8 (1949), p. 1169.

[14] AM Hillas. "Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays?" In: *Journal of Physics G: Nuclear and Particle Physics* 31.5 (2005), R95.

[15] Pasquale Blasi. "The origin of galactic cosmic rays". In: *The Astronomy and Astrophysics Review* 21.1 (2013), pp. 1–73.

[16] Katsuaki Asano and Masaaki Hayashida. "The most intensive gamma-ray flare of quasar 3C 279 with the second-order Fermi acceleration". In: *The Astrophysical Journal Letters* 808.1 (2015), p. L18.

[17] Ruth A Daly. "Black hole spin and accretion disk magnetic field strength estimates for more than 750 active galactic nuclei and multiple galactic black holes". In: *The Astrophysical Journal* 886.1 (2019), p. 37.

[18] Elliott Flowers and Malvin A Ruderman. "Evolution of pulsar magnetic fields". In: *The Astrophysical Journal* 215 (1977), pp. 302–310.

[19] FM Rieger and K Mannheim. "Particle acceleration by rotating magnetospheres in active galactic nuclei". In: *arXiv preprint astro-ph/9911082* (1999).

[20] Zaza Osmanov, Andria Rogava, and Gianluigi Bodo. "On the efficiency of particle acceleration by rotating magnetospheres in AGN". In: *Astronomy & Astrophysics* 470.2 (2007), pp. 395–400.

[21] Frank M Rieger. "Cosmic ray acceleration in active galactic nuclei-on centaurus a as a possible uhecr source". In: *arXiv preprint arXiv:0911.4004* (2009).

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

[23] AU Abeysekara et al. "Very-high-energy particle acceleration powered by the jets of the microquasar SS 433". In: *Nature* 562.7725 (2018), pp. 82–85.

[24] Aya Ishihara, IceCube Collaboration, et al. "Extremely high energy neutrinos in six years of IceCube data". In: *Journal of Physics: Conference Series*. Vol. 718. 6. IOP Publishing. 2016, p. 062027.

[25] Frank C Jones. "Inverse Compton scattering of cosmic-ray electrons". In: *Physical Review* 137.5B (1965), B1306.

[26] Enrico Fermi. "98. Galactic Magnetic Fields and the Origin of Cosmic Radiation". In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, 2013, pp. 671–676.

[27] Matthias Bartelmann. "Gravitational lensing". In: *Classical and Quantum Gravity* 27.23 (2010), p. 233001.

[28] Matthias Bartelmann and Peter Schneider. "Weak gravitational lensing". In: *Physics Reports* 340.4-5 (2001), pp. 291–472.

[29] Marijke Haverkorn. "Magnetic fields in the Milky Way". In: *Magnetic fields in diffuse media*. Springer, 2015, pp. 483–506.

[30] Pierre Auger Collaboration et al. "Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8× 1018 eV". In: *Science* 357.6357 (2017), pp. 1266–1270.

[31] J Skilling. "The diffusion of cosmic rays". In: *Monthly Notices of the Royal Astronomical Society* 147.1 (1970), pp. 1–12.

[32] Leonard Searle and Robert Zinn. "Compositions of halo clusters and the formation of the galactic halo". In: *The Astrophysical Journal* 225 (1978), pp. 357–379.

[33] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: `10.1093/ptep/ptaa104`. eprint: `https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf`. URL: `https://doi.org/10.1093/ptep/ptaa104`.

[34] Kenneth Greisen. "End to the cosmic-ray spectrum?" In: *Physical Review Letters* 16.17 (1966), p. 748.

[35] GT Zatsepin and VA Kuz'min. "J. of Exp. and Theor". In: *Phys. Lett* 4 (1966), p. 78.

[36] Vadim N Gamezo, Alexei M Khokhlov, and Elaine S Oran. "Three-dimensional delayed-detonation model of type Ia supernovae". In: *The Astrophysical Journal* 623.1 (2005), p. 337.

[37] John J Cowan and Friedrich-Karl Thielemann. "R-process nucleosynthesis in supernovae". In: *Physics Today* 57.10 (2004), pp. 47–54.

[38] M Garcia-Munoz, GM Mason, and JA Simpson. "The age of the galactic cosmic rays derived from the abundance of Be-10". In: *The Astrophysical Journal* 217 (1977), pp. 859–877.

[39] JZ Wang et al. "Measurement of cosmic-ray hydrogen and helium and their isotopic composition with the BESS experiment". In: *The Astrophysical Journal* 564.1 (2002), p. 244.

[40] Michael W Kirson. "Mutual influence of terms in a semi-empirical mass formula". In: *Nuclear Physics A* 798.1-2 (2008), pp. 29–60.

[41] Thomas K Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic rays and particle physics*. Cambridge University Press, 2016.

[42] Hans Dembinski et al. "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV". In: *arXiv preprint arXiv:1711.11432* (2017).

[43] Anatoly Klypin et al. "Structure formation with cold+ hot dark matter". In: *arXiv preprint astro-ph/9305011* (1993).

[44] Fiorenza Donato et al. "Constraints on WIMP Dark Matter from the High Energy PAMELA p̄/p data". In: *Physical review letters* 102.7 (2009), p. 071301.

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

[46] Roberto Aloisio et al. "A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays". In: *Astroparticle Physics* 27.1 (2007), pp. 76–91.

[47] Denis Allard. "Extragalactic propagation of ultrahigh energy cosmic-rays". In: *Astroparticle Physics* 39 (2012), pp. 33–43.

[48] Sow-Hsin Chen and Michael Kotlarchyk. *Interactions of photons and neutrons with matter*. World Scientific, 2007.

[49] Stefano Meroli. "The Straggling function. Energy Loss Distribution of charged particles in silicon layers". In: *Home Cern, https://meroli. web. cern. ch/lecture_-StragglingFunction. html* (2017).

[50] Walter Heitler. *The quantum theory of radiation*. Courier Corporation, 1984.

[51] Mukund Gupta. *Calculation of radiation length in materials*. Tech. rep. 2010.

[52] Koichi Kamata and Jun Nishimura. "The lateral and the angular structure functions of electron showers". In: *Progress of Theoretical Physics Supplement* 6 (1958), pp. 93–155.

[53] Kenneth Greisen. "Cosmic ray showers". In: *Annual review of nuclear science* 10.1 (1960), pp. 63–108.

[54] Gert Moliere. "Theorie der streuung schneller geladener teilchen i. einzelstreuung am abgeschirmten coulomb-feld". In: *Zeitschrift für Naturforschung A* 2.3 (1947), pp. 133–145.

[55] James Matthews. "A Heitler model of extensive air showers". In: *Astroparticle Physics* 22.5-6 (2005), pp. 387–397.

[56] Tomás Capistrán, I Torres, and L Altamirano. "New method for Gamma/Hadron separation in HAWC using neural networks". In: *arXiv preprint arXiv:1508.04370* (2015).

[57] Fabian Schmidt. *Sample Corsika showers*. `https://www.zeuthen.desy.de/~jknapp/fs/proton-showers.html`. Accessed: 08th Dec. 2022.

[58] Martin Pittermann and Paul Filip. *Simple hadronic shower simulation in C++*. `https://github.com/martin2250/showersim`. Accessed: 13th Dec. 2022.

[59] Sea Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[60] Dieter Heck et al. "CORSIKA: A Monte Carlo code to simulate extensive air showers". In: *Report fzka* 6019.11 (1998).

[61] S Ostapchenko. "QGSJET-II: towards reliable description of very high energy hadronic interactions". In: *Nuclear Physics B-Proceedings Supplements* 151.1 (2006), pp. 143–146.

[62] DJ Fox et al. "Early tests of scale invariance in high-energy muon scattering". In: *Physical Review Letters* 33.25 (1974), p. 1504.

[63] Larry C Andrews. "An analytical model for the refractive index power spectrum and its application to optical scintillations in the atmosphere". In: *Journal of Modern Optics* 39.9 (1992), pp. 1849–1853.

[64] The VERITAS Collaboration. *VERITAS Telescope Homepage*. `https://veritas.sao.arizona.edu/`. Accessed: 26th Dec. 2022.

[65] The H.E.S.S. Collaboration. *H.E.S.S. Telescope Homepage*. `https://www.mpi-hd.mpg.de/hfm/HESS/`. Accessed: 26th Dec. 2022.

[66] The Pierre Auger Collaboration. *Pierre Auger Observatory Homepage*. `https://www.auger.org/`. Accessed: 26th Dec. 2022.

[67] The HAWC Collaboration. *HAWC Observatory Homepage.* `https://www.hawc-observatory.org/`. Accessed: 26th Dec. 2022.

[68] The Kamiokande Collaboration. *Super-Kamiokande Homepage.* `https://www-sk.icrr.u-tokyo.ac.jp/en/sk/`. Accessed: 26th Dec. 2022.

[69] 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.

[70] Louis Elterman. *UV, visible, and IR attenuation for altitudes to 50 km, 1968.* Tech. rep. AIR FORCE CAMBRIDGE RESEARCH LABS HANSCOM AFB MA, 1968.

[71] Ghouti Abdellaoui et al. "EUSO-TA–First results from a ground-based EUSO telescope". In: *Astroparticle Physics* 102 (2018), pp. 98–111.

[72] M Nagano et al. "New measurement on photon yields from air and the application to the energy estimation of primary cosmic rays". In: *Astroparticle Physics* 22.3-4 (2004), pp. 235–248.

[73] Ina Holl, Eckart Lorenz, and Gikas Mageras. "A measurement of the light yield of common inorganic scintillators". In: *IEEE Transactions on Nuclear Science* 35.1 (1988), pp. 105–109.

[74] The KASCADE Collaboration. *KASCADE Homepage.* `https://www.iap.kit.edu/kascade/index.php`. Accessed: 26th Dec. 2022.

[75] Alexander Aab et al. "Observation of inclined EeV air showers with the radio detector of the Pierre Auger Observatory". In: *Journal of Cosmology and Astroparticle Physics* 2018.10 (2018), p. 026.

[76] The Pierre Auger Collaboration. *AERA - Auger Engineering Radio Array.* `https://auger.org/observatory/aera`. Accessed: 18th Jan. 2023.

[77] Jaime Álvarez-Muñiz et al. "The giant radio array for neutrino detection (GRAND): science and design". In: *Science China Physics, Mechanics & Astronomy* 63.1 (2020), pp. 1–43.

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

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

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

[81] 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.

[82] 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.

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

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

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

[86] 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.

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

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

[89] The Pierre Auger Collaboration. *Official Auger Reconstruction and Simulation Framework*. `https://gitlab.iap.kit.edu/auger-observatory/offline`. HEAD = 5f8b537e (Tue Dec 13 09:58:43).

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

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

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

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

[94] 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.

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

[96] 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.

[97] Markus Roth et al. "Constant Intensity Cut: Unbinned Estimation of the Signal Attenuation Function". GAP 2015-065.

[98] 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.

[99] 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`.

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

[101] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[102] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[103] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[104] Sepp Hochreiter. "Untersuchungen zu dynamischen neuronalen Netzen". In: *Diploma, Technische Universität München* 91.1 (1991).

[105] Cristopher Olah. *Understanding LSTM Networks*. `https://colah.github.io/posts/2015-08-Understanding-LSTMs/`. Accessed: 11th Mar. 2023. Aug. 2015.

[106] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM". In: *Neural computation* 12.10 (2000), pp. 2451–2471.

[107] Antonia Creswell et al. "Generative adversarial networks: An overview". In: *IEEE signal processing magazine* 35.1 (2018), pp. 53–65.

[108] Rahul Dey and Fathi M Salem. "Gate-variants of gated recurrent unit (GRU) neural networks". In: *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. IEEE. 2017, pp. 1597–1600.

[109] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

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

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

[112] M Boezio et al. "Measurement of the flux of atmospheric muons with the CAPRICE94 apparatus". In: *Physical Review D* 62.3 (2000), p. 032007.

[113] David Schmidt. *Background Simulation library*. personal communication. June 2022.

[114] Denis V Martynov et al. "Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy". In: *Physical Review D* 93.11 (2016), p. 112004.

[115] David Nitz. *Random trace characteristics*. personal communication. Sept. 2022.

[116] BL Burrows and DJ Colwell. "The Fourier transform of the unit step function". In: *International Journal of Mathematical Education in Science and Technology* 21.4 (1990), pp. 629–635.

[117] 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.

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

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

[120] David Nitz. "Trigger comissioning". Auger Collaboration Meeting. Mar. 2023.

[121] Quentin Luce. "Comparing UB and UUB data: station-level differences in simulated events". Auger Collaboration Meeting. Mar. 2023.

[122] Pedro Abreu et al. "The lateral trigger probability function for the ultra-high energy cosmic ray showers detected by the Pierre Auger Observatory". In: *Astroparticle physics* 35.5 (2011), pp. 266–276.

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

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

[125] Laura Molina Bueno, Pierre Billoir, and Ioana Maris. "Signal variance for the TOTd and MoPS triggers". GAP 2014-035.

[126] Mariangela Settimo et al. "Trigger probability for single stations and air-showers with the TOTd and MoPS algorithms". GAP 2013-114.

[127] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/`.

[128] *TensorFlow v2.8.4*. `https://github.com/tensorflow/tensorflow/releases`. Accessed: 24th Apr. 2023. Nov. 2022.

[129] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[130] Rachel Zhou. *ARTIX-7FPGADevelopment Board AX7102 User Manual*. English. Version Version 1.2. ALINx. 52 pp.

[131] Vladimir Nasteski. "An overview of the supervised machine learning methods". In: *Horizons. b* 4 (2017), pp. 51–62.

A. Filter and Downsample Algorithm

```python
1  import numpy as np
2
3  def apply_downsampling(pmt : np.ndarray, random_phase : int) -> np.ndarray :
4      '''Receive UUB-like ADC trace and filter/downsample to emulate UB electronics'''
5      n_bins_uub    = (len(pmt) // 3) * 3             # original trace length
6      n_bins_ub     = n_bins_uub // 3                 # downsampled trace length
7      sampled_trace = np.zeros(n_bins_ub)             # downsampled trace container
8
9      # ensure downsampling works as intended
10     # cuts away (at most) the last two bins
11     if len(pmt) % 3 != 0: pmt = pmt[0 : -(len(pmt) % 3)]
12
13     # see Framework/SDetector/UUBDownsampleFilter.h in Offline main branch for more information
14     kFirCoefficients = [ 5, 0, 12, 22, 0, -61, -96, 0, 256, 551, 681, 551, 256, 0, -96, -61, 0, 22, 12, 0, 5 ]
15     buffer_length = int(0.5 * len(kFirCoefficients))
16     kFirNormalizationBitShift = 11
17     kADCsaturation = 4095                           # maximum FADC value: 2^12 - 1
18
19     temp = np.zeros(n_bins_uub + len(kFirCoefficients))
20
21     temp[0 : buffer_length] = pmt[:: -1][-buffer_length - 1 : -1]
22     temp[-buffer_length - 1: -1] = pmt[:: -1][0 : buffer_length]
23     temp[buffer_length : -buffer_length - 1] = pmt
24
25     # perform downsampling
26     for j, coeff in enumerate(kFirCoefficients):
27         sampled_trace += [temp[k + j] * coeff for k in range(random_phase, n_bins_uub, 3)]
28
29     # clipping and bitshifting
30     sampled_trace = [int(adc) >> kFirNormalizationBitShift for adc in sampled_trace]
31
32     # Simulate saturation of PMTs at 4095 ADC counts ~ 19 VEM LG
33     return np.clip(np.array(sampled_trace), a_min = 0, a_max = kADCsaturation)
```

Listing A.1: Python implementation as used in section 7.1 and chapter 8

```cpp
1  #ifndef _sdet_UUBDownsampleFilter_h_
2  #define _sdet_UUBDownsampleFilter_h_
3
4  #include <utl/Trace.h>
5  #include <utl/TimeDistribution.h>
6  #include <utl/AugerUnits.h>
7  #include <utl/Math.h>
8
9
10 namespace sdet {
11
12   /**
13     Functions to perform the downsample algorithm used for legacy triggers in the UUB firmware
14
15     Besed on descriptions and reference code from Dave Nitz
16
17     The filter will only slightly change the scale of the trace values by factor 2048/2059,
18     since the normalization in the firmware uses a bit-shift instead of a division with the
19     exact norm. Note that the filter does not produce an UB equivalent trace, this would
20     require the division of the trace with an additional factor of 4 which is ommited to
21     the preserve dynamic range.
22
23     \author Darko Veberic
24     \date 13 Nov 2020
25   */
26
27   namespace {
28
29     // FIR coefficients from Dave Nitz's implementation in UUB firmware
30     constexpr int kFirCoefficients[] = { 5, 0, 12, 22, 0, -61, -96, 0, 256, 551, 681, 551, 256, 0, -96, -61, 0,
31       22, 12, 0, 5 };
32     // FIR normalization
33     // in firmware an 11-bit right shift is used instead, ie 2048
34     constexpr int kFirNormalizationBitShift = 11;
35     //const int kFirNormalization = 2059;  // true norm of FIR
36     //constexpr double kFirNormalization = (1 << kFirNormalizationBitShift);  // actually used
37     constexpr double kUbSampling = 25*utl::nanosecond;
```

```
37    constexpr int kADCSaturation = 4095;
38
39
40    // enforce ADC saturation, both ways
41    inline
42    constexpr
43    int
44    Clip(const int i)
45    {
46      return std::max(0, std::min(i, kADCSaturation));
47    }
48
49  }
50
51  // phase can be one only of { 0, 1, 2 }
52  inline
53  utl::TraceI
54  UUBDownsampleFilter(const utl::TraceI& trace, const int phase = 1)
55  {
56    // input trace is assumed to have 8.333ns binning
57    const int n = trace.GetSize();
58    if (!n)
59      return utl::TraceI(0, kUbSampling);
60    const int m = utl::Length(kFirCoefficients);
61    const int m2 = m / 2;
62    std::vector<int> t;
63    t.reserve(n + 2*m2);
64    // pad front with the first trace values, but backwards
65    for (int i = m2; i; --i)
66      t.push_back(trace[i]);
67    // copy the whole trace
68    for (int i = 0; i < n; ++i)
69      t.push_back(trace[i]);
70    // pad back with the last trace values, but backwards
71    for (int i = 1; i <= m2; ++i)
72      t.push_back(trace[n-1-i]);
73    const int n3 = n / 3;
74    utl::TraceI res(n3, kUbSampling);
75    for (int k = 0; k < n3; ++k) {
76      auto& v = res[k];
77      const int i = 3*k + phase;
78      for (int j = 0; j < m; ++j)
79        v += t[i + j] * kFirCoefficients[j];
80      v >>= kFirNormalizationBitShift;
81      v = Clip(v);
82    }
83    return res;
84  }
85 }
86
87
88 #endif
```

Listing A.2: Implementation in C++, as used in Offline. From [89]

B. Lateral particle probability fit parameters

$$0° \leq \theta < 33°$$

	$\epsilon^* \pm \Delta\epsilon^*$	$R_0 \pm \Delta R_0$	$\sigma_R \pm \Delta\sigma_R$
$16.0 \leq \log_{10}(E\,/\,eV) < 16.5$	2.00 ± 3.46	407 ± 743	230.57 ± 97.08
$16.5 \leq \log_{10}(E\,/\,eV) < 17.0$	2.00 ± 0.21	534 ± 46	250.02 ± 10.03
$17.0 \leq \log_{10}(E\,/\,eV) < 17.5$	2.00 ± 0.26	723 ± 71	325.04 ± 13.23
$17.5 \leq \log_{10}(E\,/\,eV) < 18.0$	2.00 ± 0.17	1003 ± 57	385.80 ± 12.29
$18.0 \leq \log_{10}(E\,/\,eV) < 18.5$	1.45 ± 0.13	1580 ± 68	393.54 ± 23.21
$18.5 \leq \log_{10}(E\,/\,eV) < 19.0$	1.65 ± 0.18	1844 ± 96	492.20 ± 24.57
$19.0 \leq \log_{10}(E\,/\,eV) < 19.5$	1.42 ± 0.08	2470 ± 50	470.81 ± 16.22

$$33° \leq \theta < 44°$$

	$\epsilon^* \pm \Delta\epsilon^*$	$R_0 \pm \Delta R_0$	$\sigma_R \pm \Delta\sigma_R$
$16.0 \leq \log_{10}(E\,/\,eV) < 16.5$	2.00 ± 1.52	321 ± 300	217.08 ± 46.44
$16.5 \leq \log_{10}(E\,/\,eV) < 17.0$	2.00 ± 0.30	488 ± 61	245.26 ± 11.35
$17.0 \leq \log_{10}(E\,/\,eV) < 17.5$	2.00 ± 0.24	648 ± 70	343.20 ± 13.82
$17.5 \leq \log_{10}(E\,/\,eV) < 18.0$	2.00 ± 0.19	930 ± 66	412.94 ± 13.59
$18.0 \leq \log_{10}(E\,/\,eV) < 18.5$	1.65 ± 0.19	1421 ± 101	464.60 ± 28.20
$18.5 \leq \log_{10}(E\,/\,eV) < 19.0$	1.36 ± 0.14	2043 ± 89	443.85 ± 28.59
$19.0 \leq \log_{10}(E\,/\,eV) < 19.5$	1.38 ± 0.07	2497 ± 52	508.05 ± 17.83

$$44° \leq \theta < 51°$$

	$\epsilon^* \pm \Delta\epsilon^*$	$R_0 \pm \Delta R_0$	$\sigma_R \pm \Delta\sigma_R$
$16.0 \leq \log_{10}(E\,/\,eV) < 16.5$	1.39 ± 0.39	405 ± 130	198.91 ± 37.41
$16.5 \leq \log_{10}(E\,/\,eV) < 17.0$	2.00 ± 0.31	402 ± 64	245.06 ± 12.93
$17.0 \leq \log_{10}(E\,/\,eV) < 17.5$	2.00 ± 0.28	537 ± 84	356.95 ± 16.55
$17.5 \leq \log_{10}(E\,/\,eV) < 18.0$	2.00 ± 0.22	843 ± 79	420.26 ± 16.51
$18.0 \leq \log_{10}(E\,/\,eV) < 18.5$	2.00 ± 0.42	1122 ± 199	550.53 ± 42.14
$18.5 \leq \log_{10}(E\,/\,eV) < 19.0$	1.34 ± 0.14	1983 ± 94	487.21 ± 32.37
$19.0 \leq \log_{10}(E\,/\,eV) < 19.5$	1.34 ± 0.08	2484 ± 57	523.70 ± 20.92

$51° \leq \theta < 56°$			
	$\epsilon^* \pm \Delta\epsilon^*$	$R_0 \pm \Delta R_0$	$\sigma_R \pm \Delta\sigma_R$
$16.0 \leq \log_{10}(E / eV) < 16.5$	2.00 ± 1.44	130 ± 239	181.54 ± 58.89
$16.5 \leq \log_{10}(E / eV) < 17.0$	2.00 ± 0.38	268 ± 88	276.20 ± 17.06
$17.0 \leq \log_{10}(E / eV) < 17.5$	2.00 ± 0.36	468 ± 106	354.45 ± 19.80
$17.5 \leq \log_{10}(E / eV) < 18.0$	2.00 ± 0.26	731 ± 98	447.75 ± 20.08
$18.0 \leq \log_{10}(E / eV) < 18.5$	1.71 ± 0.31	1203 ± 154	468.22 ± 40.67
$18.5 \leq \log_{10}(E / eV) < 19.0$	1.55 ± 0.22	1748 ± 144	548.91 ± 40.98
$19.0 \leq \log_{10}(E / eV) < 19.5$	1.34 ± 0.09	2423 ± 81	596.88 ± 28.54

$56° \leq \theta < 65°$			
	$\epsilon^* \pm \Delta\epsilon^*$	$R_0 \pm \Delta R_0$	$\sigma_R \pm \Delta\sigma_R$
$16.0 \leq \log_{10}(E / eV) < 16.5$	2.00 ± 1.03	27 ± 176	192.66 ± 41.51
$16.5 \leq \log_{10}(E / eV) < 17.0$	2.00 ± 0.36	121 ± 80	268.29 ± 14.55
$17.0 \leq \log_{10}(E / eV) < 17.5$	2.00 ± 0.25	301 ± 78	370.71 ± 15.16
$17.5 \leq \log_{10}(E / eV) < 18.0$	2.00 ± 0.21	595 ± 81	454.64 ± 15.59
$18.0 \leq \log_{10}(E / eV) < 18.5$	1.25 ± 0.07	1364 ± 49	454.27 ± 20.02
$18.5 \leq \log_{10}(E / eV) < 19.0$	1.40 ± 0.12	1788 ± 89	519.93 ± 30.00
$19.0 \leq \log_{10}(E / eV) < 19.5$	1.56 ± 0.12	2163 ± 91	628.26 ± 26.78