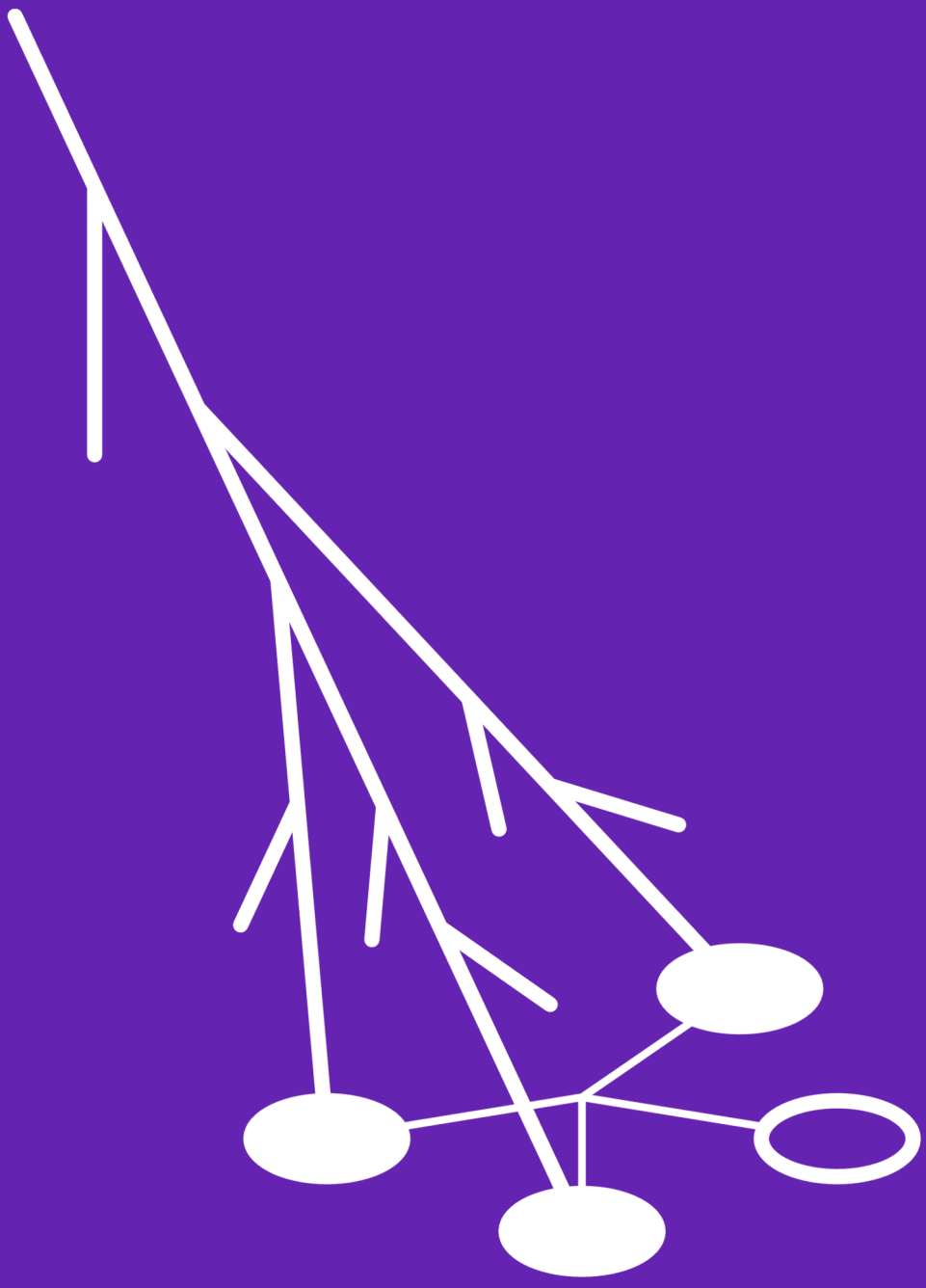
The HiSPARC Experiment

data acquisition and reconstruction of shower direction



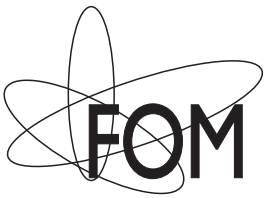
David B. R. A. Fokkema

THE HISPARC COSMIC RAY EXPERIMENT

DATA ACQUISITION AND RECONSTRUCTION OF SHOWER DIRECTION

Dutch title: *Het kosmische straling experiment HiSPARC: data-acquisitie en de reconstructie van de richting van air showers.*

ISBN: 978-90-365-3438-3
DOI: 10.3990/1.9789036534383



Netherlands Organisation for Scientific Research

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). It was carried out at the National Institute for Subatomic Physics (Nikhef).

THE HISPARC COSMIC RAY EXPERIMENT

DATA ACQUISITION AND RECONSTRUCTION OF SHOWER DIRECTION

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op woensdag 17 oktober 2012 om 14.45 uur

door

David Boudewijn Reinder Alexander Fokkema
geboren op 14 december 1978
te Alphen aan den Rijn

Dit proefschrift is goedgekeurd door:

Promotor:	prof. dr. ing. B. van Eijk
Assistent-promotor:	dr. J. Steijger

Voor Eveline

Voor je ongekende geduld met mij, en al je steun en toewijding

Voor Esther en Hannah

Niets fijners dan een kus en een knuffel van mijn meiden!

Voor mijn moeder

Die vanaf dag één ononderbroken in mij heeft geloofd

Contents

1 Cosmic Rays

1.1 A Short History

The discovery of cosmic rays has a long history. The electroscope, a device used to measure the amount of electrical charge in a body, has been crucial to understanding the origin of this radiation. As early as 1785, de Coulomb [1] had found that an electroscope would discharge spontaneously in the air due to some unknown action. He ruled out the effect of imperfect insulation. This has been further studied by e.g. Faraday [2] around 1835, and Crookes [3] in 1879.

A critical development was the discovery of X-rays in 1895 by Röntgen [4]. Röntgen performed experiments with different kinds of cathode ray tubes. At one point, he noted a faint glow from a nearby fluorescent screen which he had prepared for one of his experiments. Upon further investigation, he theorized the existence of invisible radiation emanating from the tube, which he called X-rays. Röntgen also discovered the ionizing properties of X-rays, noting that air conducts electricity when traversed by the radiation [5].

The component of X-ray tubes[1] responsible for the emission of X-rays is also a source of fluorescence. Becquerel extensively studied the possible connection between the emission of visible light and X-rays. For this, he used a uranium salt, known for its strong phosphorescence. This resulted in the discovery of spontaneous radioactivity

[1] Early designs for cathode ray tubes were later optimized to efficiently produce X-rays.

in 1896 [6]. Becquerel found that, like X-rays, the new radiation was also capable of ionizing dry air and discharging an electroscope.

Over the following years, many experiments were performed to study the emission of ionizing radiation and its properties. It remained a curious phenomenon that a completely isolated electroscope would discharge slowly, even when no X-ray tubes or radioactive materials were used in the currently-running experiment. First observed by de Coulomb, it was now suspected that the environment itself contained low levels of ionizing radiation.

In 1903 Rutherford & Cooke [7], and independently McLennan & Burton [8], enclosed their electroscopes in shields of metal which they kept free from radioactive elements. They found that the electroscopes discharged more slowly. Their conclusion was that the ionizing radiation must come from outside the vessel and that it was not spontaneously generated inside.

In 1909 Kurz [9] wrote a review on the origin of the ionizing radiation. He concluded that there were three options: a) the radiation was extra-terrestrial, b) it was an effect of radioactivity in the crust of the Earth, or c) it was an effect of radioactivity in the atmosphere. He concluded that the most likely option was the second one: radioactive elements known to be present in the crust emit the ionizing radiation measured by discharging electroscopes. From this assumption, equations were derived describing the amount of ionizing radiation as a function of height above the surface of the Earth.

Theodor Wulf, a German scientist and Jesuit priest teaching physics in the Netherlands, designed an improved electroscope. He visited Paris in 1909 and measured the intensity of the radiation at the bottom and at the top of the Eiffel Tower (300 m). He found that the intensity of the radiation decreased, but not nearly enough to confirm the hypothesis of radiation coming from the Earth's crust [10].

Pacini questioned that same hypothesis. From 1907 to 1912 he measured the intensity of radiation on both land and sea. If the radiation were coming from the Earth's crust, a body of water above it should reduce the intensity by absorption. Measurements at the surface of the sea did not show a reduction in the intensity of the radiation. Pacini then submerged an electroscope to a depth of 3 m (the total depth of the Gulf of Genoa at that location is 8 m) [11]. He measured a 20 % decrease in discharge rate, consistent with the hypotheses that the radiation was coming from above and was penetrating the water. Pacini then derived the absorption rate for water from his measurements.

In 1909 Gockel [12] ascended in a balloon to 4500 m and found no decrease in the intensity of the radiation, at variance with the then-accepted hypothesis that the

radiation was coming from the surface of the Earth. In doing so, he confirmed Pacini's results. However, many physicists were reluctant to give up the established view.

Hess believed that resolving the dispute was of the highest importance. First, he carefully measured the absorption coefficient of gamma rays (believed to be responsible for the penetrating radiation) through air [13]. His results were in agreement with earlier measurements performed by Eve [14]. Now, having obtained an accurate absorption rate, he undertook a series of balloon flights in 1911 and 1912, up to an altitude of 5200 m [15]. His final results were that there was a small decrease in the intensity of the radiation in the first few hundred meters above the surface of the Earth. Higher altitudes showed an increase in intensity. At 5200 m, the intensity of the radiation was measured to be higher than at ground level. This was not at all consistent with the observed absorption rates, under the assumption that the radiation was coming from the Earth's surface. Hess concluded that the radiation was of extra-terrestrial origin [16].

Kolhörster [17] confirmed Hess' results in 1913–1914 by performing a series of balloon flights with a maximum altitude of 9200 m. He also calculated, in the assumption that the radiation was coming from above, the absorption coefficient for air. The result was eight times smaller than the known absorption coefficient of *gamma rays* for air.

Millikan & Bowen [18] developed small and light electroscopes which could autonomously record a series of measurements on photographic film. Instrumented, unmanned balloons took flights up to 15 000 m and found intensities of only 1/4 of those measured by Hess and Kolhörster. They concluded in 1925 that at higher altitudes the intensity decreased again and therefore, that the radiation was of terrestrial origin. However, experiments performed only one year later by Millikan & Cameron [19] using electroscopes submerged in lakes at different altitudes, confirmed that the radiation was coming from above. It was Millikan who first proposed the term *cosmic rays*.

Soon thereafter, Clay [20, 21] found that ionization increases with latitude, almost certainly caused by the magnetic field of the Earth. He concluded that gamma rays could not be the (only) source of cosmic rays. His results were disputed by Millikan. When the Geiger-Müller counter became available in 1928, the particle nature of cosmic rays was established by Bothe & Kolhörster [22]. Millikan did not accept this interpretation. When, in 1932, Compton [23] performed a world-wide survey of cosmic ray intensities and firmly established the dependence on latitude and the charged particle nature of cosmic rays, Millikan attacked his views. Finally, after performing new experiments of his own, Millikan admitted in 1933 that cosmic rays indeed (mostly) consisted of charged particles.

Cosmic rays have been instrumental in developing a coherent view of the particle nature of our world. Both antimatter (positrons) and the first examples of other particle families (muons) were discovered in cosmic rays. With the discovery of *extensive air showers* by Auger *et al.* [24] in 1939 and independently by Rossi [25, p. 5] in 1934 a new field of research was established.

A recent review of the history of cosmic rays is given in [26]. The authors focus on the contributions of Pacini.

1.2 Cosmic Rays in the Solar Neighborhood

Cosmic particles continuously bombard the atmosphere of the Earth. These particles are called *primary cosmic rays*, as they are not yet altered by interactions in the atmosphere.

The composition and energy spectrum of primary cosmic rays have been studied extensively. For recent results, see e.g. [27, 28, 29, 30]. The most accurate experiments make use of direct measurements, i.e. whereby the primary cosmic rays are detected. To reduce the effect of the atmosphere, these experiments are usually carried into the stratosphere using balloons, such as [31]. The experiment is concluded when the balloon bursts. The equipment falls back to the surface, deploying a parachute to ensure a soft landing. The experiment is then recovered for analysis. It is also the case that experiments are being conducted in space, such as [32].

Charged cosmic rays consist mainly of protons (84 %) and alpha particles (12 %). Most of the remainder are heavier nuclei. Electrons make up less than 1 % of cosmic rays [33].[2] The chemical abundances are shown in Figure 1.1, along with the relative abundances of elements in the Solar System. Large similarities clearly exist. However, the light elements Li, Be and B are much more abundant, as well as the sub-iron elements Sc, Ti, V, Cr and Mn. This is attributed to *spallation*, the splitting of nuclei due to interactions in the interstellar medium (ISM). The most abundant elements in the universe are H, He, C, O and Fe. Spallation of C and O will thus increase the abundances of Li, Be and B. Similarly, the sub-iron elements are produced in spallation of Fe.

Another feature which is evident in both the solar system abundances and the cosmic ray abundances is the difference between elements with even and odd atomic numbers. An explanation for this can be found by considering the *shell model* of the nucleus. Nuclear configurations with an even number of protons and neutrons

[2] Numbers are taken from Grupen [33, pp. 78, 84], but corrected here for *total* composition, including electrons.

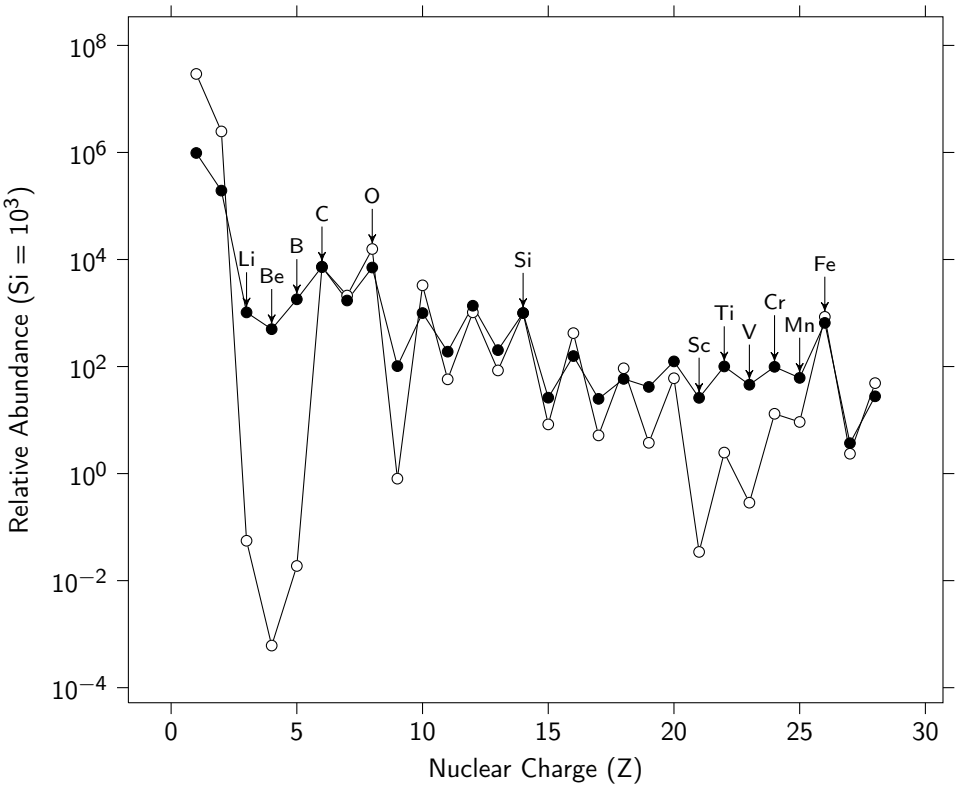


Figure 1.1 – Relative abundances of elements in the Solar System and in cosmic rays. The relative abundances are normalized, with the abundance of Si set to 1000. Two datasets are shown: relative abundances of elements in the Solar System (open circles), and relative abundances of galactic cosmic rays (GCR) measured by the CRIS instrument on the NASA/ACE satellite ($Z \geq 3$) and the BESS balloon experiment ($Z \leq 2$), during a solar minimum (closed circles). The plot is redrawn from [32], using data from [32, 34, 35, 36].

(considered separately) are more stable than odd-numbered configurations. If the total number of nucleons is a so-called *magic number*, i.e. 2, 8, 20, 28, 50, 82 or 126, then the nucleus is extremely stable.

Antiparticles are extremely rare in cosmic rays [37, 38]. This is further corroboration for the observation that the universe seems to consist only of matter, without any regions made up of antimatter. The antimatter that *is* found in cosmic rays is produced by the interaction of cosmic rays with the ISM. For example, antinucleons can be produced by proton-proton collisions. Positrons are produced by pair creation, i.e.

$$\gamma^* \rightarrow e^- + e^+. \tag{1.1}$$

Only one in every 10 000 charged nucleons is an antiproton [38]. Positrons are relatively more abundant, for every ten electrons there is one positron.

The particle flux strongly depends on the energy of the particle. Low-energy particles are deflected by the magnetic fields of the Earth and the Sun. Very low-energy particles get trapped in the Van Allen radiation belts surrounding the Earth. These particles leak away when they enter the atmosphere near the magnetic poles. Interactions between these particles and air molecules produce the spectacular aurorae. These particles stem mainly from the solar wind, the stream of particles ejected from the upper atmosphere of the Sun. Solar wind particles are predominantly protons and electrons in the MeV energy range [33] and where the solar wind encounters the magnetic field of the Earth, the interaction region is called the *magnetosphere*. The properties of the magnetosphere strongly depend on solar activity and thus follow the solar cycle [39], which lasts eleven years. Furthermore, the solar wind, as it is a *plasma*, carries the magnetic field of the Sun into the entire solar system and beyond, throughout the *heliosphere*. Therefore, the so-called *interplanetary magnetic field* also strongly depends on the solar cycle. The particle fluxes of low-energy primary cosmic rays thus follow this same cycle. When solar activity is high, fewer cosmic rays reach the Earth. For particle energies larger than 10 GeV, solar modulation is no longer apparent.

The energy spectrum of primary cosmic rays is shown in Figure 1.2. Except for the low-energy region ($E < 10\,\mathrm{GeV}$), the spectrum follows a power law, i.e. $F(E) \sim E^{-\gamma}$, with the spectral index $\gamma \sim 3$. It is assumed that this feature of the spectrum largely originates at the source, i.e. it is the result of the acceleration process.

Cosmic rays are predominantly produced inside our Galaxy and are contained by the galactic magnetic field. Above a certain energy threshold, cosmic rays start to leak away and the spectrum becomes steeper. This is a probable explanation for the so-called *knee* in the cosmic ray spectrum. At the knee ($E \approx 4\,\mathrm{PeV}$), the spectral index γ changes from 2.7 to 3.1 [41]. Another explanation is that at the knee, the maximum energy at which cosmic ray particles can be accelerated by the sources in the galaxy is reached. Both explanations may play an important role. See Hörandel [42] for a detailed overview of models explaining the knee. Since both magnetic confinement and maximum attainable energy are proportional to charge, heavier nuclei in the galaxy must be observed to have larger maximum energies than protons. It means that while protons start to leak away, heavier nuclei like iron are still bound to the galaxy. In a particular source, heavy nuclei can be accelerated to higher energies than protons. Therefore, at energies around the knee, the composition of the spectrum changes slowly from proton to iron [30]. At energies significantly above the knee, cosmic rays must originate from extra-galactic

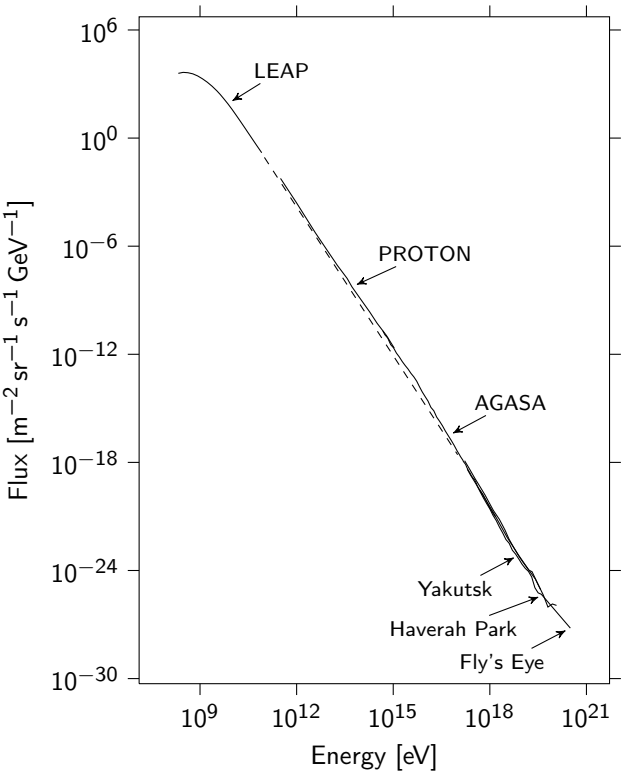


Figure 1.2 – Differential flux of primary cosmic rays as a function of particle energy. The dashed line corresponds to a spectral index $\gamma = 2.75$. The figure shows the results from several experiments, indicated by the labels. LEAP is a balloon-borne experiment; PROTON is a satellite and the other experiments are ground-based arrays. The data from several experiments overlap. Figure redrawn from Cronin *et al.* [40].

sources.

At even higher energies, the spectrum becomes less steep. This region is called the *ankle*. At these energies, the flux is so low that measurements of the spectrum show very low statistics.

Greisen, Zatsepin and Kuzmin predicted that ultra-high energy cosmic rays (UHECRs) will interact with the cosmic microwave background (CMB). At ultra-high energies, the available energy in a collision between a cosmic ray particle and a CMB photon is

above the so-called *pion production threshold*:

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow p + \pi^0, \tag{1.2}$$

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow n + \pi^+. \tag{1.3}$$

Each collision results in a loss of energy for the cosmic ray particle. The process continues until the available energy becomes lower than the pion production threshold. For UHECR protons, the GZK limit is 50 EeV. The mean free path for a proton with an energy of 60 EeV is 10 Mpc. If there are protons above this energy, they cannot have traveled farther than about 50 Mpc, since their energy would have been reduced by the GZK effect. Similar arguments hold for heavier nuclei and photons. The latter lose energy due to pair production:

$$\gamma + \gamma_{CMB} \rightarrow e^- + e^+. \tag{1.4}$$

The GZK effect should result in a pile-up of cosmic rays with an energy below the GZK limit, similar to the pile-up of sub-iron nuclei which is observed in the *composition* of cosmic rays. This pile-up is most likely responsible for the smaller slope of the spectrum around the ankle.

If cosmic ray sources above the GZK limit exist that are closer than 50 Mpc, a proportion of them should be detected. Thus, if the limit is observed, it proves that at these energies, the majority of cosmic ray sources are indeed more than 50 Mpc away. Cosmic ray particles, produced with energies above the GZK limit and originating at those sources, drop below the limit before they reach the Earth.

To put these numbers into perspective, the diameter of the Milky Way is approximately 35 kpc and the largest distance within the *local supercluster*[3] is only 30 Mpc.

1.3 Production

1.3.1 Sources

The galactic magnetic field deflects cosmic ray particles. At energies below 100 TeV, the direction of arrival is not correlated to the direction of the source. At these energies, the distribution of cosmic rays becomes highly isotropic. To try and find a source of

[3]The Local Supercluster contains the Virgo cluster and the Local Group. The latter comprises more than 50 galaxies, including the Andromeda galaxy and the Milky Way. The Local Supercluster contains more than 100 galaxy groups and clusters.

cosmic ray particles at a distance of 10 Mpc, the evolution of the galactic magnetic field over the last 50 million years would need to be known. At energies above 10^{19} eV, however, the direction of arrival should be in correlation with the direction of the source. No correlation with the galactic plane is found, however, which is further proof for extra-galactic sources. Recently, the Auger collaboration reported [43] on their current results regarding the correlation of UHECRs with active galactic nuclei (AGNs). Events with estimated energies above 55 EeV are reconstructed and their origins are compared to the positions of known AGNs within 75 Mpc, as recorded in the 12^{th} edition of the Véron-Cetty and Véron catalog [44]. If the origin of an event is closer than 3.1° to the position of an AGN, it is counted as a *correlation*. The Auger collaboration reports 21 correlations out of 55 events (38 %) [43]. The correlation fraction is down from an early estimate of 69 %, which was derived from 9 correlations out of 13 events. The correlation fraction which is expected from an isotropic distribution of events is 21 %. The collaboration emphasizes that further research should be done.

If the direction of arrival can be used to pinpoint sources, it should be noted that the small deflections caused by intergalactic magnetic fields in combination with the large distances traveled results in significant time lags when compared to the arrival time of neutral particles like photons and neutrinos. For example, when γ-ray bursts (GRB) produce charged cosmic rays, they arrive months or even years later than the photons and neutrinos produced in the same event [33, p. 85].

In order to identify the sources of cosmic rays, Hillas [45] studied the various possible acceleration mechanisms and noted that they all need strong magnetic fields and large sizes. In the case of direct acceleration, rotating magnetic fields provide electric fields which accelerate charged particles. The stronger the field and the larger its size, the more energy a particle can acquire. For statistical acceleration models, in which acceleration occurs in many small steps, the amount of energy that a particle gains depends on the ability of the accelerator to *contain* the particles. Statistical acceleration can occur at a shock wave front, resulting from e.g. supernovas. If particles are contained by magnetic fields, they can cross the front many times and slowly gain energy.

Charged particles contained in a uniform magnetic field rotate with a radius

$$r = \frac{m\beta c}{ZB}, \tag{1.5}$$

called the *Larmor radius*, with m, Z the particle mass and atomic number (or charge), βc the particle velocity perpendicular to the magnetic field, and B the strength of the

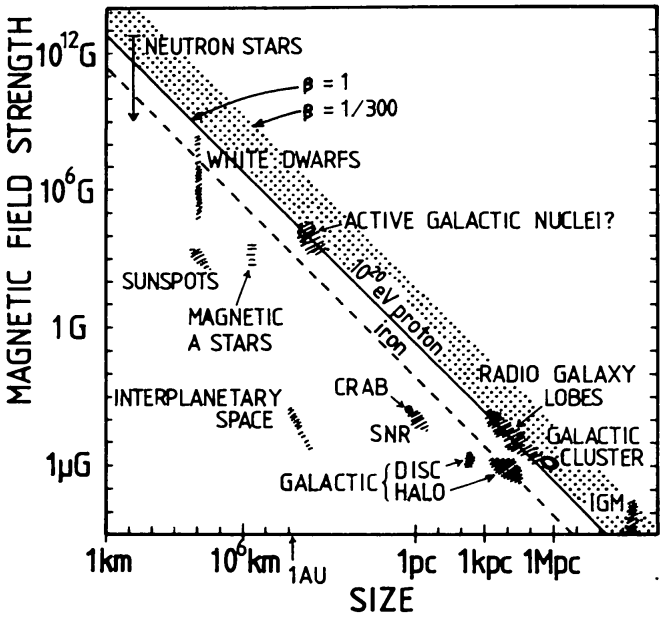


Figure 1.3 – Hillas plot. The size and magnetic field strength of possible acceleration sites are shown in the plot. The diagonal line shows the limit for 10^{20} eV protons, i.e. sources below that line can not accelerate protons to this limit. The dashed line shows the limit for 10^{20} eV iron nuclei. Reproduced from [45].

magnetic field. For particles to acquire large energies, the source should have strong magnetic fields or a large size. In particular, the source should be larger than the Larmor radius. If the source is very large, magnetic fields of the order of μG are sufficient to accelerate particles.

Hillas found [45] that for statistical acceleration models the approximate upper limit on the energy is given by

$$E_{15} \leq 0.5 \beta_s Z B_{\mu\mathrm{G}} L_{\mathrm{pc}}, \tag{1.6}$$

with E_{15} the energy in units of 10^{15} eV, $B_{\mu\mathrm{G}}$ the magnetic field of the source in μG, β_s the velocity of the shock wave relative to c and L_{pc} the size of the source in pc. When the magnetic field of the sources is plotted versus the size of the sources, one obtains the *Hillas plot* shown in Figure 1.3. It is obvious that not many sources accelerate charged particles to the highest energies.

Alternative models that do not require a large size and a strong magnetic field are so-called *top-down models*, as opposed to the bottom-up acceleration models discussed thus far. Top-down models suggest that ultra-high energy cosmic rays are produced by

the decay of highly energetic exotic objects [46]. These should be relics of the Big Bang and none of them have been observed.

Candidate sources should not only accelerate particles to the highest energies, the totality of sources should also reproduce the observed power law spectrum. Finally, sources should be able to fill the galaxy with enough cosmic rays to produce an energy density of $1\,\mathrm{eV\,m^{-3}}$ [47, p. 11].

1.3.2 Acceleration Mechanisms

Currently, acceleration is favored above decay models. This leaves open the question of *how* cosmic rays are accelerated to ultra-high energies. Such a mechanism must not only be realistic, but should also reproduce the cosmic ray energy spectrum. There are many possible acceleration sites, including sun spots and pulsars with high magnetic fields. In particular, pulsars are able to accelerate particles due to their rotating magnetic fields, which generate strong electric fields. Particles are accelerated along so-called *jets*.

Other processes work more slowly, but over long time scales. These provide an energy spectrum similar to the observed spectrum of cosmic rays up to the knee.

Two acceleration models will be discussed in the following sections. In each section, simple considerations will lead to the development of a basic model. Similar calculations can be found in the literature [33, pp. 67–68]. Those calculations, however, contain errors (a missing factor of two in parts of the equations), which have been repeated by others.

Fermi Acceleration

Two well-known candidate models are Fermi acceleration and shock acceleration. Fermi supposed [48] that the interstellar medium is filled with turbulent magnetic fields, with high fields being found mainly in interstellar clouds. These fields can act as a *magnetic mirror*. In such a configuration, charged particles encounter varying field strengths when moving along magnetic field lines. This results in the particles being deflected away from regions with strong fields. In Fermi's model, charged particles encounter magnetic clouds and eventually leave the cloud, being scattered in the process. If the particle exits the cloud traveling in its original direction, its energy remains unchanged as all scattering in a magnetic field is *elastic*. However, if the direction of the particle is *changed*, this does not need to be the case. All magnetic interactions are elastic only

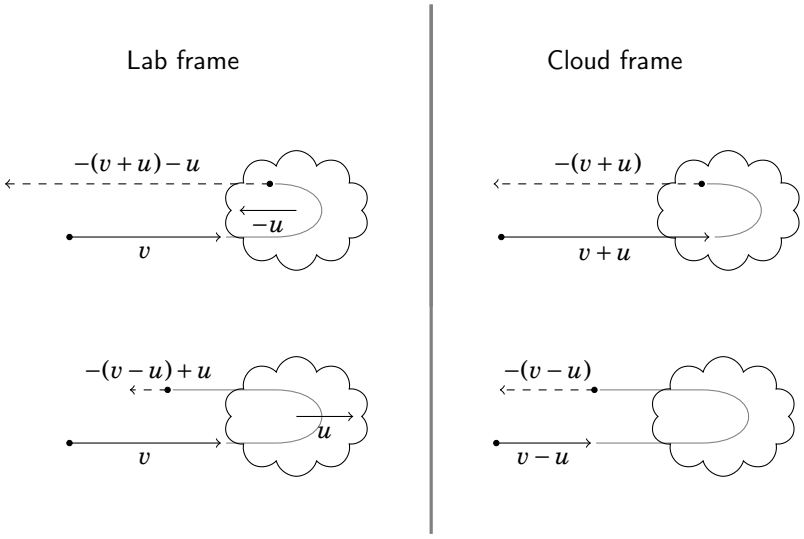


Figure 1.4 – Fermi acceleration mechanism. In the top part of the figure, a particle with velocity v moves towards a magnetic cloud with velocity $-u$. In the frame of the magnetic cloud, the velocity remains unchanged when the particle is reflected back. However, in the lab frame the particle has gained velocity, and thus kinetic energy. In the bottom part of the figure, the cloud moves with a velocity u and is overtaken by the particle. In that case, the particles *loses* energy.

in the reference frame of the cloud.[4] To determine the energy in the lab frame, two transformations must be performed: one from the lab frame to the cloud frame and one from the cloud frame back to the lab frame.

As a simple model calculation, consider a particle with velocity v traveling towards a magnetic cloud which itself is approaching with velocity $-u$. The situation is depicted in the top part of Figure 1.4. The kinetic energy of the particle is then given by $E_0 = \frac{1}{2}mv^2$. In the frame of the magnetic cloud, the particle's velocity is given by $v+u$. When reversed, this becomes $-(v+u)$. Finally, in the original lab frame, the particles velocity has become $-(v+u)-u = -v-2u$. With this, the particle's kinetic energy is now given by

$$E_1 = \frac{1}{2}m(-v-2u)^2 = \frac{1}{2}m(v^2+4u^2+4uv) \tag{1.7}$$

while the energy gain becomes

$$\Delta E_1 = E_1 - E_0 = \frac{1}{2}m(4u^2+4uv). \tag{1.8}$$

[4] In the cloud frame the velocity of the cloud is zero, and it contains only magnetic fields. In the lab frame, however, a non-zero cloud velocity will result in electric fields. Hence, the collisions will no longer be elastic.

However, it is also possible that the particle encounters a cloud moving in the same direction with velocity u, i.e. a *rear-end collision*. That situation is shown in the bottom part of Figure 1.4. In this case, the particle's velocity in the cloud frame becomes $v-u$ and reversed, $-(v-u)$. In the lab frame, this becomes $-(v-u)+u=-v+2u$. The kinetic energy is then given by

$$E_2 = \frac{1}{2}m(-v+2u)^2 = \frac{1}{2}m(v^2+4u^2-4uv), \tag{1.9}$$

and the energy gain by

$$\Delta E_2 = E_2 - E_0 = \frac{1}{2}m(4u^2-4uv). \tag{1.10}$$

If the particle's velocity is higher than the cloud's velocity, i.e. $v>u$, the energy 'gain' ΔE_2 is negative and the particle has *lost* energy. If $v<u$, it is impossible for the particle to overtake a magnetic cloud.

Under the assumption that the velocity v is much greater than u, the particle will experience approximately the same number of head-on collisions and rear-end collisions. The mean energy gain over the two types of collisions is given by:

$$\frac{\Delta E_1 + \Delta E_2}{E_0} = \frac{\frac{1}{2}m(8u^2)}{\frac{1}{2}mv^2} = 8\frac{u^2}{v^2} \tag{1.11}$$

which is proportional to the square of the velocity of the magnetic cloud. Therefore, this model of acceleration is also referred as *second order Fermi acceleration*.

If there are many clouds with randomized velocities and directions and the particles velocity v is still smaller than the mean cloud velocity u, the probability of a head-on encounter is higher than that of a rear-end encounter. This means that the probability of an energy gain is higher than that of an energy loss.

There is a probability that the particle escapes from the magnetic cloud region. Taking the energy gain per collision and the probability of a further collision, one can calculate the resulting energy spectrum, which happens to be a power law [33, p. 75]. The spectral index depends on the magnetic cloud velocities and the probability of the particle escaping the region of acceleration [33]. Fermi acceleration requires long time scales, and therefore may not work in practice.

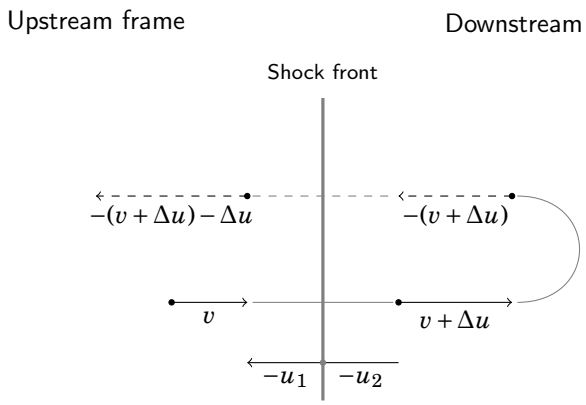


Figure 1.5 – Shock acceleration mechanism. A shock front is moving through the interstellar medium with velocity $-u_1$ in the upstream frame, and $-u_2$ in the downstream frame. A particle with velocity v in the upstream frame crosses the front. Downstream, it is reflected and in that frame the velocity remains unchanged. However, in the original upstream frame, it has gained velocity, and thus energy.

Shock Acceleration

During supernova explosions, stellar matter is ejected at very high speeds into the interstellar medium. This heats up the interstellar matter, increasing its density and pressure and driving the matter outward. The emerging *shock wave* travels faster than the supernova ejecta following it. Both stellar and interstellar matter are *plasmas* and it is the electromagnetic fields that cause the shock wave, not collisions in the plasma. In fact, the particle densities are so low that the matter is called *collisionless*.

As seen from the perspective of the shock front (Figure 1.5), upstream (*unshocked*) gas flows towards the front at a velocity of u_1 and a density of ρ_1. Since the density ρ_2 of the downstream (*shocked*) gas is higher, it is possible to calculate its velocity u_2. As the shock front is immaterial, any matter flowing into the front must also exit the front. Therefore,

$$\rho_1 u_1 = \rho_2 u_2, \tag{1.12}$$

and since $\rho_2 > \rho_1$, then $u_2 < u_1$. Let $\Delta u = u_1 - u_2$, which is a positive number. The frame of the upstream gas is shown in the left part of Figure 1.5, with the frame of the downstream gas shown in the right part.

The situation is now very similar to the situation described by second order Fermi acceleration. A particle can cross the shock front and be deflected by magnetic fields inside the shocked gas. If its velocity is larger than the velocity of the shock front, it

may cross the shock front again, resulting in

$$E_1 = \frac{1}{2}m(-v - 2\Delta u)^2 = \frac{1}{2}m(v^2 + 4\Delta u^2 + 4v\Delta u), \tag{1.13}$$

giving the energy gain

$$\Delta E_1 = E_1 - E_0 = \frac{1}{2}m(4\Delta u^2 + 4v\Delta u). \tag{1.14}$$

The difference between Shock and Fermi acceleration is that the probability of encountering a *receding* shock wave is, in this situation, zero. Unlike magnetic clouds, the shock wave travels in only one direction: outward. The relative energy gain is thus given by

$$\frac{\Delta E_1}{E_0} = \frac{\frac{1}{2}m(4\Delta u^2 + 4v\Delta u)}{\frac{1}{2}mv^2} = 4\left(\frac{\Delta u^2}{v^2} + \frac{\Delta u}{v}\right). \tag{1.15}$$

Given the velocity $v \gg \Delta u$, the linear term dominates:

$$\frac{\Delta E_1}{E_0} \sim \frac{4\Delta u}{v}. \tag{1.16}$$

Because of the magnetic fields present in the plasma on both sides of the shock front, and the fact that the particle's velocity is higher than the velocity of the shock, the particle can cross the shock front countless times. Shock acceleration is sometimes called *Fermi acceleration of the first order* and is believed to be the primary process for cosmic ray production.

1.4 Propagation

Cosmic rays propagate from their source through the interstellar medium (ISM). The ISM consists of matter, magnetic fields and radiation fields. Charged particles interact with the magnetic fields. Matter interacts with matter and produces secondary particles. Additionally, electrons emit synchrotron radiation caused by magnetic acceleration and lose energy through Brehmsstrahlung and inverse Compton scattering.

The interstellar matter mainly consists of hydrogen, the most common form of which is atomic hydrogen (H I). This can be observed by the 21 cm spectral line, resulting from the hyperfine splitting of hydrogen energies. In dense regions, such as *giant molecular clouds*, the outer parts of a cloud absorb the energetic interstellar photons. As a result, the inner parts are shielded and in such regions molecules can be formed without

breaking up. Since cold H_2 is hard to detect, carbon monoxide (CO) is usually used as a tracer for H_2 [49]. In clouds with an embedded infrared source, one can observe both H_2 and CO absorption lines, and the H_2 / CO ratio can be obtained [50]. This ratio is then assumed to be approximately correct for all molecular clouds. For a study of the relationship between H_2 and CO, see e.g. Glover & Low [51]. In these dense regions, star formation can occur. Newly formed stars break up the cloud with their radiation and stellar winds and in these regions, hydrogen is ionized (denoted as H II).

In the galactic arms, atomic hydrogen has an average density of $1\,\mathrm{atom/cm^3}$ and a scale height[5] of 100 pc to 150 pc. Between the arms, the density decreases by a factor of 2 to 3. Molecular hydrogen is concentrated within the solar circle[6] and especially in the region of the galactic center. In giant molecular clouds, the average density is of the order of $1 \times 10^2\,\mathrm{atom/cm^3}$ to $1 \times 10^5\,\mathrm{atom/cm^3}$. Ionized hydrogen only makes up a small fraction of the interstellar matter. The average interstellar matter density is $1\,\mathrm{nucleon/cm^3}$ [52, p. 75].

The large-scale structure and strength of the magnetic fields in the galaxy are unknown. Most knowledge is gathered by studying other galaxies; in particular galaxies which are perpendicular to our line of sight. By observing the Faraday rotation of linearly polarized signals from radio pulsars, which is caused by the magnetic field, the galactic distribution of the fields can be studied. It is believed that the magnetic field of the Milky Way is similar to that of other galaxies. Fluctuations in the field strength are, however, quite large. The magnetic field near the Solar System is about 1.8 µG for the regular component, which is uniform over a large region of space [53]. The total field strength, which includes random fields on smaller scales, is about 5 µG.

Ionized gas and magnetic fields carried by the gas form a *magnetohydrodynamic (MHD) fluid*. This fluid can support *Alfvén waves*. These waves can be created by cosmic rays streaming into the ISM. Alfvén waves can scatter cosmic rays and under certain conditions *self-confinement* can occur: cosmic rays stream into the ISM, create Alfvén waves which then scatter and even contain the particles to the acceleration region. The creation of such waves may even be essential to cosmic ray acceleration through shocks.

The abundance of the unstable nucleus ^{10}Be, which is created by spallation of heavy cosmic ray nuclei, can be used to calculate the confinement time of (heavy) cosmic rays in the galaxy. It is of the order of 10^7 yr [52, p. 85].

[5] Distance over which the density decreases by a factor of e. Note that the *thickness* is twice the height.
[6] The solar circle is the orbit of the Sun around the galactic center.

1.5 Cosmic Rays in the Atmosphere

On average, primary cosmic rays in the ISM traverse a column density of only a few $\mathrm{g\,cm^{-2}}$. In contrast, the atmosphere of the Earth has a column density of $1030\,\mathrm{g\,cm^{-2}}$.

When discussing cosmic rays in the atmosphere, it is common to refer to the *atmospheric depth* X, instead of the height h above sea level. The atmospheric depth is the amount of matter above the atmospheric layer at height h. Like column density, the atmospheric depth is measured in $\mathrm{g\,cm^{-2}}$ and is defined by

$$X \equiv \int_h^\infty \rho(h')\,\mathrm{d}h'. \tag{1.17}$$

For a perfect gas with a constant temperature in hydrostatic equilibrium, the profile is given by [52, p. 122]

$$X = X_0 \exp\left(\frac{-h}{h_0}\right), \tag{1.18}$$

where X_0 the atmospheric depth at sea level ($1030\,\mathrm{g\,cm^{-3}}$) and h_0 the scale height of the atmosphere.

The temperature of the atmosphere is not constant. The US Standard Atmosphere [54] is a collection of models defining temperature, pressure, density and several other observable qualities of the atmosphere over a wide range of altitudes. Using these models, the altitude-dependent atmospheric density can be approximated by dividing the atmosphere in several layers. Within each layer a linear fit of the temperature to the experimentally observed values is made. Then, the other properties of the atmosphere can be calculated.

When cosmic rays enter the atmosphere with a zenith angle θ, the amount of atmosphere traversed is called the *slant depth* and is, in the *flat Earth approximation*, given by

$$X' = X/\cos\theta, \tag{1.19}$$

with X sometimes referred to as the *vertical depth*. This approximation is valid for zenith angles less than 60°. For larger angles, the curvature of the Earth must be taken into account.

The radiation length[7] in air is $36.66\,\mathrm{g\,cm^{-2}}$ and the interaction length[8] in air is $90.0\,\mathrm{g\,cm^{-2}}$. The total atmospheric depth is therefore approximately 28 radiation lengths and 11 interaction lengths. The first interaction of primary cosmic rays with the

[7] The mean distance over which the energy of a high-energy electron is reduced to a factor $1/e$. This is approximately 7/9 of the mean free path of pair-production for a high-energy photon.

[8] The mean distance before a high-energy hadron undergoes a hadronic interaction.

atmosphere is at a height of about 15 km to 20 km.

1.5.1 Interactions

Cosmic rays traversing the atmosphere (or any substance of matter) can undergo many different types of interactions.

Photons, for example, lose energy through *Compton scattering* (photons interacting with electrons, ionizing the atoms) and *pair production* (photons creating a particle-anti-particle pair in the vicinity of a nucleus).

Charged particles lose energy through *ionization losses* (charged particles transferring energy to atomic electrons thus exciting or ionizing the atoms), *Bremsstrahlung* (charged particles radiating photons while interacting with the electromagnetic field of atomic nuclei), and *Rutherford scattering* (deflection by Coulomb forces). Smaller losses are due to *synchrotron radiation* (charged particles deflected by the Earth's magnetic field radiate photons), and the *Cherenkov effect* (charged particles traveling faster than the *phase velocity* of light in a medium emit radiation in a cone).

For hadrons, the situation is more complicated because of the many possible types of interactions. Hadronic processes include nuclear fragmentation, creation of resonances, and multiparticle production. Furthermore, these processes are harder to calculate. At very high energies, QCD perturbation theory can be used to calculate cross sections. At lower energies, perturbation theory breaks down and effective theories must be used [55]. However, at these energies experimental data is available from (collider) experiments. Approximations are made by measuring many cross sections at different energies and the resulting models describe the data very well.

When primary cosmic rays interact with the atmosphere they produce secondary particles. These particles will also interact and produce tertiary particles. This process continues and the total number of particles increases dramatically until the individual particle energies drop below the energy at which new particles can be created. Low-energy particles are absorbed in the atmosphere. The totality of particles created in this process is called a *cascade*. If a large number of secondary particles reaches ground level, the cascade is called an *extensive air shower (EAS)* which can have a footprint of several km^2.

1.5.2 Electromagnetic Cascades

Cascades are called *electromagnetic* when they consist of electrons and photons. They are initiated either by cosmic ray electrons and photons, or by electrons and photons

created as secondary particles in hadronic interactions from cosmic ray nuclei.

At high energies, Bremsstrahlung

$$\mathrm{e} \rightarrow \mathrm{e} + \gamma \tag{1.20}$$

and pair production

$$\gamma^* \rightarrow \mathrm{e}^- + \mathrm{e}^+ \tag{1.21}$$

dominate.

This process is surprisingly well described by the Heitler model. This model describes a cascade consisting of a single type of particle interacting exactly after an interaction length λ. Each interaction creates two particles with equal energy, which is half the parent energy. Thus, a primary particle with energy E_0 will interact after one length λ to create two particles, each with energy $E = E_0/2$. After two interaction lengths, there are four particles with energy $E = E_0/2^2$; after three interaction lengths, there are eight particles with energy $E = E_0/2^3$, and so on.

The Heitler model [56] describes an electromagnetic cascade qualitatively until the particle energy drops below the critical energy at which no more particles can be created. It does not describe the absorptive processes. Ionization losses mean that electrons and positrons lose energy rapidly until they annihilate (positrons with atomic electrons) or recombine (electrons with ionized atoms). Photons will be absorbed in Compton scattering and the photoelectric effect. A generalization of the Heitler model is discussed in [57], which does describe absorptive processes and can be used to describe the full longitudinal development of an electromagnetic shower.

1.5.3 Hadronic Cascades

Hadronic cascades are created by cosmic ray protons and nuclei interacting with the atmosphere. Hadronic interactions and decays mainly result in the creation of pions and kaons, for example

$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \Delta^+ \rightarrow \mathrm{p} + \mathrm{p} + \pi^0, \tag{1.22}$$

where the Δ^+ resonance can also decay into a neutron and a charged pion. The pion to kaon production ratio is approximately 9:1. Pions mainly decay to muons, electrons,

neutrinos, and photons, e.g.

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu, \\ \pi^- &\rightarrow \mu^- + \overline{\nu}_\mu, \\ \pi^0 &\rightarrow \gamma + \gamma. \end{aligned} \tag{1.23}$$

Kaons have many decay modes [58] and mainly decay to pions, muons, electrons and neutrino's. At relativistic energies, the decay of pions and kaons is retarded, in the lab frame, due to *time dilation*. At these energies, interactions with matter can occur before the particles decay. However, at lower energies, decay is the dominant process.

The photons produced by decaying neutral pions initiate electromagnetic cascades. Furthermore, electromagnetic cascades can be created by electrons and positrons resulting from decaying muons

$$\begin{aligned} \mu^- &\rightarrow \mathrm{e}^- + \overline{\nu}_\mathrm{e}, \\ \mu^+ &\rightarrow \mathrm{e}^+ + \nu_\mathrm{e}. \end{aligned} \tag{1.24}$$

The largest fraction of the primary energy ultimately goes towards the production of electromagnetic cascades.

1.5.4 Longitudinal and Lateral Shower Profiles

To obtain an accurate description of the evolution of a shower, Rossi & Greisen [59] solved a set of diffusion equations describing the development of electromagnetic cascades for various approximations. Qualitatively, the longitudinal development can be parametrized by [33, p. 157]

$$N(t) \sim t^\alpha \exp(-\beta t), \tag{1.25}$$

with $N(t)$ being the number of particles at $t = x/X_0$ the shower depth in radiation lengths, and α and β being free fit parameters parameterizing the creation and absorption of particles respectively. Figure 1.6 shows the longitudinal development of an EAS initiated by a 1 PeV proton.

EAS is spread out laterally because of multiple scattering in electromagnetic showers and transverse momenta in hadronic interactions. Figure 1.7 shows the lateral distribution of particles reaching sea level in an EAS initiated by a 1 PeV proton. The most abundant particles are photons. Near the shower core, the number of electrons is much larger than the number of muons. The lateral distribution of muons is flatter,

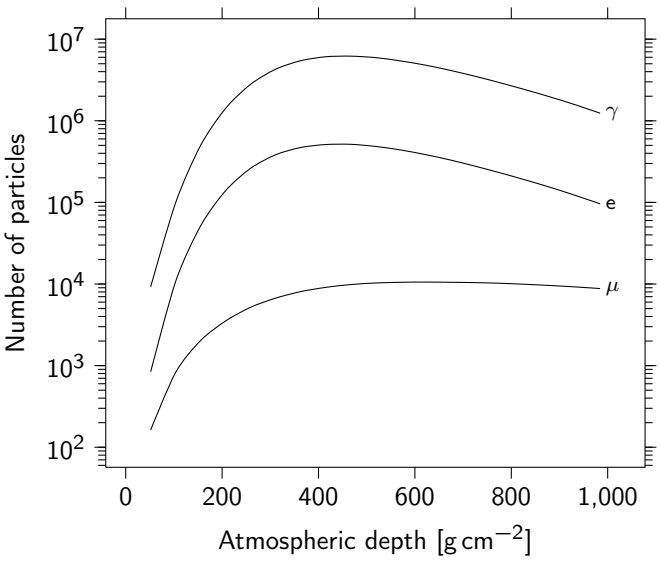


Figure 1.6 – Longitudinal development of an EAS initiated by a 1 PeV proton. Only the photon, electron (e^- and e^+), and muon (μ^- and μ^+) densities are shown.

however. At larger distances the number of muons is comparable to or even larger than the number of electrons. This is mainly due to the difference in mean free path. Muons are predominantly created high in the atmosphere and have few, if any, interactions before they reach the ground. Their expected lateral distance is proportional to the vertical distance, i.e. $\langle x \rangle \propto h$. Electrons, on the other hand, are created throughout the development of the shower and undergo many interactions. Electrons create photons which, in turn, can create electrons. As a result, the distribution of electrons is subject to the *random walk* process. The expected lateral distance is proportional to the *square root* of the vertical distance, i.e. $\langle x \rangle \propto \sqrt{h}$.

For increasing primary energy, the atmospheric depth of the shower maximum increases logarithmically and the total number of particles (the *shower size*) increases linearly. By measuring particle densities, the shower size can be estimated. This can be used to determine the primary energy from observation of the EAS.

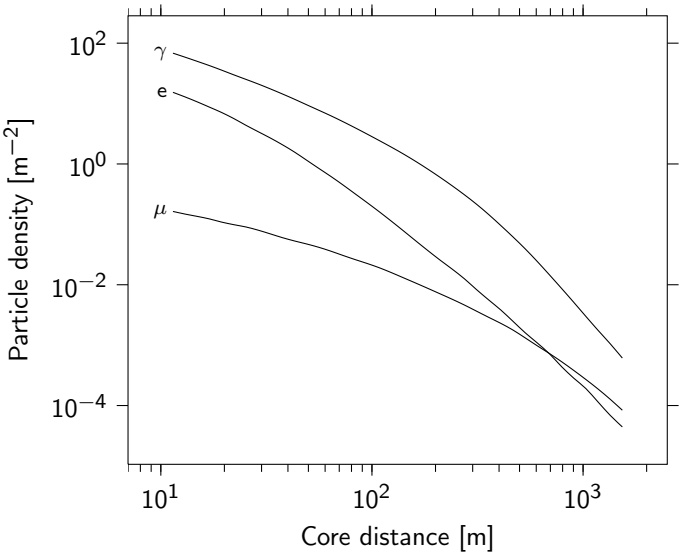


Figure 1.7 – The lateral distribution of particles at sea level of an EAS initiated by a 1 PeV proton. Only the photon, electron (e^- and e^+), and muon (μ^- and μ^+) densities are shown.

1.6 Ground-Based Detection

1.6.1 Particle Flux at Ground Level

Most primary particles do not have enough energy to generate showers which reach ground level. However, muons are created in such showers. The lifetime of muons is only 2.2 µs, but high-energy muons have a Lorentz factor sufficient to reach sea level. At 1 GeV, the Lorentz factor $\gamma = 9.4$, resulting in a *decay length* of $s \approx \gamma\tau c = 6.2\,\text{km}$. Moreover, muons lose much less energy due to Bremsstrahlung than electrons. Therefore, muons are far more likely to reach sea level than electrons. As a result, of all the charged particles at sea level, 80 % are muons, resulting in a flux of approximately $1\,\text{cm}^{-2}\,\text{min}^{-1}$ (Figure 1.8).

There is a small flux of nucleons at sea level. They are created in hadronic cascades. There is a small probability that individual nucleons of a primary cosmic ray nucleus survive and reach sea level. Electromagnetic cascades create photons, electrons and positrons and the flux of these particles is much lower than the muon flux.

The decay of pions and muons generates neutrinos, which form a large background signal for neutrino telescopes.

The particle fluxes discussed so far form a background signal for all experiments

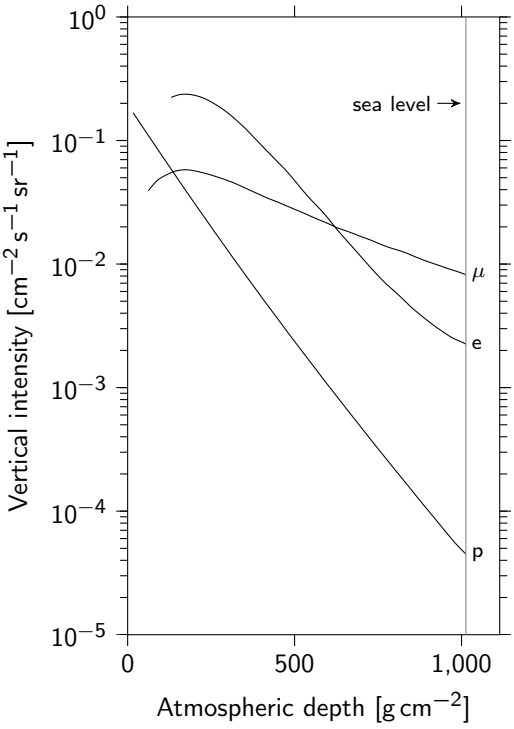


Figure 1.8 – Particle composition in the atmosphere as a function of atmospheric depth. Figure redrawn from [33, p. 145].

looking for EAS. The minimum energy needed for a primary particle to generate an EAS which can reasonably be measured on the ground, is about 100 TeV. When observing showers, the particle fluxes *in the shower* are very different. Only about 10 % of charged particles are muons. The remainder is dominated by electrons and positrons. The flux of photons is even higher (Figures 1.6 and 1.7).

1.6.2 Ground-based Experiments

In this section, an overview will be given of experimental techniques and ground-based experiments. It is not an exhaustive review.

One of the most commonly used tools to detect cosmic rays is a scintillator, which is discussed in Section 2.2.2. In short, scintillators consist of materials that emit light when charged particles traverse them. This light can then be collected by sensitive photomultiplier tubes (PMTs). These particle detectors have been used in many ground-

based experiments. Typically, a large array of detectors is used to measure as many of the particles that make up an EAS as possible. The first experiment of this kind was constructed at Volcano Ranch in New Mexico, USA. Furthermore, the KASCADE array at Karlsruhe, Germany and the AGASA experiment at Akeno, Japan have done extensive research into the structure of EAS, as well as the energy and composition of cosmic rays and their origin.

Another type of ground-based detector is the water Cherenkov detector. It consists of a large tank filled with water in which penetrating high-energy charged particles will emit Cherenkov light. High-energy photons can also be observed because they create electron-positron pairs inside the tank. The light is detected using PMTs. Cherenkov water tanks are very efficient in measuring photons and electrons as they provide few absorption lengths ($X_0 = 36\,\text{cm}$). The downside is that they must be very large, making them very unwieldy and more expensive than scintillators. Cherenkov tanks were deployed at the Haverah Park array in the UK, and are currently in use e.g. at the Pierre Auger observatory near Malargüe, Argentina.

The atmosphere effectively acts as a calorimeter for primary cosmic rays, with a thickness of 27 radiation lengths. One of the problems with using scintillators or Cherenkov water tanks to measure EAS, is that only information from one layer of this calorimeter is available, i.e. ground level. Thus, the primary energy or longitudinal development of the shower cannot be accurately measured. Moreover, the layer that *is* sampled, is usually not sampled very densely.

EAS in the atmosphere emit radiation in the form of Cherenkov light, which is emitted in a specific cone. Detectors at the ground looking into the sky can detect this light, but only on clear, moonless nights. This principle is, for instance, employed by the *High Energy Stereoscopic System (H.E.S.S.)* experiment near the Gamsberg mountain Namibia. The H.E.S.S. detector can measure showers from TeV photons and reconstructs the point of origin. The detector can spatially resolve extended gamma ray sources.

When an EAS develops in the atmosphere, light is also emitted isotropically in the form of fluorescent light from nitrogen. By observing this light (again, only on clear, moonless nights) from a lateral distance, the longitudinal development of the shower can be observed. The amount of light received from a certain direction is a measure for the number of charged particles in the shower at that point. Using this information, much more accurate estimates of the primary energy as well as the composition can be made. This technique is employed by the Fly's Eye and HiRes detectors at Dugway Proving Grounds in Utah, USA, as well as the Pierre Auger observatory. As with Cherenkov

light detection, the strict requirements on dark backgrounds restrict the use of these measurements to only 10 % duty cycle.

One can also measure radio emissions caused by the synchrotron radiation of electrons deflected by the magnetic field of the Earth. However, strong backgrounds exist in practically all wavelength ranges. This is a relatively new and very challenging field of research, which is currently being conducted as part of the LOFAR and Pierre Auger experiments. See e.g. [60].

1.6.3 Cosmic Ray Experiments and Outreach

There are many open questions in cosmic ray physics and it is therefore an interesting and challenging field of research. It involves astrophysics, particle physics and a number of topics such as the photoelectric effect and special relativity. Phenomena such as black holes, supernovae and pulsars have captured the imaginations of many. High-energy particles colliding with atmospheric nuclei create a cascade of secondary particles. They hurtle through the atmosphere and are detected by flashes of light in dark slabs of plastic, after they already should have ceased to exist. Cosmic ray physics is ideally suited to interested high school students and can serve to introduce them to the many concepts of modern physics.

High school students are mainly taught physics from the late nineteenth century and before. They are unfamiliar with topical research interests and are unable to form an accurate picture of what research and physics is about. Therefore, outreach projects are deemed essential to educate students and to interest them in a career in physics. Several outreach projects focus on cosmic rays.

James Pinfold, of the University of Alberta, was the first to propose an outreach project on cosmic ray physics [61]. ALTA, the Alberta Large-area Time-coincidence Array, consists of a sparse array of cosmic ray detection stations located at the University of Alberta and local high schools. Students build, deploy and maintain the detectors, as well as conducting basic research. Its example has been followed by several other projects in the US, such as WALTA[9] (Washington), CROP[10] (Nebraska), CHICOS[11] (California), SALTA[12] (Snowmass), VICTA[13] (Victoria, Canada) and MARIACHI[14] (New

[9] Proposed in 1999, active through the first decade of this century. The last workshop was held in summer, 2009.

[10] Started in 1999, does not seem to be active.

[11] The website seems to have been taken offline.

[12] Last activity from around 2001.

[13] Active since 2003, last activity possibly around 2009.

[14] Some activity, but live data view has not been updated since 2009.

York). As of 2006, CHICOS is the largest array in the US featuring 70 high schools over an area of 400 km^2 [62].

In 2001, NAHSA was proposed [63] in the Netherlands as high school array in the city of Nijmegen, and has been recording data since June 1, 2002. The project had a very successful start and resulted in the creation of a national project, HiSPARC, in 2002 [64]. HiSPARC is an abbreviation of High School Project on Astrophysics Research with Cosmics. SEASA[15] was proposed in 2002 [65]. It is an array in Stockholm, Sweden. The CZELTA array has been built in the Czech Republic. In a collaboration with ALTA, data is analyzed in the search for large-distance cosmic ray phenomena [66]. Other projects in Europe include SkyView (Germany), the Roland Maze Project (Poland), RELYC (France), Cosmic Rays Telescope in Portuguese High Schools (Portugal) and EEE (Italy) [62].

The HiSPARC project has two goals: to study cosmic rays and to expose high school students to the challenges and rewards of scientific research. In the framework of this project, high school students are introduced to concepts of astroparticle physics. They construct their own detectors, test them and deploy them at their school (Figure 1.9). The completed detectors are used by the students in research projects, for which they will be graded by their teachers.[16] During the following semesters, other groups of students can also do research using the detectors at their school.

Once a year, the HiSPARC project organizes a national symposium. Students present their research and the group giving the best presentation receives an award. Furthermore, students take part in hands-on analysis sessions.

Students participating in HiSPARC obtain a clearer picture of what actual scientific research involves and are more interested in pursuing a scientific career. The decision to take part in HiSPARC is, in the majority of cases, made by the teachers, not their students. This rules out the possibility that the students were more inclined towards research beforehand and joined the HiSPARC project because of that.

HiSPARC currently consists of close to a hundred stations concentrated around scientific institutes in major cities in the Netherlands. The project has stations in Denmark, the UK and even Vietnam. HiSPARC also deploys weather stations and lightning detectors. Unique to HiSPARC, cosmic ray and particle physics have been integrated in the curriculum of participating schools. There are NiNa[17] modules for cosmic ray physics [67], as well as a series of topical letters for students and teachers (RouteNet) [68].

[15] Last activity in 2007.
[16] Final-year students are required to perform research which will be graded as part of their final exams.
[17] Nieuwe Natuurkunde

Figure 1.9 – High-school students and their teacher with one of the HiSPARC detectors, at the Bonhoeffer College, in Castricum. Photo courtesy of B. van Eijk.

Recently, revised teaching materials have been certified as an approved module for the NLT (Nature, Life and Science)[18] secondary education subject. All schools in the Netherlands can now teach cosmic ray physics as part of their examination program [69].

The Dutch Foundation for Fundamental Research on Matter (FOM) has funded a program to enable teachers to conduct research at a scientific institute or university. Teachers perform research during one year for one day a week, in close collaboration with scientific staff. The HiSPARC project has been a popular choice for teachers and has worked with 6 to 7 teachers each year for the past four years [70, 71, 72]. Research projects include analyzing HiSPARC data to determine the effect of atmospheric variables and the evolution of the detector response over time. Teachers have also studied the feasibility of applying pixel detectors (MPPCs) as alternatives to PMTs.

The scientific premise underlying sparse but large detector arrays is the detection

[18] Natuur, Leven en Technologie

of large-scale correlated effects, like the Gerasimova-Zatsepin effect, short bursts of showers, or other, as yet unknown, phenomena. Deploying cosmic ray detectors at high schools naturally creates a large and sparse array with detectors situated at locations with interested and knowledgeable people willing to maintain them. As a spin-off, electronics developed for HiSPARC are now applied in the Auger Radio experiment [60] and EPR spectroscopy [73].

2

The HiSPARC Experiment

Propagating charged particles lose energy in matter. In scintillators, the energy loss can be observed in the form of light. HiSPARC employs scintillator detectors to detect EAS. In this chapter, the experimental setup is discussed in detail. In Section 2.1 the design criteria are presented. Section 2.2 discusses detector physics and the geometry of the HiSPARC station. Section 2.3 describes the signal characteristic, the trigger, and the data acquisition system. Section 2.4 discusses some features of the pulseheight spectrum. Section 2.5 discusses the timing between stations. Finally, in Section 2.6 an overview of the HiSPARC locations of stations in the Netherlands is given.

In HiSPARC the distances between clusters of HiSPARC stations ranges from tens of kilometers to hundreds of kilometers. HiSPARC is therefore ideally suited to study long-range correlations between cosmic ray showers.

2.1 Design Criteria

The purpose of the HiSPARC experiment is to detect air showers and reconstruct shower direction and energy. However, due to the collaboration with high schools, resources are limited. The majority of detectors is financed by high schools, while students are responsible for assembly and subsequent installation on the roof of their high school. Therefore, detectors should be cheap, robust and easily maintainable (e.g. scintillators). The geometry of the detector network is constrained by the geographical location of each

high school.

2.1.1 Extensive Air Showers

High energy cosmic ray particles generate cascades in the atmosphere. The number of particles in an EAS depends on the energy and nature of the primary particle, as well as the first series of particle interactions. The particle density in an EAS falls steeply with increasing distance to the center of the shower (core distance). The *lateral distribution* for a few particle types in an 1 PeV proton EAS is depicted in Figure 1.7.

For charged particles, the efficiency of scintillator detectors is close to 100 %. The probability that exactly k charged particles are detected is given by the *Poisson distribution*

$$P_k(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{2.1}$$

with λ the *expected* number of detected particles. The probability of detecting *zero* particles then becomes

$$P_0(\lambda) = e^{-\lambda}. \tag{2.2}$$

The probability of detecting at least one particle is then given by

$$P_p(\lambda) = 1 - P_0(\lambda). \tag{2.3}$$

For a scintillator surface area of 0.5 m^2, $\lambda = 0.5\rho$, with ρ the particle density. The particle density at which a single scintillator has a 50 % detection probability is then calculated to be 1.39 m^{-2}.

A single scintillator can not distinguish between charged particles that are part of a shower, or stray particles not correlated to any shower. Low-energy primary particles are much more abundant than high-energy particles. This results in a large number of showers of which only a few particles reach the ground. These spurious charged particles, mainly muons, form the background (see Section 1.6.1).

However, observation of coincident signals in two scintillators that are a few meters apart, decreases the probability that the particles are uncorrelated. Coincidences are predominantly due to EAS. For two detectors, the probability of detecting a particle in both (given a particle density ρ) is $P_p(\rho)^2$. The particle density at which the combination of two detectors has a 50 % probability of recording at least one particle in each detector becomes 2.46 m^{-2}.

Figure 2.1 shows the lateral distribution function (LDF) summed over electrons and positrons in proton-induced EAS. Primary energies ranging from 10^{14} eV to 10^{18} eV are

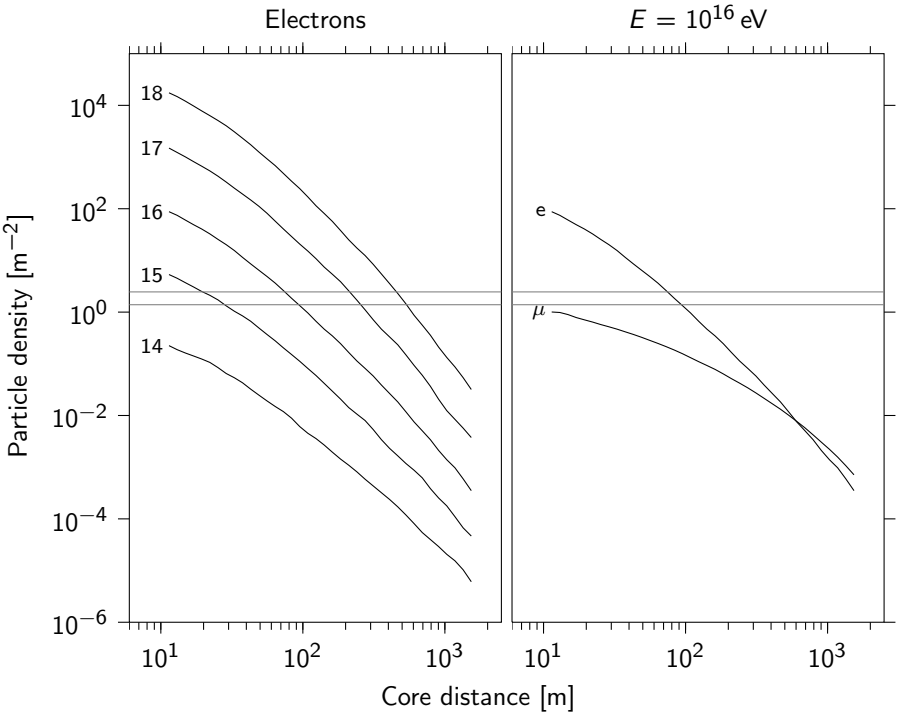


Figure 2.1 – Lateral distribution functions (LDFs) for proton-induced EAS. The LDF is summed over electrons and positrons for primary energies ranging from 10^{14} eV to 10^{18} eV (left). The two horizontal lines show the particle densities of 1.39 m^{-2} and 2.46 m^{-2}, i.e. the 50 % detection probabilities for one and two detectors, respectively. For EAS with primary energies of 10^{14} eV, the particle densities are too low to reach a detection probability of 50 % for any core distance. A two detector setup can only measure EAS of 10^{15} eV up to 20 m with probabilities higher than 50 %. EAS of 10^{17} eV, on the other hand, can be detected at distances up to 200 m. The LDF for electrons ($e^- + e^+$) and muons ($\mu^- + \mu^+$) is shown for a primary energy of 10^{16} eV (right). The muon distribution is much flatter. At 600 m, the densities are equal. The muon particle density does not contribute significantly to the charged particle density for core distances smaller than a few hundred meters.

depicted. For EAS with primary energies of 10^{14} eV, the particle densities are too low to reach a detection probability of 50 % for any core distance. A two-detector setup will only measure EAS of 10^{15} eV up to 20 m with probabilities higher than 50 %. EAS of 10^{17} eV, on the other hand, can be detected at distances up to 200 m. On the right hand side of the figure the LDF for electrons and positrons, and muons is shown for a primary energy of 10^{16} eV. The muon distribution is much flatter. At 600 m, the densities are equal. The muon particle density does not contribute significantly to the charged particle density

for core distances smaller than a few hundred meters.

The particle densities increase linearly with increasing primary energy. By determining the shower size, i.e. the number of particles in the shower, one can estimate the primary energy. Determination of the core position is essential, since the shower size can only be measured by sampling the particle density at multiple core distances.

A greater distance between stations means that the effective detection area increases. However, the detection *threshold* also increases. The result of this is that for a given energy, assuming that this energy can still be measured at the greater distance, more EAS will be observed. The typical distances between high schools in cities are of the order of 500 m to 1500 m. At the Science Park, the distance between stations is reduced to only 100 m to 240 m. With increasing primary energy, the number of EAS decreases as a power law (Figure 1.2). This implies that most of the observed showers have an energy approximately equal to the minimum detection energy threshold. Figure 2.2 shows the primary energy of proton showers that can be detected with a 50 % probability, as a function of the core distance. EAS with a primary energy of 2×10^{16} eV can be detected at a core distance of 100 m with a probability of 50 %. Two stations separated by twice this distance, i.e. 200 m, can observe the same shower with a probability $P_2 = P_{\text{station}}^2 = 0.5^2 = 0.25$, with P_2 the probability to detect an EAS with two stations in coincidence. However, the shower should then have a core position exactly between the two stations.

2.1.2 Shower Front

In an EAS, almost all particles are relativistic. Figure 2.3 shows the arrival time distribution of leptons reaching ground level, for a simulated vertical 1 PeV proton-induced shower.

While interactions introduce lateral velocity components, they are small with respect to the longitudinal velocity. As a result, the bulk of the particles travel in a thin disk, the *shower front*. The shower front can be approximated by a plane to simplify the reconstruction of shower orientation. However, this is not entirely accurate. The distribution of the arrival times is skewed; its tail may become very long. At a distance of 100 m from the shower core, 50 % of the leptons arrive within 15 ns of the time of arrival of the first particle that reaches ground level. The shower front is better described by a cone with a certain thickness.

The arrival time difference at detectors in the footprint is related to the inclination of the shower. In only one spatial dimension, the arrival time difference Δt for two

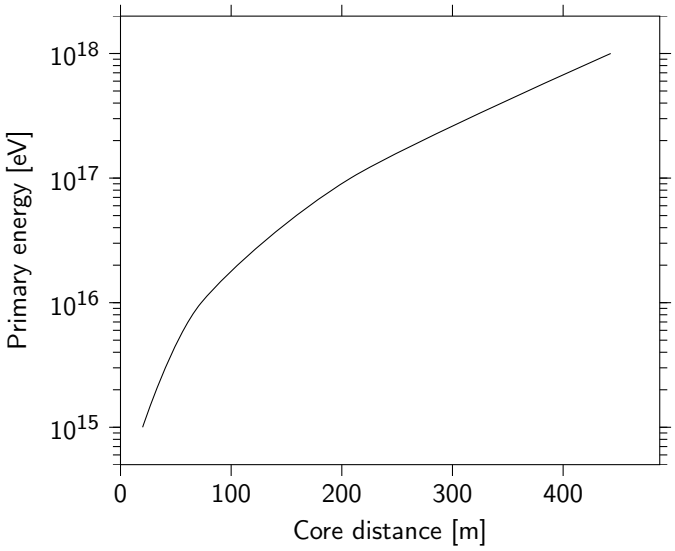


Figure 2.2 – Primary energy of proton showers that can be detected with a 50 % probability, as a function of the core distance.

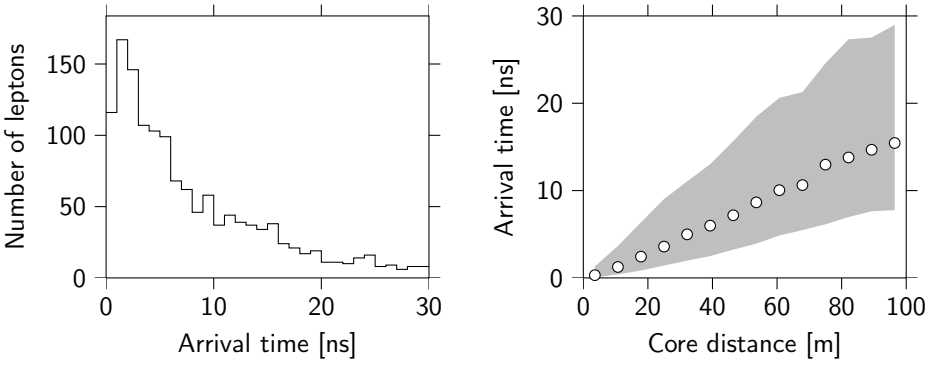


Figure 2.3 – Time structure of the shower front. Arrival times of leptons in a typical 1 PeV proton-induced vertical shower. The shower was simulated using AIRES, with default parameters but thinning disabled. In the left figure, a histogram of the arrival times is given for particles that arrive with a core distance (40 ± 2) m. The arrival time of the first particle reaching the ground is taken to be 0 ns. The most probable value of the arrival time is 1 ns to 2 ns. However, there is a long tail extending past 30 ns. The right figure shows the arrival time distribution as a function of core distance. The open circles show the median value of the arrival time. The shaded region contains 50 % of the particle arrival times, split evenly below and above the median value. At a distance of 100 m from the shower core, 50 % of the electrons arrive within 15 ns of the arrival time of the first particle to reach ground level.

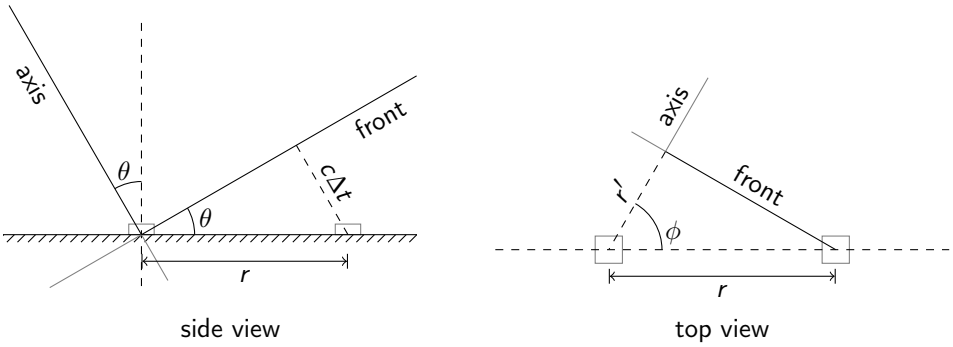


Figure 2.4 – The shower front has different arrival times in detectors separated by a distance r. The arrival time differences depend on the shower direction. The figure on the left shows a side view. The front of non-vertical showers will reach the two detectors with a time difference Δt, depending on the zenith angle θ. The figure on the right shows a top view. When a shower arrives from an azimuthal direction ϕ other than zero, the distance r is replaced by a distance r'.

stations separated by a distance r is given by (Figure 2.4):

$$\Delta t = \frac{r \sin\theta}{c}. \tag{2.4}$$

With $r = 100\,\text{m}$ and $\theta = 22°$, one obtains $\Delta t = 125\,\text{ns}$. In the two-dimensional case when the shower hits the stations more from the side, the correct equation is given by

$$\Delta t = \frac{r' \sin\theta}{c} = \frac{r \cos\phi \sin\theta}{c}, \tag{2.5}$$

and Δt becomes smaller. The timing accuracy of the experiment is crucial and an essential prerequisite for the design of the (fast) electronics.

2.2 Overview of the Experiment

2.2.1 Energy Loss in Matter

Particles traversing any material lose energy. The energy loss of high-energy charged particles is mainly due to collisions (ionization) and radiation (Bremsstrahlung). The processes are inherently statistical in nature. For example, ionization losses (due to collisions with atoms in the medium) are very high when atoms are hit centrally, and low when particles only graze the atoms. If many collisions are considered, the statistical deviations level out and a mean energy loss, or *stopping power* can be calculated.

The stopping power is the characteristic energy loss of a material. It is expressed in $\mathrm{MeV\,g^{-1}\,cm^2}$, which is the amount of energy loss in MeV if the particle traversed a column density of $1\,\mathrm{g\,cm^{-2}}$.

The first (classical) calculations of the energy loss were performed by Bohr. A much more accurate quantum-mechanical treatment was done by Bethe, with additions by many people, in particular Bloch, resulting in the *Bethe-Bloch formula* [74, p. 286]:

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{2.6}$$

with $-\langle \mathrm{d}E/\mathrm{d}x \rangle$ the stopping power in $\mathrm{MeV\,g^{-1}\,cm^2}$. That is, E is the energy of the incoming particle and x is the thickness of the impact material, in $\mathrm{g\,cm^{-2}}$. The other parameters are: z the charge of the incoming particle in units of e, Z the atomic number of the absorber, A the atomic mass of the absorber, $\beta = v/c$ the speed of the incoming particle relative to c, the electron mass m_e, the Lorentz factor $\gamma = (1-\beta^2)^{-1/2}$, I the mean excitation energy of the medium in eV, and $\delta(\beta\gamma)$ the so-called *density correction*. The constant K is given by

$$K = 4\pi N_A r_e^2 m_e c^2, \tag{2.7}$$

with N_A Avogadro's number and r_e the classical electron radius. The maximum kinetic energy transfer in one collision, $T_{\max}$, is given by

$$T_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2}, \tag{2.8}$$

with M the mass of the incoming particle.

At intermediate energies, the Bethe-Bloch formula is accurate to a few %. The energy loss in this region is by approximation only a function of $\beta\gamma$. At higher energies radiative losses come into play. At these energies, Bremsstrahlung starts to play a dominant role and the stopping power will increase steeply. For heavier particles, like muons, Bremsstrahlung only becomes dominant at very high energies.

Figure 2.5 shows the stopping power of *polyvinyltoluene*-based scintillators for electrons and muons as a function of $\beta\gamma$. The energy loss of low-energy particles is high, but falls steeply to a minimum (1 MeV for electrons and 325 MeV for muons). At these energies, the particles are called *minimum ionizing particles* (MIP). Above this energy, the stopping power increases slowly until radiative losses come into play. Note that the ionization part of the stopping power continues to increase slowly (Equation 2.6). Radiative losses are not important for scintillators, because photons from Bremsstrahlung

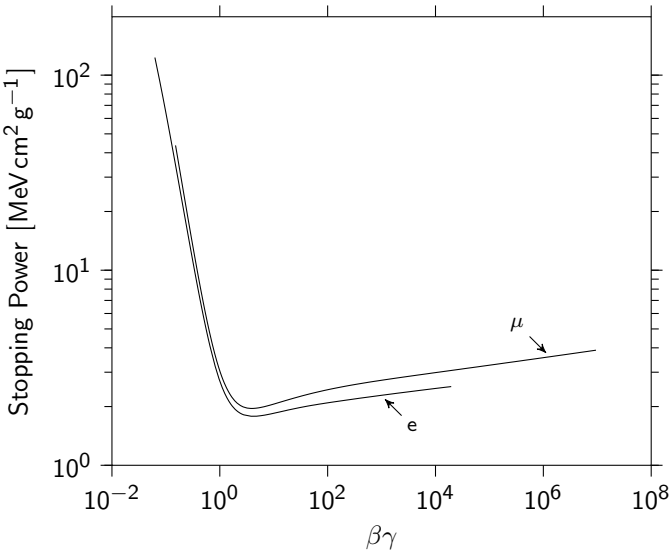


Figure 2.5 – Stopping power for muons and electrons in the plastic scintillator polyvinyltoluene. Note that only ionization losses are included, since they cause the scintillation. Data from [75, 76].

do not ionize the medium and the spectrum does not overlap with that of the visible scintillation light.

While the Bethe-Bloch equation describes the *mean* energy loss of particles, it does not describe the fluctuations in the energy loss. These can be quite large. This process is called *energy straggling*. For thin absorbers like plastic scintillator plates, this distribution was first calculated by Landau [77]. This *Landau distribution* is usually expressed in terms of the Landau probability density function $\phi(\lambda)$ which is independent of physical parameters. The energy loss distribution is given by [78]:

$$f(\Delta) = \frac{1}{\xi}\phi(\lambda), \tag{2.9}$$

with $f(\Delta)$ the relative probability of the energy loss Δ, which is related to the parameter λ by

$$\lambda = \frac{1}{\xi}[\Delta - \xi(\ln\xi - \ln\epsilon + 1 - C)], \tag{2.10}$$

with C the *Euler-Mascheroni constant*, and

$$\ln\epsilon = \ln\frac{(1-\beta^2)I^2}{2mc^2\beta^2} + \beta^2. \tag{2.11}$$

The parameter ξ depends on the medium and its thickness, and can be calculated using

$$\xi = \frac{K}{2}\frac{Z}{A}\left(\frac{z}{\beta}\right)^2 x, \tag{2.12}$$

with the same parameters as in Equation 2.6. Finally, an integral representation of $\phi(\lambda)$ is given by

$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \exp(-u\ln u - u\lambda)\sin(\pi u)\,\mathrm{d}u. \tag{2.13}$$

There are a few important assumptions made in the derivation of these equations:

- $T_{\max} \to \infty$, i.e. the maximum kinetic energy transfer in a single collision is taken to be infinite
- the energy transfer in a single collision is much larger than the binding energy of the electrons in the target material, i.e. the electrons are treated as *free particles*
- the total energy loss of the incoming particle is much smaller than the initial energy

Calculating the Landau distribution is very time-consuming. For strategies to compute $\phi(\lambda)$, see [79] and, more recently, [80].

2.2.2 Scintillator Detectors

Scintillators are cheap and highly efficient. Scintillation is the process where ionizing particles losing energy in a material cause luminescence. In other words, when a particle traverses a scintillator, it emits light. A scintillator has to be transparent to its own scintillation light. The most commonly applied scintillators are *plastic* scintillators. They are solutions of organic scintillators (aromatic hydrocarbons, the *fluors*) in a solid plastic solvent (the *base*). A very common plastic scintillator is the solvent polyvinyltoluene (the base) containing the solute *anthracene* (the fluor).

The density of the fluor is generally so low that particles traversing the scintillator only excite the base, i.e. the energy is only absorbed by the plastic. The base emits scintillation light of its own, but generally with low yields. It is also not transparent for its own scintillation light. Therefore, the addition of a fluor is required. The energy transfer of the base to the fluor is still not clearly understood. One mechanism appears to be that the fluor absorbs the scintillation light from the base, whereas non-radiative transfer also plays an important role. The scintillation of the fluor is caused by free valence electrons which occupy the so-called *π-molecular orbitals*. These electrons are

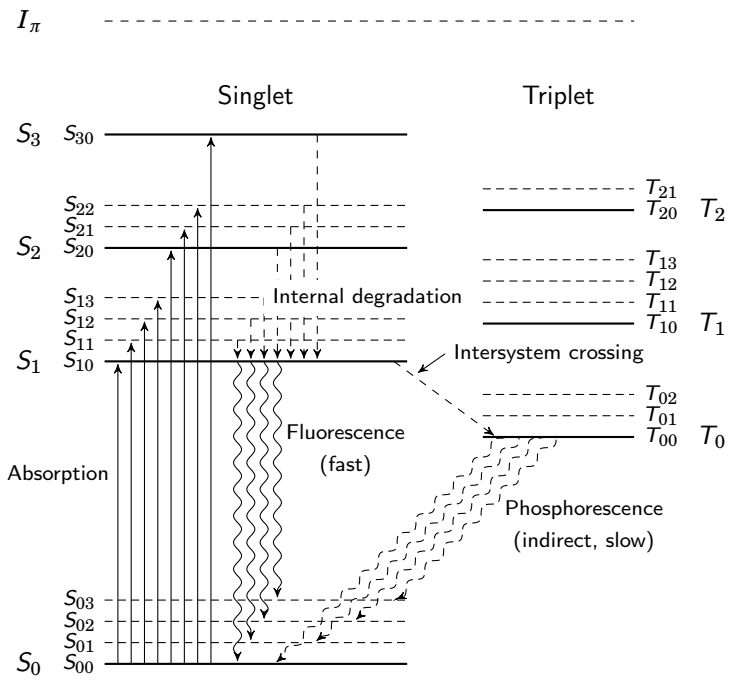


Figure 2.6 – Typical energy levels for π-orbitals of a fluor molecule. Spin singlet (S) and triplet (T) states are separated for clarity. The ionization level I_π is shown at the top. Excited states as well as vibrational sublevels (dashed horizontal lines) are shown. Internal degradation is a non-radiative process, while fluorescence and phosphorescence are radiative decays. The decay $T_0 \rightarrow S_0$, however, is indirect, by interactions with other molecules.

not associated with single atoms, but rather with the molecule as a whole. Typical energy levels are shown in Figure 2.6. After excitation, the molecule quickly decays to the S_1 level, without emitting radiation. This is called *internal degradation*. The subsequent decay of the S_1 level to the ground state and its vibrational levels occurs by emitting photons. This is called *fluorescence*. Photons from the decay of S_1 to one of the vibrational levels of the ground state ($S_{01}, S_{02}, \ldots$) will not excite other fluor molecules which are in the ground state S_0. Thus, the scintillator is transparent to these photons, which is important for the collection of the scintillation light.

Excited triplet states decay quickly to the lowest triplet state T_0. Decay of $T_0 \rightarrow S_0$ is highly forbidden by spin-selection rules, so decay usually happens through interaction with other molecules:

$$T_0 + T_0' \rightarrow S_0 + S_1 + \text{phonons}. \tag{2.14}$$

The newly created S_1 state can then decay to the ground state by emitting photons.

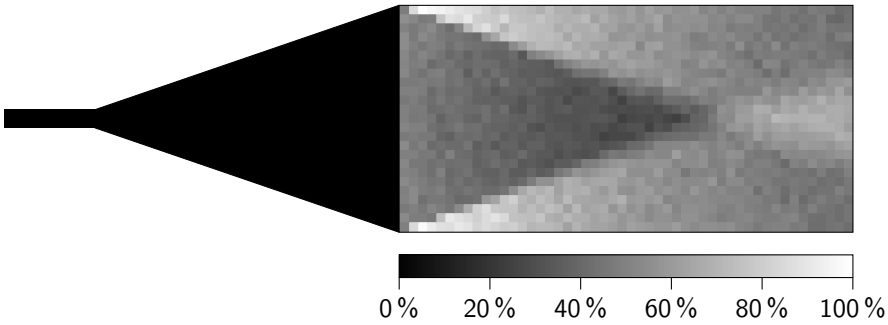


Figure 2.7 – The simulated transmission as a function of position. The percentages are relative to the maximum transmission efficiency of 2.3 %. The fishtail lightguide is located at left side of the scintillator. Attached to that is the PMT. Figure redrawn from [84].

Because this process happens rather slowly, this is called the *slow component* of the scintillator light. The decay of the excited singlet states is called the *fast component*. The fast component of the scintillator employed by HiSPARC has a decay time of 2.1 ns [81], and the slow component has a decay time of 14.2 ns [82]. The wavelength of maximum emission is 425 nm.

HiSPARC uses a rectangular 1 m × 0.5 m scintillator. A triangular lightguide is glued to the short edge of the scintillator. The scintillation light is detected by a photomultiplier tube (PMT) attached to the lightguide. The peak quantum efficiency of the PMT is 28 % at 375 nm [83]. At 425 nm, the quantum efficiency is 25 %.

2.2.3 Transmission of Scintillation Light

The transmission of scintillation light to the PMT of HiSPARC detectors has been studied [84] using Monte Carlo techniques. Over the complete scintillator area random positions have been generated. For each position, scintillation photons are emitted in random directions. Each photon is tracked until it leaves the scintillator (angle of incidence when a photon hits a boundary surface is smaller than the critical angle) or reaches the PMT. Figure 2.7 shows a two-dimensional histogram of the results. The transmission is shown for positions where a charged particle enters the scintillator. A fishtail lightguide is attached to the scintillator. The PMT is attached to this lightguide. The transmission depends strongly on the position of the origin of the scintillation light. The maximum efficiency occurs in the corners of the scintillator adjacent to the lightguide, and has a mean value of 2.3 % for one bin.

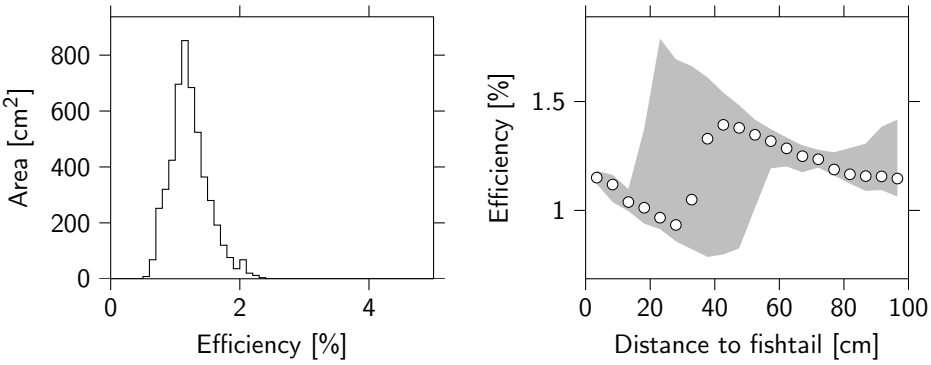


Figure 2.8 – Histogram of the area of the scintillator with certain transmission factor (left). Mean transmission as a factor of 'horizontal' distance to the read-out side of the scintillator (right). For each distance, the transmission is averaged over the width of the scintillator. Figure drawn with data from [84].

In Figure 2.8 the area of the scintillator with certain values of the transmission is shown on the left. Indeed, the bulk of the scintillator has very similar transmission characteristics. As a function of distance to the lightguide, shown on the right, it is clear that from 30 cm to 40 cm of the scintillator the transmission changes rapidly. Ref. [84] suggests replacing the fishtail with a rectangular lightguide with dimensions 50 cm × 30 cm. This lightguide basically extends the scintillator with a length of 30 cm of non-scintillating material, eliminating the feature of Figure 2.8. The scintillator itself then has an approximately uniform transmission efficiency.

The simulation has been experimentally verified in [85]. In the experiment, a trigger is constructed between a single HiSPARC detector and a probe. The probe is a 1.5 cm × 1.5 cm scintillator connected to a small PMT. The probe is placed on top of the scintillator at positions in a 3 × 5 grid. Four additional measurements are performed in a small rectangle around one of the grid points, where the simulation predicts a large gradient in the transmission efficiency (Figure 2.9).

Charged particles traversing both scintillators result in a trigger. In this way, the light yield for the known positions in the scintillator can be measured. The yield is defined as the most probable value of the signal pulseheight for charged particles (MIP peak). Experimental measurements are compared to the simulated transmission efficiency (Figure 2.10). The simulation does not include the PMT. The transmission of scintillation light must be convolved with the PMT's quantum efficiency and response function. Furthermore, when a photon exits the scintillator it is removed from the simulation. A HiSPARC detector is wrapped in aluminum foil to reflect photons back into

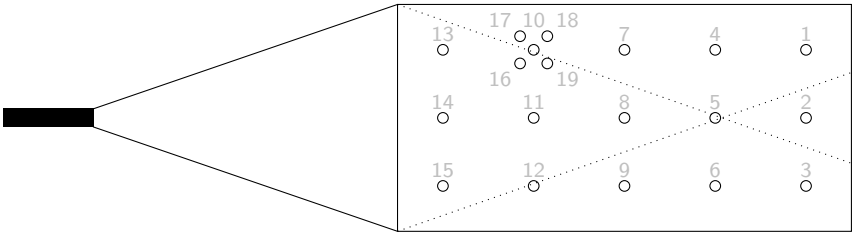


Figure 2.9 – Experimental setup for measuring the scintillator transmission efficiency. The dotted lines show the position of the largest gradients in the simulated transmission (Figure 2.7). The circles indicate the positions at which the efficiency is measured. Fifteen positions are defined on a grid, with four additional measurements performed in an area along the large transmission gradient.

the scintillator. The y-intercept in Figure 2.10 implies that photons will be detected in the experiment, even when the simulation predicts that no photons will reach the PMT. The use of the aluminum foil may explain this discrepancy [85].

2.2.4 Detectors and Stations

The current revision of the electronics is able to operate in a *master* / *slave* combination, which allows two electronics units to operate four detectors at a time. The use of four detectors, with a suitable trigger condition, improves sampling of a shower and allows to even measure the direction of the shower with a single station. This setup is favored for remote schools[1] and is used throughout the *Amsterdam Science Park Array*. Figure 2.11 shows the layout and dimensions of such a station. The station is an equilateral triangle with sides of 10 m. That size was chosen as a compromise between timing accuracy (larger is better) and ability to fit the station on a typical roof (smaller is better). Each detector is placed in a roof box for protection.

2.3 Data Acquisition

2.3.1 ADC

The PMTs of the scintillators of a HiSPARC station are connected to fast AD converters and a trigger matrix. Each PMT is connected with two cables, one for power and control voltages, the other for the (negative) analog signal. The readout unit contains four

[1] With a four-detector station the direction of the shower can be reconstructed, which provides a valuable contribution to HiSPARC.

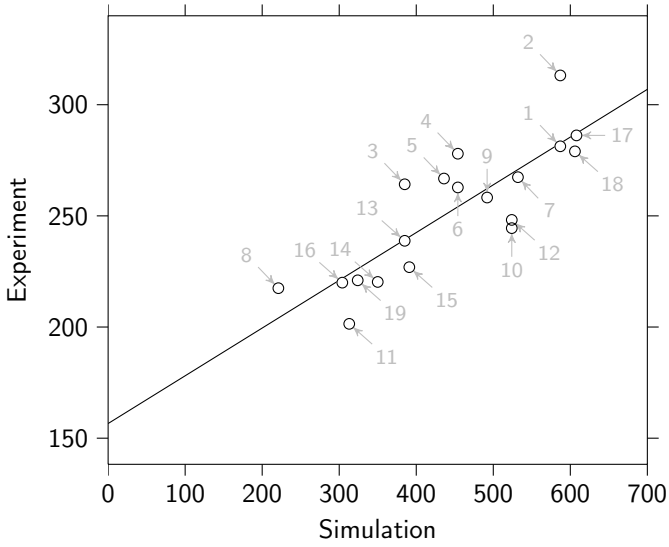


Figure 2.10 – Correlation between the light yield of a HiSPARC detector and the simulated transmission efficiency. The units are arbitrary. Each data point is a measurement at a location in a 3×5 grid on the scintillator. Four additional measurements are performed in a small rectangle around one of the grid points, where the simulation predicts a large gradient in the transmission efficiency (Figure 2.9). The y-intercept is explained by the use of aluminum foil which reflects lost photons back into the scintillator. This has not been taken into account by the simulation. Redrawn from [85].

ADCs which can be driven at 200 MHz, as well as a 200 MHz crystal. By using a very stable clock, the sampling frequency can be doubled by driving one ADC on the rising edge of the clock signal, and one ADC on the falling edge, for each channel. Thus, each channel uses two ADCs, and the PMT signal is sampled at 400 MHz, or $\Delta t = 2.5\,\text{ns}$. The maximum time window for a single event is 10 µs.

The use of two ADCs per channel means that they need to be carefully aligned because the output will consist of samples alternately provided by the first and the second ADC. Any difference between the baselines of the ADCs would result in a ragged signal, resembling a *triangle wave*. This alignment procedure can be carried out by the user after installation of the station and is performed by applying several different internal reference voltages on the input channel. The ADC gains and offsets can be controlled by software and are adjusted until the ADCs are both aligned and are providing a sampling range of +113 mV to −2222 mV. The ADC response is linear over this range. The converters provide a 12 bit output, providing a resolution of −0.57 mV per ADC count. A conversion of ADC counts to mV units is then given by $V = -.57x + 113$,

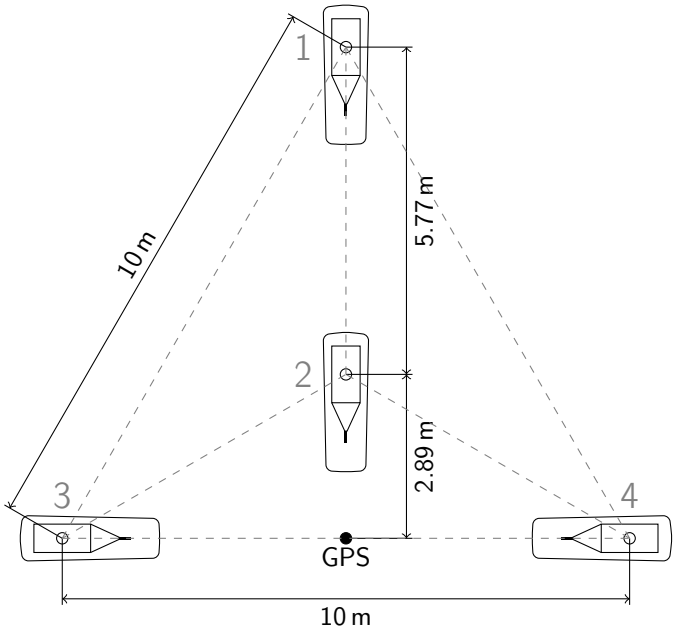


Figure 2.11 – Layout of a star-shaped four-detector HiSPARC station. The station is shaped as an equilateral triangle with sides of 10 m. Three detectors are placed on the triangle corners, with an additional detector in the geometric center of the triangle. A GPS antenna is placed in the center of the baseline.

Figure 2.12 – The four-detector HiSPARC station (Figure 2.11) at TeleCity, Amsterdam Science Park. Each detector is placed in a roof box for protection against the elements. The inset shows a close-up of the GPS antenna.

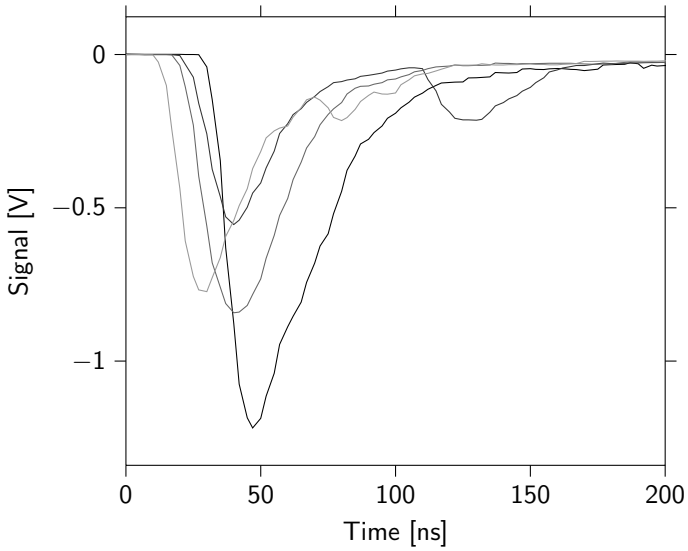


Figure 2.13 – An event measurement from a HiSPARC station, digitized by the ADCs. Note the differences in arrival time of the particles. In at least two detectors multiple particles are observed.

with x the number of ADC counts.

Figure 2.13 shows an event sampled by the ADCs. The *pulseheight* of an event is the largest negative value attained by the signal. The pulseheight is usually expressed as a positive number.

2.3.2 Trigger

For each channel, two discrimination thresholds can be set. This provides for an elaborate trigger definition. For example, if a signal goes over the *high threshold*, there is a high probability that the signal was generated by a particle in the detector. Such a signal in at least two of the four detectors is sufficient to identify an EAS. However, if this condition is not satisfied, but at least three detectors *do* detect a signal above the *low threshold*, there is a high probability that these are correlated and a trigger is generated accordingly. In practice it turns out, however, that the first trigger condition is sufficient and that a trigger is seldom generated on at least three low signals. Figure 2.14 shows the two discriminator levels in a pulseheight histogram of the HiSPARC signals. The edge at 70 mV is due to the trigger. The trigger matrix is defined in the readout unit's FPGA (Appendix A).

When a trigger is issued, the ADC output is stored in a buffer in the FPGA. The

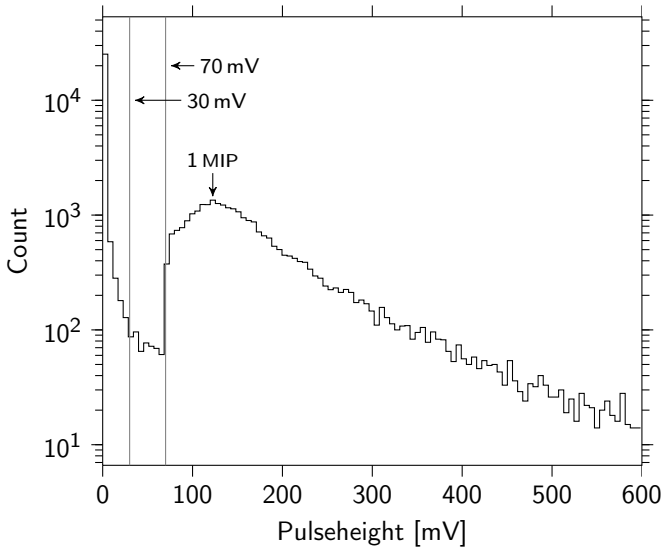


Figure 2.14 – The 30 mV and 70 mV discriminator levels for a four-detector HiSPARC station. The edge at 70 mV is due to the trigger. The peak to the right of both levels is the MIP peak.

buffer captures both pre-trigger and post-trigger signals, up to a total of 10 µs per event (Figure 2.15). The buffer size is sufficient to hold 30 µs of data. The default measurement time for HiSPARC is 6 µs: the pre-trigger is 1 µs, the coincidence window is 1.5 µs and the post-trigger is 3.5 µs. In this configuration, the buffer can hold 5 events. The buffer is capable of continuously storing event data, eliminating dead time. Events do not overlap. Particles are either observed within the previous event (if observed before the end of the post-trigger window) or are part of a new event. Figure 2.16 shows the time between triggers for a large set of events. There are no triggers for intervals shorter than 6 µs, since events do not overlap. For intervals longer than 6 µs, events are observed. No dead time is observed.

2.3.3 GPS Timestamp

The HiSPARC electronics contains a dedicated high accuracy GPS board, which provides an accurate timestamp to the data [86]. The GPS provides several modes of operation. In *self-survey mode*, the antenna's position is determined by averaging over many *fixes*[2]. For each fix, the GPS operates in *full position mode*. The signals from the satellites are used to calculate both the current time and the position of the antenna. Once

[2] In 24 h, 86 400 fixes are collected.

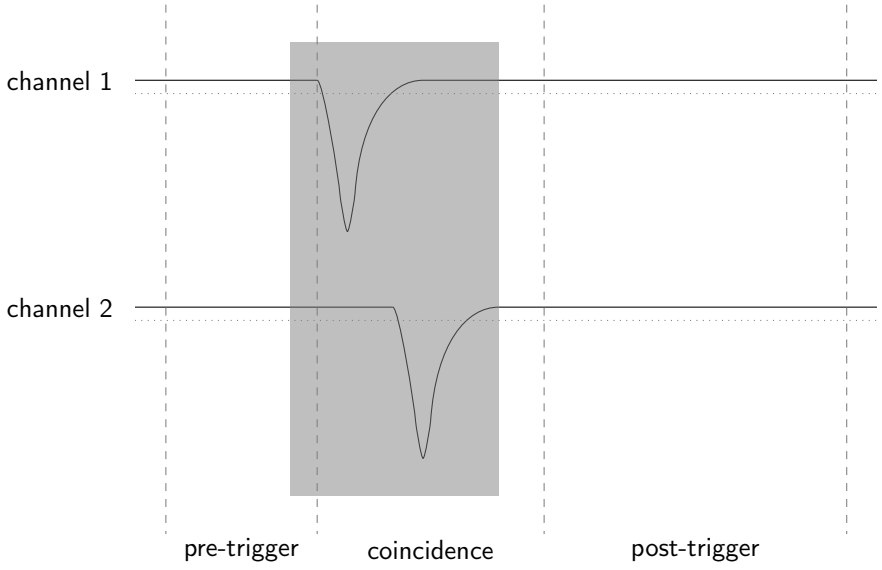


Figure 2.15 – DAQ time windows. A schematic representation of an event. Dashed vertical lines: the pre-trigger (1 µs), coincidence (1.5 µs) and post-trigger (3.5 µs) windows. Dotted horizontal line: the data reduction threshold (20 ADC). Shaded area: the data-reduction window. Data outside this window will not be stored.

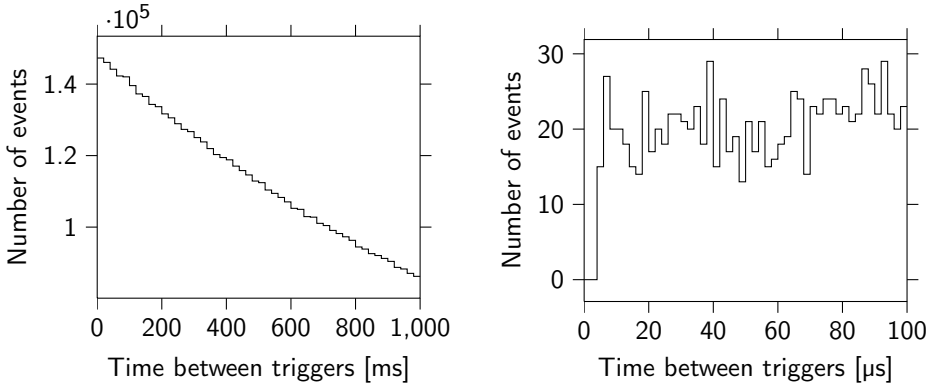


Figure 2.16 – Time between triggers. The recorded length of an event is 6 µs. There are no triggers for intervals shorter than this value, since events do not overlap. For all intervals longer than 6 µs, events are observed. Since the GPS timestamp is determined when the trigger condition is fulfilled, and not at the start of the coincidence window, it is possible to observe events with a time difference less than 6 µs. Indeed, 15 events occur in the 4 µs to 6 µs interval. There is no significant dead time.

the position is accurately known, at the end of the self-survey, the GPS switches to *overdetermined clock mode*. Since the position is static, the equations can be solved for current time only. Using many satellites, the system of equations is overdetermined and this results in a more accurate measurement of the GPS time.

The HiSPARC 200 MHz clock is disciplined with the GPS pulse-per-second (PPS) signal. The PPS signal and the clock tick counts are combined into a trigger time which is determined at 5 ns intervals. By inspecting the PMT signal the trigger time is adjusted to 2.5 ns intervals. The trigger time is then corrected by taking the GPS-provided quantization error into account, and is rounded to the nearest nanosecond.

The GPS can report time either in UTC or GPS time. Both these times are essentially equal to *international atomic time* (TAI), up to a difference in *leap seconds*. Leap seconds are introduced to better match the UTC clock to *mean solar time*, determined by the Earth's rotation. Since the rotation of the Earth is not constant in time, and is continuously slowing down, leap seconds need to be introduced to synchronize UTC to mean solar time. The introduction of leap seconds, however, means that UTC is periodically adjusted and is not strictly monotonically increasing. GPS time, however, is directly based on TAI, and thus maintains the same offset, but was designed to match UTC in 1980. Since then, GPS time and UTC have diverged. As of December 31, 2008, the difference between UTC and GPS is 15 s, GPS time being ahead of UTC. Because the GPS time is both continuous and more natural for the GPS receiver (UTC has to be calculated, after all necessary parameters are received from the GPS satellites, which can take several minutes after startup), all HiSPARC stations have their clocks set to GPS time.

The accuracy of the GPS device is important for the analysis of coincidences between multiple stations which each have their own GPS device. To test the accuracy a pulse generator was used as an external trigger source. The trigger was distributed simultaneously to stations 501 and 502, which have a separation distance of 100 m. The two cables used to connect the triggers had different lengths and the resulting time delay has been measured and corrected for. The trigger period has been set to 0.251 s, ensuring a trigger rate of approximately 4 Hz, while sampling the timing accuracy over the full subsecond range. The results of a 24 h measurement of the time differences between both stations are shown in Figure 2.17. The histogram is fitted with a normal distribution. The distribution has an offset of 18 ns and a standard deviation of 3 ns. The offset is as yet unaccounted for.

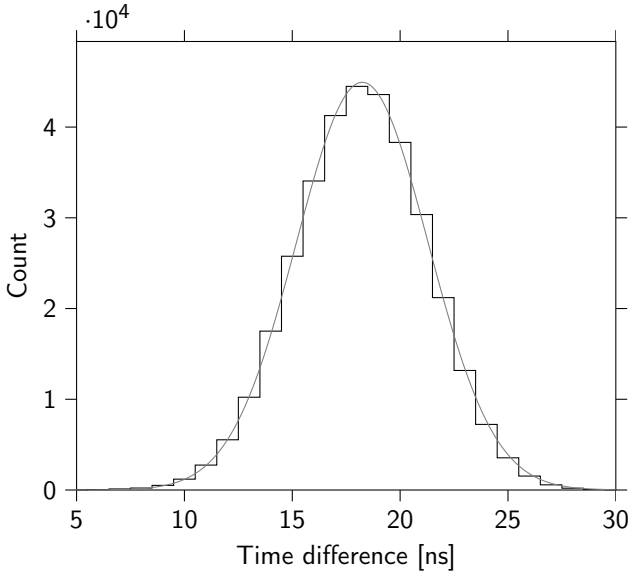


Figure 2.17 – Distribution of GPS time differences. The bin size is 1 ns, since the timestamp is reported to nanosecond accuracy. The histogram is fitted with a normal distribution with an offset of 18 ns and a standard deviation of 3 ns.

2.3.4 Data Acquisition Software

LabVIEW [87] is used to control and read out the HiSPARC electronics. It provides a graphical user interface for aligning the ADCs, changing settings like time windows, thresholds and PMT voltages, and a data display of running measurements. Upon startup, it initializes the hardware, powers on the PMTs, initializes the GPS module and starts data acquisition (Figure 2.18). The DAQ software performs a preliminary analysis. Baseline, pulseheight and pulse integral are determined. The baseline and noise level are derived from the pre-trigger window. The post-trigger signal may contain particles if the shower front is particularly thick.

Finally, the baseline is used to apply a *data reduction* algorithm. The algorithm determines the part of the signal containing the PMT pulses and removes the rest. The 6 µs signal is thus greatly reduced in length. The algorithm determines the first instant that any detector had a signal higher than 20 ADC, as well as the last instant. Only data within this time window is stored in a local event store (Figure 2.15).

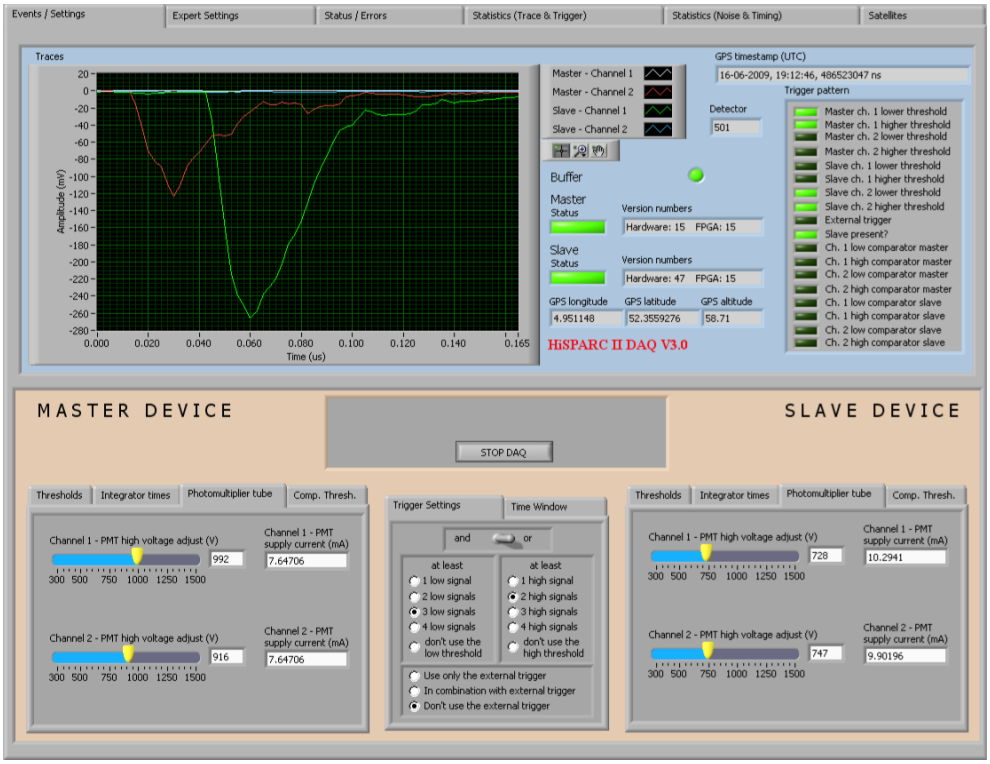


Figure 2.18 – Screenshot of the HiSPARC DAQ software. The top left part of the user interface shows the signal of the latest event. The top right shows the timestamp of the event, the trigger matrix, GPS location and hardware version information. The bottom part of the UI is dedicated to detector and trigger settings. Additional tabs (visible at the top of the screen) are available to inspect expert settings, status and error messages, and several important statistics (e.g. pulseheight histograms, trigger rates, etc.).

2.4 Pulseheight Spectrum

Figure 2.19 shows a histogram of the pulseheights of a large number of HiSPARC events. This *pulseheight spectrum* has several features. The left side of the spectrum shows a large number of events with small pulseheights, which falls steeply. This part of the spectrum is significantly above the noise level, which is measured to be only a few ADC counts. The events strongly correlate with showers, as discussed in [88]. In fact, by calculating the spectrum that would be generated by high-energy photons, [88] shows that this fits the measured spectrum well.

Photons lose energy in matter by the photo-electric effect, Compton scattering and pair production. These processes have very small cross sections, so it is hard to detect

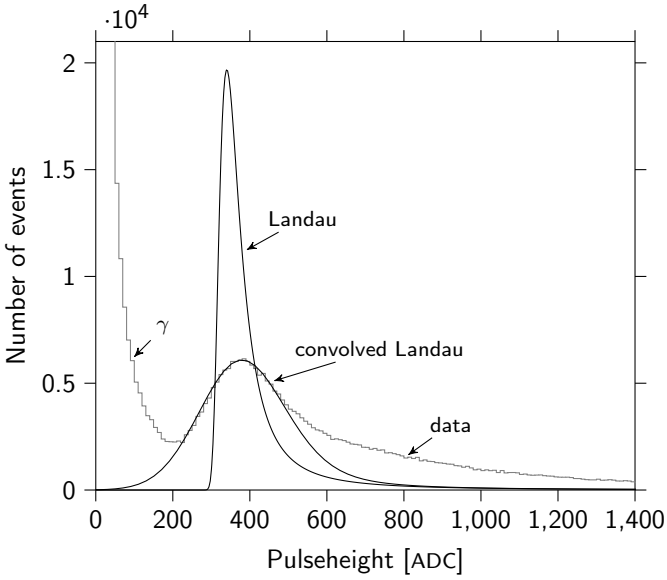


Figure 2.19 – Features of the pulseheight spectrum. The data in this figure are produced by using an external trigger to trigger the detector on showers. The data contains no artifacts resulting from triggering on the discriminator levels. The left flank in the data is due to the detection of photons, while the bump and the tail are due to the detection of charged particles. The energy loss of charged particles is described by the Landau distribution, but can only be observed as a convolution with a normal distribution describing the resolution of the detector.

photons using their energy loss in a scintillator. The photo-electric effect dominates for photon energies below 50 keV. The resulting signal is very small because so little energy is available. In fact, the signal level is below the noise level. For photons with energies in the range of 50 keV to 10 MeV the dominant interaction process is the Compton effect, in which a photon scatters off an electron and loses energy to the electron. The electron is ejected from the atom and travels through the scintillator, losing energy resulting in scintillation light, which can be detected. The amount of energy that is lost to the scintillator depends on the energy of the electron (Figure 2.5), but also on the position in the scintillator where the interaction occurred and the path the electron travels. Photons with an energy of approximately 1 MeV to 10 MeV produce a signal corresponding to about 0.15 MIP to 1 MIP.

Photons with energies above 100 MeV primarily produce electron-positron pairs when interacting with matter in the scintillator. The resulting minimum-ionizing particles travel through the scintillator. The observed signal is thus in the range of 0 MIP to 2 MIP depending on where in the scintillator the interaction occurred.

In [88] a Monte Carlo simulation is performed of photons hitting a HiSPARC scintillator. The resulting spectrum can be described by

$$N(A) = N(0)\exp(-A/A_0), \tag{2.15}$$

with $N(A)$ the number of events with pulse height A and A_0 a constant which depends on the gain of the detector. The nominal value of $A_0 \approx 80\,\text{ADC}$.

The part of the spectrum from 200 ADC to 2000 ADC is due to charged particles (electrons and muons). As discussed in Section 2.2.1 on page 36 the fluctuations in the energy loss are described by the Landau distribution. This is a relatively narrow asymmetric distribution with a long tail. Due to the resolution of the detector this distribution is only observed as a convolution with a normal distribution describing this resolution (Figure 2.19). The sum of the photon and charged particle spectra closely matches the observed spectrum up to approximately 600 ADC. Above that, the spectrum is dominated by events in which more than one charged particle hit the detector. Most charged particle events contain only one particle.

Further evidence corroborating the photon interpretation of the spectrum is given in [88] by showing a comparison between signals in different detectors of one station measuring the same event (see Figure 2.20). By comparing low pulseheight signals (mainly photons) with high pulseheight signals (mainly charged particles), it is obvious that in general, the distribution of arrival times shows a longer tail for photons than for charged particles. This observation is as yet unexplained.

2.5 Timing Between Stations

HiSPARC *stations* consist of two or four *scintillator detectors* and a *GPS antenna*. A collection of stations in the vicinity of a major city is called a *cluster*. Each cluster is centered around a scientific institute and managed by a *cluster coordinator*.

A subcluster is located at the *Amsterdam Science Park* (Figure 2.21). This subcluster has eight stations and is detecting EAS with a lower energy threshold.

An analysis of the time differences of coincidences between stations has been performed. The data contains EAS events from stations 501 and 502 during the full month of September, 2011. The time differences ($\Delta t = t_{502} - t_{501}$) are shown in Figure 2.22. A normal distribution has again been fitted to the data. The offset is −15 ns and the standard deviation is 102 ns. The width of the distribution is much larger than the distribution of GPS timing differences. This is the result of the difference in arrival

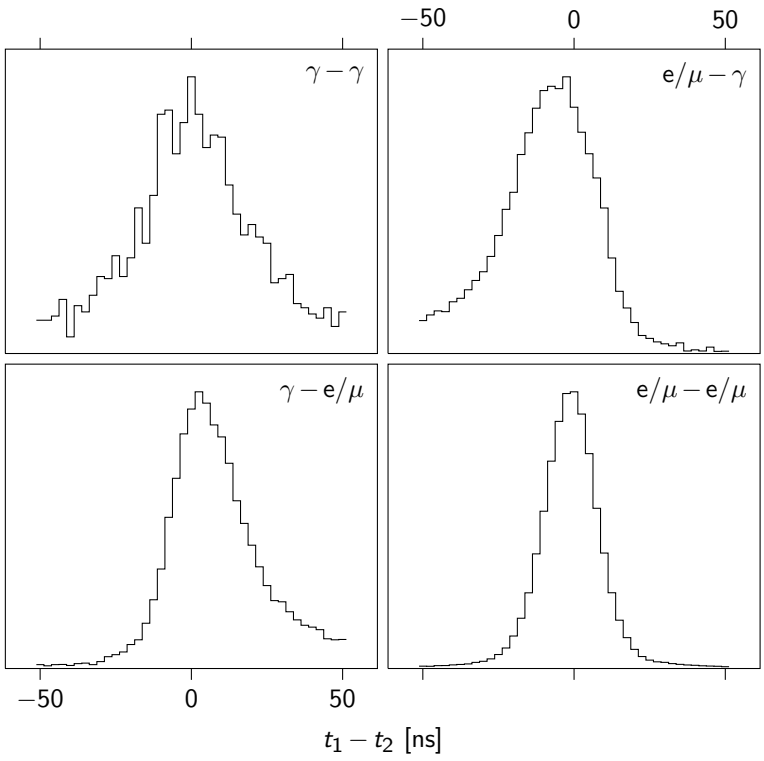


Figure 2.20 – Time spectra of coincidence events in two detectors. The time differences are given by $t_1 - t_2$ with t_1 the arrival time in detector 1 and t_2 the arrival time in detector 2. Shown are the distribution of low signals ($\gamma - \gamma$) and high signals ($e/\mu - e/\mu$) as well as the asymmetrical distributions resulting from low signals only in detector 1 or detector 2. Redrawn from [88].

time of EAS at the two stations.

The arrival time difference distributions of HiSPARC stations at the Science Park, Amsterdam have been studied [89]. In particular the relationship between station separation distance and the width of the arrival time difference distributions have been studied. The width of the distributions are expected to increase with increasing distance between stations. This results from the expectation that any pair of stations measures approximately the same distribution of shower angles. For inclined showers, the difference of arrival time of the shower front at two stations is, in the plane front approximation, linearly dependent on the distance between the stations. See also Figure 2.4. The experimental results, as well as a fit to the data, shown in Figure 2.23 have been taken from [89]. The linear function which is fitted to the data has a slope of $1\,\mathrm{ns\,m^{-1}}$, which illustrates the effect.

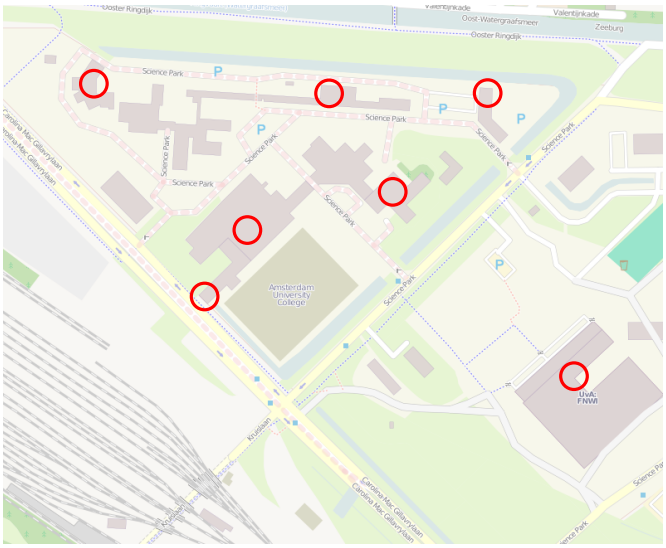


Figure 2.21 – Locations of HiSPARC stations in the Amsterdam Science Park Array. The station located inside the Nikhef lobby is not shown.

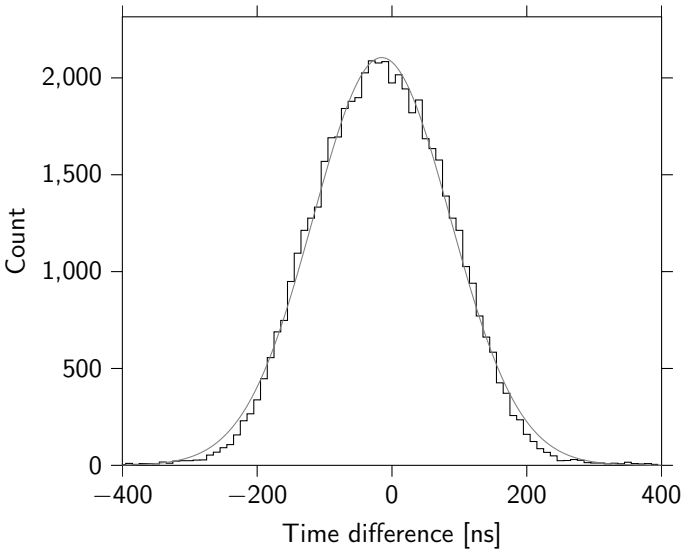


Figure 2.22 – Distribution of time differences between two stations. $\Delta t = t_{502} - t_{501}$. The histogram is fitted with a normal distribution with an offset of −15 ns and a standard deviation of 102 ns.

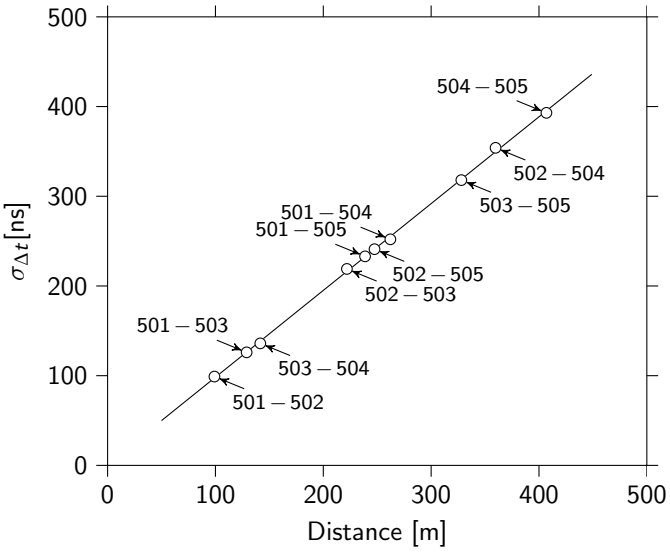


Figure 2.23 – Standard deviation of time difference distributions vs the distance between stations. The slope of the fit is $1\,\mathrm{ns\,m^{-1}}$. Figure redrawn from [89].

2.6 HiSPARC Clusters

Figure 2.24 shows the locations of HiSPARC stations in the Netherlands. Clusters cover fairly large areas. For example, the Eindhoven cluster includes Tilburg, which is 30 km from Eindhoven. When analyzing EAS, stations within Tilburg are usually analyzed separately from the stations within the city of Eindhoven: the cluster is divided into *subclusters*.

The number of active stations per cluster has increased over the years. For the period 2003 – 2011 the numbers are given in Table 2.1. Many stations have been operational for years. The project has been growing over the entire period. In 2012, a dozen additional stations are expected to come online. Over the years, the success of HiSPARC has generated interest from abroad. Overseas clusters are located in Aarhus, Denmark (since 2007), Durham, the UK (since 2011) and Hanoi, Vietnam. An additional cluster is planned in Bristol, the UK.

The use of a weather station is fully integrated in the HiSPARC software and data flow. Each participating high school can install a weather station at their location. Students can analyze correlations between cosmic rays and atmospheric conditions. About a dozen schools currently have a weather station installed. In addition, lightning detectors are also being integrated and are planned to start taking measurements in

Figure 2.24 – Locations of HiSPARC stations in the Netherlands.

Cluster	2003	2005	2007	2009	2011
Amsterdam	0	16	20	27	29
Eindhoven	0	0	0	8	18
Leiden	0	3	8	9	15
Nijmegen	5	9	11	11	11
Enschede	0	0	1	2	7
Utrecht	0	1	1	2	5
Groningen	0	4	4	4	4

Table 2.1 – The number of active stations in HiSPARC clusters, for the period of 2003 – 2011.

the second half of 2012.

As a rule, software systems do not work well until they have been used, and have failed repeatedly, in real applications.

David Parnas

3 Software Architecture

3.1 Data Management

Data from the HiSPARC experiment is stored centrally at Nikhef. The following requirements have been defined for the HiSPARC project:

1. need for a well-documented and standardized format
2. access to the *complete* dataset for all schools
3. high-bandwidth connection
4. data integrity

The first requirement is taken care of by publishing the format. The second means that there is a central location where either the data itself is stored, or metadata is stored which describes where to find the data. The HiSPARC experiment has chosen to store the data in one location. The third requirement depends on the size of a typical dataset. For large datasets, to satisfy the third requirement multiple copies of the data may exist. In such cases, (partial) copies are located as close to the researcher as possible, at a location with high-bandwidth access.[1] Methods for data access have been implemented to enable downloading of all the data. The fourth requirement is solved by backups,

[1] Including, but not limited to, the desktop or laptop of the researcher.

as well as strict integrity checks. Note that backups and integrity checks should be designed to prevent both *data loss*, as well as *duplicate* or *corrupt* data.

Storage may be divided into storage *tiers*. Storage tiers are different kind of storage which differ in at least one of four attributes: price, performance, capacity or function. For example, the bulk of the data might be stored on magnetic tape (cheap, but slow), while an array of solid state drives (fast, but expensive) contain only data currently being analyzed. The most expensive storage is usually called *tier 1*, as this contains the most critical, or recently accessed data. Lower tiers are usually cheaper, slower, but with more capacity. Three storage tiers have been implemented for HiSPARC. Tier 1 and 2 will be discussed shortly. Tier 3 is a backup storage on magnetic tape in a different geographical location from tier 1 and 2.

Data may also be divided into levels. Loosely following the definition used by NASA [90, p. 5], level 0 data is unprocessed, raw data. Level 1 data have undergone some form of processing. Data are supplemented with data like timestamps and instrument configuration. Simple processing steps which are certain not to change during the lifetime of the experiment may be performed. For example, compressed data may be decompressed and maximum signal heights (in raw units) may be recorded. In contrast, conversion of raw units to physical units is not performed, as detector calibration usually is a complex process which is likely to be done differently when the project matures. Level 2 (and higher level) data *is* converted to physical units, and may thus be subject to change.

The HiSPARC experiment consists of close to a hundred different stations in several countries, the majority of which are operated by high school teachers and support staff. The quality of internet connections varies from school to school, and connectivity depends on firewalls, proxy servers and anti-virus software, which are all implemented by the high schools at their own discretion. The data management requirements discussed previously, are much harder to satisfy as long as data remains at the station locations. It was decided to implement a system where data is offloaded from the station locations as quickly as possible and is stored in a central location, where it is easily accessed.

3.1.1 Data Flow

"Any system should be as simple as possible, but no simpler" is a variation on a quote ascribed to Einstein [91], and relevant to the design of this system. The central data storage designed for HiSPARC is called the *datastore*. The use of a central location ensures that most of the data management requirements need to be addressed only once.

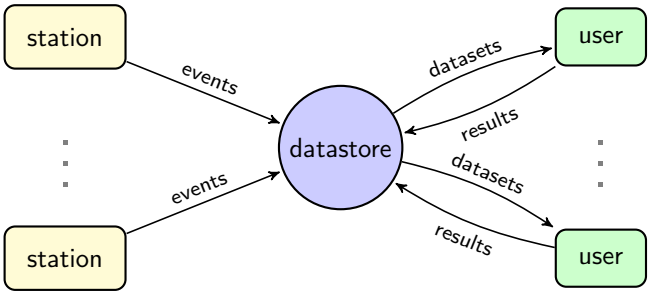


Figure 3.1 – Context-level data flow diagram (DFD) of HiSPARC. Stations pass events to the datastore, where they are accessible to users. Apart from raw event data, histogrammed data and selected datasets are also made available. Using special tools, users can even submit some results from analysis.

Figure 3.1 shows a *context-level* data flow diagram (DFD) for the HiSPARC experiment. Context-level diagrams only show the interaction between the system under consideration and all external agents, which act as data sources or data sinks. In this view, stations act as data sources which offload event data to the datastore, as quickly as possible. Teachers, students and researchers working on HiSPARC data, denoted as *users* in the diagram, act as both data sinks (they retrieve data for analysis) and data sources (results from their analysis may be stored centrally).

Of course, this is a very simplistic view. It does not specify how stations internally handle data, or which processes handle user requests. Context-level DFDs are exploded in a level 0 DFD, or *system-level* DFD, shown in Figure 3.2. The system is divided into processes and data stores. How the processes work internally is not yet shown. The *station PC* process is running on PCs connected to the detector hardware. For every station, there is such a process. The other processes are unique. There exist only single copies, and they all run on servers located at Nikhef, Amsterdam. We've strived to strike a balance between maintainability (do as much work as possible on the server, to which we have easy access), and performance (do work on the station PCs, of which there are many). The result is that raw data (level 0) is processed on the station PCs and the resulting level 1 data is sent to the datastore writer process. No further processing of the data itself is required before storage.

There are only two more processes accessing the datastore. The HiSPARC quicklook generator generates daily histograms from event data and stores them in the public database. These quicklooks can be accessed on the web, at `http://data.hisparc.nl/`. This website is served by the public database access process. This process also serves

level 1 event data to users for detailed analysis. The results from the analysis are stored in the public database.

3.1.2 Station PC

Every HiSPARC station is equipped with a dedicated PC controlling and reading out the detectors. The PC is connected to the HiSPARC electronics. The dataflow of a station PC is shown in Figure 3.3. The DAQ process communicates with the electronics unit. In the case of a master / slave combination, two raw event messages (one for each device) make up an event and three GPS one-second messages[2] are needed to calculate an accurate GPS timestamp for the event. The event data is analyzed and baseline, pulseheights and integrals are calculated and added to the event messages. These events are then stored on disk in the *buffer*, a MySQL database. The *monitor* process reads data from the buffer. The event messages are unpacked and cast in a format which can be easily parsed by the datastore processes. After processing, these events are stored on disk, in the *storage*, an SQLite database. When a certain number of events are stored[3] the data is uploaded to the datastore. If the datastore is temporarily unavailable, e.g. due to the loss of an internet connection at the school, the storage is designed to hold a substantial backlog. Once the datastore is available again the backlog is uploaded in batches.[4]

3.1.3 Tier 1: Datastore

The datastore is the central data storage created for HiSPARC. It receives data from all stations and provides data to all users. The datastore is designed to be simple, and efficient.

Stations upload data using HTTP POST requests to the datastore server. Several worker processes handle the request. First, the data's origin is verified by checking the station number and password. If the combination is correct, a checksum is calculated for the payload and compared to the checksum sent in a POST variable. If the checksum is verified, the flat data stream is converted to a data structure. If that succeeded, the data is assumed to be correct and written to a file in the `incoming` directory in the datastore's file system. Finally, a return code is sent back to the station. In the event of an error, an error code is returned.

[2] Once a second, the GPS module sends out a data message containing the current time and several data values needed for correcting quantization errors. In total, three one-second messages are needed to obtain all the necessary values for correcting the timestamp of a particular event.

[3] User-configurable, but the default value is 100 events

[4] Again user-configurable, but the default maximum batch size is 1000 events

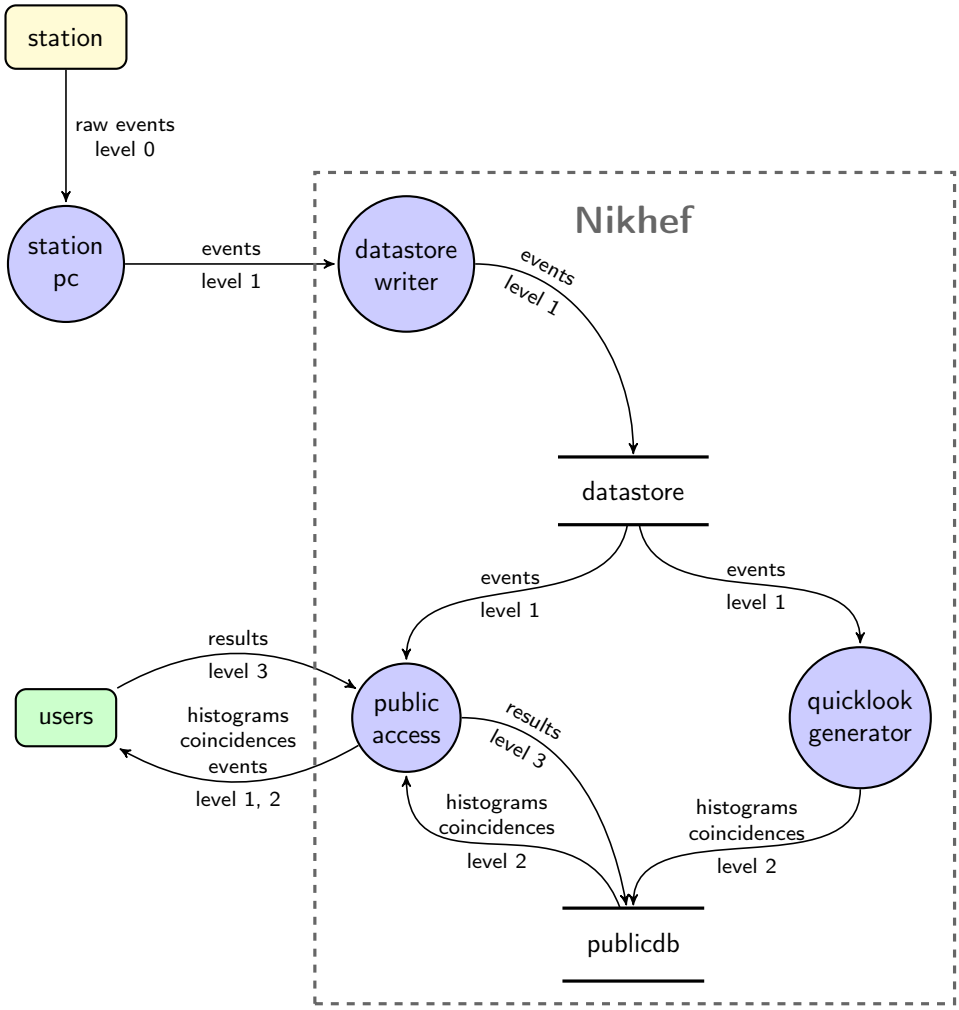


Figure 3.2 – Level 0 data flow diagram. This diagram shows some more detail. Each station has a dedicated PC running the data acquisition software. Event data are sanitized and some preliminary analysis results and configuration settings are added. The results are then sent to the datastore. The process handling public data access reads events from the datastore and stores selected events and histogrammed data in the public database. Users can then access this data.

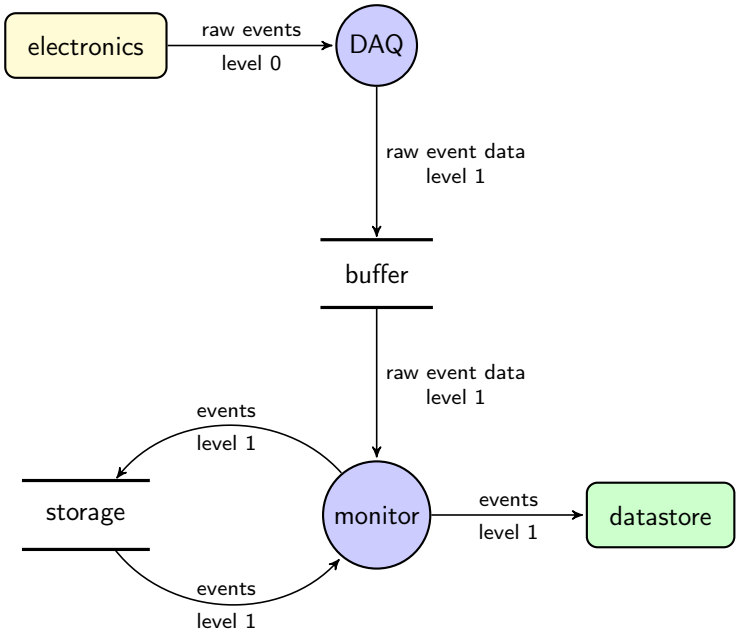


Figure 3.3 – Data flow diagram of a HiSPARC station PC. A LabVIEW program communicates with the HiSPARC II electronics. Data is sanitized and, along with preliminary analysis results and configuration settings, sent to the *buffer*. The monitor program retrieves raw event data from the buffer and creates structured array data. The events are then stored in the *storage*. When a batch of events is ready, it is retrieved from the storage and uploaded over the internet to the datastore.

A single writer process is monitoring the `incoming` directory. The use of a single process removes the need for locking or concurrent write access. Several conventional approaches to locking have proven to be unreliable on networked filesystems. The writer process opens each file sequentially and the payload is processed. The payload is a list of events with a header. The process writes the event list to disk in binary files. Data from all stations is aggregated into single files for each day. To keep the number of files per directory small, in the top-level directory there is a subdirectory for each year, containing subdirectories for each month. In these directories are the binary files. See Figure 3.4.

Binary files have several benefits: they are compact (each day's worth of data is just *one file*, instead of several files in some directory structure, which is typical for databases), they can easily be backed up (the files do not change once all data for a single day is received), and they scale linearly. Of course, there are also downsides. Most

```
2004/
2005/
...
2011/
    1/
    2/
    ...
    12/
        2011_12_1.h5
        2011_12_2.h5
        ...
        2011_12_31.h5
incoming/
    tmp048r_J
    tmpF7b3se
    tmpj22L8s
    tmpnVBoSh
    tmpZdlj_H
partial/
    tmpFG3c0C
tmp/
    tmpuE0EAM
```

Figure 3.4 – Directory structure of the datastore. The datastore contains directories for each year. Each year contains directories for each month. Finally, each month directory contains individual data files. The name of each file is simply the date with an extension. The datastore also contains directories for temporary files. The HTTPD daemon writes data files in the `tmp/` directory and moves them to the `incoming/` directory when finished. The `incoming` directory is watched by the datastore writer process and each file is converted and stored in the final binary data files in the year/month directories. During conversion, the writer process keeps the file in the `partial/` directory.

importantly, the inability to use a single query[5] to, for example, retrieve several days' worth of data from all stations in a single cluster. The downsides are easily mitigated by providing the users with a higher-level abstraction of the datastore.

The binary files are stored in the open *Hierarchical Data Format, Version 5 (HDF5)* format. This data model is widely used by research organizations and academia, and is primarily intended for very large and complex datasets. The format is also extremely portable[6] and can hold data types ranging from simple arrays to complex gridded data and even pictures. It scales very well from individual analysis results to data for an entire array of stations.

A higher-level library, *PyTables* [92], is used to access the HDF5 files, both for

[5] An SQL query on a relational database, for example.

[6] The HDF5 libraries are available for almost any computing platform and bindings exist for many languages, including C, C++, Java and Python.

```
hisparc/
    cluster_aarhus/
        station_20002/
        station_20003/
    cluster_amsterdam/
        station_2/
        station_3/
        ...
        station_501/
            blobs
            configs
            events
            errors
            weather
        station_502/
        ...
    cluster_eindhoven/
        station_8001/
        ...
        station_8101/
        ...
    cluster_enschede/
    cluster_groningen/
    cluster_leiden/
    cluster_nijmegen/
    cluster_utrecht/
```

Figure 3.5 – File structure of a datastore file. A data file first contains a `hisparc` group. In that group, all clusters are represented by their own group, named after their location. In each cluster group, there is a group for each station, named after their *number*, not their location. Inside a station group are *tables*, containing the actual data. The `blobs` table contains (compressed) binary data from HiSPARC events, or error messages. The `events` table contains event data, `errors` contains error logs and `weather` contains data from weather stations located at HiSPARC stations.

reading and writing.

The structure of an HDF5 file is shown in Figure 3.5. The top level HiSPARC group contains *groups* for each cluster. Each cluster contains groups for each station. And each station contains several *tables*: `events` (HiSPARC events), `errors` (detector software error messages), `configs` (detector configuration settings, like photomultiplier voltage and thresholds), `weather` (optional: weather station data) and `blobs` (binary, variable length data, like raw event traces).

The `events` table structure is shown in Table 3.1. Each event has a unique[7] identifier,

[7]Unique in this table. When data is downloaded for analysis and combined with other data into one

Field	Type	Description
event_id	unsigned 32-bit integer	Unique number
timestamp	32-bit timestamp	Unix timestamp (GPS) [s]
nanoseconds	unsigned 32-bit integer	Sub-second part of timestamp
ext_timestamp	unsigned 64-bit integer	Full timestamp (GPS) [ns]
data_reduction	boolean	Are traces reduced?
trigger_pattern	unsigned 32-bit integer	Hardware trigger condition
baseline	signed 16-bit integer (4x)	Baseline of traces [ADC]
std_dev	signed 16-bit integer (4x)	Standard deviation of the baseline [ADC]
n_peaks	signed 16-bit integer (4x)	Number of peaks in data
pulseheights	signed 16-bit integer (4x)	Pulse height [ADC]
integrals	signed 32-bit integer (4x)	Pulse integral [ADC ns]
traces	signed 32-bit integer (4x)	Indexes into blobs array
event_rate	32-bit floating point	Trigger rate [Hz]

Table 3.1 – Structure of the events table in the datastore HDF5 files. Each event is assigned a unique event_id which is only unique within that table. Stations with only two detectors substitute values of −1 for the missing quantities. If the baseline cannot be determined, a value of −999 is substituted for all derived quantities.

event_id. Each event has a Unix timestamp in GPS time, *not* UTC. The sub-second part of the timestamp is given in nanoseconds. The ext_timestamp is the full timestamp in ns. Since there cannot exist another event with the same timestamp, this field in combination with the station number uniquely identifies the event. The data_reduction flag signifies whether the full PMT trace (*no* reduction) has been stored, or just the PMT pulse (*reduced*, or *zero suppression*). The trigger_pattern is a binary value containing the exact trigger condition at the time of the event. The baseline, std_dev, n_peaks, pulseheights and integrals fields are values derived from the PMT traces. Each field contains four values, one for each detector. If a station only has two detectors, the last two values for each field are –1. If the baseline cannot be determined, all these values are –999. The event_rate is the trigger rate at the time of the event.

3.1.4 Tier 2: Public Database

The *HiSPARC public database* is not a database as such, but rather the complete software solution providing users with data. It is a Django [93] application living at http://data.hisparc.nl/.

table, the event_id will be different. To uniquely define an event, use a station number / ext_timestamp combination.

The public database provides several services to users:

1. administration of clusters, stations, passwords, contact information, etc.
2. data quicklooks
3. histogrammed data downloads
4. API[8] for analysis programs to download data and submit results
5. analysis results pages

The quicklooks are, for example, useful for inspecting the quality of the data for a given station (Figure 3.6) or quickly observing whether stations in the same area have seen an increase in event count during a thunderstorm. These pages also provide links to download the histogram data.

Each night, a cron job runs on the public database server to determine which stations have received new data since the last time it was run. Then, the event data is read from the datastore using a read-only data connection and histograms and datasets are generated for the quicklook pages. This data is stored in a MySQL database running on the public database server. This database further acts as the backend storage for the complete Django application.

Several analysis tools use the public database API to fetch events and submit results. Work is underway to make a few of them available at `http://data.hisparc.nl/` to be used in a classroom setting. The public database API further defines the only public interface to accessing HiSPARC event data. Since the public database itself only has a read-only connection to the datastore, enforced by the OS of the datastore server, *not* the OS of the public database server, data integrity is assured.

3.2 Monitoring and Control

The HiSPARC experiment is a large distributed network with stations being maintained by high school teachers, support staff and students. Since students usually only work on HiSPARC for a few months before graduating, and even teachers move on to teaching other grades or at another location, HiSPARC stations need to run reliably without supervision for extended periods of time. To make sure that they do, a system of monitoring is employed which ensures that the HiSPARC support team has access to information both

[8] Application Programming Interface, a set of rules for software programs to communicate with each other.

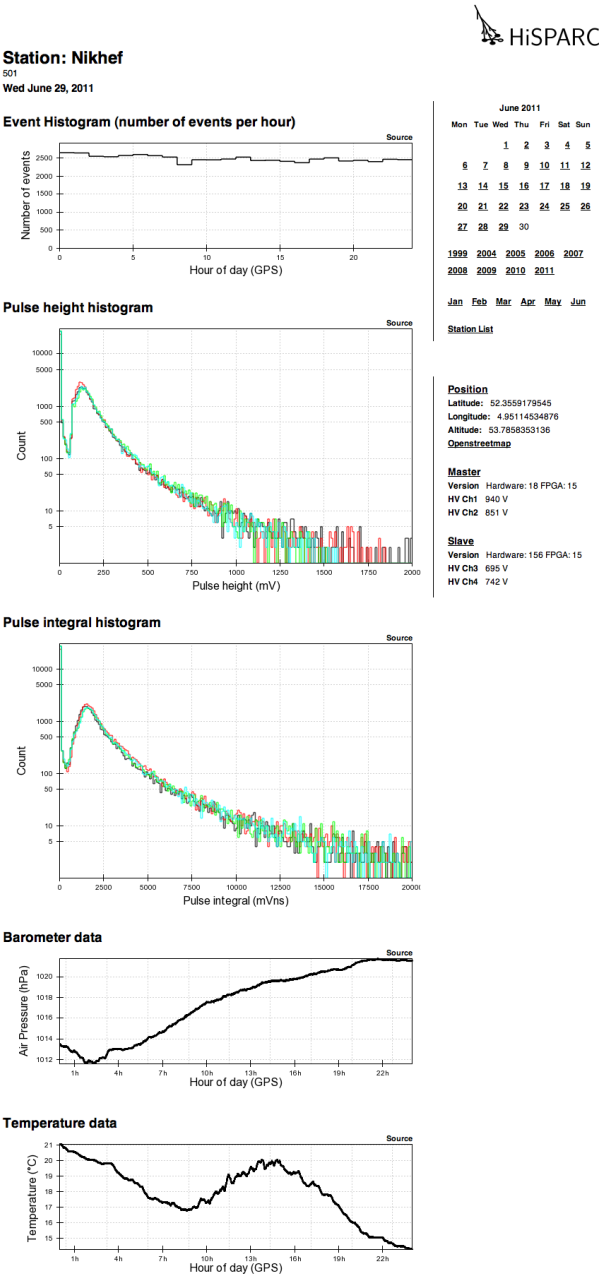


Figure 3.6 – Example of a quicklook page served by the public database. This particular page can be found at `http://data.hisparc.nl/django/show/stations/501/2011/6/29/`.

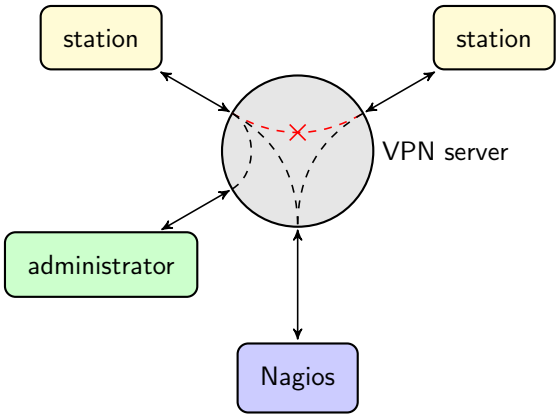


Figure 3.7 – Allowed and forbidden VPN connections. Stations and administrators connect to the VPN server using strong encrypted connections. The Nagios server connects to the VPN server on the local network. Strict firewall rules allow (black, dashed lines) or forbid (red, dashed line) connections. Connections between the Nagios server and the stations are allowed, as well as connections from administrators to stations. Connections between stations are strictly forbidden.

offline (from analyzing the data) and online (from the running station). In case of problems, accessing the station PC and making adjustments, or restarting the detector, may become necessary. This, too, is realized within this framework.

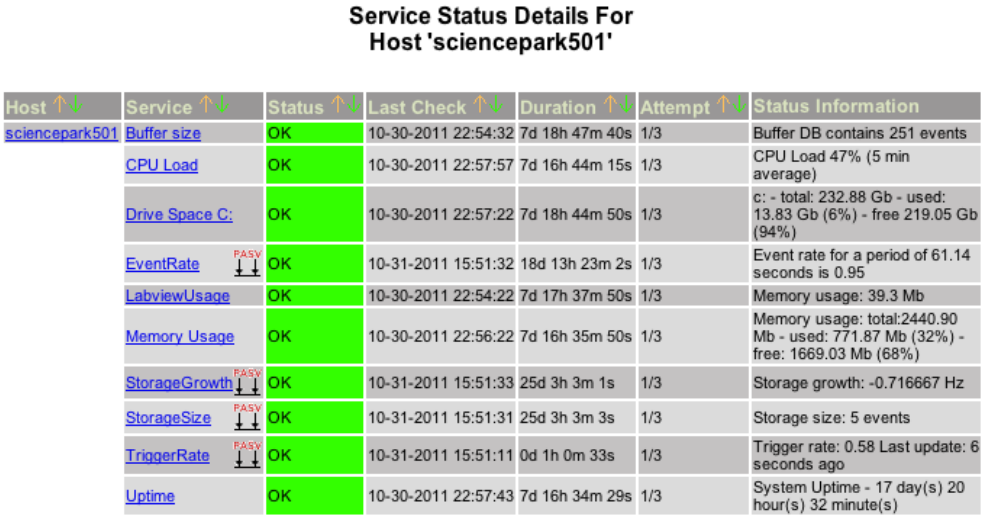
Each station receives an X.509v3 certificate which uniquely defines the station PC, along with copies of the public certificates of the *Virtual Private Network (VPN)* server running OpenVPN [94]. Each station contacts the VPN server on start up and is connected to the HiSPARC VPN network using strong encryption. Strict firewall rules make sure that no data can be exchanged between different station PCs. A very restricted set of individuals also receive special-purpose certificates and can connect to the HiSPARC *administration VPN*, which is logically different from the station VPN network.

There are only two types of connection allowed over the VPN:

1. diagnostic data between a station PC and the VPN server
2. remote desktop connections from an administrator PC to a station PC

Figure 3.7 shows the connections which are allowed over the VPN (dashed black lines) and the connections which are forbidden (dashed red lines).

The VPN server runs an instance of the *Nagios monitoring system* [95] which can exchange diagnostic messages with the station PCs. Basic checks like uptime and disk and memory usage are performed, but also more specific checks like the number of



Service Status Details For Host 'sciencepark501'

Host	Service	Status	Last Check	Duration	Attempt	Status Information
sciencepark501	Buffer size	OK	10-30-2011 22:54:32	7d 18h 47m 40s	1/3	Buffer DB contains 251 events
	CPU Load	OK	10-30-2011 22:57:57	7d 16h 44m 15s	1/3	CPU Load 47% (5 min average)
	Drive Space C:	OK	10-30-2011 22:57:22	7d 18h 44m 50s	1/3	c: - total: 232.88 Gb - used: 13.83 Gb (6%) - free 219.05 Gb (94%)
	EventRate	OK	10-31-2011 15:51:32	18d 13h 23m 2s	1/3	Event rate for a period of 61.14 seconds is 0.95
	LabviewUsage	OK	10-30-2011 22:54:22	7d 17h 37m 50s	1/3	Memory usage: 39.3 Mb
	Memory Usage	OK	10-30-2011 22:56:22	7d 16h 35m 50s	1/3	Memory usage: total:2440.90 Mb - used: 771.87 Mb (32%) - free: 1669.03 Mb (68%)
	StorageGrowth	OK	10-31-2011 15:51:33	25d 3h 3m 1s	1/3	Storage growth: -0.716667 Hz
	StorageSize	OK	10-31-2011 15:51:31	25d 3h 3m 3s	1/3	Storage size: 5 events
	TriggerRate	OK	10-31-2011 15:51:11	0d 1h 0m 33s	1/3	Trigger rate: 0.58 Last update: 6 seconds ago
	Uptime	OK	10-30-2011 22:57:43	7d 16h 34m 29s	1/3	System Uptime - 17 day(s) 20 hour(s) 32 minute(s)

Figure 3.8 – Screenshot of part of the Nagios user interface. This screenshot shows the status of station 501 (Nikhef).

events residing in the Storage on the station PC. This information can be accessed using a web interface provided by Nagios, and is a more rapid, albeit more basic, alternative to using the data quicklooks to check on station health (Figure 3.8).

When an administrator wants to access a station PC to perform maintenance, he connects to the administration VPN and can initiate a remote desktop connection to any station PC. The remote desktop software chosen for this purpose is *TightVNC* [96]. Its main advantage over other solutions, like RDP[9], is that it takes over the currently running session, instead of creating a new session by way of login. This way, remote assistance can be provided and the software running on the current desktop can be accessed.

3.3 Station Software

The software running on the HiSPARC station PC is installed by one installer package. This package not only installs software developed by HiSPARC, but also all third-party software packages which are in use. The installer consists of three components: the *main installer*, the *admin updater* and the *user updater*.

The main installer prepares the system for the station software and presents a

[9]Remote Desktop Protocol (RDP) can be used to initiate a new session on a remote Windows PC.

graphical user interface during installation. Then, it unpacks the admin and user updater packages and runs them one after another. The admin updater contains all the software which must be installed and run using administrator rights. This includes:

niruntime The National Instruments LabVIEW runtime engine. This is needed to run the DAQ programs.

openvpn The virtual private network, running as a service.

tightvnc The remote desktop, running as a service.

nsclient++ The Nagios client, providing check results to the Nagios server, running as a service.

drivers Various drivers necessary for communication with the hardware, and providing a connection from LabVIEW to MySQL.

The user updater can be installed and run without administrator rights. It is thus possible to distribute user updates and have the software installed by regular users, or by software running without administrator rights. The user updater includes:

diagnostictool A troubleshooting tool for performing diagnostic checks.

dspmon A program used to read out and program the GPS receiver in the HiSPARC electronics unit.

hisparcdaq The HiSPARC DAQ software.

hisparcweather The weather station DAQ software

hsmonitor The HiSPARC monitor application, collecting and uploading HiSPARC and weather station data. It also provides periodic check results to Nagios.

mysql The MySQL server

python The Python interpreter and standard libraries

updater The updater program, which periodically checks for software updates.

The updater downloads software updates. If an admin update is available, the user is notified to log in as an administrator to run the update. If a user update is available, all user software is shut down, the update is performed, and the software is restarted. This mechanism proved necessary to make sure that all station PCs are running up-to-date software.

3.4 SAPPHiRE

The results of many experiments are compared to simulations of the physics behind the experiment. Often, simulations guide the design of the experiment. The final results are documented in *Technical Design Reports*, e.g. [97, p. 29]. The bulk of the simulation code is usually shared between scientists, with different implementations of, and additions to, various parts of the code. The same holds true for the analysis of experimental results.

Generally, the problem of managing the code is solved by creating a *framework*. A framework enables researchers to use the exact same code, while being able to extend the code with additions and reimplementations of various parts. If useful, these contributions can then be added to the framework, to the benefit of all users.

3.4.1 Frameworks

A framework is different from a software *library*, which is a collection of functions that can be used in programs to perform various tasks. However, a framework can *evolve* from a library. The differences can be summarized as follows [98]:

1. modularity
2. reusability
3. extensibility
4. inversion of control

The *modularity* of a framework ensures that key concepts and implementations of a certain aspect of the functionality is kept together, in one part of the framework. This can be accomplished by designing classes with limited functionality which do not depend on implementation details of other classes. Furthermore, classes can be grouped together depending on their function and a hierarchy can be defined by using *namespaces*. For example, a `simulations.detector.ScintillatorSimulation` class does not depend on the implementation details of `simulations.detector.PMTSimulation`, but may share a common interface. Such common interfaces promote *reusability*, where components of the framework can be easily reused in user applications. If a framework is highly modular, reusing a class has no side-effects. For example, a user can be ensured that using a `DirectionReconstruction` class does not require a simulation to be performed first. Common interfaces result in the possibility of exchanging a `FullDetectorSimulation` with a `ToyMonteCarloDetectorSimulation` by changing only one line of code.

Extensibility is a necessary feature of a framework. Scientists need to override parts of the framework, like simulation algorithms, with custom versions to investigate their behavior. The framework ensures that common code is shared, while custom code can be incorporated. It is important to note that this custom code is not defined *in the framework*, but is defined only in the user application. If useful, new algorithms can be submitted and incorporated in the framework itself for use by other users.

Another important aspect of frameworks is *inversion of control*. In applications using libraries the control flow of a program is defined by the application code. The code calls functions and methods of the library when it sees fit. Complex tasks which are handled by the library can't be extended by the program. Frameworks allow users to extend and alter the framework. The framework contains a skeleton of a complex flow, e.g. an analysis procedure, and allow user programs to supply classes which handle certain tasks. The flow of control is handled by the framework, not the program. The program does not call methods of the framework one after another to complete a task. Rather, the framework calls parts of the user program to perform aspects of the task at hand. The flow is under control of the framework, and thus the control is *inverted*.

Modularity, reusability, extensibility and inversion of control are closely related and each of them enables the others.

3.4.2 The HiSPARC Framework

Reusing common parts of simulation and analysis programs can result in a library shared by a group of scientists. This library can be both modular and reusable. Only when program *flow* needs to be shared in an extensible way, a framework can result as an evolution of the library. This is contrary to the situation in which a framework is designed up-front by first specifying all requirements and then implementing them. It is common for a framework to be the result of both.

The code used by the HiSPARC experiment started life as a library of functions for downloading data and processing events. Early simulations were added as functions as well. While the library had multiple, modular, components, it was not easily extensible. The requirement of more complex analysis procedures which needed to work on both simulated and experimental data resulted in common interfaces and inversion of control. Finally, extensibility was incorporated into the design. The resulting framework was clearly an evolution of the library, with some later parts designed up-front.

SAPPHiRE stands for Simulation and Analysis Program Package for HiSPARC Research and Education. It's purpose is to be used heavily by researchers performing simulation

and analysis tasks, as well as high school and university students who need easy access to the data and analysis tools. Light reflections in blue sapphires are reminiscent of the blue / violet color of scintillation light in the common scintillator used by HiSPARC.

As the name suggests, the two main components of the framework are simulation and analysis. The simulation component includes classes to perform event simulation based on the results from full Monte Carlo EAS simulations, or lateral distribution functions. Also, these simulations can be performed running on a local workstation or a cluster like the Nikhef Stoomboot cluster, just by switching classes in one line of code. An example of code using SAPPHiRE is shown in Figure 3.9. An event simulation of the Science Park Array is run by simulating 1 000 000 showers, making use of a lateral distribution function derived from analysis of data in the KASCADE experiment. For each simulated event the core position is reconstructed using a new algorithm defined in the user code. This algorithm is a subclass of the framework-supplied algorithm and possibly overrides only a small part of the algorithm. Storage of the reconstruction results is still fully handled by the framework and thus compatible with all existing visualization tools.

SAPPHiRE will be extensively documented in [99].

3.5 Software Management

Software development practices can be grouped in various schools of thought which sometimes are at odds with each other. For example, the *waterfall model* [100] defines a sequential design process. This process is defined in the following order: requirements, design, implementation, verification and maintenance. The requirements and design of the software are completed before work is started on the implementation. On the other hand, *agile software development* [101, 102] advocates processes and methods based on iterative and incremental development, resulting in an evolution of requirements and solutions.

The one thing that all schools of thought agree on is the need for *revision control*, or *version control*. The software changes substantially during development and it usually becomes necessary at some point to know exactly which changes were introduced during a certain development phase. Possible reasons include uncovering bugs which were introduced after a certain date. Ad hoc solutions like copying code to directories with versioned names like `code-2008-02-19` are very fragile. Especially when working in teams, merging code changes from various members can introduce subtle side effects.

Revision control systems solve the problem of managing an evolving code base by

```
import tables

from sapphire.clusters import SciencePark
from sapphire.simulations.ldf import KascadeLdfSimulation
from sapphire.analysis.core_reconstruction import CoreReconstruction, \
                                                  CorePositionSolver


class MyCorePositionSolver(CorePositionSolver):
    """Implements some new algorithm"""
    #...
    pass


def plot_reconstruction_results(results_table):
    """Generates a set of plots for review"""
    #...
    pass


if __name__ == '__main__':
    data = tables.openFile('my_data.h5', 'w')

    # simulation
    cluster = SciencePark()
    simulation_results = '/simulations/ldf/run_1'
    sim = KascadeLdfSimulation(data, cluster, simulation_results, N=1e6)
    sim.run()

    # reconstruction
    reconstruction_results = '/reconstructions/ldf'
    rec = CoreReconstruction(data, reconstruction_results)
    rec.reconstruct_core_positions(simulation_results,
                                   solver=MyCorePositionSolver())

    plot_reconstruction_results(data, reconstruction_results)
```

Figure 3.9 – SAPPHiRE example code. This code performs a simulation using lateral distribution functions of 1 000 000 showers and reconstructs core positions using a user-supplied algorithm. Storing simulation and reconstruction results inside a group hierarchy is fully handled by the framework. This code makes use of default values of the framework, e.g. the shower size.

storing many versions of the code in one location, called a *repository*, and providing various tools to explore the differences between versions and reverting parts of the code to earlier versions. Over the years, many such systems were created. At first, such systems stored all the code and history in one central location, and are thus called *centralized*. Popular systems include *CVS* and *Subversion*. More recently, *distributed* systems were created, like *Bazaar, Mercurial* and *Git*. Distributed systems store all the history and code locally and each users repository is on equal footing. Usually a distributed workflow is combined with a central location containing the copy used to created software releases. This copy is updated by all team members when new features are implemented.

Revision control systems often include tools to create *branches* of the code, in which a certain feature can be developed independently of changes to the *master*, or *trunk*, branch. When such a feature is ready for inclusion, it is *merged* back into the master branch. Systems like Git encourage heavy use of branching and include powerful tools for merging the changes. Distributed systems are more compatible with agile development methods in that they evolve along with the changing work flow of software projects. For example, there is no need to set up a central repository. A single developer issues a `git init` inside his project directory and starts developing and *committing* changes to the repository. This repository can then be shared with other developers when the need arises and changes can flow both ways. If the team grows, it may become preferable to create a central location which is kept in sync with all developers' copies.

The HiSPARC project has used revision control since 2008. At first, Bazaar [103] was used because of its excellent documentation. Recently, all repositories were converted to use Git [104] and the repositories are now hosted on Github [105]. Github simplifies the hosting of code and includes an *issue tracker*, which allows anyone to report a bug. The problem and possible solutions can then be discussed in one central location. Git (and Github) give all team members a powerful work flow by easily visualizing changes, develop new features or fix bugs in isolation, and merge the changes into the master branch. By making development public, accountability is increased, which promotes the use of *clean code*. Furthermore, any user can *fork*[10] the repository and fix bugs or add features. They can then issue a *pull request*, which asks the team developing the main repository to include the changes. Bachelor or Master students can thus immediately use and alter the HiSPARC software, without giving them any rights to the main repositories.

The home page for HiSPARC on Github is located at `http://github.com/HiSPARC/`.

[10] Essentially making a copy of the repository to start development independently of the parent repository.

It includes the following repositories:

datastore The code running the central datastore. It is responsible for first accepting the data and for aggregating the data and storing it in binary HDF5 files.

labview The code for the HiSPARC DAQ software.

publicdb The code for the public database, which is a Django application running at `http://data.hisparc.nl/`.

sapphire The code of the SAPPHiRE framework, used for analyzing HiSPARC data and simulating events.

station-software The code running on station PCs, including the code for building the installer package, with the exception of LabVIEW sources. The HiSPARC DAQ software, as well as the weather station DAQ software, have their own repositories. All other code, including third-party components, is located in this repository. The most important piece of code in this package is the *monitor*, responsible for uploading data and sending check results to the Nagios instance running on `http://vpn.hisparc.nl/`.

vhdl The source of the VHDL code used to program the FPGA unit inside the HiSPARC electronics module.

weather The code for the weather station DAQ software.

All repositories have their own issue tracker and most contain documentation describing the code base and instructions on how to set up a development environment. Access rights can be specified by including developers into *teams*, which can be assigned to one or multiple repositories. Each repository can grant write access to multiple teams. Development is completely open, however, and everyone has read access.

Using Github, the HiSPARC experiment has extended its outreach character to include not only cosmic ray research, but also all aspects of software development.

4

Single Station Event Simulation and Reconstruction

A HiSPARC station consists of two or four detectors. The four-detector stations have the layout shown in Figure 2.11. An inclined shower will not reach the detectors within the station simultaneously. The arrival time differences depend on the direction of the shower. From the time measurements of the individual detectors, it should be possible to reconstruct the direction of the shower.

To calculate the arrival time differences, a model of the shower front has to be defined. The shower direction, i.e. the zenith and azimuthal angles, are parameters in this model. The conventional method for reconstructing the shower direction is to fit the results of the model to the arrival times measured by the stations of a (dense) array. An accurate estimate for the direction of the shower can then be determined using a minimization method. Some of the algorithms developed in the literature do not make use of a minimization procedure [106], but require a large number of detectors. Both methods work best for extended arrays with knowledge of the shower core position.

In the case of a single HiSPARC station, a method has been developed using direct calculation of the shower direction based on the arrival time measurements of three out of the four detectors.

In this chapter, this algorithm will be tested on simulated air shower data. First, the simulations will be discussed. Then, the algorithm for the reconstruction of the direction

of the shower will be developed and the statistical uncertainties will be evaluated. The uncertainties will be propagated through the analysis yielding equations for the accuracy of the reconstructed azimuthal and zenith angles. Correlations between experimental errors are not considered and errors on errors are not determined. Finally, the reconstruction results of the simulated data will be discussed and errors in the reconstruction will be compared to the calculated uncertainties.

4.1 Event Simulation

The simulation contains three parts. The first part is a full Monte Carlo simulation of the shower development. The second part is a simulation which uses the results of the first part to effectively simulate many showers occurring in the neighborhood of a HiSPARC station. This part tracks which of the shower particles hit the detectors. The third part is a simulation of the detector response.

4.1.1 Extensive Air Showers

AIRES [107], such as the CORSIKA [108] program, is a simulation package for studying extensive air showers. Given an input file containing parameters defining the primary particle and the accuracy of the calculations, AIRES simulates interactions and tracks particles through the atmosphere.

CORSIKA makes extensive use of third-party interaction models. This allows CORSIKA to be based on proven models. While AIRES also links to high-energy hadronic interaction models such as SIBYLL and QGSJET, other interactions such as electrodynamical processes, particle decays and the propagation of charged particles, is handled by custom algorithms. This allows AIRES to be tuned to be fast. In fact, it is about 3.5 times faster than CORSIKA [109]. The input file syntax of AIRES is less complex than the one used by CORSIKA. A comparison between the results of AIRES and CORSIKA can be found in [109]. Differences in the results are mainly due to different simulation parameters and details in the types of interactions considered. When the simulations are rerun with identical parameters and interactions for both packages, no significant differences are observed. Overall, the authors conclude that CORSIKA and AIRES agree to better than 20 % for the basic shower parameters observed by the Pierre Auger Observatory.

Both programs include *thinning algorithms*. These algorithms are designed to reduce the number of particles that have to be fully tracked. This reduces computation time, memory usage and the size of a simulated event considerably. AIRES makes use of

an extended version of the *Hillas thinning algorithm*. Hillas [110, 111] proposed the following procedure to reduce the number of particles. Consider the process in which a particle A generates a number of secondary particles:

$$A \rightarrow B_1, B_2, \ldots, B_n, \qquad \text{with } n \geq 1. \tag{4.1}$$

Let E_A be the energy of particle A, E_{B_i} be the energy of particle B_i, and E_{th} the so-called *thinning energy*. Then, if $E_A \geq E_{th}$, the secondaries are considered separately and included in the remainder of the simulation with probability P_i, given by

$$P_i = \begin{cases} 1 & \text{if } E_{B_i} \geq E_{th}, \\ \frac{E_{B_i}}{E_{th}} & \text{if } E_{B_i} < E_{th}. \end{cases} \tag{4.2}$$

That is, if a particle has an energy *greater than* the thinning energy, it is kept in the simulation. If the energy is *less than* the thinning energy, there is a probability that the particle is removed from the simulation. Particles with a larger energy fraction have a larger probability to be kept in the simulation. If the energy $E_A < E_{th}$, then particle A is the result of a previous thinning operation. Only one of the secondary particles is kept. In other words, the number of particles with energies below the thinning energy E_{th} is never increased. The secondary particle to keep in the simulation is selected among all n particles with the probability given by

$$P_i = \frac{E_{B_i}}{\sum_{j=1}^{n} E_{B_j}}. \tag{4.3}$$

The remaining particles are given a statistical weight $w_{B_i} = w_A / P_i$, with w_A the weight of particle A. This enables faster simulations while keeping the results statistically correct. When taking the statistical weights into account during analysis of the shower, the global shower observables are almost unaffected. Fluctuations increase, however.

To reduce the fluctuations, AIRES incorporates an extension to the original splitting algorithm. The new algorithm monitors the statistical weights of the particles. If the weights become too large, the thinning algorithm will keep more particles in the simulation.

It is essential to track individual particles in a detector to simulate detector responses. To accomplish this, it becomes necessary to not use or reverse the thinning. A *resampling* algorithm to obtain individual particle properties for use in simulations for the Auger Observatory is proposed in [112]. A similar method for application in

the Telescope Array, including validation studies, is discussed in [113]. However, the validation studies are only done using the same thinned showers *before* and *after* the resampling procedure. Comparison of a *dethinned* shower with a shower *without* thinning is carried out in [114], using the CORSIKA program. A shower induced by a proton with an energy of 10^{19} eV is simulated with and without thinning. The thinned shower is subsequently resampled to obtain individual particles. The study shows that fluctuations of the number of particles in the detectors are significantly larger for dethinned showers than for showers without thinning. The fluctuations in the simulation are compared to the expected Poissonian fluctuations. Showers are simulated with thinning level $E_{th} = 10^{-6}E_p$, with E_p the primary energy. This is a commonly used setting. For electrons, the ratio $\sigma/\sigma_{\text{Poisson}} \approx 80$ [114].

A realistic simulation of the shower front is essential to study the effect of the accuracy of the time-of-arrival measurements on the reconstruction of the shower direction. The simulation of showers with $E_p \leq 10^{16}$ eV is still feasible without thinning. Therefore, to obtain the most realistic results, thinning and resampling is not performed in this work.

A single HiSPARC station is sensitive to extensive air showers starting from primary energies of approximately 1 PeV, see Section 2.1.1 for details. Since the number of showers falls steeply with increasing primary energy, only an energy of 1 PeV is taken.

Showers have been simulated for a series of discrete zenith angles: 0°, 5°, 10°, 15°, 22.5°, 30°, 35° and 45°. For each angle, 10 showers are generated. All showers are simulated with an incoming proton with an energy of 1 PeV.

4.1.2 Detector Response

Since the shower footprint is very large compared to the size of the station, it is justified to change the coordinate system such that a single shower can be used multiple times. A station is then positioned at various core distances inside the same shower and is given a random rotation (Figure 4.1).

For each position, the charged leptons that traverse the detectors are identified. Leptons are by far the most numerous charged particles in a shower. Hadrons appear only at small core distances and are ignored since even where they appear, they are far outnumbered by electrons.

The (cartesian) x, y coordinate system is used by AIRES and places the shower core at the origin O (Figure 4.2). The azimuthal angle of the shower, ϕ, is always equal to zero. For each event, a random position inside the shower is chosen for the location of

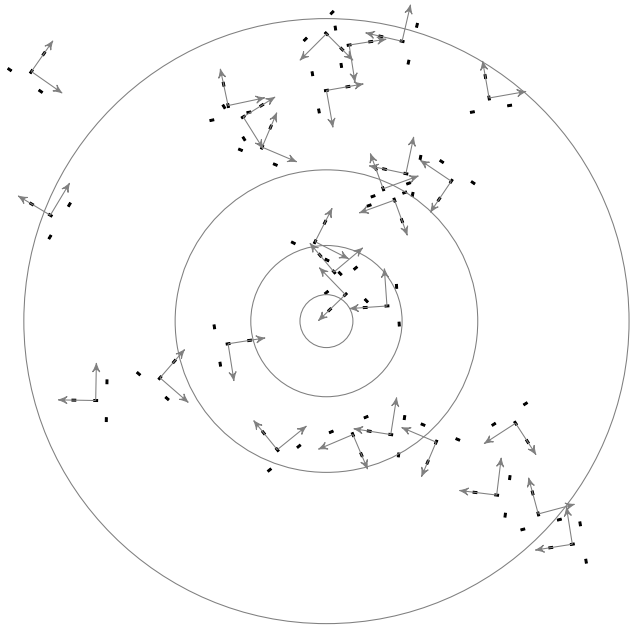


Figure 4.1 – A station is placed at various locations inside an inclined shower and rotated over random angles. This is equivalent to showers arriving at the station from random azimuthal angles. In this way, a single simulated shower can be reused many times.

the station. The coordinate system x', y' is introduced such that the station is always at the origin O'. This position is defined by a core distance r and polar angle Φ, with respect to O. An angle α is randomly chosen and the station is rotated over that angle. The coordinate system x'', y'' is defined as the rotation of x', y' over $-\alpha$, such that in the system x'', y'' the positions of the detectors are the same for all events. This coordinate system describes a stationary station which detects showers with random core positions and random azimuthal angles ϕ''. The transformation of x, y coordinates to x'', y'' coordinates is given by

$$r'' = r, \qquad \Phi'' = \Phi + \pi - \alpha, \qquad \phi'' = -\alpha. \tag{4.4}$$

The shower core position x'', y'' and direction θ'', ϕ'' is given by

$$x'' = -r\cos(\Phi - \alpha), \qquad y'' = -r\sin(\Phi - \alpha), \qquad \theta'' = \theta, \qquad \phi'' = -\alpha. \tag{4.5}$$

Photons generated by the charged particles propagate through the scintillator. A photon may reflect multiple times before it reaches the PMT, or it may escape the

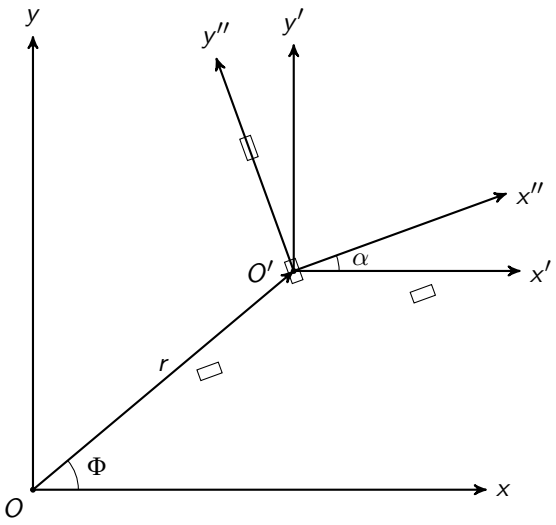


Figure 4.2 – Coordinate system used in the simulation. The x, y axes are the natural coordinates for the shower, as used by AIRES, with the shower core at the origin O. The four rectangles are the four HiSPARC detectors. In polar coordinates, the center of the station is defined as $O' = (r, \Phi)$, with the x', y' coordinates defined as a translation. The angle α is defined as the angle over which the station is rotated with respect to the x', y' coordinates. The x'', y'' coordinate system is defined as the natural coordinates for the station in which the detectors are always stationary. This coordinate system is used for the simulation output.

scintillator. In Section 2.2.3 a simulation is discussed in which the propagation of photons is tracked to measure the transmission efficiency of the scintillator. The simulation is extended to include the propagation time of photons [115]. The transport time of the photons from the point of impact to the PMT depends on the location of the particle path in the scintillator, as well as the initial directions of the photons. If a photon reaches the PMT there is a probability Q, the PMT's *quantum efficiency*, that it will create a photoelectron. The photoelectron creates an avalanche in the PMT, resulting in a pulse of $\mathcal{O}(10^6)$ electrons. For a signal above the 70 mV threshold, 15 photoelectrons are required. Figure 4.3 shows the distribution of the time of arrival of the 15^{th} *detected* photon. The transport time distribution is approximated by the straight lines. The mean value of the distribution is 4.4 ns, with a standard deviation of 1.2 ns.

In the detector response simulation, for each charged particle traversing the detector a random propagation time is chosen according to the transport time distribution. This value is added to the arrival time of the particle in the scintillator, resulting in the

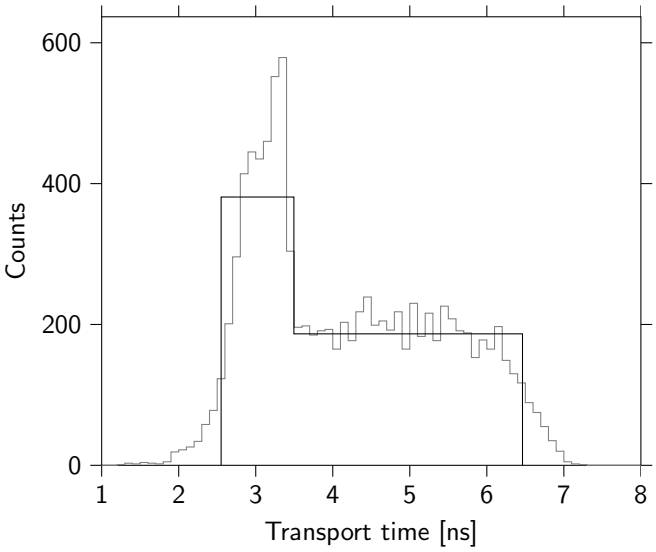


Figure 4.3 – Transport time of photons from the point of impact of the detected charged particles to the PMT (gray). The time for the 15^{th} photoelectron is recorded. This is the first photon over the detection threshold and will result in the arrival time measurement. The mean value is 4.4 ns and the standard deviation is 1.2 ns. An approximation for the transport time distribution, used in the detector simulation, is shown in the same graph (black). Redrawn from [115].

measurement of the arrival time:

$$t_{\text{measured}} = t_{\text{arrival}} + \Delta t_{\text{transport}}. \tag{4.6}$$

More than one particle may traverse a detector. In this case, the detector simulation is performed for each particle independently. From the list of measured arrival times, the smallest value of t_{measured} is selected as the time measurement. The simulation records the total number of particles which are detected in each scintillator.

4.2 Reconstruction of Shower Direction

To reconstruct the direction of the shower, it is sufficient to determine the arrival time in three detectors. The procedure is discussed below. The shower front is approximated by a plane. This means that all particles in a vertical shower should arrive at the same time, i.e. there is no curvature of the shower front, and the front is infinitely thin. These assumptions will be justified in the following discussion.

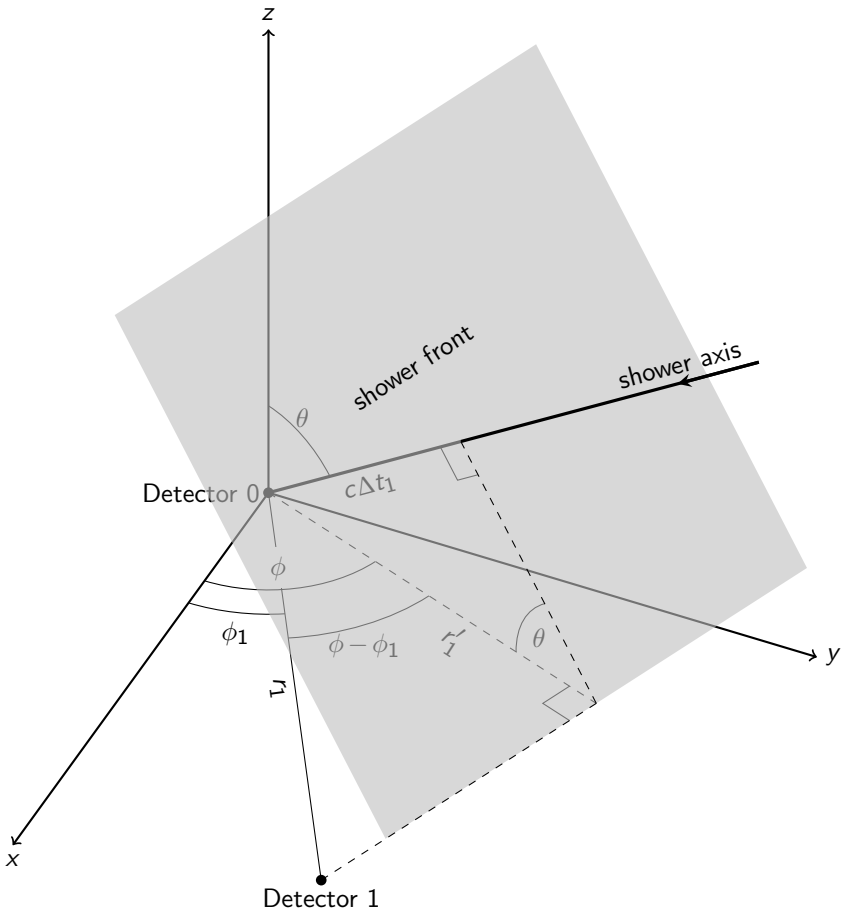


Figure 4.4 – Three-dimensional representation of the coordinate system used in reconstructing the direction of a shower. Refer to Figure 2.4 for a two-dimensional representation.

In Figure 4.4 a three-dimensional representation is given of the coordinate system used in the reconstruction of the direction of a shower. For a slightly less general two-dimensional representation, refer to Figure 2.4. The plane shower front is depicted as a surface traveling in the direction of the shower axis. The shower axis has an angle θ with respect to the vertical direction (z-axis), and an angle ϕ with respect to the positive-x direction. In the figure, the shower front is at the location of detector 1. To reach detector 0, it has to travel a distance $c\Delta t_1$, where the velocity of the shower front is approximated by c, and $\Delta t_1 \equiv t_1 - t_0$ is the arrival time difference between the detectors 0 and 1. The projection of the shower front on the ground, the xy-plane, travels

a distance r'_1. It can be seen that

$$\sin\theta = \frac{c\Delta t_1}{r'_1}, \tag{4.7}$$

and

$$\cos(\phi - \phi_1) = \frac{r'_1}{r_1}. \tag{4.8}$$

Combining Equations 4.7 and 4.8 results in

$$c\Delta t_1 = r_1 \cos(\phi - \phi_1)\sin\theta. \tag{4.9}$$

Then ϕ and θ are the two unknowns that can be resolved by two measurements of the time differences (Δt_1, $\Delta t_2 \equiv t_2 - t_0$):

$$\left.\begin{aligned} c\Delta t_1 &= r_1 \cos(\phi - \phi_1)\sin\theta \\ c\Delta t_2 &= r_2 \cos(\phi - \phi_2)\sin\theta \end{aligned}\right\}, \tag{4.10}$$

where ϕ_2 is the azimuthal angle of detector 2 with respect to detector 0. From Equations 4.10 one obtains

$$\frac{\Delta t_1}{\Delta t_2} = \frac{r_1 \cos(\phi - \phi_1)}{r_2 \cos(\phi - \phi_2)}, \tag{4.11}$$

which, after using the trigonometric identity $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$, turns into

$$\frac{r_2\Delta t_1}{r_1\Delta t_2} = \frac{\cos\phi\cos\phi_1 + \sin\phi\sin\phi_1}{\cos\phi\cos\phi_2 + \sin\phi\sin\phi_2} = \frac{\cos\phi_1 + \tan\phi\sin\phi_1}{\cos\phi_2 + \tan\phi\sin\phi_2}. \tag{4.12}$$

By reordering, one arrives at

$$\tan\phi = -\frac{r_1\Delta t_2\cos\phi_1 - r_2\Delta t_1\cos\phi_2}{r_1\Delta t_2\sin\phi_1 - r_2\Delta t_1\sin\phi_2}. \tag{4.13}$$

Once the angle ϕ is determined, either one of the Equations 4.10 can be solved for θ:

$$\sin\theta = \frac{c\Delta t_1}{r_1\cos(\phi - \phi_1)}, \tag{4.14a}$$

$$\sin\theta = \frac{c\Delta t_2}{r_2\cos(\phi - \phi_2)}. \tag{4.14b}$$

4.3 Measurement Uncertainties

4.3.1 Timing Uncertainties

Experimental uncertainties in the determination of the arrival time are introduced by the decay time of the scintillator (Section 2.2.2), time-of-flight of scintillation photons from the point of impact of the particle to the PMT, transit time spread of the electrons through the PMT, and the timing resolution of the ADCs. The GPS timing system has no influence on the reconstruction from a station, since it is only used to generate an overall timestamp for the event.

The decay constant of the scintillator is 2.1 ns [81]. On average, approximately 440 photons reach the PMT, while only 15 photoelectrons are required to cross the detection threshold. Most of these photons will therefore have a vanishingly small delay due to the decay time.

When a charged particle traverses the scintillator, the time for scintillation photons to reach the phototube depends on the distance between the phototube and the point of impact. This is determined by the geometry of the scintillator and the lightguide. When the scintillator is uniformly illuminated, the mean time-of-flight is 4.4 ns with a standard deviation of 1.2 ns [115], see Figure 4.3. The sampling time of the ADCs is 2.5 ns.

4.3.2 Model Uncertainties

The shower front is not a flat plane. Also, the shower front has a certain thickness. It is therefore unknown whether a particle is measured very near the causal front[1] or lagging behind, which leads to an uncertainty in the timing measurement. The arrival time distributions of simulated vertical 1 PeV proton showers are shown in Figure 4.5. The arrival time is defined as the amount of time a particle arrives after the causal front has passed. Only events that have at least one charged particle in all corner detectors are considered. The median arrival time delay increases with increasing core distance. 50 % of the particles are contained within the gray band. The width of the band, i.e. the thickness of the shower front, also increases with core distance.

[1] When the primary cosmic ray particle interacts, information from that event can at best be transmitted at the speed of light. All possible light signals from the event form an expanding sphere. The part of the sphere in the direction of the primary particle momentum is called the *causal front*. The shower front, containing the secondary particles, travels behind the causal front.

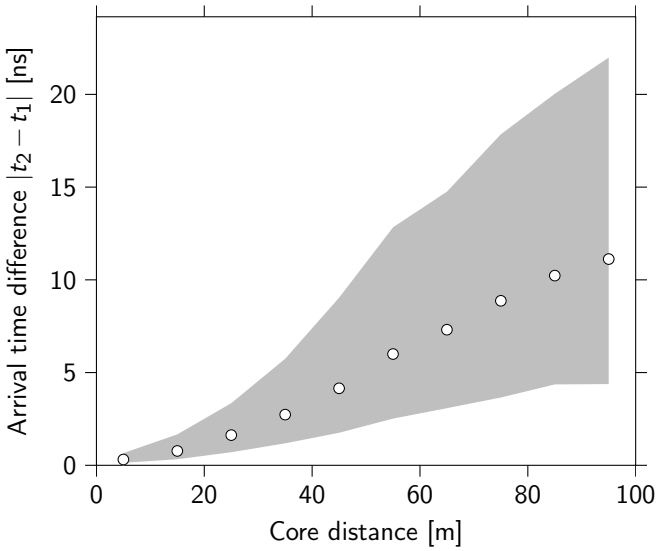


Figure 4.5 – The measured arrival time distributions of vertical showers. The difference in arrival time in two detectors is graphed. The showers are generated by a 1 PeV proton. Only measurements with at least one charged particle in all three corner detectors are taken into account. The dots show the median arrival time and the gray bands contain 50 % of the events, evenly distributed around the median. This figure can be compared to Figure 2.3, which shows the actual simulated arrival time distributions. The data in this graph is an approximation derived from *measuring* the arrival times in two detectors.

4.3.3 Propagation through Analysis

The uncertainty in reconstructing the azimuthal angle introduced by timing uncertainties is given by

$$\sigma_\phi^2 = \sigma_t^2 \left(\left| \frac{\partial \phi}{\partial t_0} \right|^2 + \left| \frac{\partial \phi}{\partial t_1} \right|^2 + \left| \frac{\partial \phi}{\partial t_2} \right|^2 \right), \tag{4.15}$$

with σ_ϕ, σ_t the *standard deviation*, of ϕ and t respectively. The first-order derivatives of $\phi(t_0, t_1, t_2)$ are:

$$\frac{\partial \phi}{\partial t_0} = \frac{1}{1+\tan^2\phi} \cdot \frac{r_2\cos\phi_2 - r_1\cos\phi_1 + (r_2\sin\phi_2 - r_1\sin\phi_1)\tan\phi}{\frac{r_1 r_2}{c}\sin\theta\left(\sin\phi_2\cos(\phi-\phi_1) - \sin\phi_1\cos(\phi-\phi_2)\right)}, \tag{4.16}$$

$$\frac{\partial \phi}{\partial t_1} = \frac{1}{1+\tan^2\phi} \cdot \frac{-r_2(\sin\phi_2\tan\phi + \cos\phi_2)}{\frac{r_1 r_2}{c}\sin\theta\left(\sin\phi_2\cos(\phi-\phi_1) - \sin\phi_1\cos(\phi-\phi_2)\right)}, \tag{4.17}$$

$$\frac{\partial \phi}{\partial t_2} = \frac{1}{1+\tan^2\phi} \cdot \frac{r_1(\sin\phi_1\tan\phi + \cos\phi_1)}{\frac{r_1 r_2}{c}\sin\theta\left(\sin\phi_2\cos(\phi-\phi_1) - \sin\phi_1\cos(\phi-\phi_2)\right)}. \tag{4.18}$$

The time differences Δt remaining after the differentiation procedure have been rewritten in terms of θ and ϕ using Equations 4.14. Applying the trigonometric identities, these equations reduce to

$$\frac{\partial \phi}{\partial t_0} = -\left(\frac{\partial \phi}{\partial t_1} + \frac{\partial \phi}{\partial t_2}\right), \tag{4.19}$$

$$\frac{\partial \phi}{\partial t_1} = \frac{-c\cos(\phi - \phi_2)}{r_1 \sin\theta \sin(\phi_2 - \phi_1)}, \tag{4.20}$$

$$\frac{\partial \phi}{\partial t_2} = \frac{c\cos(\phi - \phi_1)}{r_2 \sin\theta \sin(\phi_2 - \phi_1)}. \tag{4.21}$$

Singularities occur for $r_1 = 0$, $r_2 = 0$, $\theta = 0$, and $\phi_2 - \phi_1 \in \{-\pi, 0, \pi\}$. For $r_1 = 0$ or $r_2 = 0$, the direction cannot be reconstructed since two detectors will be on top of each other. The same holds for $\phi_2 - \phi_1 \in \{-\pi, 0, \pi\}$: the three detectors are on a straight line. For $\theta = 0$, the azimuthal angle ϕ is also undefined and the uncertainty becomes infinite.

The full expression for the uncertainty in ϕ is then given by

$$\begin{aligned}\sigma_\phi^2 &= 2\sigma_t^2 \left(\left|\frac{\partial \phi}{\partial t_1}\right|^2 + \left|\frac{\partial \phi}{\partial t_2}\right|^2 + \frac{\partial \phi}{\partial t_1}\frac{\partial \phi}{\partial t_2} \right) \\ &= \frac{2c^2\sigma_t^2}{\sin^2\theta \sin^2(\phi_2 - \phi_1)} \left(\frac{\cos^2(\phi - \phi_2)}{r_1^2} + \frac{\cos^2(\phi - \phi_1)}{r_2^2} - \frac{\cos(\phi - \phi_1)\cos(\phi - \phi_2)}{r_1 r_2} \right).\end{aligned} \tag{4.22}$$

Similarly, the uncertainty in the zenith angle in Equation 4.14a becomes:

$$\begin{aligned}\sigma_\theta^2 &= \sigma_t^2 \left(\left|\frac{\partial \theta}{\partial t_0}\right|^2 + \left|\frac{\partial \theta}{\partial t_1}\right|^2 + \left|\frac{\partial \theta}{\partial t_2}\right|^2 \right) \\ &= \sigma_t^2 \frac{A\sin^2\theta + B\sin\theta + C}{r_1^2(1-\sin^2\theta)\cos^2(\phi - \phi_1)},\end{aligned} \tag{4.23}$$

with

$$A = r_1^2 \sin^2(\phi - \phi_1) \left[\left|\frac{\partial \phi}{\partial t_0}\right|^2 + \left|\frac{\partial \phi}{\partial t_1}\right|^2 + \left|\frac{\partial \phi}{\partial t_2}\right|^2 \right], \tag{4.24}$$

$$B = -2r_1 c \sin(\phi - \phi_1) \left[\frac{\partial \phi}{\partial t_0} - \frac{\partial \phi}{\partial t_1} \right], \tag{4.25}$$

$$C = 2c^2. \tag{4.26}$$

For horizontal showers ($\theta = 90°$) the acceptance becomes zero and the uncertainty becomes infinite. Furthermore, when a shower hits detector 0 and detector 1 (Figure 4.4)

from a direction perpendicular to the line connecting the two detectors, $\left(\phi - \phi_1 = \pm\frac{\pi}{2}\right)$, it is impossible to determine the zenith angle since all angles result in the same time difference $\Delta t_1 = 0$. In this situation, it is required to switch from Equation 4.14a to Equation 4.14b. The uncertainty in that equation can be calculated as shown above. The end result is identical for the substitutions $r_1 \to r_2, \phi_1 \to \phi_2, t_1 \to t_2$ and $t_2 \to t_1$. Denoting the angle and uncertainties from Equations 4.14 as $(\theta_1, \sigma_{\theta_1})$ and $(\theta_2, \sigma_{\theta_2})$, respectively, one obtain an expression for the zenith angle:

$$\theta = \frac{\frac{1}{\sigma_{\theta_1}}\theta_1 + \frac{1}{\sigma_{\theta_2}}\theta_2}{\frac{1}{\sigma_{\theta_1}} + \frac{1}{\sigma_{\theta_2}}}. \tag{4.27}$$

This will lead to a robust calculation which is only susceptible to large uncertainties for (near-) horizontal showers. Obviously, for horizontal showers the acceptance of the detectors is extremely small.

4.4 Performance of a Single Station

4.4.1 Reconstruction Efficiency

Three independent measurements are required to reconstruct the shower direction. The best result is obtained when the distance between the detectors is large. Then the difference between arrival times is large and the uncertainty is subsequently smaller. Therefore, the three corner detectors of the station (Figure 2.11) are chosen to reconstruct the shower direction.

Particle densities in a shower are highest near the shower core and fall off steeply (Figure 2.1). Therefore, the probability that a detector is traversed by a particle decreases with increasing core distance. Figure 4.6 shows the fraction of simulated showers detected by all three corner detectors. Three zenith angles are shown. If the shower core is close to the station, almost all showers are detected. At a distance of approximately 40 m, the detection efficiency is 50 % for vertical showers. For inclined showers, the detection efficiency is much lower. Since the distance particles travel through the atmosphere is larger, fewer particles reach the ground. Moreover, the acceptance of the detectors is less for inclined showers, since the effective detection area is decreased.

A detected shower yields three independent arrival time measurements. Thus, the equations given in Section 4.2 can be used to reconstruct the direction of the shower.

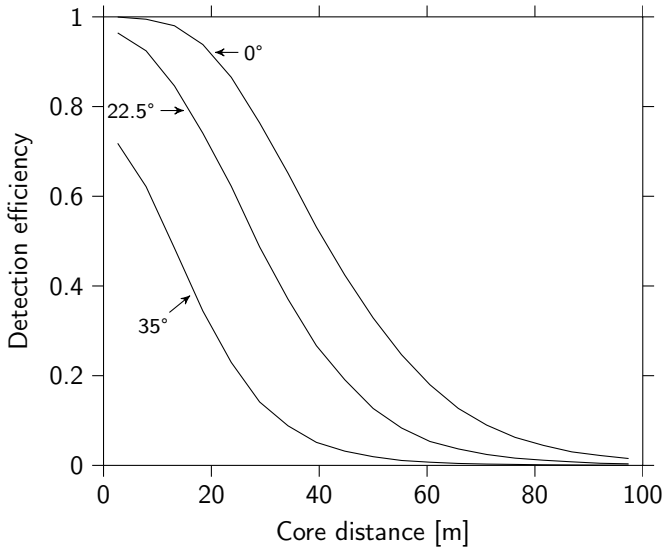


Figure 4.6 – Fraction of showers that are detected with at least 1 charged particle in the three corner detectors as a function of core distance. The showers are initiated by 1 PeV protons for a series of zenith angles. If the shower core is close to the station, almost all showers are detected. At a distance of approximately 40 m, the detection efficiency is 50 % for vertical showers. For inclined showers, the detection efficiency is much lower, as fewer particles reach the ground.

However, if the arrival time differences become too large, the reconstruction results in an unphysical solution and the reconstruction is rejected. This occurs when the arrival time difference between two detectors is larger than the *light time* ($\Delta t = d/c$) between the detectors with distance d. Since the velocity of charged particles in a shower is approximated by the speed of light, the light time between two detectors is the largest time difference allowed in the reconstruction. For a HiSPARC station, $d = 10\,\text{m}$, thus $\Delta t = 33\,\text{ns}$.

Figure 4.7 shows the fraction of *detected* 1 PeV proton showers for which the direction can be reconstructed as a function of the core distance. The reconstruction efficiency decreases with increasing core distance, but is up to 90 m well above 50 %. The shape and thickness of the shower front differs from the approximated thin, flat plane. This leads to larger arrival time delays for increasing core distance (Figure 4.5). Approximately 25 % of the vertical EAS events with a core distance of 70 m, arrive later than the light time for the distance between two detectors in a HiSPARC station and are thus rejected. This fraction increases for inclined showers.

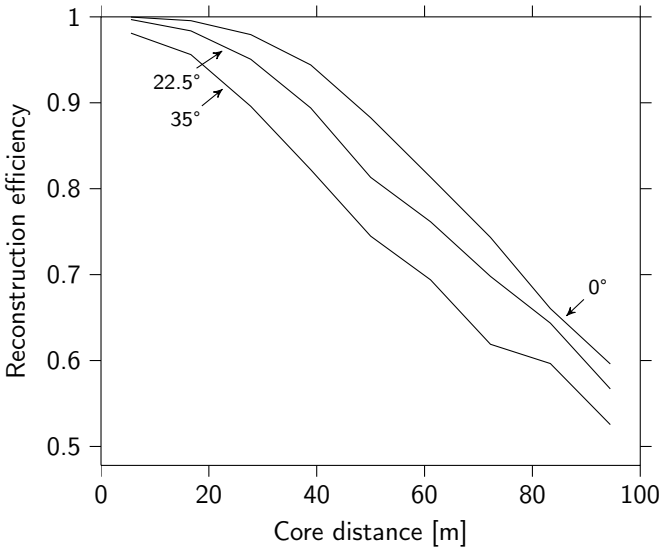


Figure 4.7 – Fraction of *detected* 1 PeV proton showers for which the direction can be reconstructed as function of the core distance. The requirement for detection is at least 1 charged particle in the three corner detectors. The reconstruction is successful if the zenith and azimuthal angles can be calculated using the equations given in Section 4.2. The reconstruction efficiency decreases with increasing core distance, but is well above 50 % up to 90 m.

4.4.2 Shower Direction

Figure 4.8 shows a strong correlation between simulated and reconstructed azimuthal angles for 1 PeV proton showers with a zenith angle $\theta = 22.5°$. Reconstruction uncertainties are depicted in Figure 4.9. For the left-hand figure at least one particle in all three corner detectors is required. On the right, however, more strict conditions are applied: at least two particles in each of the three corner detectors, which results in a more accurate azimuthal angle.

The uncertainty in the reconstruction of the zenith angle is shown in Figure 4.10, for $0° \leq \theta \leq 45°$. With increasing θ, the arrival time difference between detectors increases (Equation 4.9) while the absolute uncertainty in the measurements, due to instrumentation and the structure of the shower front, is not changed. The result is that the *relative* timing uncertainty decreases and the reconstruction becomes more accurate. The right hand figure shows more stringent reconstruction requirements. When more particles traverse the detectors, the reconstruction uncertainty reduces.

The timing accuracy resulting from the plane front approximation and the transport

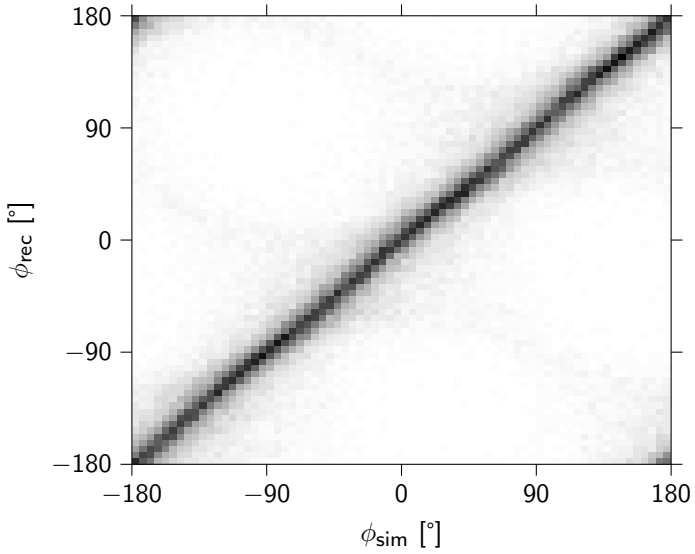


Figure 4.8 – Simulated versus reconstructed azimuthal angle for showers of 1 PeV protons with θ = 22.5°. Only events with at least 1 MIP in all corner detectors are included. There is a strong correlation between simulated and reconstructed azimuthal angles.

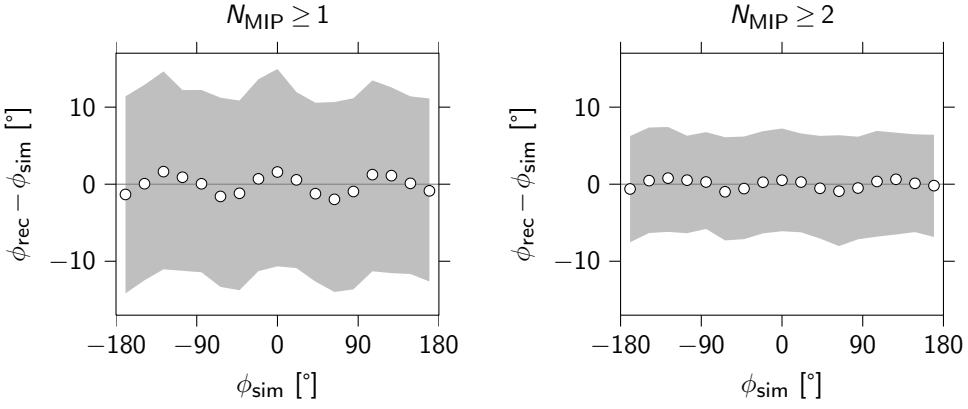


Figure 4.9 – The uncertainty of the reconstruction of the azimuthal angle as a function of the simulated angle, for θ = 22.5°. The dots show the median values and the shaded region contains 50 % of the events, evenly distributed around the median. On the left, minimum reconstruction requirements are used. That is, at least 1 particle in all three corner detectors. On the right, at least 2 particles in all three corner detectors are selected. The more stringent requirements result in a more accurate reconstruction.

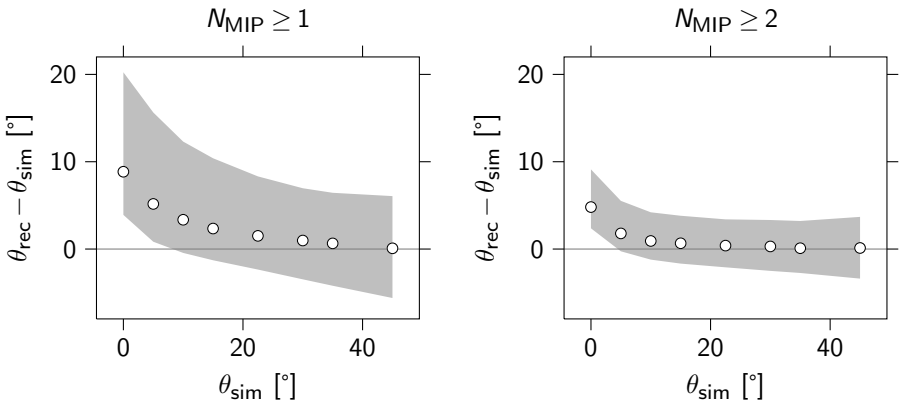


Figure 4.10 – The uncertainty of the reconstruction of the zenith angle as a function of the simulated angle. The dots show the median values and the shaded region contains 50 % of the events, evenly distributed around the median. On the left, minimum reconstruction requirements are used. That is, at least 1 particle in all three corner detectors. On the right, at least 2 particles in all three corner detectors are selected. The more stringent requirements result in more accurate reconstructions. With increasing zenith angle, the reconstruction uncertainty decreases.

time of scintillation photons in the detector is given by

$$\sigma_t = \sqrt{\sigma^2_{t,\mathrm{front}} + \sigma^2_{t,\mathrm{transport}}}. \tag{4.28}$$

For all events, the difference in azimuthal angles $\phi_{\mathrm{sim}} - \phi_{\mathrm{rec}}$ is determined. Then, the value for the angle difference containing 66 % of the events will be taken as the reconstruction uncertainty. This estimator behaves similarly to the standard deviation, but is more robust for distributions with long tails. The same procedure is followed for the zenith angle. The uncertainties will be compared to the results from Equations 4.22 and 4.23.

The first particle that traverses the detector determines the measured arrival time. In [116] the distribution of the arrival times of particles in the shower front is approximated by a normal distribution. The standard deviation of the distribution that arises from taking N samples is calculated and is equivalent to detecting N particles in a detector. The standard deviation is taken as an estimate for the uncertainty on the arrival time. The angular uncertainty can then be calculated using Equations 4.22 and 4.23.

Figure 4.11 presents the zenith and azimuthal reconstruction uncertainties as a function of the *minimum* number of particles in the three corner detectors, for 1 PeV

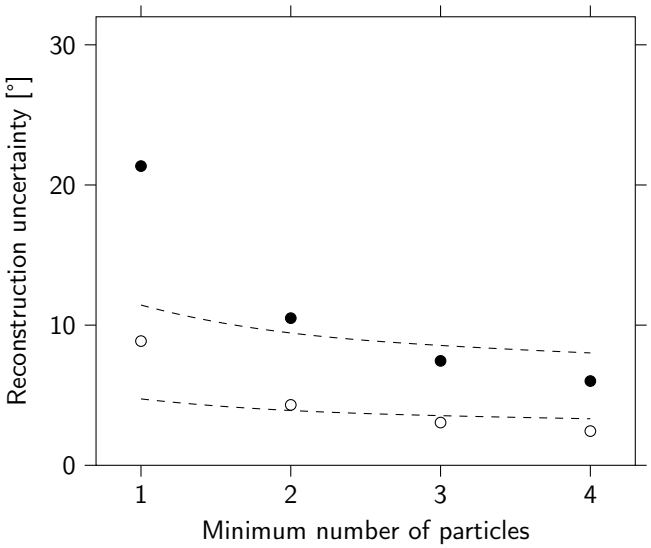


Figure 4.11 – Dependence of reconstruction accuracy on the minimum number of particles in all three corner detectors. Data for the most probable zenith angle ($\theta = 22.5°$) is shown. The open circles are the uncertainties for the zenith angle, the closed circles for the azimuthal angle. The dashed lines show the estimates for the uncertainties as calculated in [116], with a standard deviation of 1.8 ns.

proton showers with a zenith angle of 22.5°. The open circles are the results for the zenith angle reconstruction, and the closed circles depict the azimuthal angle. The dashed line is the calculation performed according to [116]. The results clearly show an increase in accuracy for larger particle numbers. When more than one particle hits a detector, the shower front is better defined. The probability that the first particle travels close to the causal front increases with the number of particles considered. However, the data points fall off much steeper than the calculated uncertainty. This is attributed to the fact that the shower front time structure can not be approximated by a normal distribution. As seen in Figure 2.3, the distribution has long tails. To better understand the effect of the front structure on the presented results, the simulated arrival time distribution is used as the basis for a Monte Carlo procedure. First, consider that the time structure of the shower front is a function of the core distance (Figure 4.5). Therefore, for each value of the minimum number of particles (N_{MIP}) in the detectors, the median core distance R_0 is determined (Figure 4.12). The arrival time of leptons in the simulated 1 PeV proton showers ($\theta = 22.5°$) is determined for core distances $R = R_0 \pm \Delta R$, with $\Delta R = 5\,\mathrm{m}$ chosen in accordance with the size of the HiSPARC station.

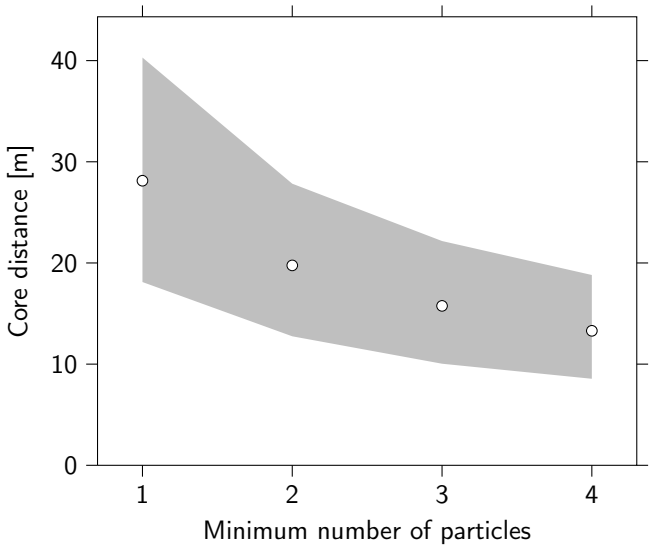


Figure 4.12 – Distribution of core distances as a function of the minimum number of particles in all three corner detectors. Events with a larger number of detected particles are generally found closer to the shower core. The results are obtained from showers produced by 1 PeV protons with $\theta = 22.5°$.

The arrival times are corrected for the passage of the causal front (Equation 4.9). The arrival times of particles in different parts of the shower can thus be compared to each other. For events with at least two particles in the three corner detectors, the median core distance is $R \approx 20\,\mathrm{m}$. The arrival time distribution is given in Figure 4.13 (gray line). Using this distribution as the basis for a Monte Carlo, we can simulate the arrival times of any number of particles in a detector. Drawing two random numbers from this distribution and taking the smallest value as the arrival time measurement in the detector, and repeating this process, one obtains the distribution depicted as the black line in Figure 4.13. The median value of this distribution is used as the uncertainty in the timing measurements. For $N_{\mathrm{MIP}} \geq 2$, one finds $\sigma_{t,\mathrm{front}} = 1.4\,\mathrm{ns}$. The total timing uncertainty is thus given by:

$$\sigma_t = \sqrt{\sigma_{t,\mathrm{front}}^2 + \sigma_{t,\mathrm{transport}}^2} = \sqrt{1.4^2 + 1.2^2} = 1.8\,\mathrm{ns}. \tag{4.29}$$

Repeating this procedure, the solid line in Figure 4.14 is obtained. These results reproduce the uncertainties more accurately than the approximation discussed in [116]. For $N_{\mathrm{MIP}} \geq 1$, however, the reconstruction uncertainties are smaller than estimated.

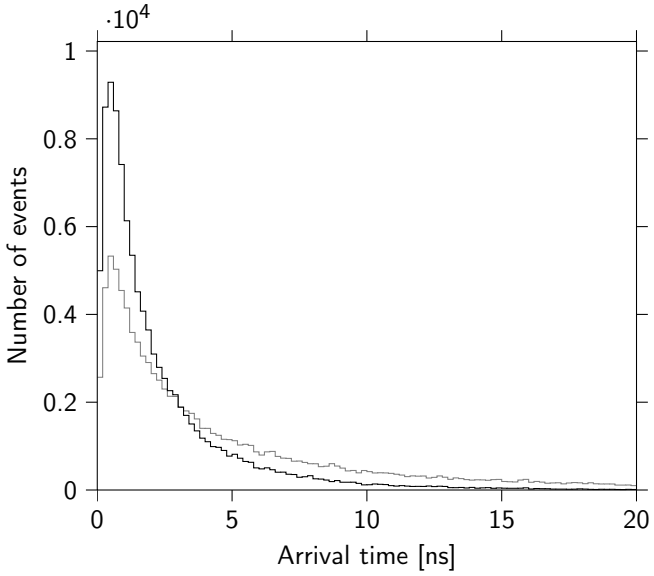


Figure 4.13 – Time structure of the shower front (gray). This distribution is used as the basis for a Monte Carlo. The black line shows the distribution resulting from taking the first particle (out of two) arriving in a detector.

This is attributed to the fact that events with more than one particle in some (but not all) detectors have a better timing accuracy, which is not taken into account in the estimation. The discrepancy is large for $N_{\mathrm{MIP}} \geq 1$, since the slope in the uncertainty is very large for small numbers of particles.

In the following discussion the reconstruction will be restricted to at least 2 particles in each detector ($N_{\mathrm{MIP}} \geq 2$). This reduces the effect of fluctuations in the shower front for large core distances. The timing uncertainty will thus be taken to be $\sigma_t = 1.8\,\mathrm{ns}$. This value is of the order of the uncertainties discussed in Section 4.3.1. It is therefore justified to use the plane front approximation.

Figure 4.15 shows the reconstruction uncertainty as a function of zenith angle, for 1 PeV proton showers. While the uncertainty in the zenith angle does not differ significantly for vertical and inclined showers, the uncertainty in the azimuthal angle diverges. It should be noted that for small zenith angles, two points with a large difference in azimuthal angle may actually be close together in terms of angular distance. Estimates for the reconstruction uncertainties are calculated using Equations 4.22 and 4.23. The estimates are calculated by averaging over all azimuthal angles. Data and simulation are in good agreement.

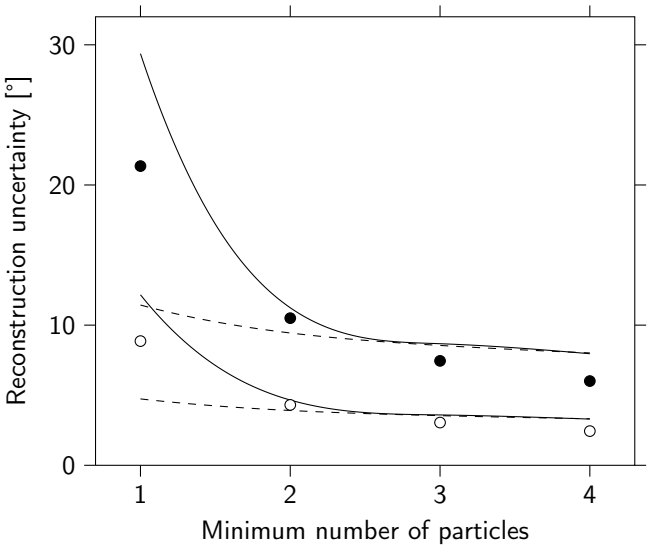


Figure 4.14 – Dependence of reconstruction accuracy on the minimum number of particles in all three corner detectors. Data for the most probable zenith angle ($\theta = 22.5°$) is shown. The open circles are the uncertainties for the zenith angle, the closed circles for the azimuthal angle. The dashed lines show the estimates for the uncertainties as calculated in [116], with a standard deviation of 1.8 ns. The solid lines show the estimates from the Monte Carlo distributions.

Figure 4.16 shows the dependence of the angular resolution on station size for 1 PeV proton showers with $\theta = 22.5°$. The resolution increases with increasing size of the station. Data and calculations are in agreement to better than 20 %. The results show that a detector distance smaller than 10 m would have a detrimental effect on the reconstruction uncertainties. The calculated uncertainty distribution flattens.

Sampling frequency limits the accuracy with which the analog signal can be digitized. To simulate the effect of sampling, discrete arrival times have been chosen corresponding with the sampling rate. The following sampling times are used: 0 ns (no sampling), 1 ns, 2.5 ns and 5 ns. The reconstruction uncertainty is shown in Figure 4.17. The width of the bins introduce an uncertainty $\sigma^2_{\text{bin}} = \frac{w^2}{12}$, with w the width of the bin. The total uncertainty is then given by

$$\sigma_t = \sqrt{\sigma^2_{t,\text{front}} + \sigma^2_{t,\text{transport}} + \frac{w^2}{12}}. \tag{4.30}$$

For the 2.5 ns resolution of the HiSPARC hardware, one obtains $\sigma_t = 2.0$ ns, only slightly larger than the total uncertainty with perfect time resolution. Consequently, the

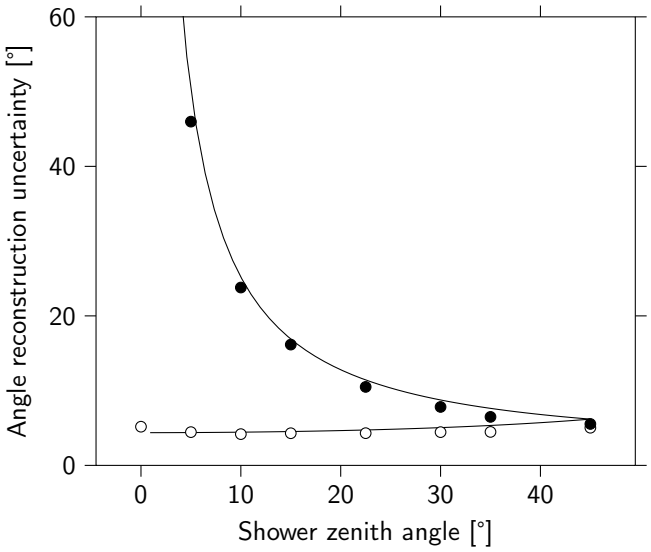


Figure 4.15 – Dependence of reconstruction uncertainty on zenith angle, for 1 PeV proton showers. A minimum of 2 particles is required in all three corner detectors. The data points show the uncertainties in the reconstruction of zenith angle (open circles) and azimuthal angle (closed circles). Estimates for the reconstruction uncertainties (solid lines) are calculated using Equations 4.22 and 4.23, by averaging over all azimuthal angles.

reconstruction uncertainties are only slightly affected. The results show that the 2.5 ns sampling time is an excellent choice.

4.5 Discussion and Conclusions

A 1 PeV proton shower has particle densities higher than $1\,\mathrm{m}^{-2}$ up to 30 m from the core (Figure 2.1). Defining the *footprint* of an EAS as the area containing particle densities higher than $1\,\mathrm{m}^{-2}$, such a shower has a footprint of $2.8 \times 10^3\,\mathrm{m}^2$. A HiSPARC station has an area of $43\,\mathrm{m}^2$, which is 1.5 % of this footprint. The detection area of the three corner detectors is only $1.5\,\mathrm{m}^2$ (0.053 % of the footprint). Using the limited information available, the reconstruction of shower direction is performed with surprisingly good accuracy. In some cases, the direction can be reconstructed for core distances significantly larger than 30 m.

It is shown that the uncertainty in the reconstruction of the azimuthal angle can be quite large. It is important to realize, however, that while the azimuthal angle may not be very accurately known, the effect of this on the position on the celestial sphere

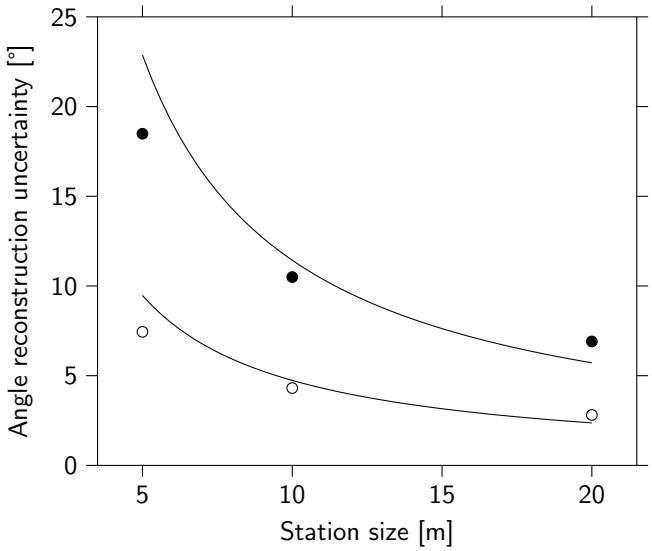


Figure 4.16 – Reconstruction uncertainty as a function of station size. A minimum of 2 particles is required in all three corner detectors. Data for the most probable zenith angle ($\theta = 22.5°$) is shown. The data points show the uncertainties in the reconstruction of zenith angle (open circles) and azimuthal angle (closed circles). The estimates for the uncertainties (solid lines) are calculated by averaging over all azimuthal angles.

decreases with decreasing θ. For small zenith angles, all possible directions on the celestial sphere are close together. A small uncertainty in the direction can thus result in a large uncertainty of the azimuthal angle. For two vectors which only differ in their azimuthal angles, $\Delta\phi = \phi_1 - \phi_2$, the angular distance is given by

$$\phi_{\text{dist}} = \Delta\phi \sin\theta. \tag{4.31}$$

For 1 PeV proton showers with a zenith angle of 22.5°, and $N_{\text{MIP}} \geq 2$, the uncertainty in the reconstruction of shower direction is determined to be $\sigma_\theta = 4.3°$ and $\sigma_\phi = 11°$. Using Equation 4.31, the angular distance between two directions with $\theta = 22.5°$ and separated by $\Delta\phi = 11°$ is only $\phi_{\text{dist}} = 4.0°$. The angular distance between two directions separated by $\Delta\phi = 11°$ and $\Delta\theta = 4.3°$, then becomes $d = 5.9°$. In other words, for $\theta = 22.5°$ and $N_{\text{MIP}} \geq 2$, 66 % of all EAS directions are accurately reconstructed to within 5.9° on the celestial sphere. Figure 4.18 shows the angular distance as a function of zenith angle, with the direction uncertainties calculated using Equations 4.22 and 4.23 with $\sigma_t = 1.8\,\text{ns}$. The results from the simulation are more accurate than the calculated estimates.

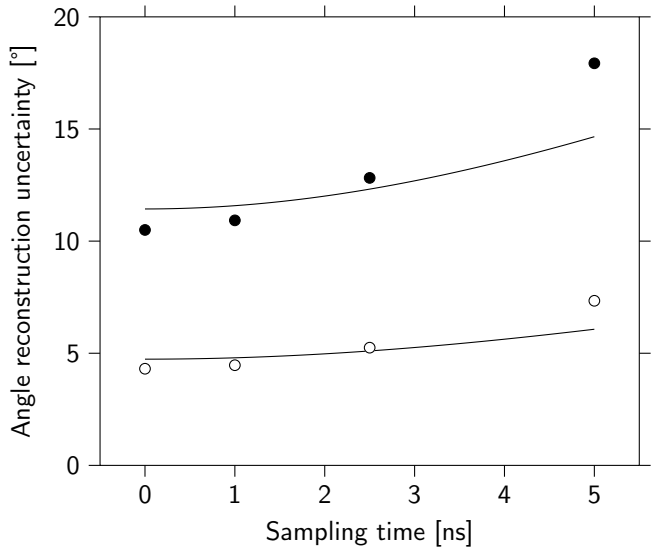


Figure 4.17 – Reconstruction uncertainty as a function of bin size. A minimum of 2 particles is required in all three corner detectors. Data for the most probable zenith angle ($\theta = 22.5°$) is shown. The data points show the uncertainties in the reconstruction of zenith angle (open circles) and azimuthal angle (closed circles). Estimates (solid lines) are calculated using $\sigma^2_{\text{bin}} = \frac{w^2}{12}$, with w the width of the bin.

There is a slight quantitative difference between calculations and simulations which suggests that a detailed analysis of the shower front is required. The arrival time uncertainty is assumed to be 1.8 ns for $N_{\text{MIP}} \geq 2$. In reality, this value depends on the distance of the detector to the shower core, as the shower front becomes thicker farther away from the core. To improve the understanding of the reconstruction uncertainties, the shower front arrival time distribution should be parametrized for several zenith angles, as a function of the core distance. This distribution should then be propagated using the detection efficiency to obtain an estimate which is approximately correct for all core distances and also yields a correct average over all events (i.e. detected showers). Figure 4.11 shows the reconstruction uncertainty as a function of the minimum number of particles in the detectors. In the analysis presented in this chapter, a value for the timing accuracy is determined for a median core distance R_0, which is taken from the simulation results. This value is used for the calculations. The discrepancy between calculation and simulation demonstrates that the uncertainties in the data cannot be fully described in this way and indicates that there must be another contribution which depends on the minimum number of particles per detector. This is most likely the

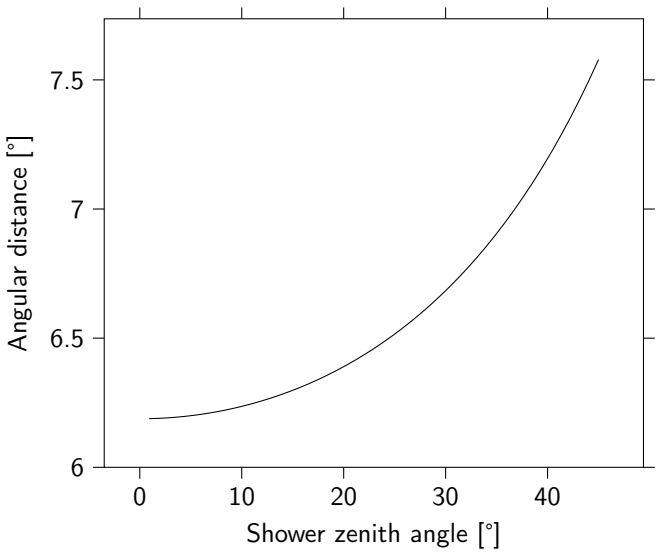


Figure 4.18 – Calculated estimate of EAS direction accuracy expressed as angular distance, as a function of zenith angle. The reconstructions are restricted to $N_{\mathrm{MIP}} \geq 2$, and all showers are generated by 1 PeV protons. 66 % of all EAS directions are accurately reconstructed to within the angular distance shown in the figure.

core distance, which is, on average, smaller for large numbers of particles per detector (Figure 4.12). This explains the discrepancy for larger numbers of particles. The first datapoint, at $N_{\mathrm{MIP}} \geq 1$, is also better than expected from calculations, but this was explained to be the result of more than one particle in some, but not all, detectors. This results in a more accurate reconstruction, not taken into account by the calculations.

The energy of the primary particle is also important. The particle density on the ground scales linearly with the primary energy and thus the maximum (and median) core distance for which EAS are detected depends on this energy. Furthermore, the particle density determines the number of particles in the detectors. It has been demonstrated that this is of great importance to the accuracy of the reconstructions.

The nature of the primary particle is also expected to play a role. The early development of EAS differs for protons and heavy nuclei. The effect of this on the arrival time distribution of particles in the shower front has not been studied. However, as discussed in Section 1.2, 84 % of the primary particles are protons. An additional 12 % is made up of helium nuclei. Therefore, the effect is expected to be minor.

Restricting the analysis to at least two particles in all three corner detectors significantly reduces the uncertainty in the arrival time distribution. The reconstruction

uncertainty as a function of zenith angle shows an excellent result (Figure 4.15).

The 10 m detector separation distance is a proper choice to obtain a good accuracy (Figure 4.16). Increasing the size of the station would make it difficult to fit the station on a typical roof, while the accuracy only slowly improves with size. A smaller station size, however, would be detrimental.

The 400 MHz sampling rate of the analog signal is also a proper choice. Sampling accuracy better than 2.5 ns only decreases the uncertainties slightly. However, 5 ns resolution is significantly worse.

A single HiSPARC station can be used to reconstruct the direction of EAS. In this chapter, an algorithm was developed to calculate the direction of EAS from relative differences in the measured particle arrival times in the individual detectors. The accuracy of the presented algorithm has been determined from the simulation results. To better understand the accuracy of the reconstruction, several sources of measurement uncertainties have been identified. The resulting timing uncertainties have been propagated through the analysis to obtain an estimate for the size of the reconstruction uncertainties. These calculations have been compared to the uncertainties determined using the simulation. Errors on errors were not considered. Howerver, the uncertainties are reasonably well understood.

Considering the cost of a single HiSPARC station (€ 10 000), the precision with which it determines the direction of EAS can compete with an air shower array. The reconstruction performs well for all azimuthal angles and all zenith angles considered in this analysis ($\theta < 45°$).

The next chapters will elaborate on the present analysis. Integrating a HiSPARC station in the KASCADE array enables a detailed performance study. In Chapter 5 simulation results of the direction reconstruction accuracy are compared to experimental data. This analysis will be extended in Chapter 6, in which a triangular cluster of three stations in the Science Park Array will be discussed.

5

HiSPARC at KASCADE

On July 1, 2008 a four-detector HiSPARC station was deployed at the KASCADE array at the Forschungszentrum Karlsruhe. In this chapter, a study on the performance of this HiSPARC station is presented. The HiSPARC station receives a trigger signal from the KASCADE array whenever the array detects an EAS. This enables the HiSPARC station to observe the same showers and the HiSPARC measurements are compared to the large and well-calibrated KASCADE array. Detector efficiency and the reconstruction of shower direction will be discussed. The dataset analyzed in this chapter was taken from July 1 to Aug 6, 2008, and contains more than 5×10^5 events.

5.1 KASCADE

The KASCADE experiment [117] is a well-studied and calibrated cosmic ray experiment which has taken data continuously since 1996. Officially, the experiment closed down early 2009, but it continues to provide data to guest experiments. In addition to HiSPARC, the LOPES [118] and TAUWER [119] experiments are external users of the KASCADE facility.

The array consists of 252 detector stations spaced 13 m apart on a square grid 195 m on a side. KASCADE has an area of 38 025 m^2 with an active detector area of 490 m^2 for e/γ and 622 m^2 for μ detection. For practical purposes, the array is divided into 16 clusters, each with its own DAQ electronics. The four inner clusters contain 15 stations

and each station has four e/γ detectors without a muon detector. The twelve outer clusters contain 16 stations each with two e/γ detectors and one muon detector. The e/γ detectors consist of a liquid scintillator (5 cm thickness) to measure the electromagnetic component of the shower. In the outer clusters, these detectors are placed on a lead/iron plate (10 cm Pb and 4 cm Fe, corresponding to 20 radiation lengths). A scintillator (3 cm thickness) is placed below this absorber to measure the muonic component. For details, see [117]. For a map of the KASCADE array, including the position of the HiSPARC station, see Figure 5.1. The station is shown in Figure 5.2.

All detector signals in the same cluster are read out by the DAQ electronics. These electronics are located in a shelter in the center of the cluster. If a signal goes over threshold, it is temporarily stored in a buffer. When a preset number of detectors go over threshold, a trigger is generated. For the inner clusters this number is 20 out of 60 e/γ detectors, whereas for the outer clusters a minimum of 10 out of 32 e/γ detectors is required. A trigger is distributed throughout the experiment and then the entire array is read out. All buffered detector signals are transferred to the central electronics. Each detector signal receives a timestamp from the cluster electronics. All clocks within the experiment are synchronized using fiber optic cables, carrying a 1 Hz and a 5 MHz signal. The timestamps are generated using these signals at 200 ns accuracy.

Showers are reconstructed in three stages using different algorithms [120]. First, the shower core position is determined by the center of gravity of the e/γ detector signals. The shower direction is determined by assuming a plane shower front. The shower size (number of electrons N_e and muons N_μ) is estimated by summing the detector signals (weighted by a factor which depends on the core distance). In the next step, the core position and electron shower size are used as a first guess in a fitting procedure. In this procedure, a revised core position and electron shower size are determined, as well as a shape parameter s, by fitting the lateral distribution function to the detector signals. The arrival times (for detectors within a distance of 70 m from the shower core) are fitted to a conical shower front. In the final step the detector signals are adjusted for expected contributions from other particles, and the fitting procedure is repeated.

Once the shower parameters have been determined, the particle densities can then be calculated at arbitrary locations and this is used to provide guest experiments with estimated particle densities at the location of their detectors. The particle densities are calculated on the shower front, not on the ground. This is crucial for inclined showers, see Figure 5.3. From the figure, it becomes clear that the core distance r that is being sampled by a detector is not the distance r' of the detector to the shower core on the ground. Furthermore, particles reaching the ground are distributed over a larger area

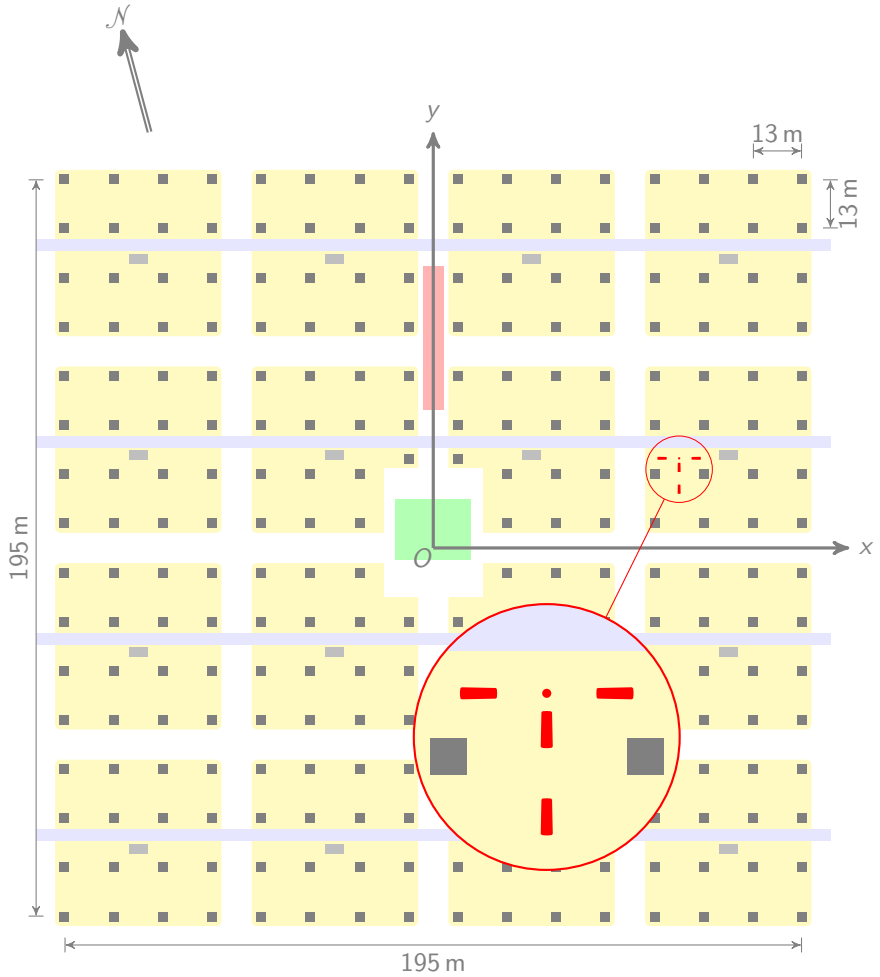


Figure 5.1 – Schematic layout of the KASCADE experiment, including the HiSPARC guest experiment. Green: the calorimeter. Light red: the muon tracker. Yellow: the sixteen array clusters. Dark gray: the array detectors. Light gray: the electronic huts. Bright red: the four HiSPARC detectors. Pale blue: access roads. The coordinate system used by KASCADE is shown in gray. For the present analysis, the KASCADE x, y coordinate system is used. Azimuthal angles are between $-180°$ and $180°$ with the positive x-direction at $0°$. The positive y-direction points to $15°$ North.

Figure 5.2 – HiSPARC station at the KASCADE experiment. The roof boxes contain the HiSPARC detectors. Several KASCADE detector huts are visible in the background. The large building in the background on the right is the central calorimeter and electronics building.

and thus the density ρ' is lower than the particle density on the shower front ρ. The zenith angle of the shower only has a small effect on the distribution of particle density on the shower front.

A HiSPARC station is installed inside the KASCADE array. The station consists of four scintillator detectors in the standard triangle setup shown in Figure 2.11. The KASCADE cluster in which the HiSPARC station is located will be referred to as the *local cluster*.

The HiSPARC station is not used in self-trigger mode. To observe the same showers, the cluster electronics provide a pulse whenever the KASCADE array is triggered. This signal is used by the HiSPARC electronics to read out the detectors. For each trigger, the HiSPARC GPS receiver provides a timestamp. This timestamp is used to synchronize with the KASCADE clock. The events detected by the HiSPARC station are reconstructed using the algorithm described in Chapter 4. The KASCADE array provides estimates of the particle densities at the location of the HiSPARC detectors, as well as the direction and primary energy of the shower. These measurements are used to determine the efficiency and accuracy of the reconstruction performed by HiSPARC.

5.1.1 Trigger Synchronization

While HiSPARC and KASCADE use different clocks, both clocks derive from the same time standard. By determining the offset between the two clocks, the corresponding events

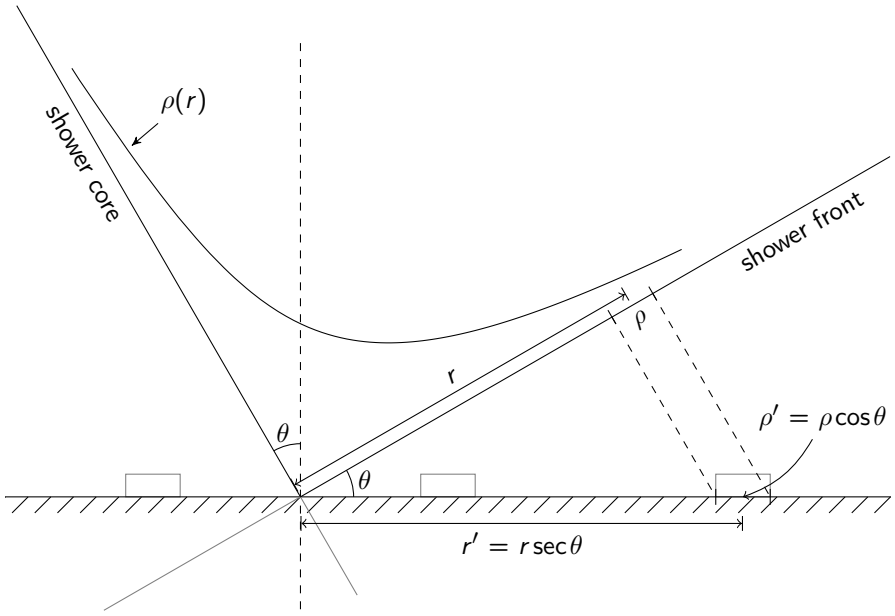


Figure 5.3 – Mapping of the lateral distribution of particle density of an angled shower to ground level density measurements. In the figure, ρ and ρ' illustrate the measurement of the particle density in an area occupied by a detector.

in the datasets can be found. Note that the KASCADE clock expresses time in UTC, while the HiSPARC clock uses GPS time.

The showers in the reconstructed dataset are a subset of the showers seen by the KASCADE array. In the following, the HiSPARC events will be referred to as *triggered events*, and the dataset provided by KASCADE, containing only reconstructed showers, will be referred to as *reconstructed events*. The occurence of triggered events follows the Poisson statistics. In particular, the time differences between consecutive events are independent random variables which are exponentially distributed.

To determine the clock offset the HiSPARC dataset, i.e. the triggered events, is shifted in time. For each event, the nearest-neighbor triggered event is determined. Figure 5.4 shows histograms of the resulting time differences for various timeshifts. When different

events are compared in the two datasets the resulting time differences are random and determined by the statistical nature of the triggered events. The time differences should thus follow an exponential distribution with rate parameter λ. This rate parameter is not equal to the rate of the triggered events. Since the procedure uses nearest neighbors, both preceding and following the reconstructed events, the mean time differences are halved and the rate parameter is *twice* the trigger rate. The figure shows that $\lambda = 7.2\,\text{Hz}$ which is consistent with the observed trigger rate of 3.6 Hz.

When the timeshift is close to the true offset between the clocks the correlation shows itself by a spike in the histogram of the time differences, resulting from correctly synchronized events. Uncorrelated events with time differences smaller than the remaining offset will still be selected as nearest neighbor. Hence, the data is random and follows Poisson statistics up to the observed spike. Uncorrelated events with larger time differences will not be selected as the nearest neighbor. Thus, the distribution is cut off after the spike. Optimizing the timeshift results in a single spike.

Figure 5.5 shows the residual time differences. The distribution on the left has a standard deviation of approximately 0.5 µs, which is not yet fully understood but is an unwanted property of the KASCADE hardware [121]. The shape of this distribution is already visible in a few minutes worth of data and does not result from changes in the clock offset. The clock offset slowly moves back-and-forth over the five-week period with a largest observed deviation of 8 µs. Since this happens slowly, it is possible to correct for this shift. The time difference distribution over the whole period results from a drifting offset of the KASCADE clock hardware. The offset is stable for periods of time, and then changes again.

The probability of one or more random triggered events occuring in a time window of 1 µs is 7.0×10^{-6}. Therefore, it can be concluded that triggered and reconstructed events are correctly synchronized and that the probability of incorrect matches is very low.

For this particular dataset, the timeshift which results in a mean time difference closest to zero, is found to be $\overline{t_K - t_H} = -13.180\,\text{s}$. The time difference between the HiSPARC clock (GPS) and the KASCADE clock (UTC) should be 14 s due to leap seconds.[1] The 820 ms difference is caused by the clock offset in the KASCADE experiment [121].

5.1.2 KASCADE Data

The KASCADE experiment provides a list of reconstructed events in which the local cluster took part. As previously mentioned, the test facility provides estimates of the

[1] Between January 1, 2006 and January 1, 2009 the offset between GPS and UTC clocks was 14 s, with UTC lagging behind GPS. At the end of December 31, 2008 a leap second was introduced to bring the offset to 15 s.

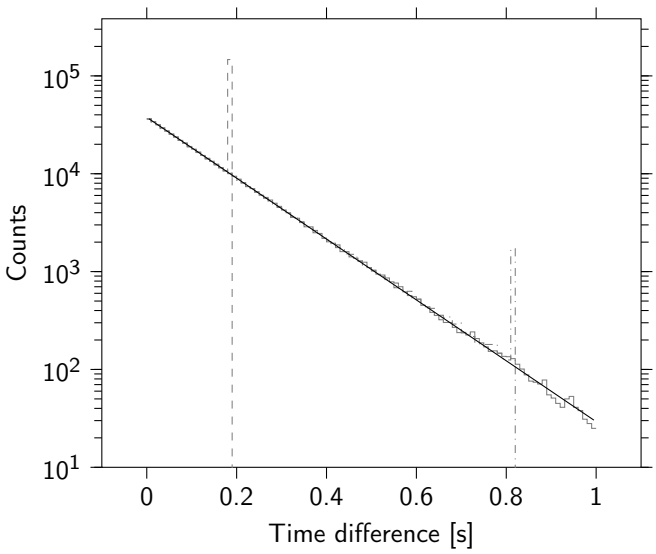


Figure 5.4 – The triggered data (HiSPARC clock) is shifted in time to synchronize with the reconstructed data (KASCADE clock). For each reconstructed event, the time difference with the nearest triggered event is determined. The time shifts are −12 s (solid gray line), −13 s (dashed gray line) and −14 s (dashdotted gray line). For uncorrelated events, the time differences follow Poisson statistics and hence give rise to an exponential distribution. The slope of the distribution is 7.2 Hz (solid black line), which is equal to twice the observed trigger rate of 3.6 Hz. If the time shift differs only slightly from the clock offset, partial matches give rise to spikes in the graph. These occur at the residual time difference to the correct clock offset. In that case, for each reconstructed event, there is a triggered event at precisely this time difference. There are uncorrelated events randomly occuring with smaller time differences, resulting in an exponential distribution up to the spike.

particle densities at the location of the detectors. In addition to that, shower direction, primary energy, core position and various atmospheric observables are provided. For details of the data format, see Table 5.1. The list is provided offline, and events in this list must be synchronized with the corresponding HiSPARC events.

5.2 Detector Efficiency

The dataset of synchronized events contains all showers on which the local KASCADE cluster has triggered. That does not mean that the HiSPARC station should have detected the shower. Many showers are very small. Larger showers may have a core position outside the cluster, with low particle densities at the position of the HiSPARC detectors. The HiSPARC station only covers 43 m^2 of the cluster, which has an area of 1521 m^2.

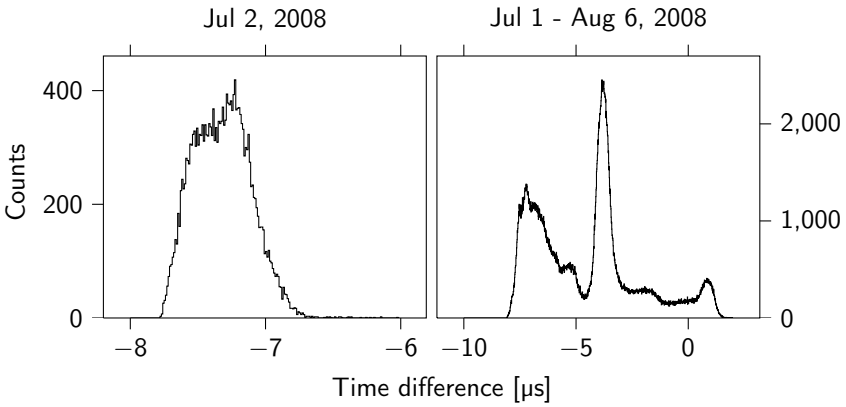


Figure 5.5 – Time differences of synchronized events showing the residual offset between HiSPARC and KASCADE clocks during one day (left) and over the whole five-week period (right). The distribution in the left graph does *not* result from a drifting clock offset. Every sample of events throughout the day shows this distribution.

Column name	Column description
IRUN	KASCADE run number
IEVE	KASCADE event number
GT	Timestamp (UTC)
MMN	Nanosecond part of timestamp
ENERGYARRAY	Primary particle energy estimation (eV)
XC	X coordinate of the shower core position
YC	Y coordinate of the shower core position
ZE	Zenith angle (rad)
AZ	Azimuth angle (rad)
SIZE	Total number of electrons in the shower
NMU	Total number of muons in the shower
HE0, . . . , HE3	Density of electrons at HiSPARC detector 1, . . . , 4 (m^{-2})
HMU0, . . . , HMU3	Density of muons at HiSPARC detector 1, . . . , 4 (m^{-2})
T200	Temperature at 200m above ground) (°C)
P200	Atmospheric pressure at 200 m above ground (hPa)

Table 5.1 – Format of the text file provided by KASCADE. Each row consists of the columns shown, in listed order. The temperature and atmospheric pressure are measured quantities, while all other physical observables are reconstructed. The particle densities are given on the shower front.

The efficiency of the detectors is determined by measuring their response to known particle inputs. On an event by event basis, the number of particles traversing each detector are unknown. However, estimated particle densities in the shower are provided by KASCADE. Given a particle density, the probabilities for an exact number of particles traversing a detector follow Poisson statistics. Let $\lambda = \rho A$ be the expected number of particles, with the particle density ρ and the detector area A. Then the probability of k particles hitting the detector is given by

$$P_k(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \tag{5.1}$$

It is very difficult to disentangle the contributions for 1, 2, 3, ... particles. It is much easier to distinguish between *no* particles and *any number* of particles. Given the particle densities ρ_e (for $e^{\pm}$) and ρ_μ (for $\mu^{\pm}$) on the shower front, the probability of any number of *charged* particles in a HiSPARC detector is given by

$$P_p(\rho_e, \rho_\mu, \theta) = 1 - e^{-A\cos\theta(\rho_e+\rho_\mu)}, \tag{5.2}$$

with θ the zenith angle of the shower and A the detector area ($0.5\,\mathrm{m}^2$).

Given a sufficiently large number of events, the fraction of events with charged particle content is determined. This fraction is the *probability* of finding charged particles in the data. By making cuts based on the particle density, the data can be compared to the probability distribution from Equation 5.2.

First, some parameters are determined from the complete dataset. Figure 5.6 shows the distribution of pulse integrals of a detector. Analogous to Section 2.4, the γ and charged particle contributions are determined. The γ contribution is fitted with a parametrized distribution inspired by the Monte Carlo described in [88]. The γ distribution is given by

$$N_\gamma(S) = a_\gamma S^{k_\gamma}, \tag{5.3}$$

with $N_\gamma(S)$ the number of events with the value of the integral equal to S, $a_\gamma = 4.477 \times 10^6$ and $k_\gamma = -1.106$. The distribution is cut off at 3 MIP. This value was chosen because the simulation shows that the γ contribution for higher energies can be neglected. In this range, Equation 5.3 is no longer valid. The charged particle contribution is determined by first subtracting the γ contribution from the spectrum. Then, the remaining spectrum is fitted with a Landau distribution convolved with a

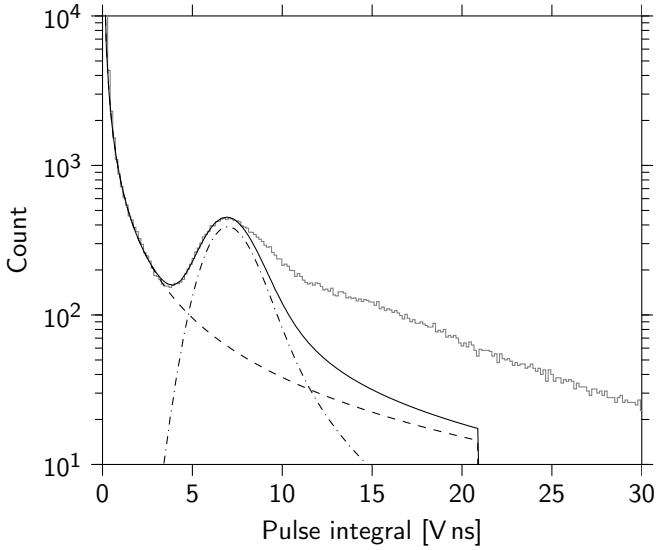


Figure 5.6 – Integral spectrum for five weeks of data (gray). The γ contribution is fitted with a power law, which is cut off at 3 MIP (dashed). The single charged particle contribution is fitted with a Landau distribution convolved with a normal distribution (dashdotted). The sum of the two distributions describes the data well up to the MIP peak (solid line).

normal distribution to account for the resolution of the detector:

$$N_{e/\mu}(S) = a_{e/\mu}(f \otimes g(\sigma_{\text{res}}))(C_{\text{MeV}}S), \tag{5.4}$$

with $f(\Delta)$ the Landau probability density function for an energy loss Δ, $g(\sigma_{\text{res}})$ the normal distribution with σ_{res} the resolution of the detector, and C_{MeV} the scale factor relating the signal S to the energy loss in MeV, i.e. $\Delta = C_{\text{MeV}}S$. The fit gives $a_{e/\mu} = 5.958 \times 10^3$, $C_{\text{MeV}} = 7.389 \times 10^{-4}\,\text{MeV}\,\text{nV}^{-1}\,\text{s}^{-1}$ and $\sigma_{\text{res}} = 6.767 \times 10^{-1}$. The γ and charged particle contributions taken together, explain the data up to the MIP peak. For higher energies, the contributions from multiple charged particles in the detector start to dominate.

Using $\rho'_{\text{charged}} = \cos\theta(\rho_e + \rho_\mu)$, with ρ'_{charged} the charged particle density *on the ground*, cuts are imposed on the data for a series of densities. The data surviving each cut is fitted with Equations 5.3 and 5.4. Only the count scale parameters a_γ and $a_{e/\mu}$ are left as free parameters in the fits on partial data. The number of charged particle events in the dataset is parametrized by Equation 5.4 for $S \leq 1\,\text{MIP}$. For $S > 1\,\text{MIP}$, one has to account for multiple charged particles in the detector. In this region, the number

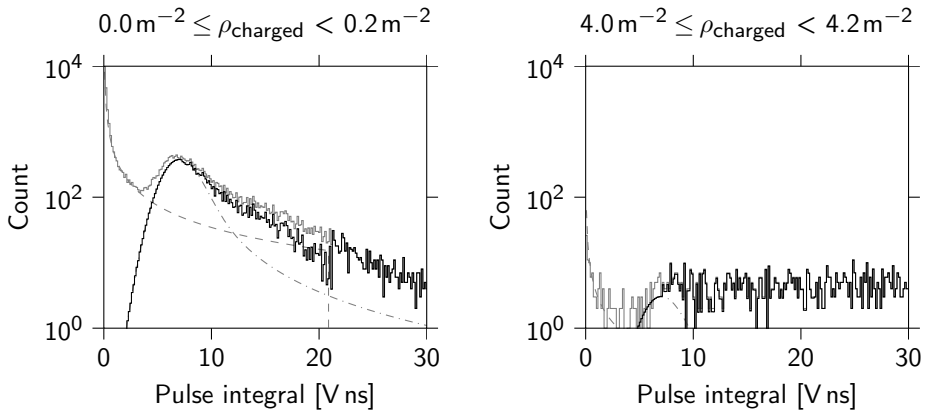


Figure 5.7 – The charged particle contribution to the spectrum (black solid line) is determined for various particle density ranges. Left: $0.0\,\mathrm{m}^{-2} \leq \rho_{\mathrm{charged}} < 0.2\,\mathrm{m}^{-2}$. Right: $4.0\,\mathrm{m}^{-2} \leq \rho_{\mathrm{charged}} < 4.2\,\mathrm{m}^{-2}$. The γ contribution (dashed) and single charged particle contribution (dashdotted) are shown for comparison. The charged particle contribution is parametrized by Equation 5.4 for $S \leq 1\,\mathrm{MIP}$. The data (gray solid line) with the γ contribution (Equation 5.3) subtracted, is used for $1\,\mathrm{MIP} < S \leq 3\,\mathrm{MIP}$. Finally, for $S > 3\,\mathrm{MIP}$, only the data is used to determine the charged particle contribution.

of charged particles are determined by the data, with the γ contribution (Equation 5.3) subtracted. Since the γ contribution is cut off at 3 MIP, the number of charged particle events higher than the cut-off is taken to be equal to the number of events in the data. See Figure 5.7. The probability of particles in the detector, for a given charged particle density ρ'_{charged}, is estimated by dividing the number of charged particle events by the total number of events.

Figure 5.8 shows the detection efficiency of a HiSPARC detector. The data are compared with the Poisson probability from Equation 5.2. At densities less than $0.5\,\mathrm{m}^{-2}$, the data points lie above the Poisson expectation. This suggests that there *are* charged particles, where KASCADE only reports few. Convoluting the Poisson probability with a normal distribution ($\sigma = 0.40$) accounts for uncertainties in the KASCADE particle densities. The actual value of this uncertainty is not communicated by the KASCADE experiment. While the data at low densities suggest the uncertainties are larger, the data at densities of $2\,\mathrm{m}^{-2}$ to $4\,\mathrm{m}^{-2}$ suggests that the uncertainties are quite small. Furthermore, the data at these densities shows that the present analysis overestimates the fraction of charged particles by a few percent. Overall, however, the data fits the Poisson curves well and the conclusion is justified that the detection of charged particles by the HiSPARC detectors is well understood.

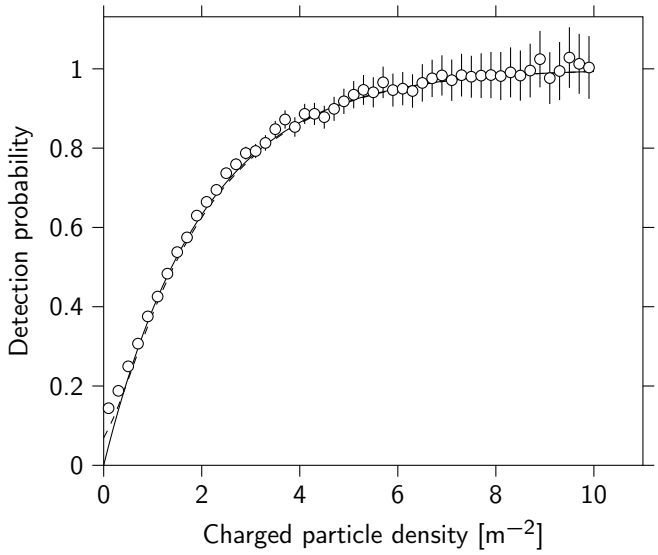


Figure 5.8 – Detection efficiency of a HiSPARC detector. Results from data (circles) are compared with Poisson probability (solid line), and Poisson probability convolved with a normal distribution ($\sigma = 0.40$) describing KASCADE density uncertainties (dashed line). For low densities, the data points are too high, suggesting that the density uncertainty is higher. For intermediate and high particle densities, the uncertainties seem to be lower, as the probability curve would go down with higher uncertainties.

5.3 Reconstruction of Shower Direction

To reconstruct the shower direction the algorithms and uncertainty estimation developed in Section 4.2 will be used. For this, the particle arrival times measured by the detectors are determined. Two different methods were used to obtain a value for the start of the pulse. The first method consists of simply taking the timing value from the first sample which goes over threshold (*first sample over threshold*, denoted by FSOT), where a low threshold is chosen (20 ADC, which is approx. 6 % of the MIP peak, but more than $28\,\sigma$ above the noise level). The second method interpolates linearly between the last sample below threshold and the first sample above threshold (*linear interpolation*, denoted by LINT), as shown in Figure 5.9. The latter method is expected to give a better estimate for the arrival time with a standard deviation smaller than $\sqrt{\frac{2.5^2}{12}} = 0.72\,\text{ns}$, which is the uncertainty introduced by the 2.5 ns sampling time.

Due to differences in the voltage applied to the phototubes a systematic time shift will be introduced since a higher voltage causes electrons to travel faster through the phototube. Furthermore, the transit time of the cables (30 m) can differ slightly. It is

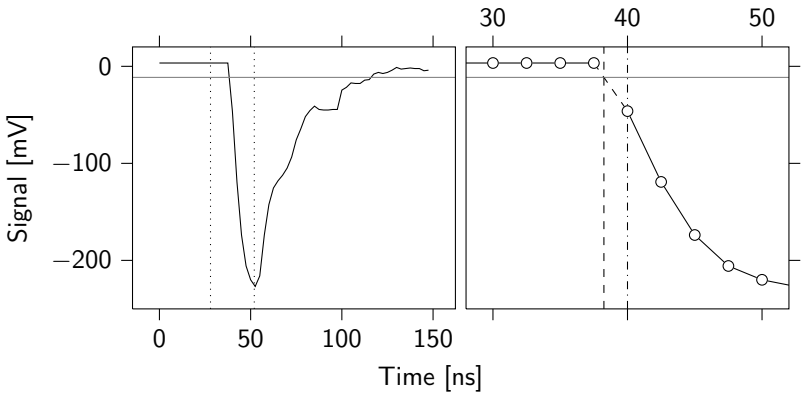


Figure 5.9 – Reconstruction of start-of-pulse time. The figure on the left shows the full signal, where the area between the dotted lines is enlarged and shown in the figure on the right. The individual samples are graphed as open circles. The threshold is indicated by the gray line. Two methods are used for determining the start-of-pulse time: first sample over threshold (dashdotted line) and linear interpolation of samples (dashed line). The difference between both values in this particular event is 1.6 ns.

Detector Pair	Offset [ns]
1 - 2	-0.36 ± 0.02
3 - 2	-1.43 ± 0.02
4 - 2	0.27 ± 0.02

Table 5.2 – Systematic time shifts in detector data, most probably due to transit time differences in the phototubes and signal cables. These values are obtained from the complete dataset (about five weeks), by fitting the data to a normal distribution.

possible to correct for this by plotting time differences between the central detector and all other detectors and measure the mean time difference (Table 5.2). The systematic errors are very small and certainly much smaller than the 2.5 ns sampling time of the electronics.

The distribution of propagation times of scintillation light photons to the PMT depends on the position of the particle in the scintillator. In [115] simulations of photon paths in a uniformly illuminated HiSPARC scintillator are performed. The standard deviation of the transport time for the first photon over threshold is found to be 1.2 ns (Section 4.1.2). Further uncertanties are introduced by the propagation time jitter of the PMT. The datasheet quotes 4 ns for a single photoelectron [83]. Since the trigger threshold is approximately equal to 15 photoelectrons ($\sigma_{t,\text{transport}} = 1.2\,\text{ns}$) [115], the

effect on the timing measurement uncertainty is expected to be smaller than 4 ns. The uncertainty in the arrival time measured by a single detector is taken to be:

$$\sigma_t = \sqrt{\sigma^2_{t,\text{front}} + \sigma^2_{t,\text{transport}} + \sigma^2_{t,\text{sampling}} + \sigma^2_{t,\text{other}}}, \tag{5.5}$$

with $\sigma_{t,\text{other}}$ the unknown uncertainties, including the PMT jitter and the thickness of the shower front. An unknown (with respect to the simulations) contribution of the shower front is expected since the HiSPARC station at KASCADE also measures EAS from primary particles other than protons, and at other energies than 1 PeV. Also, a full detector simulation with GEANT or FLUKA was not performed. The effects are expected to be minor. This unknown contribution will be determined from data, with the other contributions taken from Chapter 4.

The HiSPARC reconstructed shower direction is compared to the KASCADE data on an event by event basis. In the following analysis, experimental uncertainties (resolution) are determined by the difference in HiSPARC and KASCADE angle containing 66 % of the events, unless otherwise noted.

Figure 5.10 shows the azimuthal angle reconstruction for events which have at least a 1 MIP signal in all corner detectors. An excellent correlation between HiSPARC reconstruction and the KASCADE reference is observed. The resolution does not depend on the azimuthal angle. The two clusters in the upper-left and lower-right corners are over- and underflows at the $\phi = -180°/180°$ boundaries. Similarly, Figure 5.11 shows the comparison between zenith angles.

Systematic errors are shown in Figures 5.12 and 5.13. There is a bias in the azimuthal reconstruction towards 0°. For $\phi < 0$, the azimuthal angle tends to be overestimated, bringing it closer to zero. For $\phi > 0$, the azimuthal angle tends to be underestimated, again bringing it closer to zero. The explanation may be found in Table 5.2. It is clear that there is a slight bias towards *earlier* timings from detector 3. Referring to Figure 2.11 and Figure 5.1, it follows that earlier arrival times in detector 3 would suggest an azimuthal angle closer to 0°. Correcting for these time differences by subtracting the mean values from the measured timings, results in the right plot in the figure. The systematic effect is much reduced.

The systematic error in the zenith angle reconstruction shows a bias towards more inclined showers, especially for small zenith angles. This is a direct result from timing uncertainties. For a shower with zero inclination one should expect to have no time differences between the detectors. In reality, small time differences will exist and therefore, the reconstruction will give a (slightly) inclined shower. For larger zenith

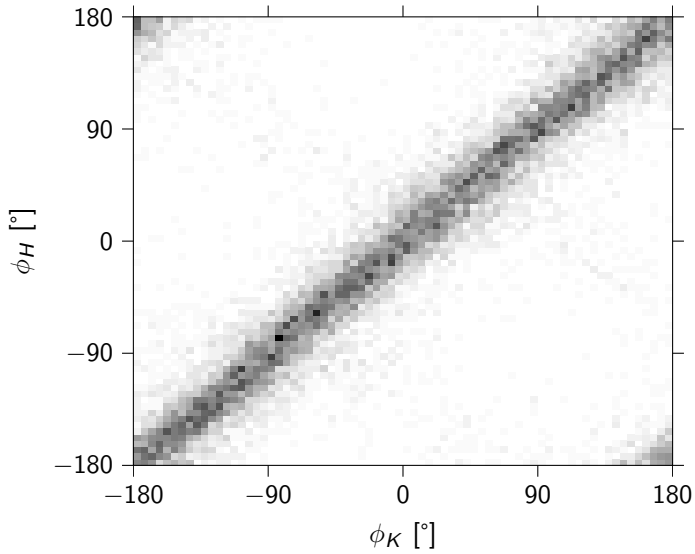


Figure 5.10 – Two-dimensional histogram of the azimuthal angle reconstruction. A single HiSPARC station is able to reconstruct the azimuthal angle of a shower. Only events with at least 1 MIP in all corner detectors are shown. ϕ_K is the azimuthal angle reconstructed by KASCADE, and ϕ_H is the HiSPARC reconstruction. For this plot, LINT timings were used to avoid artifacts from sampling.

angles, the zenith angle is overestimated by less than 5°. The reconstruction algorithm is based on the assumption that the shower front is a flat plane. The additional time lag introduced by the curvature is of the order of $0.14\,\mathrm{ns\,m^{-1}}$ [122], consistent with the simulations (Figure 4.5). This results in an error of a few degrees [122].

Next, following Section 4.4.2, the analysis will consider $N_{\mathrm{MIP}} \geq 2$. The uncertainty $\sigma_{t,\mathrm{other}}$ will first be determined from data.

The uncertainty in angle reconstruction as a function of zenith angle is shown in Figure 5.14. The zenith reconstruction does not depend much on zenith angle. While the azimuthal reconstruction appears to depend heavily on zenith angle, this is not entirely accurate. For small zenith angles, two points with very different azimuthal angles may still be close together, in terms of angular distance.

The solid lines follow from the calculations discussed in Section 4.3. The value of σ_t is determined to be equal to 2.4 ns by means of a fit. Using Equation 5.5 and the values from Equations 4.29 and 4.30 one obtains $\sigma_{t,\mathrm{other}} = 1.6\,\mathrm{ns}$. This contribution is large compared to the other contributions, but smaller than the 4 ns single-electron jitter reported for the PMT.

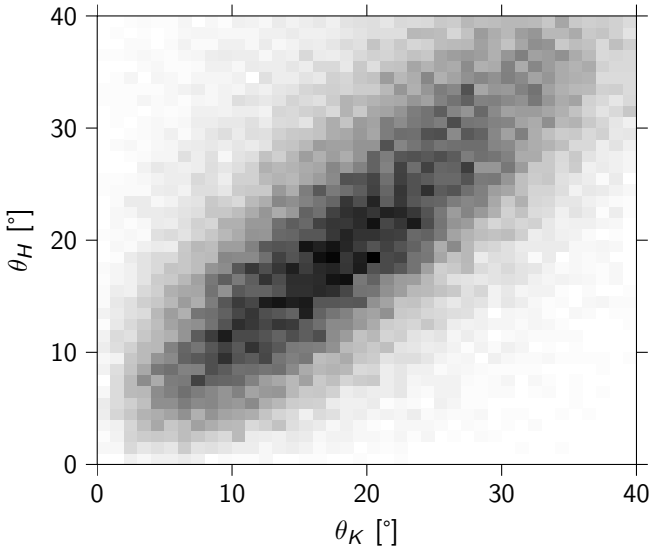


Figure 5.11 – Two-dimensional histogram of the zenith angle reconstruction. A single HiSPARC station is able to reconstruct the zenith angle. Only events with at least 1 MIP in all corner detectors are shown. θ_K is the zenith angle reconstructed by KASCADE, and θ_H is the HiSPARC reconstruction. For this plot, LINT timings were used to avoid artifacts from sampling.

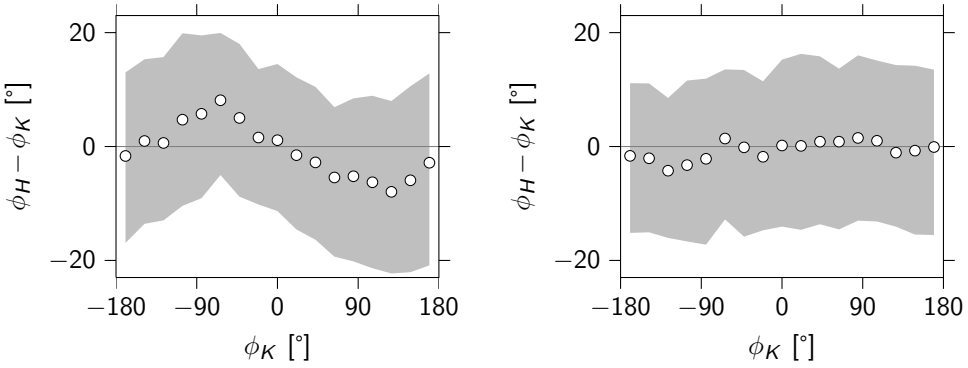


Figure 5.12 – The median uncertainty (open circles) of the azimuthal angle reconstruction as a function of the azimuthal angle clearly shows some systematic effects. The shaded region contains 50 % of the events equally distributed around the median. Left: FSOT timings. Right: FSOT timings, corrected for the observed offsets from Table 5.2.

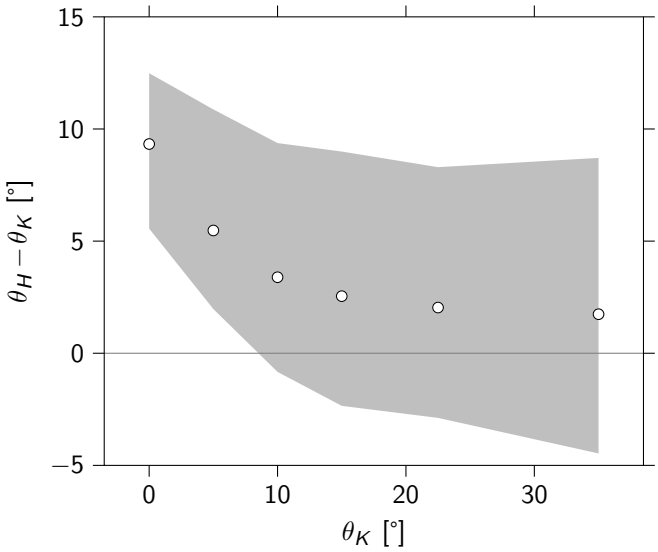


Figure 5.13 – The median uncertainty (open circles) of the zenith angle reconstruction as a function of the zenith angle shows a bias towards more inclined showers. The shaded region contains 50 % of the events equally distributed around the median.

The estimated uncertainty describes the data well over the range of zenith angles. This is true for both the zenith and azimuthal angles. The data from the simulation (Figure 4.15) does not include the experimental uncertainty of $\sigma_{t,\text{other}} = 1.6\,\text{ns}$. Therefore, the uncertainties are smaller than those from the HiSPARC/KASCADE comparison.

Figure 5.15 shows the uncertainties as a function of number of particles in the HiSPARC detectors. Data points from this analysis and the simulations are shown for comparison. The solid lines show the calculated uncertainties. The simulation underestimates the reconstruction uncertainties.

Figure 5.16 depicts the relation between N_{MIP} and shower core distance. Large numbers of particles in a detector are usually observed close to the shower core. Experiment and simulation show the same dependence. However, the median core distance is smaller in the experiment. This may be due to the experimental cuts imposed on the KASCADE analysis, in particular the requirement that the EAS should be observed by the local cluster. Not all particles are found close to the causal front and the arrival time spread increases with increasing core distance (Figure 5.17). The arrival time spread has been estimated by taking the arrival time differences between detectors 1 and 2. Data and simulation are in agreement for core distances larger than 30 m. For smaller core distances, the effect of the transport time of photons in the detector dominates

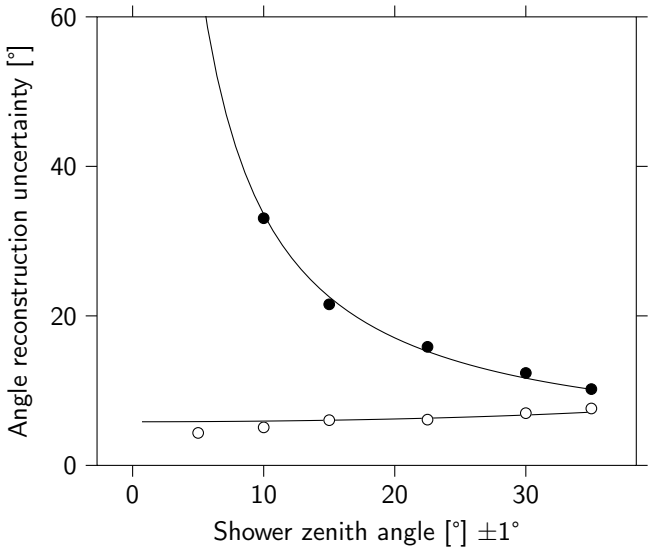


Figure 5.14 – Uncertainty in angle reconstruction as a function of the zenith angle, for $N_{\mathrm{MIP}} \geq 2$ and $14.5 \leq \lg E \leq 15.5$. The data points show the uncertainties in the reconstruction of zenith angle (open circles) and azimuthal angle (closed circles). Estimates for the reconstruction uncertainties are represented by the solid lines.

the measured time differences. The larger spread in arrival time with increasing core distance means a larger uncertainty in arrival time measurements and thus a larger uncertainty in angle reconstruction (Figure 5.18).

Two methods for determining the arrival time of a particle in a HiSPARC detector are applied. In the previous analysis, the FSOT timings are used. In Figure 5.19 the analysis is repeated using LINT timings. Differences are very small and no significant improvements of using linear interpolation are observed.

5.4 Discussion and Conclusions

The integration of a HiSPARC station into the KASCADE array allows for detailed performance studies. The detection efficiency is consistent with expectations from Poisson probability distributions for particle densities larger than $0.5\,\mathrm{m}^{-2}$. At lower particle densities, the detection efficiency is better than expected. KASCADE does not provide an estimate of the uncertainty on the particle densities, and thus the effects of local density fluctuations can not be estimated. The KASCADE collaboration has agreed to release a dataset of the individual KASCADE detector measurements for future studies [123].

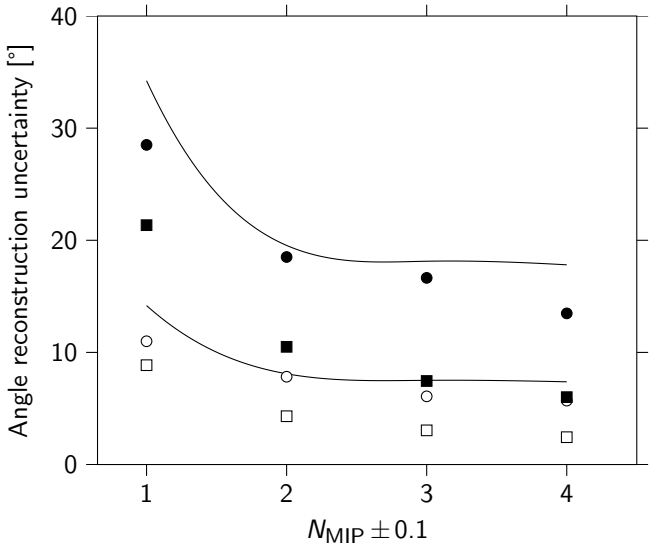


Figure 5.15 – Uncertainty in angle reconstruction as a function of the number of particles in the HiSPARC detectors, for $\theta = 22.5° \pm 5$ and $14.5 \leq \lg E \leq 15.5$. Data points from experiment (circles) are shown next to simulation (squares). The symbols show the uncertainties in the reconstruction of zenith angle (open) and azimuthal angle (closed). The calculated uncertainties are depicted by the solid lines. The timing uncertainty is larger in the experiment than in the simulation.

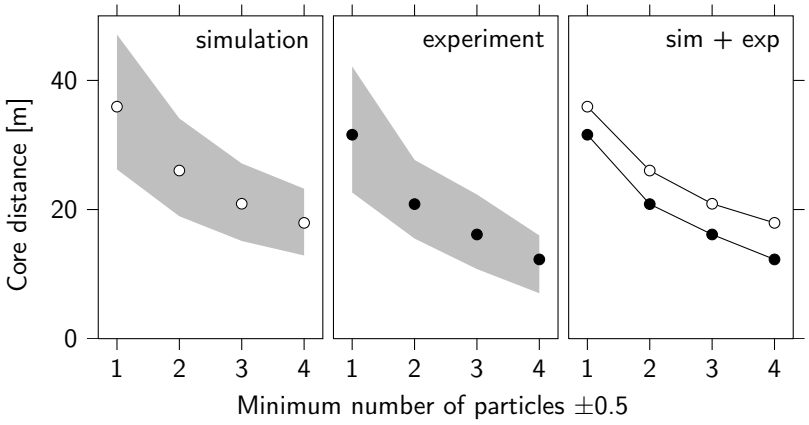


Figure 5.16 – Core distance as a function of the number of particles in the HiSPARC detectors for the simulation (open circles), and KASCADE data (closed circles). The dots show the median values. The shaded region contains 50 % of the events equally distributed around the median. The third panel is included for easy comparison of the median values. A larger number of particles implies a smaller core distance and thus increased timing accuracy due to a thinner shower front.

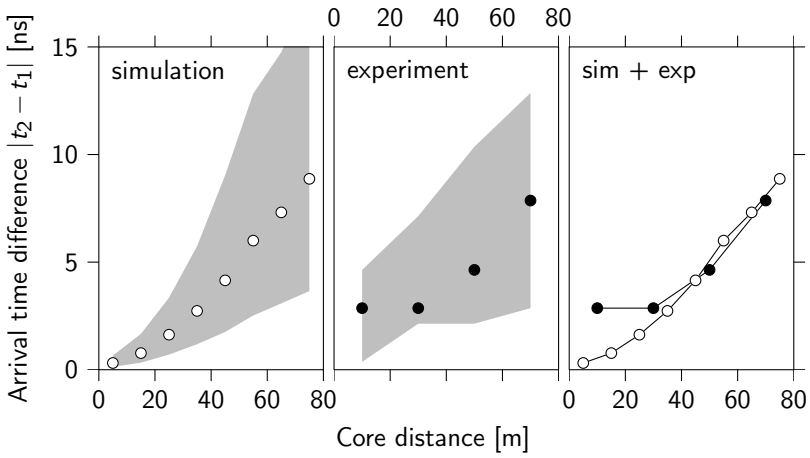


Figure 5.17 – The spread in particle arrival times increases with increasing shower core distance. In other words, the shower front is thicker further away from the shower core. For both simulation and KASCADE data, data points are taken by calculating detected particle arrival time relative to the station's center detector. The dots show the median values. The shaded region contains 50 % of the events equally distributed around the median. The third panel is included for easy comparison of the median values.

Direction reconstruction of EAS using a single HiSPARC station is surprisingly accurate, when taking the size of a single station into account. Systematic reconstruction errors are explained by transit time differences in the PMTs and cables. These errors have been corrected for (Figure 5.12). Statistical errors on the data points are inconsequential.

Uncertainties in angular reconstruction are due to timing uncertainties ($\sigma_t = 2.4\,\text{ns}$), caused by the thickness of the shower front, and by an as yet unexplained contribution with a standard deviation of 1.6 ns. The PMT jitter may explain part of this uncertainty, while a full detector simulation using a simulation package such as GEANT or FLUKA may further explain this discrepancy.

Uncertainties are compared to calculations. HiSPARC and KASCADE reconstruction of the EAS direction is in agreement. Deviations from calculated uncertainties are explained by taking into account secondary effects of the shower front structure. Figure 5.17 demonstrates that the shower front thickness (increasing with core distance) is evident in the HiSPARC data, and is consistent with the simulations. Using linear interpolation (LINT) to determine the arrival time of particles yields no significant improvement over first sample over threshold (FSOT) timings.

For the most probable zenith angle $\theta = 22.5°$ and $N_{\text{MIP}} \geq 2$ the observed uncertainties

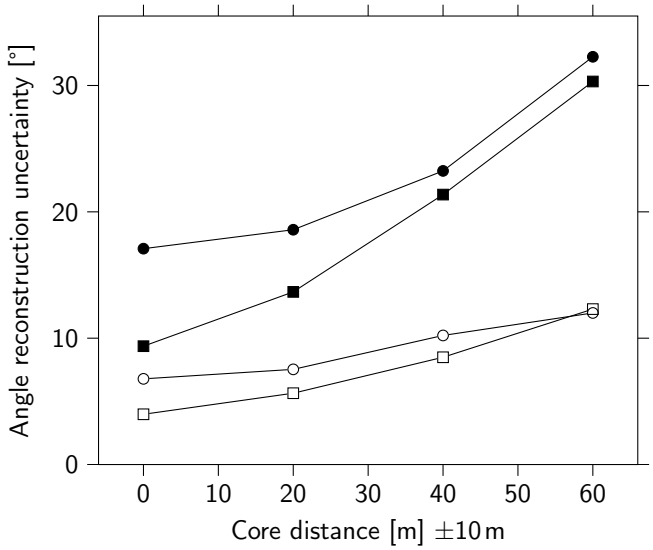


Figure 5.18 – Uncertainty in angle reconstruction as a function of shower core distance, for $N_{\mathrm{MIP}} = 2 \pm 0.5$, $\theta = 22.5^\circ \pm 5^\circ$ and $14.5 \leq \lg E \leq 15.5$. Data points from experiment (circles) are shown next to simulation (squares). The symbols show the uncertainties in the reconstruction of zenith angle (open) and azimuthal angle (closed). The data points are connected to guide the eye. A larger core distance implies a thicker shower front and thus a larger uncertainty in the reconstruction.

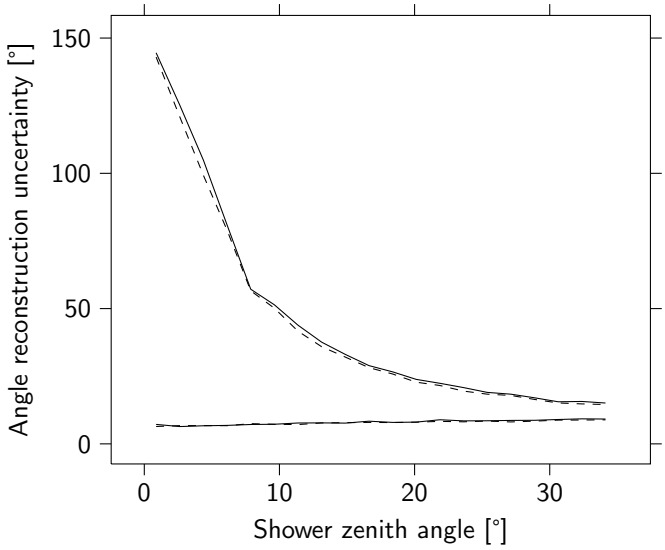


Figure 5.19 – Comparison of results using *first sample over threshold (FSOT)* timings (solid) and *linear interpolation (LINT)* timings (dashed). Note that the differences are very small.

are $\sigma_\theta \simeq 6.1°$ and $\sigma_\phi \simeq 15.9°$ (Figure 5.14). Following the discussion in Section 4.5, the angular uncertainties are calculated to be $\phi_{\text{dist}} = 6.1°$, and $d = 8.6°$. Thus, under these conditions, 66 % of all reconstructed EAS directions are accurate to within 8.6° on the celestial sphere, which demonstrates the excellent performance of the HiSPARC station.

The performance of a single HiSPARC station is well understood. This performance will now be used in the study of a triangular setup (equilateral triangle with sides of $\approx 130\,\text{m}$) of three stations within the Science Park array (Chapter 6). The direction reconstruction of single stations will be compared to the direction reconstruction of the large triangle.

6

The Amsterdam Science Park Array

6.1 Introduction

The *Amsterdam Science Park Array* is a cluster of eight HiSPARC stations located on the roofs of several institutes and businesses. One station is located inside Nikhef and is prominently displayed in the main atrium. The largest distance inside the cluster is 407 m. Each station has four detectors. The stations are placed in the configuration shown in Figure 2.11. Figure 6.1 shows the locations of the stations in the array.

6.2 Coincidences

The stations in the array have been configured to use the trigger conditions discussed in Section 2.3.2:

- low trigger threshold at −30 mV (requiring *three* detectors over threshold)
- high trigger threshold at −70 mV (requiring *two* detectors over threshold)

Either condition, when satisfied, will result in a trigger. The high threshold corresponds to approximately 0.5 MIP. The stations have a trigger rate of 0.6 Hz to 0.8 Hz. All recorded events are sent to the datastore at Nikhef.

From Figure 2.1 it follows that extensive air showers with energies of 10 PeV or more have a sufficient footprint to be observed by stations with a separation distance of

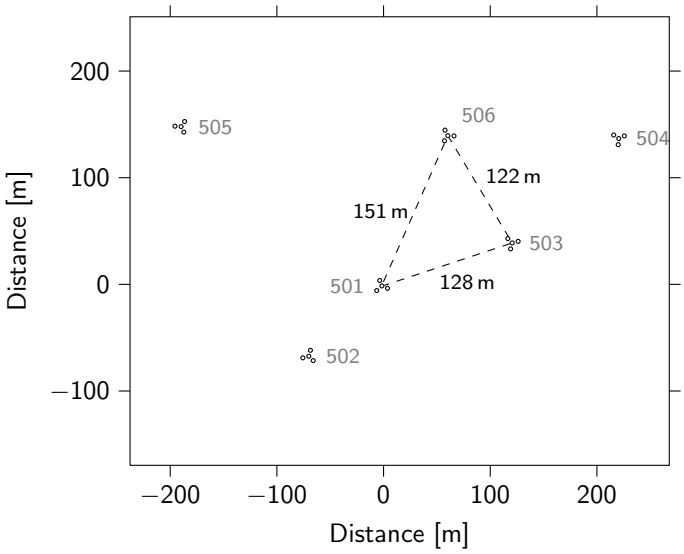


Figure 6.1 – The Science Park Array. Each station has four detectors. Individual detectors are represented by dots. The stations are placed in the configuration shown in Figure 2.11. The analysis presented in this chapter is based on data taken with three stations (501, 503 and 506) which form an almost equilateral triangle with sides ranging from 122 m to 151 m. The station at the University of Amsterdam is located outside the figure.

the order of 100 m. Searching for these *coincidences* is based on the event timestamps generated by the GPS timing modules. The time difference resulting from a horizontal shower in the direction of the largest distance in the cluster is given by $\Delta t = s/c = 407\,\text{m}/c = 1.36\,\mu\text{s}$. Events are considered part of a coincidence if their time difference is less than 1.36 µs.

The cosmic ray spectrum is a power law of the primary energy (Figure 1.2), and the footprint of EAS is a function of this energy. Therefore, the number of EAS of given size (or energy) observed by a pair of stations is a function of the distance between the stations. The cosmic ray flux decreases with energy, so EAS with small footprints are more abundant than EAS with large footprints. Thus, more coincidences should be observed for pairs of stations which are in the vicinity of each other. Figure 6.2 shows the number of coincidences for all pairs of stations in the array. The expected number of coincidences (depicted as the solid line in the figure) is calculated as follows. First, the flux of EAS is approximated by $F = E^{-2.7}$, with E the energy of the primary particle. Then, particle densities for a range of core distances are estimated using a lateral distribution function, given in [124]. The detection probability is finally estimated using

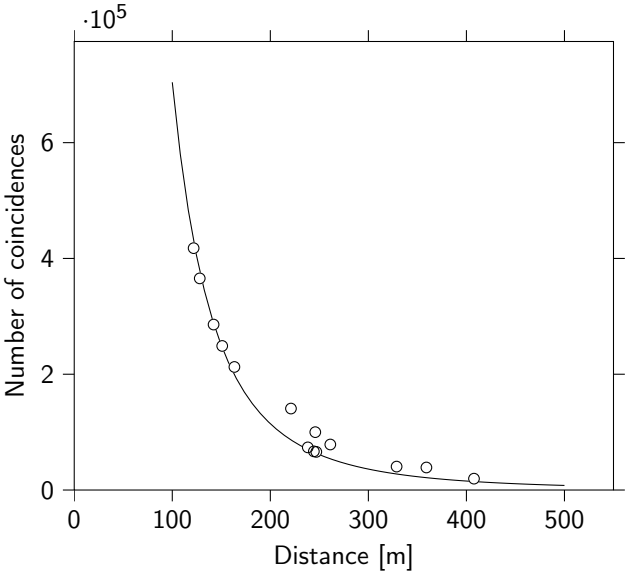


Figure 6.2 – The number of coincidences between pairs of stations as a function of their separation distance. The solid line is an estimate for the number of coincidences, summed over all energies. The cosmic ray flux decreases with energy, so EAS with small footprints are more abundant than EAS with large footprints. This is reflected in the above results. More coincidences are observed when the distances are small, due to the many showers that are not sufficient in size to trigger stations farther away.

Poisson statistics. The probability of detection is then weighted with the flux, resulting in an estimate summed over all energies.

There is no large background of random coincidences obvious in Figure 6.2. The expected number of random coincidences between two stations can be calculated using

$$N_r = 2 f_1 f_2 \tau T, \tag{6.1}$$

with f_1 and f_2 the trigger rate of the two stations, τ the coincidence time window and T the duration of the measurement. For a trigger rate $f_1 = f_2 = 0.8\,\mathrm{Hz}$, $\tau = 1.36\,\mu\mathrm{s}$ and $T = 1\,\mathrm{d}$, we have $N_r = 0.15$ random coincidences per pair of stations for a day, which amounts to approximately 2 random coincidences per day for the six stations (15 pairs) considered in this analysis.

The GPS locations of the stations have to be accurately determined. Firstly, the locations define the relative positions of all detectors in the analysis. Secondly, inaccurate positions inhibit the ability of the GPS receiver to accurately determine the current

station	latitude [°]	longitude [°]	altitude [m]	orientation [°]
501	52.355 924	4.951 144	56.1	135
502	52.355 293	4.950 105	56.0	345
503	52.356 255	4.952 944	51.6	45
504	52.357 179	4.954 384	54.6	175
505	52.357 252	4.948 401	47.7	86
506	52.357 179	4.951 986	43.9	267

Table 6.1 – GPS locations of the stations in the Amsterdam Science Park Array. These are the results of a self-survey of 86 400 fixes during a full day, given to 10 cm accuracy. The orientation of a station with respect to "North" is measured using a compass to an accuracy of 1°.

time. All stations have performed an automatic self-survey of 1 d (86 400 fixes). The location of the stations are listed in Table 6.1.

The accuracy of the GPS timing module has been discussed in Section 2.3.3. Since there is an offset of tens of nanoseconds between the GPS setups, a correction must be made. First, the offset has to be determined from the data. Station 501 is chosen as a reference. For each station, the coincidences with station 501 are analyzed. The differences of the timestamps are the result of both the arrival time differences of particles in the EAS, and the differences between the GPS modules. Summed over a large number of EAS, the distribution should be centered around zero, since the arrival directions are isotropic. The offset of the GPS modules, however, causes an equal offset in the timestamps. Figure 2.22 shows the distribution of time differences between stations 501 and 502. The GPS offset is measured to be 15 ns.

6.3 Reconstruction of Shower Direction

The direction of an EAS is reconstructed using the same method and algorithm discussed in Section 4.2. This implies that for each reconstruction, three measurements are used. In the case of a cluster, this means that three stations are used to reconstruct the direction of the shower.

The analysis presented in this section is based on data taken from January 1, 2012 to April 1, 2012, with three stations (501, 503 and 506) which form an approximately equilateral triangle with sides ranging from 122 m to 151 m.

First, the performance of the single stations is checked. EAS directions are reconstructed using data from a single station only. Figure 6.3 shows the reconstructions of shower direction for the stations 501, 503 and 506. The azimuthal distribution is flat,

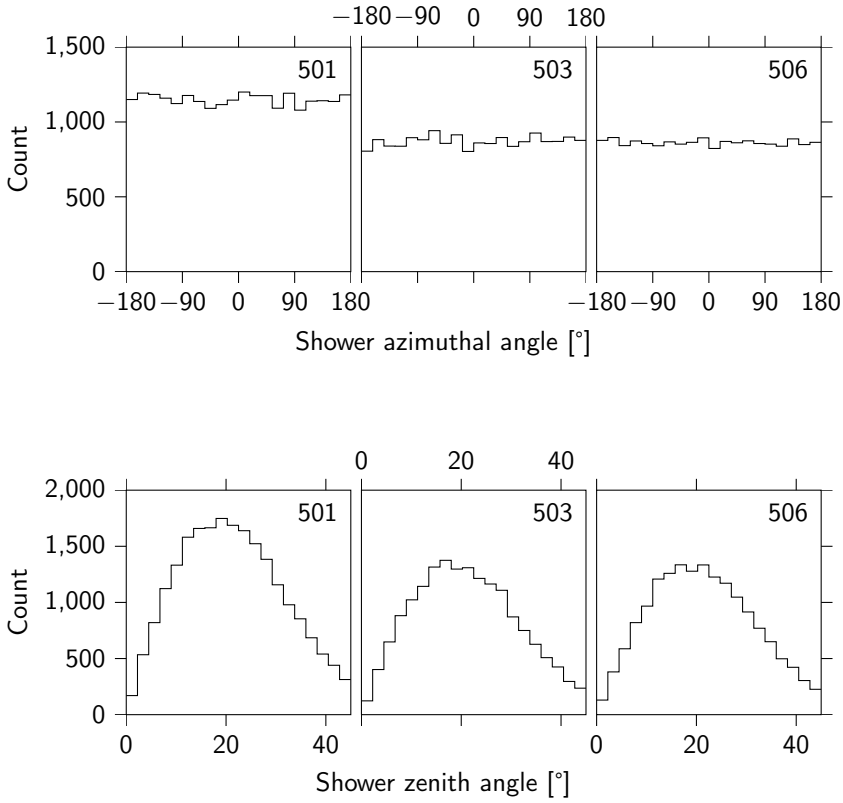


Figure 6.3 – Reconstruction of shower direction for single stations. Results are shown for stations 501, 503 and 506. The azimuthal angle ϕ shows a flat distribution, while the zenith angle θ shows a distribution peaked around 20°.

consistent with an isotropic distribution of the direction of EAS. Deviations from the flat distribution can be explained by statistical uncertainties in the distribution of arrival directions, and are not the result of anisotropy. The zenith distribution shows few showers with small zenith angles, and peaks around 20°. For larger zenith angles, the acceptance of the detectors decreases. Furthermore, since slanted showers traverse a larger atmospheric depth, fewer particles reach the ground. This reduces the probability of detection. The difference in the number of events is due to differences in PMT voltage settings. As a result of this, the trigger levels are at different values relative to the MIP peak. We conclude that the stations perform well.

Coincidences of three or more stations are selected from the complete dataset. Each event in the coincidence is analyzed using a three-step process to obtain an accurate

arrival time for the first particle in the station. Firstly, the linear interpolation method (LINT), discussed in Section 5.3, is used to determine the arrival time of particles in the detectors. Secondly, the event timestamp is corrected so that it no longer defines the moment in time that the trigger condition was met, but rather the instant that the first particle was detected by the station. Thirdly, the timestamp is corrected for the GPS timing offset. When a coincidence occurs between four or more stations, multiple reconstructions are performed. All unique combinations of three stations are reconstructed separately.

To investigate the accuracy of the reconstructions, the direction of EAS are also reconstructed using the single stations, when possible. The results of the reconstructions can then be compared. Stations 501, 503 and 506 form a triangle with sides of 122 m, 128 m and 151 m. Reconstructions from this subcluster are compared to reconstructions from the individual stations.

To verify that the single station reconstructions can be used as a reference, EAS are selected which have been reconstructed by more than one station. Each station is required to have at least two particles in all corner detectors ($N_{\mathrm{MIP}} \geq 2$). Figure 6.4 shows the comparison of reconstructed shower directions by single stations. The plots are laid out as a matrix. From top to bottom, and from left to right, data is presented from stations 501, 503 and 506. Each datapoint represents an EAS which was reconstructed by two stations, simultaneously. The three plots in the top right show a comparison of the azimuthal direction ϕ, whereas the bottom left plots show a comparison of the zenith direction θ. The reconstructions performed by each pair of stations are in agreement.

Figure 6.5 shows the uncertainties in the reconstructions. The plots are again laid out as a matrix. The uncertainties are defined as the difference in angle which contains 66 % of the reconstructions, and are estimated using the equations developed in Section 4.3. The uncertainty in the arrival time of a single station reconstruction is taken to be 2.4 ns (Section 5.3). The total uncertainty in the azimuthal angle reconstruction is then given by the quadratic sum of the uncertainties for a single station, i.e.:

$$\sigma_{\phi,\mathrm{total}} = \sqrt{\sigma^2_{\phi,j} + \sigma^2_{\phi,k}}, \tag{6.2}$$

with j,k indexes for the stations. Similar equations describe the uncertainties in the zenith angles. The reconstruction accuracy is shown as a function of the zenith angle. The three plots in the top right show the uncertainties of the azimuthal direction ϕ, whereas the bottom left plots show the uncertainties of the zenith direction θ. The points represent the experimental data. The solid lines are estimates for the uncertainties.

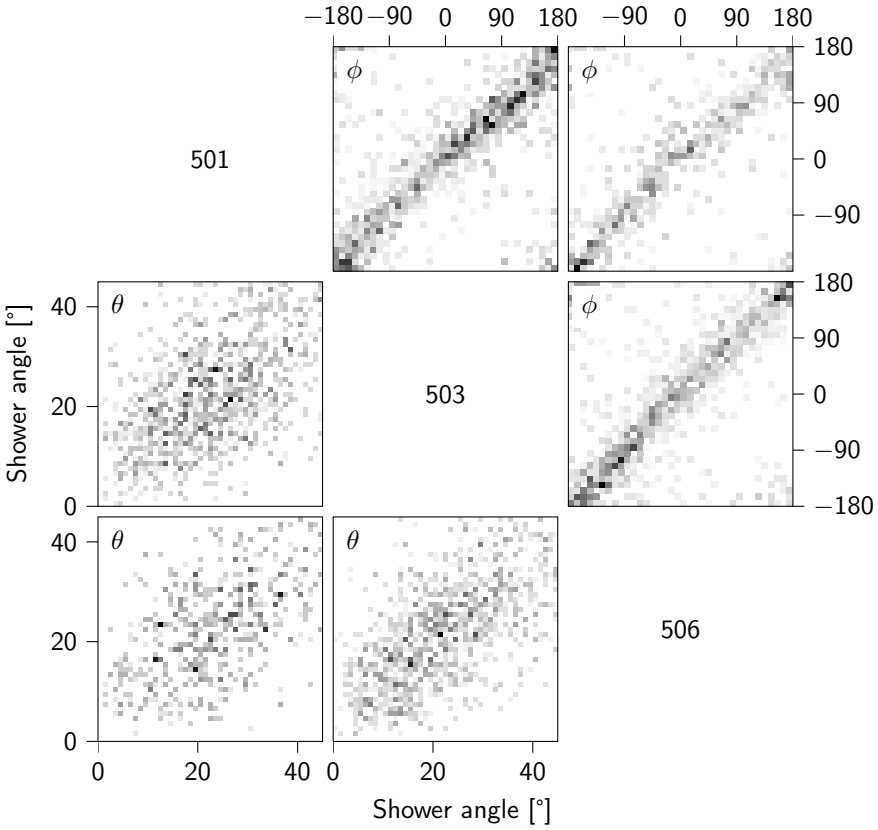


Figure 6.4 – Comparison of reconstructed shower directions by single stations, using two-dimensional histograms. From top to bottom, and from left to right, data is presented from stations 501, 503 and 506, respectively. For example, the plot in the upper right corner shows EAS reconstructed by stations 501 (vertical axis) and 506 (horizontal axis). The three plots in the top right show a comparison of the azimuthal direction ϕ, whereas the bottom left plots show a comparison of the zenith direction θ. Each bin represents EAS which were reconstructed by two stations simultaneously. Each station is required to have at least two particles in all corner detectors ($N_{\mathrm{MIP}} \geq 2$). The reconstructions performed by each pair of stations are in agreement.

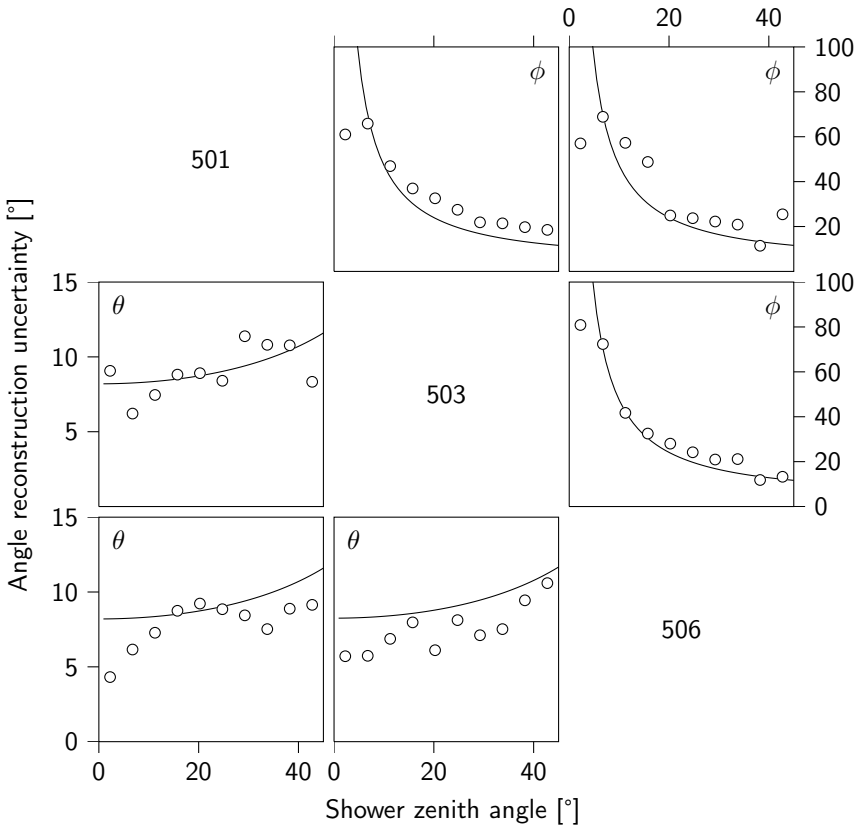


Figure 6.5 – Uncertainties in the reconstruction of EAS simultaneously performed by a pair of single stations. Each station is required to have at least two particles in all corner detectors ($N_{\mathrm{MIP}} \geq 2$). From top to bottom, and from left to right, data is presented from stations 501, 503 and 506, respectively. The uncertainties are defined as the difference in angle which contains 66 % of the reconstructions. The reconstruction accuracy is shown as a function of the zenith angle. The three plots in the top right show the uncertainties of the azimuthal direction ϕ, whereas the bottom left plots show the uncertainties of the zenith direction θ. The experimental data is depicted by circles. The solid lines show estimates for the uncertainties obtained by propagating the experimental timing uncertainties through the analysis. The observed reconstruction differences are close to the calculated estimates.

Data and calculations agree for the azimuthal and zenith angle reconstructions. Small differences in the accuracy of the stations can be seen in the data. We conclude that the single station reconstructions in the Science Park Array meet expectations and thus can be used as a reference.

The direction reconstructions using single stations are compared to the reconstructions using the subcluster. Firstly, coincidences are selected containing stations 501,

503 and 506. Secondly, only those events are selected for which the shower direction could be reconstructed using the subcluster. No cut on the number of particles in the detectors is imposed. Thirdly, for each station, a cut is made on the data to contain only those coincidences which were also reconstructed by the single station, and which contained at least two particles in the corner detectors ($N_{\mathrm{MIP}} \geq 2$). This cut is imposed on the single-station reconstruction only, so the two other stations taking part in the coincidence are allowed to have less particles in the detectors. The resulting dataset now contains EAS which were simultaneously reconstructed by the subcluster and a single station.

Figure 6.6 shows the azimuthal and zenith reconstructions as scatter plots. A clear correlation is visible. However, small systematic uncertainties of the reconstruction of individual stations are visible. These effects are most prominent for station 503 and 506. This is most likely the result of uncertainties in the position of the detectors. In [89] the observed effect is explained by simulating the results for a slightly deformed station, where the reconstruction assumes a regular station. The positions of the detectors of station 501 have been measured to less than 10 cm and it is found that the station deviates from the ideal layout. Similar measurements need to be performed for all stations.

The experimental uncertainties are determined by taking the differences of the reconstructions and calculating the angle difference which contains 66 % of the events. The uncertainties are estimated using the equations developed in Section 4.3. The uncertainty in the arrival time of the single station reconstruction is taken to be 2.4 ns, while the uncertainty in the timing within a cluster is dominated by the GPS timing accuracy (jitter) of approximately 5 ns [86]. The arrival time uncertainty in for the cluster is given $\sigma_{t,\mathrm{cluster}} = \sqrt{2.4^2 + 5^2} = 5.5\,\mathrm{ns}$. The total uncertainty in the azimuthal angle reconstruction is given by

$$\sigma_\phi = \sqrt{\sigma^2_{\phi,\mathrm{detector}} + \sigma^2_{\phi,\mathrm{cluster}}}, \tag{6.3}$$

and similarly for the zenith angles. The contributions must be calculated separately since both the single station reconstruction and cluster reconstruction have uncertainties which depend on the geometry of the detectors or stations used in the reconstruction. The calculations are performed using Equations 4.22 and 4.23.

As a function of the zenith angle, the reconstruction uncertainties are shown in Figure 6.7. The uncertainties deviate from the expected value by approximately 10 %. This can be understood as an additional experimental uncertainty either of the station,

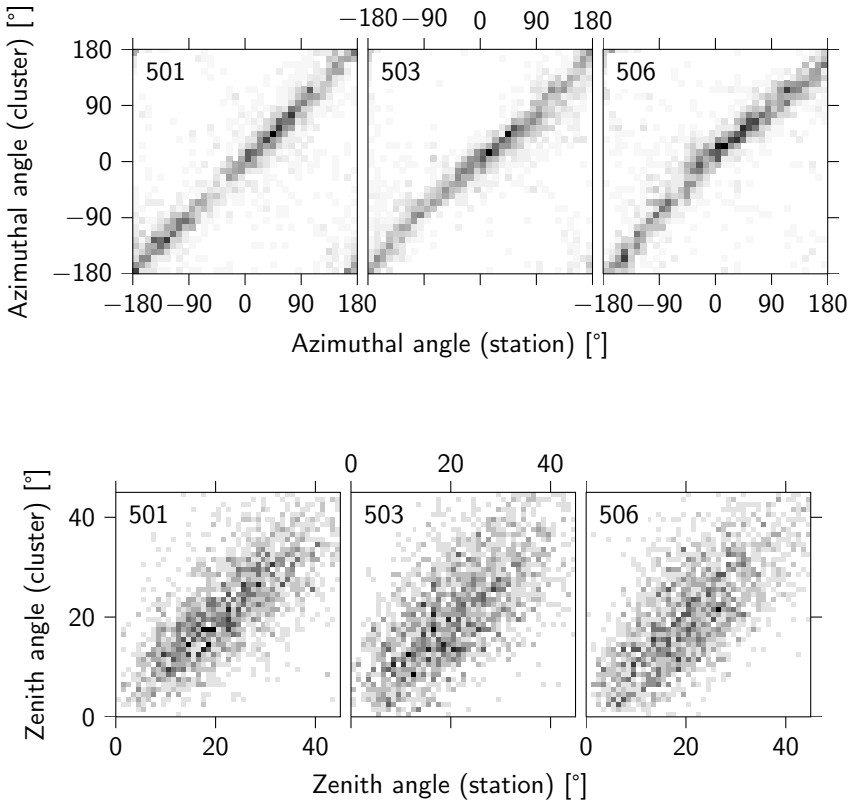


Figure 6.6 – Reconstruction of EAS simultaneously performed using a subcluster and a single station. Reconstructions from the single stations require $N_{\mathrm{MIP}} \geq 2$. The subcluster contains the stations 501, 503 and 506. The left, middle and right panel show the results of the subcluster compared to the results of the single stations 501, 503 and 506, respectively. Stations 503 and 506 show systematic errors in the reconstruction, most obvious for the azimuthal angles.

the cluster, or both. This uncertainty is very small, however, and may well be understood by further study of the GPS offsets, the positions of the detectors within the stations, and the orientation of the station.

6.4 Discussion and Conclusions

The reconstruction of the direction of a shower by different stations in a cluster has been presented. Without an external reference for the shower direction, the reconstructions from single stations have been compared to the reconstruction of a subcluster of three stations. First, the performance of the single stations was investigated. The distribution

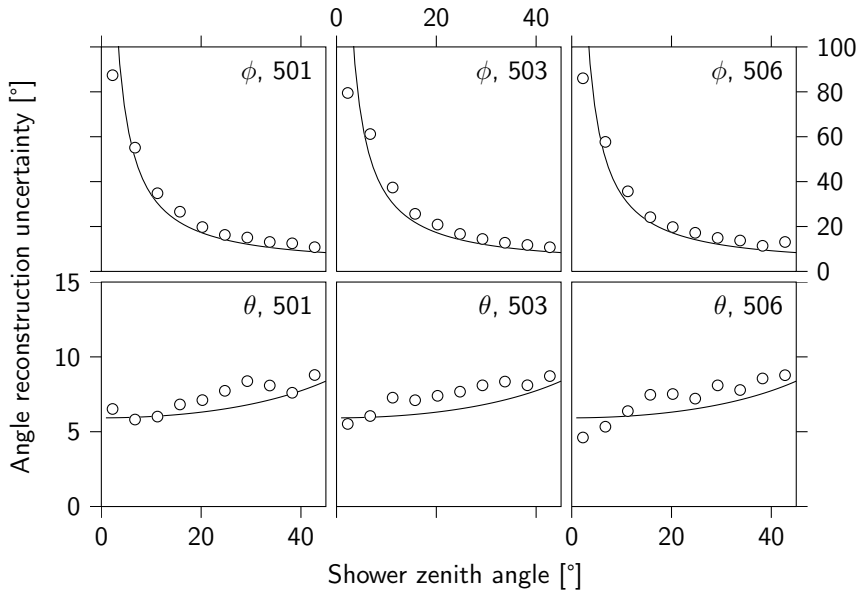


Figure 6.7 – Uncertainties in the reconstruction of EAS simultaneously performed using a cluster and a single station. The cluster contains the stations 501, 503 and 506. Reconstructions from the single stations require $N_{\mathrm{MIP}} \geq 2$. The uncertainties are defined as the difference in angle which contains 66 % of the reconstructions. The results are shown as a function of the zenith angle. The experimental data is depicted by circles. The solid lines show estimates for the uncertainties obtained by propagating the experimental timing uncertainties through the analysis. The observed reconstruction differences are close to the calculated estimates.

of arrival directions of cosmic rays is assumed to be isotropic. This is reflected in the azimuthal directions, as reconstructed by the single stations (Figure 6.3). The data show a flat distribution in ϕ. The distribution of the zenith angle is more complicated. First, for a given zenith angle θ, the area on the celestial sphere which is observed is proportional to $\sin\theta$. However, the acceptance of the detectors is proportional to $\cos\theta$. Furthermore, inclined showers traverse more atmosphere and the probability of detection further decreases with larger zenith angles. The exact relationship has not been investigated in this analysis, but the measured distributions conform to expectations [33].

The performance of the stations is further investigated by comparing the direction of EAS which have been simultaneously reconstructed by more than one station. The uncertainties in the reconstruction agree with expectations (Figure 6.5). The single stations can be used as a reference (with the stated uncertainties) for reconstructing the direction of EAS using a subcluster of three stations.

Direction reconstructions performed by the subcluster of stations 501, 503 and 506 are compared to the reconstructions of the single stations, whenever such information is availabe (Figure 6.7). The comparison shows small systematic uncertainties which are visible in the comparison with stations 503 and 506. The experimentally determined uncertainties are approximately 10 % larger than the calculated values. The uncertainty in the arrival time measurements has a standard deviation $\sigma = 5.5\,\mathrm{ns}$. This is a very nice result. By further study of the GPS offsets, the positions of the detectors within the stations, and the orientation of the station, the systematic uncertainties may be understood. Also, simulations using unthinned 10 PeV showers may be performed.

Since the reconstruction uncertainty (Equations 4.22 and 4.23) is inversely proportional to the distance between detectors (or stations), the uncertainty in the results is dominated by the single station contribution. The distance between the stations 501, 503 and 506 is of the order of 130 m, which is more than ten times the distance between detectors. Calculation of the precision of the shower direction observed by the subcluster of 501, 503 and 506, yields $\sigma_\phi = 2.7°$ and $\sigma_\theta = 1.1°$ for $\theta = 22.5°$ and $N_{\mathrm{MIP}} \geq 2$. Expressed as an angular distance (Section 4.5), the accuracy then becomes 1.5°.

7

Conclusions and Outlook

The performance of a single four-detector HiSPARC station has been investigated using Monte Carlo simulation techniques. The energy threshold for the detection of EAS is $\sim 1\,\mathrm{PeV}$. Since the energy spectrum of cosmic rays follows a power law with a spectral index γ approximately equal to 2.7, most of the observed EAS had a primary energy near the detection threshold. A library of simulated unthinned 1 PeV proton showers has been compiled for a series of zenith angles. Using this library, the detection performance of the station has been investigated. A 1 PeV proton shower with a zenith angle of 22.5° has a 70 % probability to be observed at a core distance of 20 m. At 40 m, this probability is reduced to 30 %.

With the set of simulated showers the direction sensitivity has been analyzed. The uncertainty in the directions have been studied and compared to calculations which were developed for this purpose. The reconstruction of the shower direction becomes more accurate for higher particle multiplicities in the detectors. To understand this behavior quantitatively, the arrival times of particles have been analyzed with the purpose of investigating the shower front structure. The arrival time distribution has been determined from shower data and is highly asymmetrical. It can not be approximated by a normal distribution. Averaging over many showers, a model for the shower front structure has been developed. This has enabled the explanation of the uncertainties as a function of particle multiplicity. When requiring a minimum of two particles in each corner detector ($N_{\mathrm{MIP}} \geq 2$), as opposed to only requiring one, the

accuracy of the reconstructed shower direction is significantly improved (Figure 4.11).

The precision of the direction determination of EAS has been investigated as a function of zenith angle, station size and the digitization frequency. For the design of the station, a nominal detector distance of 10 m and an ADC sampling time of 2.5 ns were an excellent choice. The arrival time of the shower front, as measured by the detectors, has an uncertainty of only 1.8 ns. For simulated 1 PeV proton showers with a zenith angle of 22.5°, and $N_{\mathrm{MIP}} \geq 2$, the uncertainty in the shower direction has been predicted to be $\sigma_\theta = 4.3°$ and $\sigma_\phi = 11°$. However, the physical observable of interest is the angular distance between the reconstructed and the simulated direction of the shower. For small zenith angles, all possible directions on the celestial sphere are close together. A small uncertainty in the direction can thus result in a large uncertainty of the azimuthal angle. For two vectors which only differ in their azimuthal angles, $\Delta\phi = \phi_1 - \phi_2$, the angular distance is given by Equation 4.31. The angular distance between two directions separated by $\Delta\phi = 11°$ and $\Delta\theta = 4.3°$, for $\theta = 22.5°$, is only 5.9°, which is beyond expectations.

The performance of a single station has been analyzed by integrating the station into the KASCADE array. The KASCADE experiment provided a trigger for the station and a dataset of fully reconstructed EAS. The detector efficiency has been verified by studying the detector response for a series of particle densities. The response has been described by Poisson probabilities, and the efficiency has been found to be close to 100 % (Figure 5.8). The shower front structure has been investigated and compared to the model developed using the simulations.

Taking an as yet unexplained experimental uncertainty of 1.6 ns into account, the data from simulations (Chapter 4) closely match the data from a single four-detector HiSPARC station (Chapter 5) and identical conclusions hold for both analyses. The uncertainties have been described and are well understood as a function of particle multiplicity, core distance and zenith angle, in both the simulation and the experiment. The total uncertainty in the measurement has been estimated to be

$$\begin{aligned} \sigma_t &= \sqrt{\sigma^2_{t,\mathrm{front}} + \sigma^2_{t,\mathrm{transport}} + \sigma^2_{t,\mathrm{sampling}} + \sigma^2_{t,\mathrm{other}}} \\ &= \sqrt{1.4^2 + 1.2^2 + \frac{2.5^2}{12} + 1.6^2} = 2.4\,\mathrm{ns}. \end{aligned} \tag{7.1}$$

For a 1 PeV shower with a zenith angle of $\theta = 22.5°$, and $N_{\mathrm{MIP}} \geq 2$, the resulting accuracy in the direction of EAS has been found to be $\sigma_\theta = 6.1°$ and $\sigma_\phi = 15.9°$. The direction of a 1 PeV shower reconstructed by KASCADE is accurate to less than 0.3° [117].

The angular distance between the direction determined by the HiSPARC station and the direction observed by KASCADE, is less than 8.6° for 66 % of the reconstructed events at $\theta \approx 22.5°$. The uncertainties in the reconstruction must be viewed in light of the size and cost of a single station. A single station has an active detector surface of only $2\,m^2$ and covers an area of $43\,m^2$. Its cost is approximately € 10 000.

The reconstruction of shower direction has been performed with a subcluster of three stations in the Amsterdam Science Park Array, which resembles an equilateral triangle with sides ranging from 122 m to 151 m. First, the performance of the individual stations has been studied. The observed cosmic ray arrival directions are indeed isotropically distributed. The zenith angle distribution peaks around 20°. The directions could not be compared to an external reference (simulated direction, or the direction as provided for a single station in KASCADE). For a single shower, the reconstructed directions obtained by the individual stations agree within an accuracy which is consistent with the calculations used in Chapters 4 and 5. Systematic discrepancies are absent. The accuracy of the shower direction determined by a single station has been verified, and it is used as a reference for the (sub)cluster.

For EAS which have been reconstructed by the subcluster and by a single station simultaneously, the directions agree. Moreover, the uncertainties are well described by the calculations (Figure 6.7). This is a critical observation since this demonstrates that the calculations for the uncertainties in the direction of EAS describe the data within a timing uncertainty of 5.5 ns (66 % of events) in the cluster. These calculations can then be used to estimate the precision of the shower direction observed by the subcluster of 501, 503 and 506. The accuracy is estimated to be $\sigma_\phi = 2.7°$ and $\sigma_\theta = 1.1°$ at $\theta = 22.5°$. Expressed as an angular distance, the precision then becomes 1.5°.

7.1 Outlook: Towards Energy Determination of EAS

The reconstruction of the size of an EAS is necessary to determine the energy of the primary particle. This can be achieved by measuring the lateral distribution of the shower particles. Then, a theoretical distribution of the lateral density can be fitted to the experimentally observed distribution. The fit parameters are the core position and the shower size. The latter can be used to estimate the primary energy. In this section, we will give a description of the procedure to determine the size of an EAS.

There are several lateral distribution functions given in the literature. Some are theoretically derived or motivated, others are empirical equations describing the experimentally observed lateral densities. The LDF originally used by the KASCADE

experiment is the Nishimura-Kamata-Greisen (NKG) function, given by [124]:

$$\rho(r) = N_e\, c(s) \left(\frac{r}{R_0}\right)^{s-2} \left(1+\frac{r}{R_0}\right)^{s-4.5}, \tag{7.2}$$

with $\rho(r)$ the particle density as a function of the core distance, N_e the number of electrons in the shower, R_0 the Molière radius, s the shower age parameter, and $c(s)$ given by:

$$c(s) = \frac{\Gamma(4.5-s)}{2\pi r_0^2\,\Gamma(s)\,\Gamma(4.5-2s)}. \tag{7.3}$$

This theoretically motivated function did not describe the KASCADE data accurately. By redefining several parameters, the data are better explained. This is achieved by writing:

$$\rho(r) = N_e\, \tilde{c}(s) \left(\frac{r}{R_0}\right)^{s-\alpha} \left(1+\frac{r}{R_0}\right)^{s-\beta}, \tag{7.4}$$

with α and β free parameters. Then, $\tilde{c}(s)$ is defined by:

$$\tilde{c}(s) = \frac{\Gamma(\beta-s)}{2\pi r_0^2\,\Gamma(s-\alpha+2)\,\Gamma(\alpha+\beta-2s-2)}. \tag{7.5}$$

The best fit to KASCADE data is obtained with $\alpha = 1.5$, $\beta = 3.6$ and $R_0 = 40$.

The parameter s is now interpreted as a *shape parameter* and does not vary significantly for most observed EAS. By fixing s the number of degrees of freedom are reduced. A fit of Equation 7.4 to lateral densities observed in the simulation of 1 PeV proton-induced vertical showers results in a *shower size* $\lg N_e = 4.8$ and $s = 0.94$. Equation 7.4 can be written as

$$\rho(r) = N_e\, \tilde{c}(s) R(r), \tag{7.6}$$

with

$$R(r) = \left(\frac{r}{R_0}\right)^{s-\alpha} \left(1+\frac{r}{R_0}\right)^{s-\beta}. \tag{7.7}$$

Several independent measurements of the particle density are required to determine the values of the remaining parameters, i.e. the shower size N_e and the shower core position x, y. These measurements are obtained if an EAS is observed by multiple HiSPARC stations. The determination of the parameters can be simplified by considering *relative* particle densities:

$$\tilde{\rho}_i(r_i) = \frac{\rho_i(r_i)}{\rho_0(r_0)} = \frac{R(r_i)}{R(r_0)}, \tag{7.8}$$

with the subscripts denoting the station index. Here, station 0 is chosen as a reference

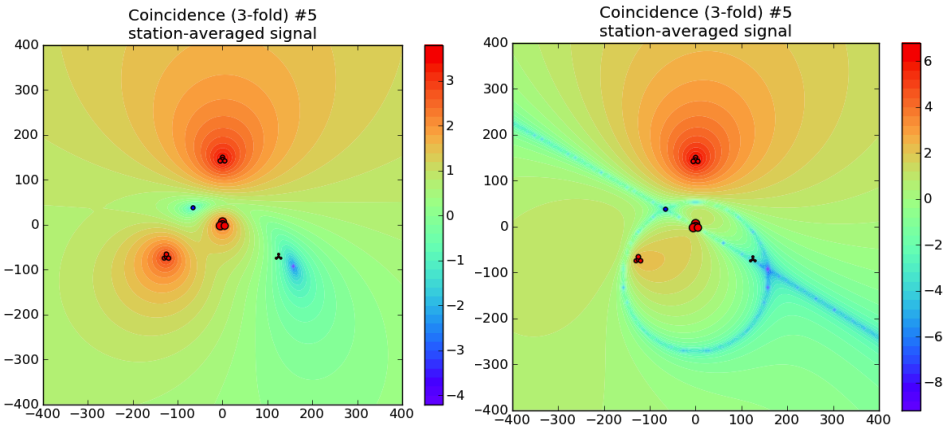


Figure 7.1 – Contour plot of the χ^2-distribution (left) used in the reconstruction of the core position. The detectors in this simulation have precise knowledge of the particle densities in the shower. The algorithm correctly finds the global minimum, denoted by the blue circle. The positions of the detectors are depicted by the red circles, with the size of the circle proportional to the observed particle density. Right: the terms in the χ^2-distribution are multiplied instead of summed, resulting in a geometric interpretation of the solution, located at the intersection of the two circles. One circle is very large, resulting in a nearly straight line in the plot.

to which all other stations (i) will be compared. Defining the χ^2-distribution by

$$\chi^2 = \sum_{i \neq 0} \left(\frac{\tilde{\rho}_{i,\mathrm{ldf}} - \tilde{\rho}_{i,\mathrm{obs}}}{\sigma_i} \right)^2 , \tag{7.9}$$

a minimization procedure provides values for the core position. For low particle densities, the uncertainties in the particle densities are not normally distributed. Therefore, the χ^2-distribution can only be used as an approximation. In Figure 7.1, contour plots of the χ^2-distribution are presented. The detectors in this simulation have precise knowledge of the particle densities in the shower. The algorithm correctly finds the global minimum, denoted by the blue circle. The positions of the detectors are depicted by the red circles, with the size of the circle proportional to the observed particle density. The figure on the right shows a geometric interpretation of the χ^2-distribution by multiplying, instead of summing, the terms. Using this interpretation, analytical approximations can be attempted [125].

Figure 7.2 shows simulated and reconstructed shower core positions for a 1 PeV vertical proton. The detectors in this simulation have precise knowledge of the particle densities in the shower. The reconstructed core position is, in general, very close to the simulated position. In Figure 7.3 the detector signals have been included using

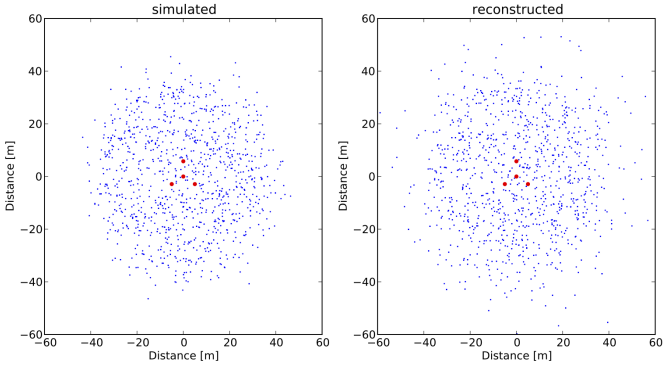


Figure 7.2 – Simulated (left) and reconstructed (right) shower core positions for a 1 PeV vertical proton. The detectors in this simulation have exact knowledge about the particle densities in the shower. The reconstructed core positions are very close (to machine precision) to the simulated positions, proving that the algorithm works.

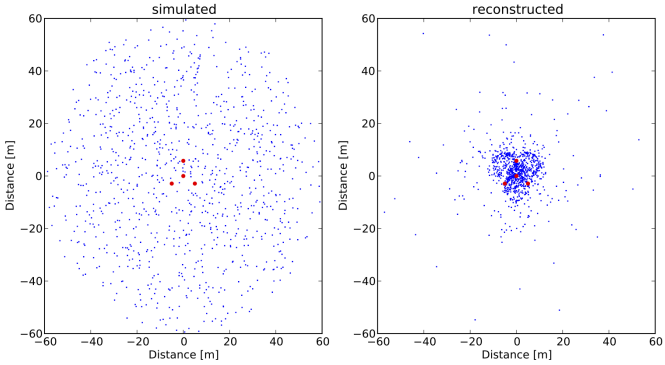


Figure 7.3 – Simulated (left) and reconstructed (right) shower core positions for a 1 PeV vertical proton. The detector signals have been simulated using the simulations discussed in Chapter 4. The detectors no longer have knowledge about the particle densities in the shower, but can only observe the few particles that traverse the detectors. This severely limits the accuracy of the reconstruction of the core position.

the simulations discussed in Chapter 4. The detectors no longer have knowledge of the particle densities in the shower (e.g. $1.3\,\mathrm{m}^{-2}$, but can only observe the integer number of particles that traverse the detectors (e.g. 2). This severely limits the accuracy of the reconstruction of the core position.

Once the core position is determined, the full lateral distribution function (Equation 7.4) can be fitted to the approximate densities to obtain the shower size. Figure 7.4

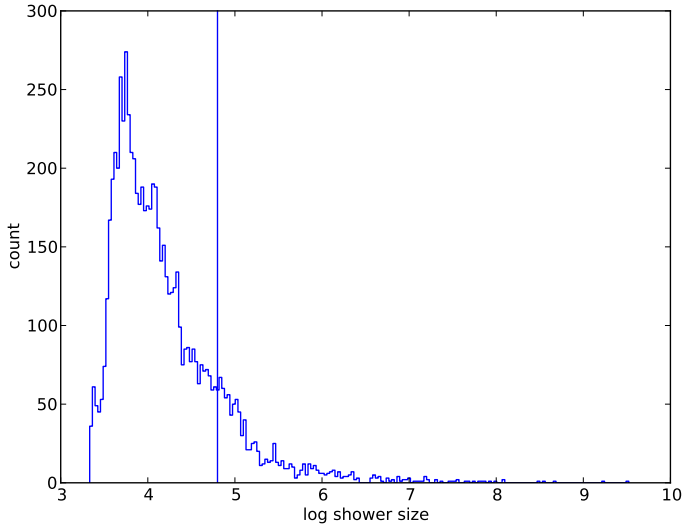


Figure 7.4 – Reconstruction of shower size. The vertical showers have been created by protons with an energy of 1 PeV, equivalent to a shower size of $\lg N_e = 4.8$ particles. The shower size is underestimated for most events.

shows the size of the shower obtained by analyzing simulated detector signals. The vertical showers have been created by protons with an energy of 1 PeV, equivalent to a shower size of $\lg N_e = 4.8$ particles. The shower size is underestimated for most events.

Thorough investigations are required to make a rigorous treatment of the shower size reconstruction. Steijger [126] has suggested the use of statistical tests to reject events which can not be accurately reconstructed. Ongoing work by Bosboom will be published in [127].

A

HiSPARC Electronics

In the early years of HiSPARC, the readout electronics comprised three separate modules: a DAQ module, a GPS module and a custom HiSPARC module. The DAQ and GPS modules were connected to the PC using the legacy RS-232 (COM) and IEEE 1284 (LPT) interfaces, respectively. The HiSPARC module contained the trigger logic and a pulse stretcher, necessary because the DAQ/ADC module had a low sampling frequency. Using the HiSPARC module the trigger threshold and the PMT voltages could be set manually.

A new version of the electronics was developed and has been in use for several years. The HiSPARC II hardware (Figure A.1) integrates the DAQ and GPS modules, the trigger logic and the PMT controller into one unit (Figure A.2). The unit is connected to the station PC using a USB 1.1 interface. The DAQ consists of a total of four AD

Figure A.1 – Front of the HiSPARC II electronics with connectors for the PMT control and signal cables for two channels.

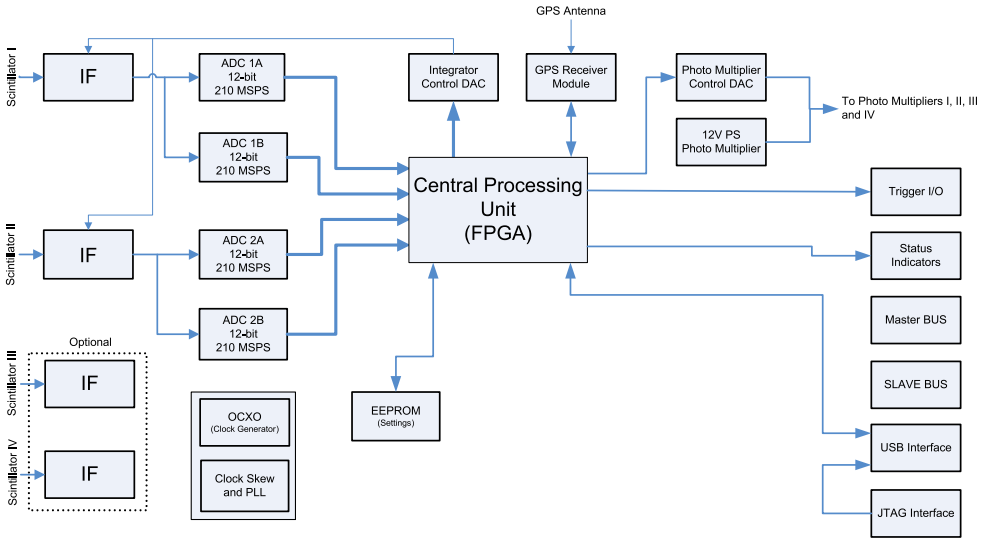


Figure A.2 – Overview of the HiSPARC electronics components. Each analog PMT signal is digitized by two ADCs and the digital signals are fed into the central processing unit, which is implemented in a FPGA. The FPGA communicates with the GPS unit and the PC. The trigger and PMT voltage control are also implemented in the FPGA. On power up, the EEPROM contents are used to initialize the FPGA. The EEPROM itself can be flashed using the on-board JTAG interface.

converters to digitize the analog signal from two PMTs. Thus, two ADCs are used per channel. The electronics contains a 200 MHz crystal to drive the ADCs. By clocking one ADC at the rising edge of the clock, and one ADC at the falling edge, the signal can be sampled with 400 MHz, i.e. a sampling time of only 2.5 ns. This way, the samples from the signal are alternately provided by the two converters. This requires that the ADCs be carefully aligned. If the baselines are not carefully aligned, a ragged signal will result (resembling a triangular wave). This alignment procedure can be carried out by the user after installation of the station and is performed by applying several different internal reference voltages on the input channel. The ADC gains and offsets can be controlled by software and are adjusted until the ADCs are both aligned and are providing a sampling range of +113 mV to −2222 mV. The process is fully automated and the ADC response is highly linear over this range. The converters provide a 12 bit output, corresponding to a resolution of −0.57 mV per ADC count. A conversion of ADC counts to mV units is then given by $V = -.57x + 113$, with x the number of ADC counts.

The trigger logic is implemented in a FPGA module. The firmware is designed to communicate with a *slave* module, which is an identical HiSPARC II unit without the

GPS board. Each readout unit has its own clock and information on signal levels are communicated to the master's FPGA. For each channel, two comparator thresholds can be set: the *low threshold*, and the *high threshold*. The trigger conditions can be set as follows: the number of required channels over the low threshold, AND or OR the number of required channels over the high threshold. For example, the default trigger condition for a four-detector station is 3 low OR 2 high signals, with the thresholds set at −30 mV and −70 mV. This means that over all four channels, at least 3 signals higher than −30 mV will result in a trigger. Additionally, two signals higher than −70 mV will also generate a trigger. Once a trigger occurs, the master unit instructs the slave unit to store and send the event messages to the station PC.

For each event, the maximum time frame is 10 µs, which is equal to 4000 samples. Each sample is 12 bit, i.e. 1.5 B. Since the electronics has two channels, the total size of an event is 12 kB. If a trigger is generated, the event is kept in a buffer (36 kB), waiting for transfer to the PC over the USB connection. If the buffer has insufficient space to store a new event, the new event is discarded. USB 1.1 has a specified data rate of 12 Mbit s^{-1}. Theoretically, this would allow for a trigger rate of up to 125 Hz. Since the Windows operating system is not a real-time OS, the USB buffers are polled. The frequency with which this polling is done reduces the maximum attainable trigger frequency substantially (typically ∼ 30 Hz).

By default, the pre-trigger window is set to 1 µs, the coincidence window to 1.5 µs and the post-trigger window to 3.5 µs, making a total of 6 µs. In this configuration, the event buffer can hold up to five events, and the maximum attainable trigger rate increases.

Samenvatting

De ontdekking van kosmische straling wordt algemeen toegeschreven aan Victor Francis Hess, die in 1911 en 1912 een serie ballonvluchten ondernam om voor eens en altijd vast te stellen of een mysterieuze en overal aanwezige vorm van straling *vanuit* de Aarde, of van *buiten* de Aarde afkomstig was. Tijdens deze vluchten, tot een hoogte van ruim 5 kilometer, stelde hij vast dat de straling sterker werd naarmate je hoger in de atmosfeer kwam. Het bleek dat de straling daadwerkelijk vanuit de ruimte kwam en daarvoor ontving hij in 1936 de Nobelprijs. Het was echter Robert Andrews Millikan die de term *kosmische straling* bedacht. Een term die vandaag de dag nog gangbaar is, ondanks het feit dat het een onjuiste benaming betreft. Wat we aanduiden met kosmische straling zijn feitelijk deeltjes.

De aard van de deeltjes is vergelijkbaar met de aard van de elementen waaruit ons zonnestelsel is opgebouwd (Figuur 1.1 op pagina 5). Er zijn verschillen die verklaard kunnen worden door aan te nemen dat de deeltjes een lange weg hebben afgelegd voordat ze de Aarde bereiken. De deeltjes in de kosmische straling hebben echter een enorme hoeveelheid bewegingsenergie. Het betreft hier dus voornamelijk 'gewone' deeltjes, die op de één of andere manier versneld zijn door natuurlijke deeltjesversnellers.

De bekendste deeltjesversneller is wellicht de *Large Hadron Collider (LHC)* bij Genève. Deze cirkelvormige versneller is 27 kilometer lang en versnelt deeltjes door middel van sterke elektrische velden en houdt ze in een cirkelvormige baan door middel van sterke magnetische velden. Dit geeft de LHC de mogelijkheid de deeltjes rondje na rondje, en beetje bij beetje, meer energie te geven, waarna de energie constant blijft bij een waarde van 7 TeV. De LHC stelt fysici in staat om door middel van botsingen het gedrag van materie te bestuderen bij extreem hoge energieën, zoals die ook kort na de oerknal bestaan moeten hebben. En, niet te vergeten, nog dagelijks vóórkomen búiten het laboratorium, een kilometer of twintig boven ons hoofd. Maar dan duizendmaal sterker, wanneer kosmische straling onze atmosfeer binnendringt en botst op luchtmoleculen.

Het is op zich niet verwonderlijk dat de natuurlijke deeltjesversnellers veel hogere energieën bereiken dan mogelijk zijn met de LHC. De LHC is slechts (!) 27 kilometer lang. Elektrische en magnetische velden komen ook voor in het heelal, maar dan op schalen van miljoenen of zelfs miljarden kilometers. In veel gevallen zijn deze velden

veel minder sterk dan we kunnen opwekken in laboratoria, maar de grootte maakt dat meer dan goed. Elektrisch geladen deeltjes kunnen lange tijd grillige, maar min-of-meer cirkelvormige, banen beschrijven en langzaam, beetje bij beetje, versneld worden tot ongekende energieën. Mogelijke kandidaten zijn de uitgestrekte resten van sterren die in de vorm van een supernova aan hun eind zijn gekomen. Zeer waarschijnlijk zijn er plaatsen in het heelal die niet alleen veel groter zijn dan de LHC, maar ook veel sterkere elektromagnetische velden bezitten: pulsars en zwarte gaten, bijvoorbeeld.

Wanneer een deeltje uit de kosmische straling de Aarde op haar pad vindt, vindt er onherroepelijk een botsing plaats tussen het deeltje en moleculen uit de atmosfeer. Bij deze botsing wordt er, volgens $E = mc^2$, een deel van de bewegingsenergie omgezet in massa, in nieuwe deeltjes. Deze deeltjes hebben nog steeds erg veel energie, en botsen weer. En weer. Er ontstaat zo een soort lawine (*air shower*) waarbij het aantal deeltjes in eerste instantie blijft toenemen. Wanneer de energie per deeltje te laag wordt om nieuwe deeltjes te maken, kan het aantal deeltjes niet langer toenemen. De deeltjes in de lawine verliezen langzaamaan hun energie en worden geabsorbeerd in de atmosfeer, of vervallen tot andere deeltjes. Bij hoog-energetische kosmische straling is het aantal deeltjes echter zo groot dat een aanzienlijk deel de grond bereikt, over een oppervlak dat wel enkele vierkante kilometers groot kan zijn. Het HiSPARC experiment bestudeert de resten van air showers zodra de deeltjes hiervan de grond bereiken.

VWO-leerlingen bouwen, onder leiding van onderzoekers, detectoren bestaande uit een scintillator en een fotobuis en installeren deze in een skibox op het dak van hun school. De scintillator bestaat uit een materiaal dat zwak oplicht zodra elektrisch geladen deeltjes (of ook hoog-energetische fotonen) uit de air shower door de plaat gaan. De fotobuis kan dit zwakke lichtsignaal versterken tot een meetbaar elektrisch signaal, dat door middel van een kabel naar een meetcomputer wordt gebracht. Iedere school beschikt over twee of vier detectoren en wanneer meerdere detectoren tegelijkertijd deeltjes waarnemen, is de kans groot dat dit een air shower betreft. De metingen worden dan via internet verstuurd naar het Nikhef. Leerlingen en onderzoekers kunnen vervolgens analyseren of de vermoedelijke air shower tegelijkertijd is geobserveerd door meerdere naburige scholen. Voor de benodigde precieze tijdmetingen wordt gebruik gemaakt van het GPS-systeem. GPS is vooral bekend als navigatiemiddel, maar naast positie geeft het systeem juist ook een heel nauwkeurige tijd.

Dit proefschrift geeft in de eerste hoofdstukken een beschrijving van kosmische straling (Hoofdstuk 1) en van het HiSPARC experiment (Hoofdstuk 2). Vervolgens richt het zich op de software die de gegevens opstuurt en beheert, en op de software die de 'gezondheid' van het HiSPARC netwerk in de gaten houdt (Hoofdstuk 3). Ook wordt

daar kort gesproken over SAPPHiRE, een zogeheten *software framework*. Het stelt onderzoekers en leerlingen in staat om eenvoudig simulaties en analyses te ontwikkelen binnen een uniforme opzet. Alle software binnen HiSPARC wordt op een open en voor iedereen toegankelijke manier ontwikkeld.

Dit proefschrift richt zich verder op de reconstructie van de richting, de herkomst, van een kosmisch deeltje dat een air shower veroorzaakt. In Hoofdstuk 4 wordt de mogelijkheid onderzocht dat één enkel vier-plaats station de herkomstrichting van een air shower bepaalt. De deeltjes in een air shower reizen als het ware in een 'platte pannenkoek', het *shower front*, naar de grond. Dat betekent dat een shower recht van boven de verschillende detectoren in een station vrijwel gelijktijdig raakt. Komt de shower echter onder een hoek binnen, dan zullen er tijdverschillen optreden in de metingen. Naast vergelijkingen die uit de tijdverschillen de hoek reconstrueren worden er ook vergelijkingen afgeleid die iets zeggen over de *nauwkeurigheid*[1] waarmee dat kan. Deze vergelijkingen worden getoetst aan een simulatie van air showers die een HiSPARC station raken. Het blijkt dat het mogelijk moet zijn om met slechts één station iets zinnigs te zeggen over de herkomst van een kosmisch deeltje. Sterker, de resultaten worden ondersteund door de vergelijkingen die de nauwkeurigheden beschrijven, wat aantoont dat we het resultaat grotendeels begrijpen. Protonen met een energie van 1 PeV die binnenkomen onder een hoek van 22.5° genereren showers die een detectiekans hebben van meer dan 50 % tot een afstand van 30 m. Worden deze showers gedetecteerd, dan geeft de reconstructie een fysisch resultaat in meer dan 95 % van de gevallen.

De dikte van het shower front introduceert onzekerheden in de tijdmeting. Dit leidt tot onnauwkeurigheden in de reconstructie. Aangezien de deeltjes met bijna de lichtsnelheid reizen en niets sneller kan dan het licht, worden de onzekerheden alléén veroorzaakt door 'trage' deeltjes. Wanneer meerdere deeltjes een detector doorkruisen wordt de onzekerheid echter kleiner. Dit is puur statistiek: neem het *eerste* deeltje dat de detector doorkruist en de tijdmeting zit per definitie dichter op de werkelijke waarde. Dit effect vlakt echter af bij meer dan twee deeltjes per detector. De tijdsonzekerheid voor 2 deeltjes is door simulatie bepaalt op 1.4 ns. Een tweede meetonzekerheid, veroorzaakt door de tijd die fotonen nodig hebben om vanuit de detector naar de fotobuis te reizen is door simulatie bepaalt op 1.2 ns. De totale onzekerheid bedraagt dan 1.8 ns. Voor 1 PeV proton showers die binnenkomen onder een hoek van 22.5° en in alle hoekdetectoren 2 of meer deeltjes deponeren is de onzekerheid in de zenith hoek $\sigma_\theta = 4.3°$ en in de azimuthale hoek $\sigma_\phi = 11°$. De onnauwkeurigheid in de azimuthale hoek is getalmatig

[1] Het is natuurkundig beter om te spreken over de *onzekerheid* van de resultaten.

groter, maar dit komt alleen door de keuze van het coördinatensysteem. De werkelijke hoekafstand is kleiner en de totale onzekerheid (de hoekafstand tussen de werkelijke en de gereconstrueerde richting) komt uit op 5.9°.

In Hoofdstuk 5 worden de methodes uit Hoofdstuk 4 getoetst aan een experiment. Eén HiSPARC station is geïnstalleerd binnen de KASCADE detector in Karlsruhe, Duitsland. Deze detector bestaat uit 252 afzonderlijke subdetectoren die ieder vergelijkbaar zijn met een HiSPARC detector. Dit veel grotere (en duurdere) experiment verzorgt voor ons een onafhankelijke meting van de richting van iedere air shower die zowel door KASCADE als door HiSPARC waargenomen wordt. De resultaten worden besproken en ook hier blijkt, na correctie voor experimentele onzekerheden die niet voorkomen in de simulatie, dat we de resultaten goed begrijpen. De totale onzekerheid in de tijdmeting wordt beschreven door:

$$\sigma_t = \sqrt{\sigma^2_{t,\text{front}} + \sigma^2_{t,\text{transport}} + \sigma^2_{t,\text{sampling}} + \sigma^2_{t,\text{other}}},$$

waarin $\sigma_{t,\text{front}} = 1.4\,\text{ns}$ veroorzaakt wordt door de dikte van het shower front, zoals eerder beschreven, $\sigma_{t,\text{transport}} = 1.2\,\text{ns}$ een beschrijving geeft van de fotonen die onderweg zijn naar de fotobuis, $\sigma_{t,\text{sampling}} = \frac{2.5\,\text{ns}}{\sqrt{12}}$ de onzekerheid veroorzaakt door de digitalisering van het signaal weergeeft en $\sigma_{t,\text{other}} = 1.6\,\text{ns}$ een onbekende experimentele bijdrage beschrijft. Deze laatste onzekerheid is bepaald uit de data. De totale onzekerheid komt dan uit op 2.4 ns. Dit is iets meer dan de waarde die verklaard werd vanuit de simulatie, en de experimentele resolutie komt dan uit op een hoekafstand van 8.6° voor 1 PeV proton showers onder een hoek van 22.5°, met $N_{\text{MIP}} \geq 2$.

In Hoofdstuk 6 wordt de methode uitgebreid naar een combinatie van stations in het *Science Park Array* in Amsterdam. In theorie moet een netwerk met grotere afstanden en meer detectoren een beter resultaat geven voor de reconstructie van de richting van een air shower. Een complicatie is nu dat ieder station zijn eigen GPS-tijd moet meten en dat daar een onnauwkeurigheid in zit. Omdat we een onafhankelijke tweede meting missen (er is geen 'KASCADE' experiment op het Science Park) wordt er eerst een aantal controles uitgevoerd tussen de stations onderling. We begrijpen immers wél de nauwkeurigheid van individuele stations! De resultaten komen overeen met de verwachtingen. Vervolgens worden de richtingen gereconstrueerd door een cluster van drie stations vergeleken met de richtingen gereconstrueerd door een enkel station. Hoewel de resultaten grotendeels voldoen aan de verwachtingen en een optimistisch beeld geven van de nauwkeurigheid van de reconstructies, is er een aantal *systematische onzekerheden* die verder moeten worden onderzocht. Door de onnauwkeurigheid in de

tijdsmeting (GPS) komt de totale tijdsonzekerheid uit op 5.5 ns. De grotere afstanden tussen de stations (122 m tot 151 m) maken deze waarde meer dan goed. Het gevonden resultaat betekent dat we, weer voor 1 PeV proton showers onder 22.5° met $N_{\mathrm{MIP}} \geq 2$, de richting van een air shower kunnen reconstrueren binnen een cirkel met een straal van 1.5° aan de hemelboog. Dat is slechts een paar keer de grootte van de maan! Dit is een mooi resultaat, zeker met de beperkte kosten van HiSPARC stations in het achterhoofd. Aangezien het Science Park Array bestaat uit méér dan drie stations, kan dit resultaat nog worden verbeterd.

Het laatste hoofdstuk, Hoofdstuk 7, vormt de conclusies van dit proefschrift. Tevens wordt daarin kort vooruit geblikt naar de mogelijkheid de energie van het oorspronkelijke kosmische deeltje te bepalen. Dit blijkt echter bijzonder moeilijk wanneer slechts een beperkt aantal detectoren voorhanden is.

Het doel van het HiSPARC experiment is tweeledig. Enerzijds is het een 'traditionele' onderzoeksgroep bestaande uit stafleden, promovendi en studenten, dat onderzoek doet naar verschillende facetten van kosmische straling. Anderzijds is het experiment ook juist bedoeld om leerlingen uit het voortgezet onderwijs deel te laten nemen aan wetenschappelijk onderzoek. Jaarlijks komt een groot aantal scholen bij elkaar voor het landelijke HiSPARC symposium, waar onderzoekers en leerlingen hun werk presenteren.

Bibliography

1. de Coulomb, C. A. Troisième mémoire sur l'Electricité et le Magnétisme. *Histoire de l'Académie Royale des Sciences,* 612–638 (1785).
2. Faraday, M. *Experimental Researches in Electricity* **2** (1844).
3. Crookes, W. On Electrical Insulation in High Vacua. *Proc. Roy. Soc. of London* **28,** 347–352 (1879).
4. Röntgen, W. Über eine Art von Strahlen. *Sitzungsberichten der Würzburger Physik-medic. Gesellschaft,* 132 (1895).
5. Flakus, F. Detecting and Measuring Ionizing Radiation – a Short History. *IAEA Bulletin* **23** (4 1981).
6. Becquerel, H. Emission de radiations nouvelles par l'uranium metalique. *Comptes Rendus de l'Acad. des Sciences* **122,** 1086 (1896).
7. Rutherford, E. & Cooke, H. L. A penetrating radiation from the Earth's surface. Minutes of the Eighteenth Meeting of the American Physical Society. *Phys. Rev.* **16,** 183 (1903).
8. McLennan, J. C. & Burton, F. Some experiments on the electrical conductivity of atmospheric air. Minutes of the Eighteenth Meeting of the American Physical Society. *Phys. Rev.* **16,** 184–192 (1903).
9. Kurz, K. The radioactive materials of the earth and air as the origin of the invasive radiation in the atmosphere. *Phys. Zeit.* **10,** 834–845 (1909).
10. Wulf, T. On the radiation of high penetrating power that exists in the atmosphere. *Phys. Zeit* **1,** 152–157 (1909).
11. Pacini, D. La radiazione penetrante alla superficie ed in seno alle acque. *Nuovo Cim.* **VI/3,** 93 (1912).
12. Gockel, A. Air electrical observations in a balloon journey. *Phys. Zeit.* **11,** 280–282 (1910).
13. Hess, V. F. On the absorption of gamma-radiation in the atmosphere. *Phys. Zeit.* **12,** 998–1001 (1911).

14. Eve, A. S. On the ionization of the atmosphere due to radioactive matter. *Phil. Mag.* **21,** 26–40 (1911).

15. Hess, V. F. Observations in low level radiation during seven free balloon flights. *Phys. Zeit.* **13,** 1084–1091 (1912).

16. Hess, V. F. The origins of penetrating radiation. *Phys. Zeit* **14,** 6120–617 (1913).

17. Kolhörster, W. Measurements of the penetrating radiation in the free balloon at high altitudes. *Phys. Zeit.* **14,** 1153–1156 (1913).

18. Millikan, R. A. & Bowen, I. S. Penetrating radiation at high altitudes. Minutes of the Pasadena Meeting, May 5, 1923. *Phys. Rev.* **22,** 198 (1923).

19. Millikan, R. A. & Cameron, G. H. High frequency rays of cosmic origin III. Measurements in snow-fed lakes at high altitudes. *Phys. Rev.* **28,** 851–868 (1926).

20. Clay, J. Penetrating radiation. *Acad. Amsterdam Proc.* **30,** 1115–1127 (1927).

21. Clay, J. Penetrating radiation II. *Acad. Amsterdam Proc.* **31,** 1091–1097 (1928).

22. Bothe, W. & Kolhörster, W. Das Wesen des Höhenstrahlung. *Zeit. f. Phys.* **56,** 751–777 (1929).

23. Compton, A. H. A geographic study of cosmic-rays. *Phys. Rev.* **43,** 387–403 (1933).

24. Auger, P., Ehrenfest, P., Maze, R., Daudin, J. & Fréon, R. A. Extensive Cosmic-Ray Showers. *Rev. Mod. Phys.* **11,** 288–291 (3-4 July 1939).

25. Rao, M. V. S. & Sreekantan, B. V. *Extensive Air Showers* (1998).

26. Carlson, P. & De Angelis, A. Nationalism and internationalism in science: the case of the discovery of cosmic rays. *The European Physical Journal H* **35,** 309–329. ISSN: 2102-6459 (4 2011).

27. Takeda, M., Sakaki, N., Honda, K., *et al.* Energy determination in the Akeno Giant Air Shower Array experiment. *Astroparticle Physics* **19,** 447–462. ISSN: 0927-6505 (2003).

28. Apel, W. *et al.* Energy spectra of elemental groups of cosmic rays: Update on the KASCADE unfolding analysis. *Astroparticle Physics* **31,** 86–91. ISSN: 0927-6505 (2009).

29. Abraham, J. *et al.* Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory. *Phys.Lett.* **B685,** 239–246 (2010).

30. Bird, D. J. *et al.* Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies. *Phys. Rev. Lett.* **71,** 3401–3404 (21 Nov. 1993).

31. Boyle, P. Cosmic ray composition at high energies: The TRACER project. *Advances in Space Research* **42,** 409–416. ISSN: 0273-1177 (2008).

32. George, J., Lave, K., Wiedenbeck, M., *et al.* Elemental Composition and Energy Spectra of Galactic Cosmic Rays During Solar Cycle 23. *Astroph. J.* **698,** 1666–1681 (June 2009).

33. Grupen, C. *Astroparticle physics* ISBN: 9783540253129 (Springer, 2005).

34. Lodders, K., Palme, H. & Gail, H.-P. in *Landolt-Börnstein - Group VI Astronomy and Astrophysics Numerical Data and Functional Relationships in Science and Technology Volume 4B: Solar System.* (ed J. E. Trümper) (2009), 44–+. doi:10.1007/978-3-540-88055-4_34. arXiv:0901.1149 [astro-ph.EP].

35. Wang, J. Z., Seo, E. S., *et al.* Measurement of Cosmic-Ray Hydrogen and Helium and Their Isotopic Composition with the BESS Experiment. *The Astrophysical Journal* **564,** 244 (2002).

36. De Nolfo, G., Moskalenko, I., *et al.* Observations of the Li, Be, and B isotopes and constraints on cosmic-ray propagation. *Advances in Space Research* **38,** 1558–1564. ISSN: 0273-1177 (2006).

37. Smoot, G. F., Buffington, A. & Orth, C. D. Search for Cosmic-Ray Antimatter. *Phys. Rev. Lett.* **35,** 258–261 (4 July 1975).

38. Picozza, P. & Morselli, A. Antimatter research in space. *Journal of Physics G: Nuclear and Particle Physics* **29,** 903 (2003).

39. Nagashima, K., Sakakibara, S., Murakami, K. & Morishita, I. Response and yield functions of neutron monitor, galactic cosmic-ray spectrum and its solar modulation, derived from all the available world-wide surveys. *Il Nuovo Cimento C* **12,** 173–209. ISSN: 1124-1896 (2 1989).

40. Cronin, J., Swordy, S. & Gaisser, T. Cosmic rays at the energy frontier. *Sci.Am.* **276,** 32–37 (1997).

41. Amenomori, M. *et al.* New Estimation of the Spectral Index of High-Energy Cosmic Rays as Determined by the Compton-Getting Anisotropy. *The Astrophysical Journal Letters* **672,** L53 (2008).

42. Hörandel, J. R. Models of the knee in the energy spectrum of cosmic rays. *Astroparticle Physics* **21,** 241–265. ISSN: 0927-6505 (2004).

43. Abreu, P. *et al.* Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astroparticle Physics* **34,** 314–326. ISSN: 0927-6505 (2010).

44. M.-P. Véron-Cetty & P. Véron. A catalogue of quasars and active nuclei: 12th edition. *A&A* **455,** 773–777 (2006).

45. Hillas, A. M. The Origin of Ultra-High-Energy Cosmic Rays. *Annual Review of Astronomy and Astrophysics* **22,** 425–444 (1984).

46. Olinto, A. Ultra high energy cosmic rays: the theoretical challenge. *Physics Reports* **333–334,** 329–348. ISSN: 0370-1573 (2000).

47. Gaisser, T. K. *Cosmic Rays and Particle Physics* ISBN: 0 521 33931 6 (Cambridge University Press, 1990).

48. Fermi, E. *Phys. Rev.* **75,** 1169 (1949).

49. Combes, F. & Pfenniger, D. Perspectives for detecting cold H_2 in outer galactic disks. *Astronomy and Astrophysics* **327,** 453–466 (Nov. 1997).

50. Lacy, J. *et al.* Detection of absorption by H_2 in molecular clouds: a direct measurement of the H_2:CO ratio. *Astrophysical Journal* **428,** 69–72 (1994).

51. Glover, S. C. O. & Low, M.-M. M. On the relationship between molecular hydrogen and carbon monoxide abundances in molecular clouds. *Monthly Notices of the Royal Astronomical Society* **412,** 337–350. ISSN: 1365-2966 (2011).

52. Stanev, T. *High Energy Cosmic Rays* ISBN: 3 540 40653 0 (Springer-Praxis, 2004).

53. Han, J. Magnetic Fields in our Galaxy: How Much Do We Know? *Astrophysics and Space Science* **278,** 181–184. ISSN: 0004-640X (1 2001).

54. United States Committee on Extension to the Standard Atmosphere (COESA). *U.S. Standard Atmosphere, 1976* <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19770009539_1977009539.pdf> (U.S. Government Printing Office, 1976).

55. Sjöstrand, T. & van Zijl, M. A multiple-interaction model for the event structure in hadron collisions. *Phys. Rev. D* **36,** 2019–2041 (7 Oct. 1987).

56. Heitler, W. *The Quantum Theory of Radiation* ISBN: 9780486645582 (Dover Publications, 1954).

57. Montanus, J. Intermediate models for longitudinal profiles of cosmic showers. *Astroparticle Physics* **35,** 651–659. ISSN: 0927-6505 (2012).

58. Cirigliano, V., Ecker, G., Neufeld, H., Pich, A. & Portolés, J. Kaon decays in the standard model. *Rev. Mod. Phys.* **84,** 399–447 (1 Mar. 2012).

59. Rossi, B. & Greisen, K. Cosmic-Ray Theory. *Rev. Mod. Phys.* **13,** 240–309 (4 Oct. 1941).

60. Coppens, J. Observation of radio signals from air showers at the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **604,** S41–S43. ISSN: 0168-9002 (2009).

61. Pinfold, J. L. <http://web.me.com/jamespinfold/Site/ALTA-NALTA-etc..html>.

62. Pinfold, J. L. *High schools focus on the extreme universe* <http://cerncourier.com/cws/article/cern/29604> (2006).

63. Timmermans, C., Schotanus, J., de Jong, S. & Kuijpers, J. *NAHSA: Nijmegen Area High School Array* (Katholieke Universiteit Nijmegen, Afdeling Experimentele Hoge Energie Fysica, 2001). <http://www.hef.ru.nl/~timmer/nahsa/doc/project.ps>.

64. <http://www.nikhef.nl/~henkjan/HISPARC.html>.

65. Pearce, M. *SEASA: The Stockholm Educational Air Shower Array* 2002. <http://www.particle.kth.se/SEASA/project.pdf>.

66. Smolek, K., Blaschke, F., Cermák, J., Lichard, P., Pinfold, J., *et al.* in *Proceedings of the 31*st *ICRC* (Łódź, 2009).

67. (ed Nieuwe Natuurkunde) *Domein E3. Kern- en deeltjesprocessen* <http://www.nieuwenatuurkunde.nl/disclaimer/28>.

68. Schultheiss, N. & Kortland, K. *RouteNet* <http://www.hisparc.nl/docent-student/lesmateriaal/routenet-lesbrieven/>.

69. (ed Natuur, Leven en Technologie) *Kosmische straling* <http://betavak-nlt.nl/les/modules_v/gecertificeerd/Kosmische_straling/>.

70. *FOM verslag LIO's 2008 – 2009* (ed van Eijk, B.) (FOM, 2009). <http://www.hisparc.nl/fileadmin/HiSPARC/documenten/Leraar_in_Onderzoek/HiSPARC_LIO_document_151009.pdf>.

71. *FOM verslag LIO's 2009 – 2010* (ed van Eijk, B.) (FOM, 2010). <http://www.hisparc.nl/fileadmin/HiSPARC/documenten/Leraar_in_Onderzoek/HiSPARC_LiO_document_071210.pdf>.

72. *FOM verslag LIO's 2010 – 2011* (ed van Eijk, B.) (FOM, 2011). <http://www.hisparc.nl/fileadmin/HiSPARC/documenten/Leraar_in_Onderzoek/HiSPARC_LiO_document_271011.pdf>.

73. Molecular and Biophysics group, Radboud University. *FLARE — Free-electron Laser for Advanced spectroscopy and high-Resolution Experiments* <http://www.ru.nl/flare/>.

74. Nakamura, K. *et al.* Review of Particle Physics. *Journal of Physics G: Nuclear and Particle Physics* **37,** 075021 (2010).

75. National Institute of Standards and Technology. *ESTAR. Stopping-Power and Range Tables for Electrons* <http://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html>.

76. Particle Data Group. *Atomic and Nuclear Properties of Materials* <http://pdg.lbl.gov/2011/AtomicNuclearProperties/>.

77. Landau, L. On the energy loss of fast particles by ionization. *J. Exp. Phys. (USSR)* **8,** 201–205 (1944).

78. Leo, W. R. *Techniques for nuclear and particle physics experiments. A how-to approach* (Springer, 1994).

79. Kölbig, K. & Schorr, B. A program package for the Landau distribution. *Computer Physics Communications* **31,** 97–111. ISSN: 0010-4655 (1984).

80. Marucho, M., Canal, C. A. G. & Fanchiotti, H. The Landau Distribution for Charged Particles Traversing Thin Films. *Int.J.Mod.Phys.* **C17,** 1461–1476 (2006).

81. Saint-Gobain Crystals. *Organic Scintillation Materials (BC-408)* (2011). <http://www.detectors.saint-gobain.com/uploadedFiles/SGdetectors/Documents/Brochures/Organics-Brochure.pdf>.

82. Lee, S., Park, S., Yea, K.-h. & Cha, H. A time-gating scintillation detector for the measurement of laser-induced fast neutrons. *Review of Scientific Instruments* **80,** 063303 (2009).

83. ET Enterprises, Ltd. *9107B series data sheet* (2010). <http://my.et-enterprises.com/pdf/9107B.pdf>.

84. Steijger, J. HiSPARC Internal note. Jan. 2010.

85. Buisman, H. in *FOM verslag LIO's 2010 – 2011* (FOM, 2011).

86. Trimble. *Resolution T GPS Timing Receiver* <http://trl.trimble.com/docushare/dsweb/Get/Document-550777/>.

87. National Instruments. *LabVIEW System Design Software* <http://www.ni.com/labview/>.

88. Pennink, D. in *FOM verslag LIO's 2009 – 2010* (FOM, 2010).

89. Bosboom, N. in *FOM verslag LIO's 2010 – 2011* (FOM, 2011).

90. National Aeronautics and Space Administration. *Report of the EOS data panel* Earth Observing System, Data and Information System, Data Panel Report, Vol. IIa. NASA Technical Memorandum 87777 (June 1986). <http://hdl.handle.net/2060/19860021622>.

91. O'Toole, G. <http://quoteinvestigator.com/2011/05/13/einstein-simple/> (2011).

92. Alted, F., Vilata, I., *et al. PyTables: Hierarchical Datasets in Python* 2002–. <http://www.pytables.org/>.

93. Django core team. *Django: A Web framework for the Python programming language* Lawrence, Kansas, U.S.A.: Django Software Foundation, 2005–. <http://www.djangoproject.com>.

94. OpenVPN project. *OpenVPN: Providing Secure Access Anywhere in the World* 2002–. <http://openvpn.net/>.

95. Galstad, E. *et al. Nagios: The Industry Standard in IT Infrastructure Monitoring* 1999–. <http://www.nagios.org/>.

96. Kaplinsky, C. *et al. TightVNC: VNC-Compatible Free Remote Control | Remote Desktop Software* 2001–. <http://www.tightvnc.com/>.

97. ATLAS Collaboration. *ATLAS Detector and Physics Performance, Volume I* Technical Design Report 14 (May 1999). <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html>.

98. Fayad, M. & Schmidt, D. C. Object-oriented application frameworks. *Commun. ACM* **40,** 32–38. ISSN: 0001-0782 (10 Oct. 1997).

99. Fokkema, D. B. R. A. *SAPPHiRE Documentation. A Framework for HiSPARC* (2012). <http://hisparc.github.com/sapphire>.

100. Royce, W. Managing the Development of Large Software Systems. *Proceedings of IEEE WESCON* **26** (1970).

101. *Agile software development* <http://en.wikipedia.org/wiki/Agile_software_development>.

102. Beck, Kent, *et al. Manifesto for Agile Software Development* <http://agilemanifesto.org/> (2001).

103. *Bazaar* <http://bazaar.canonical.com/en/>.

104. *Git, the fast version control system* <http://git-scm.com/>.

105. *Github* <http://github.com/>.

106. Mayer, H. A fast reconstruntion method for shower direction at large extended air shower arrays. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **330,** 254–258. ISSN: 0168-9002 (1993).

107. Sciutto, S. *AIRES: A System for air shower simulations. User's guide and reference manual. Version 2.2.0* (1999). arXiv:astro-ph/9911331 [astro-ph].

108. Heck, D., Knapp, J., Capdevielle, J., Schatz, G. & Thouw, T. *Report FZKA 6019* (1998). <http://www-ik.fzk.de/corsika/physics_description/corsika_phys.html>.

109. Knapp, J., Heck, D., Sciutto, S., Dova, M. & Risse, M. Extensive air shower simulations at the highest energies. *Astropart.Phys.* **19,** 77–99 (2003).

110. Hillas, A. in *International Cosmic Ray Conference* **8** (1981), 193–196.

111. Hillas, A. Shower simulation: lessons from MOCCA. *Nuclear Physics B - Proceedings Supplements* **52,** 29 –42. ISSN: 0920-5632 (1997).

112. Billoir, P. A sampling procedure to regenerate particles in a ground detector from a 'thinned' air shower simulation output. *Astroparticle Physics* **30,** 270–285. ISSN: 0927-6505 (2008).

113. Stokes, B., Cady, R., Ivanov, D., Matthews, J. & Thomson, G. Dethinning Extensive Air Shower Simulations. arXiv:1104.3182 [astro-ph.IM] (2011).

114. Bruijn, R., Schmidt, F., Ilee, J. & Knapp, J. in *Proceedings of the 31*st *ICRC* (Łódź, 2009).

115. Steijger, J. HiSPARC Internal note. Oct. 2010.

116. Steijger, J. *Time of first hit* HiSPARC Internal note. Aug. 2010.

117. Antoni, T. *et al.* The Cosmic ray experiment KASCADE. *Nucl. Instrum. Meth.* **A513,** 490–510 (2003).

118. Badea, A. *et al.* The KASCADE-Grande experiment and the LOPES project. *Nucl.Phys.Proc.Suppl.* **136,** 384–389 (2004).

119. Iori, M. *et al.* in *Proceedings of the 31*st *ICRC* (Łódź, 2009).

120. Antoni, T. *et al.* Electron, muon, and hadron lateral distributions measured in air-showers by the KASCADE experiment. *Astropart.Phys.* **14,** 245–260 (2001).

121. Schieler, H. Private communication. Dec. 2010.

122. Meyhandan, R. *et al.* in *International Cosmic Ray Conference* **2** (Aug. 1991), 712–+.

123. Haungs, A. Private communication. Jan. 2012.

124. Apel, W. *et al.* Comparison of measured and simulated lateral distributions for electrons and muons with KASCADE. *Astroparticle Physics* **24,** 467–483. ISSN: 0927-6505 (2006).

125. Montanus, H. in *FOM verslag LIO's 2009–2010* 45–58 (FOM, 2010).

126. Steijger, J. *Energy reconstruction in HiSPARC* HiSPARC Internal note. Apr. 2012.

127. Bosboom, N. in *FOM verslag LIO's 2011 – 2012* (FOM, 2012).

Dankwoord

Het was een lange weg. Er zijn niet veel promovendi die hun werk in deeltijd doen. Het blijkt, na onderzoek, dat er ook niet veel promovendi op tijd klaar zijn. De afgelopen vijf jaar (en daarvoor) heb ik voldoende mensen ontmoet die uiteindelijk mijn proefschrift mede hebben gevormd.

Allereerst wil ik Gerrit bedanken, die mij, hoewel ik in eerste instantie koos voor het onderwijs, aangemoedigd heeft verder te blijven kijken en het onderzoek niet de rug toe te keren. Je belde mij zodra je de advertentie onder ogen had gekregen (ik had hem ook al gezien) en je moedigde me aan contact op te nemen met Bob. Bob was gek genoeg om een aanstaande vader aan te nemen die a) ruim drie jaar eerder was afgestudeerd en in eerste instantie niet had gekozen voor onderzoek, b) niet meteen kon beginnen vanwege lopende examenklassen en c) enkel in deeltijd wilde werken. Bob, ik ben blij dat we er uit gekomen zijn en ik heb de afgelopen jaren mogen zien hoe je je vol vuur inzette voor HiSPARC. Gepassioneerd vertel je over natuurkunde, of het nu mij, studenten, LIO's of VWO-leerlingen betreft. Je ziet altijd mogelijkheden en 'potjes' en zodoende is HiSPARC de afgelopen jaren flink gegroeid. Als man van het eerste uur is je kennis van HiSPARC ongekend. Charles, jij was er zelfs nóg eerder bij als mede-oprichter van NAHSA, de voorloper van HiSPARC. Ik wil je danken voor je hulp bij het leggen van de contacten met KASCADE en het opzetten van de eerste analyse. Ook wil ik je danken voor de nauwgezetheid waarmee je mijn manuscript hebt gelezen.

I'd like to thank Andreas Haungs and Harald Schieler for their hospitality and help in setting up the HiSPARC station at KASCADE, rebooting our computer from time to time (damn those Windows updates) and keeping the lawnmowers off the grass lest they be cutting up our cables. Harald, thank you for driving me through awful weather and rush hour to catch my train, because the last bus leaves early on Fridays. Vitor de Souza and Jürgen Wochele, thank you for providing us with the data needed for the analysis in Chapter 5.

Jos, hoe kan ik je ooit danken voor de hulp, sturing, het lezen van alle versies van pamfletjes en hoofdstukken, goede ideeën, inzichten, weekendmailtjes, bemoedigende woorden en al die andere dingen. Hoe je er telkens weer in slaagt om op maandagochtend een pamfletje te produceren dat een uitgewerkte mini-Monte Carlo beschrijft met figuren en grafieken, en dat op vrijdag nog niet bestond, zal voor mij wel altijd een raadsel

blijven. Ik heb ontzettend veel aan onze gesprekken gehad! De resultaten van jouw pamfletjes worden her en der in dit proefschrift gebruikt.

Ook veel anderen die betrokken zijn bij HiSPARC ben ik veel dank verschuldigd. Jan-Willem, dank je voor het lezen van mijn manuscript en je opmerkingen. Hans, dank je voor je tijd wanneer ik weer eens vragen had over de elektronica. Ik wil alle leraren-in-onderzoek bedanken voor hun inzet en voor de fijne samenwerking. In het bijzonder wil ik noemen Dorrith en Niels. Dorrith, jouw vasthoudendheid in het leren van Python, de werking van de detectoren en de onderliggende natuurkunde heb ik altijd bewonderd. Je bent een onzettend harde werker en ik heb twee jaar op de maandagen van jouw gezelschap mogen genieten. Je hebt veel bereikt in die twee jaar en jouw resultaten hebben ook in dit proefschrift een plek gevonden in Hoofdstuk 2 en Hoofdstuk 5. Niels, ook jij hebt je twee jaar flink ingezet voor HiSPARC door het schrijven van simulaties en het uitvoeren van analyses op de data van het Science Park cluster. We hebben heel wat code uitgewisseld en soms gingen we dezelfde richting op, om daarna weer onze eigen weg in te slaan. Het was fijn om met je samen te werken! Jouw resultaten hebben ook een plek gevonden, en wel in Hoofdstukken 2, 6 en 7. Niek, ook jij bent een trouwe kamergenoot en ik waardeer je inzet om de HiSPARC data dicht bij de leerlingen te brengen door middel van interactieve analyses in een web browser. De overige LiO's hebben ook veel werk verzet voor HiSPARC en de afgelopen jaren heb ik verder nog samen mogen werken met Machiel, Arjan, Wim, Guus, Hans, Remon, Henk, Daniël, Paul, Wytse en Sjoerd, bedankt voor jullie werk! Ik heb met nog veel meer mensen samengewerkt die als student of in een andere hoedanigheid betrokken zijn of waren bij HiSPARC. In het bijzonder wil ik nog noemen Aart-Jan (lesmateriaal muonbalken) en natuurlijk Loran (weerdata, scripts voor leerlingen). Arne, jij neemt het stokje van mij over. In een jaar tijd heb je laten zien dat je niet alleen Python onder de knie hebt, maar dat je een grotere PEP-8 extremist bent dan ik (hou vol!).

Een groot deel van mijn eerste jaren bestond uit het helpen opzetten van een nieuw softwaresysteem voor HiSPARC. Dan kan ik natuurlijk niet voorbijgaan aan Tristan en Karel. Zonder jullie werk gingen we nog elke dag op de fiets een paar scholen langs om pc's af te regelen. Jeroen, onze *resident LabVIEW expert*, heeft over een langere periode aan de DAQ software gewerkt dan ik heb gedaan over mijn promotie. Bedankt! Het OOTI team uit Eindhoven, een groep post-doctorale Master studenten van de TU/E heeft het beheer van de Windows software sterk vereenvoudigd. Nog vele andere studenten hebben hun steen bijgedragen.

Dan de CT-groep: zonder jullie zat HiSPARC allang aan de grond. Dank aan Robert voor de overname van het beheer van de Windows software en het vernieuwen van de

installer. Bart, dank voor je werk aan de publieke database en betere wifi in H343c. André, jij hebt veel van onze servers geïnstalleerd en was, vóór de helpdesk-dienst bestond, ons eerste aanspreekpunt. Ton, als UN*X administrator heb je niet alleen onze servers gereanimeerd wanneer dat nodig bleek, maar heb je met veel geduld onze hardware crashes opgelost. Schijven, SCSI interfaces, geheugen, we hebben het geloof ik allemaal wel zien stukgaan. Je creatieve constructie via twee servers houdt op dit moment onze RAID al een jaar in de lucht. Dank! Wim, bedankt dat je altijd weer iemand wist die voor ons aan de slag kon.

Ik wil mijn oud-en-nieuw collega's op het Kaj Munk bedanken voor hun afscheid (vijf jaar terug), aanhoudende interesse, gevoel van welkom wanneer ik een keer op school was, het rotsvaste vertrouwen (in ieder geval van sommigen) dat ik geen leukere baan zou kunnen vinden om de simpele reden dat die niet bestaat, en het warme welkom toen dat het geval bleek en ik graag weer terugkwam op mijn oude stekkie. Albert, dank voor de inspiratie tijdens mijn eigen schooljaren (uiteindelijk mede resulterend in dit proefschrift) en dat ik nu (weer) jouw collega mag zijn om te leren hoe je die perfectie van goed voorbereide lessen bereikt.

Mijn vrienden en familie hebben de afgelopen vijf jaar behoorlijk last van me gehad. Nooit tijd hebbend en altijd druk, werd mijn wereldje tamelijk klein: werk en gezin(s-uitbreiding x2). Jullie steun en vertrouwen is voor mij al die jaren ontzettend belangrijk geweest. Ivo, ooit brachten wij uren per dag in elkaars gezelschap door, nu is dat anders. Maar je bent altijd mijn broer gebleven en staat nog steeds voor mij klaar! Dank je dat je mijn paranimf wilt zijn. Bart, de nieuwste aanwinst in de familie ('eindelijk' heb je Karen ten huwelijk gevraagd!), ik kan het goed met je vinden en ik ben blij dat ook jij mijn paranimf wilt zijn.

Esther en Hannah, jullie hele leven heb ik aan mijn proefschrift gewerkt. En jullie zitten allebei al op school! Twee grote meiden. Ik ben trots op jullie! Ik ben blij dat ik niets van jullie heb hoeven missen. Twee dagen per week was ik helemaal alleen met jullie thuis en de flessen werden pap en fruithapjes, en boterhammen en nu gewoon wat de pot schaft. Jullie passen niet meer in het kommetje van mijn arm, maar jullie knuffelen nog steeds even fijn en niets is beter tegen de stress van alledag dan jullie boekjes voorlezen of samen met de Duplo-trein te spelen. Ik hou van jullie!

Papa en mama, jullie hebben mij altijd het gevoel gegeven dat ik bijzonder was en hebben nooit geaarzeld al mijn vragen te beantwoorden (en dat waren er nogal wat). Ik herinner me het verhaal dat jullie ons vergeleken met een melkfles (met smalle hals). Giet al die antwoorden er maar in. Op jonge leeftijd gaat het grootste deel er naast, maar het deel dat er wél in komt is mooi meegenomen. En dus is het beter om al die

vragen maar te beantwoorden (op het vermoeiende af) dan op jonge leeftijd het vragen al af te leren. Het is gelukt, en ik ben mijn hele leven vragen blijven stellen. Op een aantal heb ik een antwoord gevonden.

> Voici mon secret. Il est très simple: on ne voit bien qu'avec le cœur.
> L'essentiel est invisible pour les yeux.
>
> *Hier is mijn geheim. Het is zeer eenvoudig: je kunt slechts werkelijk goed zien met het hart. Het wezenlijke is onzichtbaar voor het oog.*
>
> Antoine de Saint Exupéry
> *Le Petit Prince*

Sommige mensen zijn nu eenmaal voor elkaar geschapen. Jij, Eveline, bent voor mij geschapen. Je brengt orde in mijn chaos, bent een onuitputtelijke bron van liefde en steun, en hebt er letterlijk voor gezorgd dat ik de afgelopen jaren heb kunnen volbrengen. Er is niemand die mij beter kent en je neemt mijn zwakheden voor lief. De verwachtingen die je van mij hebt maken me een beter mens. Ik ben blij dat we de zorg voor onze meiden gelijk verdeelden en dat ik zodoende vijf jaar lang vier dagen per week met ze doorbracht, waarvan twee met elkaar als gezin. Ik had het voor geen goud willen missen! In de dertien jaar die we samen zijn hebben we al aardig wat meegemaakt. Soms sleepte jij mij er door heen. Soms sleepte ik jou er door heen. Meestal sleepten we elkaar er door heen. Je hebt je altijd vol overgave aan mij en de meiden gewijd, en het is heerlijk om te zien hoe dol de meiden op je zijn. Ik ook! Ik ken niemand die zo hard werkt en overal de schouders onder zet. Je weet wat je wilt en gaat er voor. Je daadkracht en zelfvertrouwen zijn dan om jaloers op te worden. Ik ben dankbaar dat hoewel we zo verschillend zijn, we zo ontzettend hetzelfde denken over alles wat er écht toe doet. Als ik het even niet meer weet, weet jij het wél. Ik ben trots op je. Je bent mijn maatje, mijn partner, mijn vrouw. Ik hou van je!

David Fokkema,
September 2012