Reconstructing the cosmic ray energy spectrum with HiSPARC

K. van Dam [a,b,*], B. van Eijk [a,c], J.J.M. Steijger [a]

[a] *Dutch National Institute for Subatomic Physics (Nikhef), Science Park 105, 1098 XG Amsterdam, The Netherlands*
[b] *Royal Netherlands Meteorological Institute, Utrechtseweg 297, 3731 GA De Bilt, The Netherlands*
[c] *Faculty of Science and Technology, University of Twente, 7500 AE Enschede, The Netherlands*

ARTICLE INFO

ABSTRACT

The high school project on astrophysics research with cosmics (HiSPARC) employs a large number of small detection stations that sample the footprint of extensive cosmic ray air showers. The cosmic ray energy spectrum between 10^{16}–10^{19} eV is reconstructed by combining data from multiple four-detector stations located at the Amsterdam Science Park. The obtained cosmic ray flux values agree well with data from other experiments. Also the value of the slope (-3.08 ± 0.02) matches the literature value. Additionally, a new method is presented in which a single two-detector station is used to probe the cosmic ray energy spectrum at a complementary energy regime between 10^{12}–10^{16} eV. The flux values and its slope value (-2.71 ± 0.07) are in good agreement with literature data. However, numerous simplifying assumptions bring about large systematic uncertainties which are not well quantified. The possibility to reconstruct the cosmic ray energy spectrum augments HiSPARC's primary goal, which is to engage high school students with modern physics and let them contribute to real scientific research.

1. Introduction

Cosmic rays are charged particles from space with energies ranging from GeV to hundreds of EeV. The majority of cosmic rays in the GeV-TeV range are protons (90%) and helium nuclei (9%) with a remainder of heavier elements [1]. The cosmic ray flux drops rapidly with increasing energy. The flux as function of energy (E) in the range from GeV to 10^{14} eV is approximately given by [2]:

$$I(E) \approx 1.8 \cdot 10^4 \ (E/1 \ \text{GeV})^{-\alpha} \frac{\text{\# of nucleons}}{\text{m}^2 \ \text{s sr GeV}} \tag{1}$$

with $\alpha \approx 2.7$. At an energy of $\sim 10^{15.5}$ eV the spectrum steepens. This region is known as 'the knee'. At $\sim 10^{18.5}$ eV the slope of the spectrum becomes less steep, this is called 'the ankle'. Cosmic rays with energies below $\sim 10^{14}$ eV can be measured using high altitude balloons or spacecraft. For higher energies this approach becomes unfeasible because of the low cosmic ray flux, the limited collection area and lifetime of space based instruments. Above $\sim 10^{14}$ eV cosmic rays are studied using the Earth's atmosphere. When a cosmic ray enters the atmosphere it will most likely interact with a nitrogen or oxygen nucleus. In this interaction new particles may be produced which, in turn, will interact with other atmospheric particles. The particle generation continues until there is insufficient energy left for the creation of additional particles. This phenomenon is known as the development of an extensive air shower (EAS). The particle footprint of an EAS at the ground is sampled using detector arrays. The direction and energy of the cosmic ray can be reconstructed from particle arrival times and multiplicities. Also atmospheric radiation generated by the EAS, such as atmospheric Cherenkov and fluorescence, and radio waves, can be used to reconstruct the shower size and direction [1]. Cosmic rays with energies below $\sim 10^{14}$ eV do not result in large air shower footprints. The particle densities are too small to reconstruct the direction and energy of the primary cosmic ray.

The high school project on astrophysics research with cosmics (HiSPARC [3]) employs a large number of small EAS detection stations. The stations are predominantly hosted by high schools. Each station consists of two or four detectors. A HiSPARC detector consists of a rectangular scintillator (100 cm × 50 cm × 2 cm) glued to a light-guide which is attached to a photo-multiplier tube (PMT). The detector is made light tight with a thin aluminum foil and pond liner. The detectors are placed inside a roof box and mounted on top of a roof. Together with a GPS receiver for accurate timing and position information the detectors are connected to an electronics unit. The distance between the detectors in a two-detector station varies per setup. In this paper a detector separation of 4.95 m is used. In a four-detector station the direction can be triangulated with a resolution of $\sim 6°$. An estimate of the shower energy can also be made. An extensive description of HiSPARC detectors, stations and network is presented in [3].
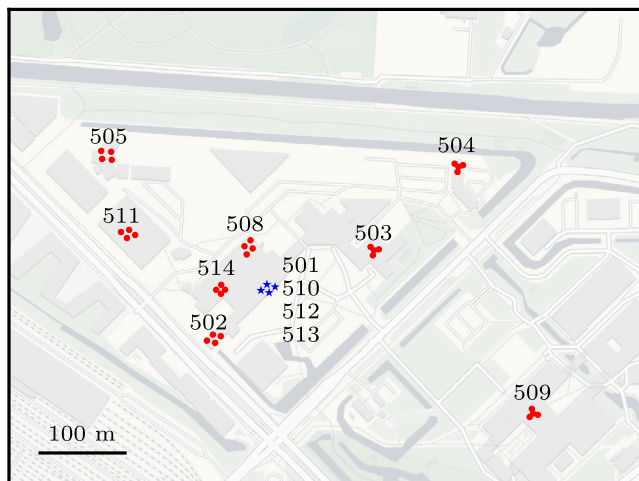

Fig. 1. Location of stations at the Amsterdam Science Park cluster since late 2018. Each red dot represents a detector and each combination of four dots forms a station. The station numbers are also denoted. At one location, four 'diamond' stations are essentially positioned 'on top of each other' (stations are shifted ~1 m with respect to each other) and only a single station is displayed (blue stars).

Table 1
Station roof heights with respect to station 501. The height of station 509 is obtained from GPS measurements (less precise), other heights come from a technical map provided by Nikhef (cm precision).

Station	Height [m]
501	0
502	0
503	−6.00
504	−0.94
505	−7.29
508	−5.33
509	0.7
510	0
511	−0.77
512	0
513	0
514	3.60

In this article the cosmic ray energy spectrum between $10^{16}-10^{19}$ eV is reconstructed by combining data from multiple four-detector stations located at the Amsterdam Science Park (Section 2). Additionally, in Section 3 a new method is presented in which a single two-detector station is used to probe the cosmic ray energy spectrum at a complementary energy regime between $10^{12}-10^{16}$ eV. The results of both methods are discussed in Section 4.

2. Science Park cluster

At the Amsterdam Science Park a collection of 12 HiSPARC stations is placed on the roofs of Nikhef and other scientific institutions. A thirteenth station is placed inside the Nikhef building. These stations are located close enough to each other such that they can sample the same EAS. Fig. 1 shows the locations of the 12 roof stations since late 2018. The station numbers are also denoted. At one location, four stations are essentially positioned 'on top of each other' (equivalent detectors are ~1 m apart) and only a single station is displayed (blue stars). The roof – and station – heights differ substantially. Table 1 lists the roof heights obtained from a technical map provided by Nikhef (cm precision). The height of station 509 is obtained from GPS measurements and is somewhat less precise.

2.1. Data selection and quality

Station coincidences between stations 501, 502, 503, 504, 505, 508, 509, 511 and 514 are searched for in the database from November 2018 to September 2020. During this period the stations are located as displayed in Fig. 1. Station coincidences are events in which multiple stations detect the same EAS and trigger within 10 μs. The trigger time is the moment at which the station trigger conditions are met (see [3]). The probability that a station coincidence is caused by two uncorrelated EASs is negligible. The detection rate (r) for EASs beyond 10^{14} eV in a single station is approximately 0.2 Hz. This rate rapidly decreases for higher energy showers. A HiSPARC station detects two EASs within a 10 μs time window (τ) every $1/(2\tau r^2) = 1.25 \cdot 10^6$ s or every 2 weeks.

The stations were fully operational most of the time. Unfortunately, problems with the absolute timing of station 508 rendered a large part of that data useless for comparison with other stations. Issues with the computer of station 511 result in quite some red periods. Station 509 was fully operational the entire period.

Only station coincidences in which stations 501, 502, 503 and 514 have triggered are considered. Requiring all Science Park stations to have detected the EAS results in a small selection of very high energy EASs. On the other hand, if only two or three stations are required to have sampled the shower, the number of EAS samples is rather limited for energy reconstruction. A minimum of four stations was deemed the best trade-off between including low energy EASs ($\sim 10^{16}$ eV) and having a sufficient number of samples to reconstruct the EASs. Stations 501, 502, 503 and 514 were chosen because of their central location and proximity to each other. Station 508, despite its central location, is not included in this selection criterion because of the large period of missing data. Additionally, as the size of the EAS is typically much larger than the distance between these four stations, the core positions of the detected EASs lie approximately within a disk around these central stations. This simplifies the determination of the effective surface area of the Science Park cluster from experimental data. A more loose requirement of, for example, at least four stations (and it does not matter which ones) results in a distribution of core positions that does not approximate a disk and subsequently complicates the calculation of the effective surface area. Only one station (501) of the four stations 'on top of each other' is considered to avoid oversampling one part of the shower footprint.

When a charged particle traverses the scintillator, it creates a light pulse which is converted into an electric pulse by the PMT. This pulse is sampled and digitized at 400 MHz. Also the pulse integral (area under the curve) is determined [4]. The PMT is calibrated such that single charged minimum ionizing particles (MIPs) generate pulses with an amplitude of ~150 mV (150 mV = 1 MIP). A 30 mV threshold rejects PMT noise and the response to low energy gamma rays. Each day the most probable energy loss of a single particle (value at which Landau distribution peaks, here referred to as 'MIP peak') is determined from the pulse integrals for each detector [4]. The number of particles traversing the detector is obtained by dividing the pulse integral value by the MIP peak value of that day. This way, large scale fluctuations in signal strength due to PMT temperature fluctuations or changes to the gain settings, are corrected for. Detections with a particle count smaller than 0.5 MIP are discarded. This way random coincidences caused by gamma rays from nearby radioactive decays (see [5]) or other small signals are excluded. Random coincident muons typically generate signals above 0.5 MIP and may still be included as it is impossible to distinguish a late shower particle from an uncorrelated muon. The probability for this to occur is small (1.5 μs trigger window per station with a muon rate of ~81 Hz, see [4]). Moreover, as the uncertainty of the particle count measured by a detector is typically several MIPs, one extra muon does not significantly affect that EAS footprint sample. Other non-EAS signals (e.g. electromagnetic radiation from nearby telecommunication antennas picked up by the cables, etc.) with a pulse height above 0.5 MIP are not excluded either. However,

as the rate for these kinds of events is extremely small, the probability that such a signal coincides with an EAS is negligible. Distortion of the signal, for example due to incorrect alignment of the HiSPARC electronics (see [3]), also may occur. The pulse integral is in these cases expected to still represent the number of particles traversing the detector as the distortion does not modify the underlying signal.

Two PMT bases are used, a commercial one and an in-house developed one [3]. Both base types have different characteristics. For PMT assemblies in which the Nikhef base is used, the PMT pulses start to exceed the ADC range (~2.2 V) if more than ~15 MIP (~150 mV × 15 = 2.25 V) traverse the scintillator. For assemblies with a commercial base the PMT pulses get spread out in time; inadvertently keeping the pulse heights within ADC range [4]. The determination of the pulse integral does not account for values exceeding the ADC range. Therefore, the number of large pulse integrals detected by detectors with a Nikhef PMT base is underestimated. Especially above 30 MIP this effect becomes apparent. In this analysis, the underestimation of the pulse integral by PMTs with Nikhef base is limited by clipping the particle counts to 30 MIP. Only 7 of the 36 PMTs in the 9 stations employ a Nikhef PMT base. For all detectors, no pulse integrals over 100 MIP are detected whereas, as the shower core sometimes falls on a detector, the number of particles traversing the detector could be well over 1000. This indicates that for very large particle numbers the determination of the number of particles traversing the scintillator is inaccurate, irrespective of the PMT base used.

Finally, only events with a reconstructed zenith angle smaller than 30° are considered. This way the difference in particle densities in EAS footprints due to extra attenuation because of larger path lengths through the atmosphere is limited.

2.2. Timing offsets

In order to reconstruct the direction of the EAS the timing offsets between the stations and the detectors within a station need to be corrected for. The timing offsets are calculated from the distribution of arrival time differences. Also the height difference (in light-time) between the stations is taken into account. This calculation is based on the assumption that the distribution of the angle of incidence of EASs is isotropic, and that each detector or station in a combination of two has equal probability to be hit first [4]. Fig. 2 shows the change of the timing offsets of the Science Park stations over time with respect to station 501. The time period required to determine the offset depends on the distance to station 501 as only high energy EASs (rarer) have sufficiently large footprints to cause coincidences. This makes that the time step at which the offsets can be determined differs per station (e.g. one data point every 8 days for station 509). For most stations the timing offsets are reasonably constant and only significantly change if there was a station reconfiguration (GPS reinitialization, hardware change, etc.). These moments are indicated with green vertical lines. The red lines show the mean timing offsets. These mean values are used to correct for the offsets. Stations 510, 512 and 513 are placed 'on top of' station 501 (blue stars in Fig. 1). Because of their proximity the offsets can be determined more frequently. Especially station 510 shows that the timing offset is not constant but varies slightly within ~10 ns. These small GPS timing offsets do not affect the direction reconstruction as stations typically have a separation of ~100 m. Moreover, only station 501 of the four station 'on top of each other' is considered.

The timing offsets of the detectors within a station are also determined. Fig. 3 shows the change of the detector timing offsets with respect to detector 2 for station 503 over time. The offsets are determined every 24 h with a 2.5 ns precision. The gradual change during the summer months could be due to temperature changes of the PMT as they differently affect the rise time of each PMT and therefore the time at which the signal passes the threshold. Contrary to the station timing offset, no mean values are calculated. Instead, for each coincidence, the determined timing offset that is closest in time is used.

2.3. Direction reconstruction

The direction of the shower is reconstructed assuming a flat shower front using the (triangulation) algorithm described in [6]. Only events having at least two MIPs in the detector are considered. This way, random coincidences caused by muons, gamma rays from nearby radioactive decays or other signals are excluded. The precision of the direction reconstruction is investigated by comparing the reconstructed direction of the cluster to that of individual stations in the coincidence. Also for the station reconstructions only events with two MIPs in all detectors are considered. Fig. 4 shows the distribution of angles between the reconstructed direction from station 501 alone and the direction as reconstructed from the entire Science Park cluster (blue interpolated dots). The increasing solid angle for larger angles has been corrected for (division by $\sin\theta$). The differences do not follow a Gaussian or Lorentzian distribution. The width at half maximum (WHM) is 7.4°. This roughly agrees with the uncertainties found in [3]. Similar values are found for the other stations indicating that the station timing offsets are correctly taken into account. Further away from the shower core the uncertainty increases. The red crosses show the same distribution for events with the shower core at least 100 m away from station 501 (see next section for core determination). A Gaussian is fitted to the distribution (red line) with a σ of 13.3°. The lower precision could be caused by the larger spread in the arrival time of the EAS particles further away from the shower core.

2.4. Energy reconstruction

From the reconstructed zenith and azimuth a vector is defined pointing towards the arrival direction of the cosmic ray. Two unit vectors, $\vec{e}_1$ and $\vec{e}_2$ are defined perpendicular to the direction vector and to each other. The detector positions are projected to the plane spanned by $\vec{e}_1$ and $\vec{e}_2$. The distances from the detector positions to their projected positions on the plane are calculated and used to adjust the arrival times (speed of light). The left plot in Fig. 5 shows an example of the projected detector positions (dots) on the $(\vec{e}_1, \vec{e}_2)$ plane. The size of the dots scales with the number of particles traversing the detector. The colour of the dots indicate the arrival time with respect to the first projected arrival time.

A modification of the Nishimura–Kamata–Greisen formula [1, and references therein] has been used as LDF. In this modification the number of free parameters in the original NKG formula has been reduced. The age parameter which is related to the shower maximum is effectively taken as a constant. In the simplified NKG formula the number of particles N at distance r is given by:

$$N(r) = A \left(\frac{r}{r_0} \right)^a \left(1 + \frac{r}{r_0} \right)^b \tag{2}$$

with $r_0 = 29.6$, $a = -0.566$, $b = -2.57$ and A the fit parameter related to the energy. These values have been obtained from CORSIKA simulations assuming a pure proton composition [4]. For inclined showers the particle numbers are reduced due to the increase in path length through the atmosphere. This is corrected for by using:

$$A_\perp = A \cdot \exp \left(p \left(\frac{1}{\cos\theta} - 1 \right) \right) \tag{3}$$

with θ the zenith angle and $p = 6.937$. The energy (eV) of the primary cosmic ray is then calculated with:

$$\log(E/\text{eV}) = c \cdot (\log(A_\perp) + d) \tag{4}$$

here $c = 0.797$ and $d = 17.62$. The values of p, c and d have been obtained from CORSIKA simulations assuming a pure proton composition [4]. The modified NKG formula is circle symmetric and two parameters are required to determine the core position.

The LDF is fitted to the projected particle numbers to find the position of the shower core and energy of the EAS. If stations were fully operational but did not trigger within the 10 μs trigger window,
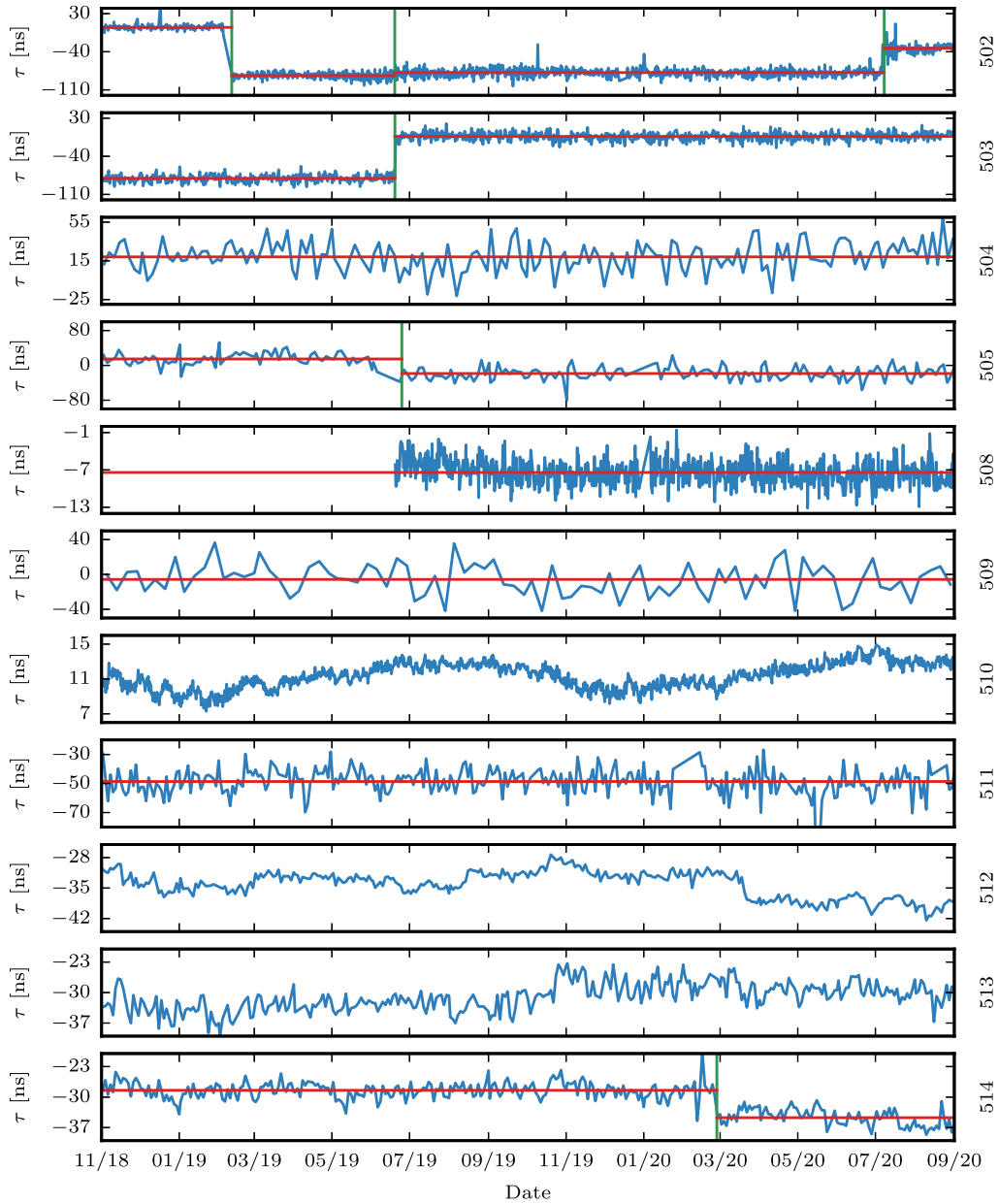
Fig. 2. Timing offsets per station with respect to station 501 over time. The time period required to determine the offset depends on the distance to station 501 and thus differs per station (e.g. one measurement every 8 days for station 509). For most stations the timing offsets are constant and only significantly change if there was a station reconfiguration (GPS reinitialization, hardware change, etc.). These moments are indicated with green vertical lines. The red lines show the mean timing offsets.

a particle count of 0 (with an uncertainty of 1) at those projected detector positions is included in the fit. Stations located at the edge of the shower where the particle density is low but non-zero may not trigger because of Poisson statistics (e.g. no or only 1 detector was hit). For these events the true particle density is underestimated as a particle count of 0 is assumed. However, this approach was deemed the best way to treat the non-trigger stations. Excluding the non-trigger stations in the fit results in unconstrained best fit solutions with core positions far away from the triggered stations. Assuming a non-zero particle count (e.g. 1) would also affect the fit as stations far away from the shower would then be assumed to still have detected the EAS footprint. The cross in the left plot of Fig. 5 indicates the best fit core position. The upper right plot shows the particle counts as a function of distance to the core position (dots) together with the fitted LDF (red line).

In first approximation the EAS shower front is a thin disk. However, at larger core distances there is an increasingly large delay on the

first arrival time. The average energy of the particles decreases with increasing radial distance. Because of larger scattering angles and lower Lorentz factors, this also leads to a larger thickness of the shower front further from the core. The shower front can be approximated by a catenary [7], i.e. a curve that an idealized hanging chain or cable assumes under its own weight when supported only at its ends:

$$t = \frac{a}{2c} \left(e^{\frac{r}{a}} + e^{-\frac{r}{a}} \right) \tag{5}$$

with r the core distance, c the speed of light and a a fit parameter. The bottom right plot in Fig. 5 shows the projected arrival times as a function of distance to the shower core (dots). The red line displays the catenary function fitted to the data points.

2.5. LDF fit results

The direction of ∼91% of the station coincidences that include stations 501, 502, 503 and 514 could be reconstructed. For the remaining
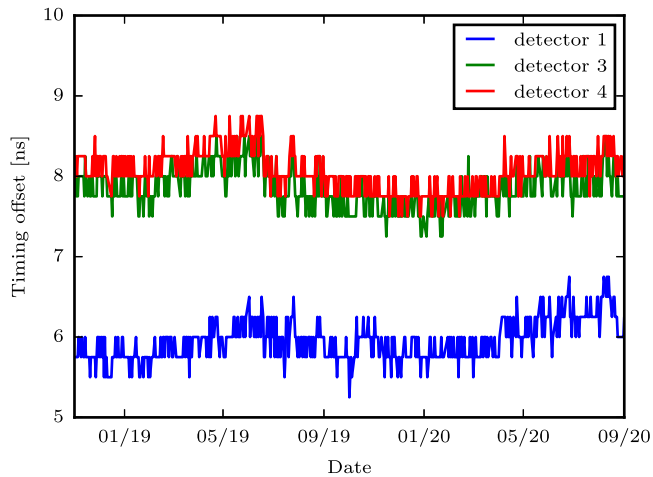
Fig. 3. Detector timing offsets with respect to detector 2 for station 503 over time. The offsets are determined every 24 h with a 2.5 ns precision. The gradual change during the summer months could be due to temperature changes of the PMT.
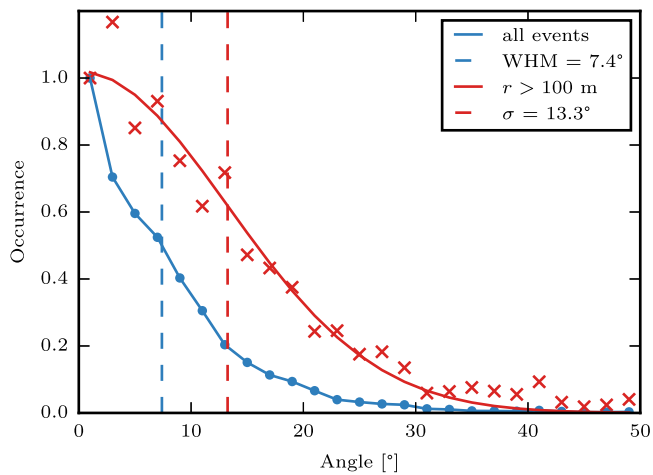

Fig. 4. Distribution of angles between the reconstructed direction from station 501 alone and the direction as reconstructed from the entire Science Park cluster (blue interpolated dots). The increasing solid angle for larger angles has been corrected for (division by $\sin \theta$). The differences do not follow a Gaussian or Lorentzian distribution. The width at half maximum (WHM) is 7.4°. This roughly agrees with the uncertainties found in [3]. Further away from the shower core the uncertainty increases. The red crosses shows the same distribution for events with the shower core at least 100 m away from station 501. A Gaussian is fitted to the distribution (red line) resulting in a σ of 13.3°. Similar values are found for the other stations.

events the arrival time data was inconclusive, for example due to a too limited number of signals above 2 MIP. Of the reconstructed EASs ~28% had a zenith angle larger than 30°. These events were discarded because of an increased systematic error due to Eq. (3) for larger zenith angles. In over 99.9% of the cases the energy could also be determined once the direction reconstruction was successful and the zenith angle was below 30°.

Fig. 6 shows the distribution of the reconstructed EAS energies detected by the Science Park cluster between November 2018 and September 2020. Below ~10^{16} eV the detection efficiency starts to decrease as the particle densities in shower footprints of lower energy EASs are often too low to trigger stations 501, 502, 503 and 514. The non smooth decrease of events with lower energies (bump at ~$10^{15.2}$ eV) is believed to be caused by the station trigger conditions [4]. In the 22 months time period only one event with an estimated energy above 10^{19} eV was detected.

2.6. Effective solid angle

In order to calculate the cosmic ray energy flux from the distribution of reconstructed EAS energies, the effective surface area and effective solid angle of the Science Park cluster are needed. The development and attenuation of an EAS is affected by its path length through the atmosphere. With increasing zenith angle, the path length and thus attenuation increases. Because of this, the acceptance of the Science Park cluster decreases for larger zenith angles. Since only zenith angles between 0° and 30° are considered, this effect is reasonably limited (path length increases with $1/\cos(30°) \approx 1.15$). Fig. 7 shows the distribution of reconstructed zenith angles for six energy bins (width $\log_{10}(E/\text{eV}) = 0.25$). For most energies, the distributions start to flatten at zenith angles approaching 30°. The EAS attenuation rises approximately exponentially with increasing path length. This makes that the EAS detection efficiency as function of zenith angle can be described with:

$$I(\theta, a, b) = a \cdot \exp[-b \cdot (\sec \theta - 1)] \tag{6}$$

with θ the zenith angle and a and b fit parameters. A zenith angle bin in Fig. 7 spans a 'ring' with a solid angle that depends on the zenith angle. This effect ($\sin \theta$) needs to be taken into account in order to fit Eq. (6) to the distributions:

$$I_{\text{fit}}(\theta, a, b) = \sin \theta \times I(\theta, a, b) \tag{7}$$

The red lines in Fig. 7 show the best fits to the data. At 10^{16} eV the fit does not correctly describe the data. It is not clear why this is the case. For high energy EASs (beyond 10^{17} eV) the shower energy is sufficient for the footprint to reach the Earth's surface irrespective of zenith angle. Because of this the lines for the two bottom plots are approximately linear.

The effective surface area is obtained by integrating the detection efficiency (Eq. (6)) over the solid angle of the circular field of view subtended by a rotated zenith angle:

$$\Omega = \int_0^{2\pi} \int_0^{\pi/6} \sin \theta \, \frac{I(\theta)}{I(0)} \, d\theta d\phi \tag{8}$$

with ϕ the azimuthal angle. The detection efficiency is set to 1 at $\theta = 0°$. Fig. 8 shows the effective solid angles as a function of EAS energy. The dashed line indicates the maximum solid angle at 30° ($\Omega = 2\pi(1 - \cos 30°) = 0.84$ sr). Above 10^{17} eV the efficiency is 100%. The red line shows a linear fit excluding the point at 10^{16} eV at which the efficiency fit was inadequate (Fig. 7). This linear fit is used for the effective surface area with a maximum of 0.84 sr.

2.7. Effective surface area

The effective surface area of the Science Park array is obtained by investigating the spatial distribution of reconstructed core positions. Fig. 9 shows the two dimensional distribution of core positions for six energy bins (width $\log_{10}(E/\text{eV}) = 0.25$). The red dots indicate the station positions. As expected, the maximum distance of the detected EAS shower cores to the central stations increases with energy. For lower energies (e.g. 10^{16} eV) there seems to be an overdensity of core positions somewhat right below station 501. As the four stations in the selection criterion (501, 502, 503 and 514) do not form a regular array, some areas are preferred in the fitting procedure as no data points constrain the fit at those positions. For higher energies the core positions lie reasonably homogeneous in a disk around the central stations.

By determining the centre of mass of the two dimensional distributions, the distribution of core distances r can be investigated. Fig. 10 shows these distributions for the six energies. To correct for the geometrical effect of increasing surface area at larger r, the counts in each bin are divided by the surface area of each bin (rings of $2\pi r$). The peaks in the two upper plots at ~50 m are the overdensities in Fig. 9 at
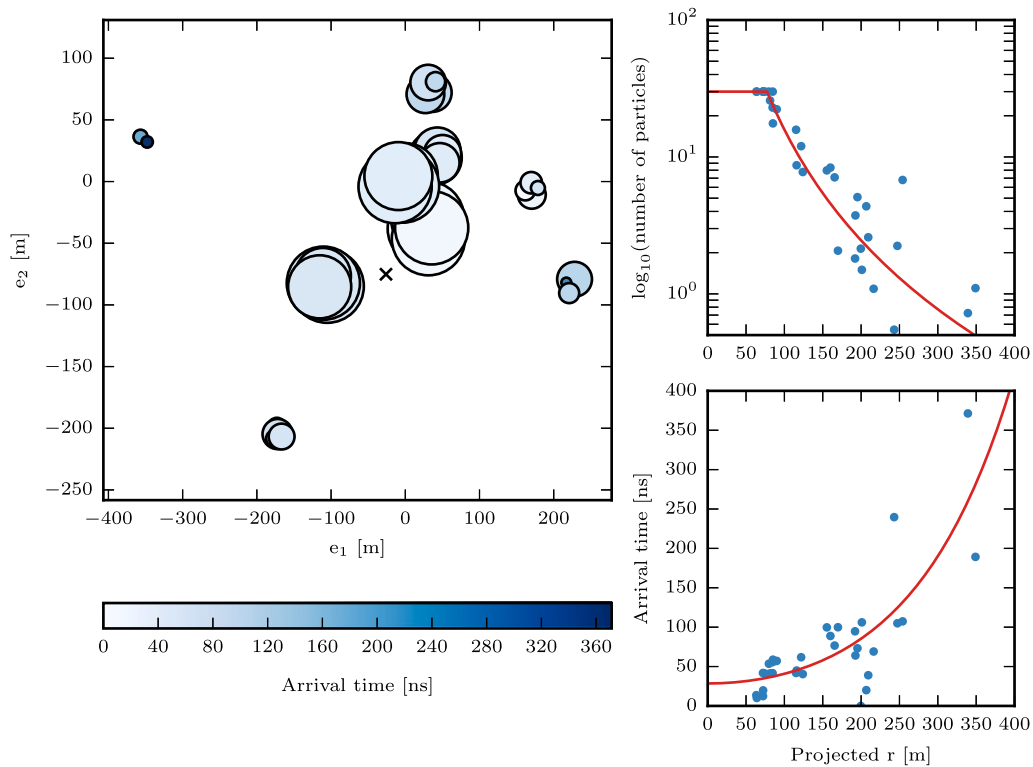
Fig. 5. The sampled particle footprint of a $10^{16.6}$ eV EAS on the $(\vec{e}_1, \vec{e}_2)$ plane perpendicular to the arrival direction of the shower (left) combined with the lateral density function (top right) and shower front arrival times (bottom right). The dots indicate the detector positions on the $(\vec{e}_1, \vec{e}_2)$ plane. The size of the dots scales with the number of particles traversing the detector. The colour of the dots indicate the arrival time with respect to the first particle. The core position (black cross) is obtained by fitting the 2D LDF (Eq. (2)) to the particle counts. Also the energy is derived from this fit (Eq. (4)). The particle counts and the arrival times as function of the distance to the shower core are shown in the upper and lower right figures resp. Also the LDF and catenary fits are shown (red lines).
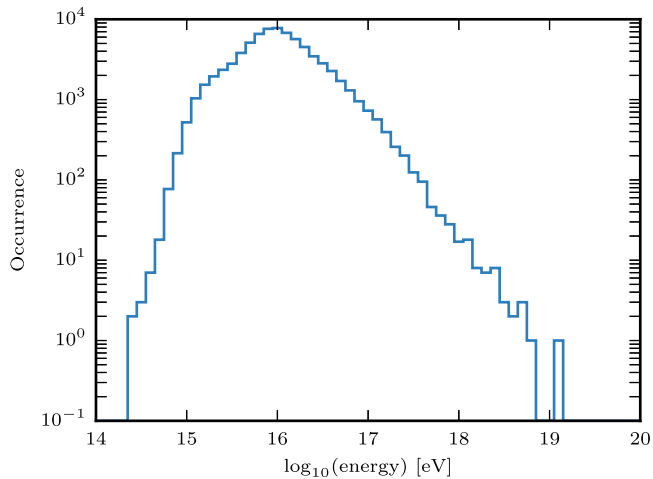

Fig. 6. Histogram of the reconstructed EAS energies. The condition that stations 501, 502, 503 and 514 have triggered imposes a lower limit on the EAS energies of $\sim 10^{16}$ eV. The non smooth decrease of events with lower energies ('bump' at $\sim 10^{15.2}$ eV) is believed to be caused by the trigger conditions. A trigger caused by two signals over the high threshold typically detects a lower particle density than a trigger caused by three signals over the low threshold.

10^{16} and $10^{16.25}$ eV. This representation in terms of core distance shows that the detection efficiency is approximately constant up to a certain radius at which it starts to decrease (e.g. see the fourth plot in Fig. 10). The distributions can be described reasonably well with the following

parametrization:

$$g(r) = \begin{cases} f(r_m, \alpha, \mu, \sigma, \lambda) & \text{for } r < r_m, \\ f(r; \alpha, \mu, \sigma, \lambda) & \text{for } r \geq r_m. \end{cases} \quad (9)$$

Here, all values below r_m have a constant value which is equal to the function $f(r; \alpha, \mu, \sigma, \lambda)$ evaluated at that r_m. All other values follow the exponentially modified Gaussian distribution:

$$f(r, \alpha, \mu, \sigma, \lambda) = \alpha \exp\left[\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2r)\right]$$
$$\times \operatorname{erfc}\left(\frac{\mu + \lambda\sigma^2 - r}{\sqrt{2}\sigma}\right) \quad (10)$$

Here α is a scaling parameter, μ and σ are the mean and standard deviation of the Gaussian part of the distribution and λ is the rate of the exponential distribution. The $\operatorname{erfc}(x)$ factor is the complementary error function and is given by:

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \quad (11)$$

Unlike in the case of a Gaussian distribution, the mode is not equal to μ. For the exponentially modified Gaussian distribution the mode is given by:

$$r_m = \mu - \operatorname{sgn}(1/\lambda)\sqrt{2}\sigma \operatorname{erfcxinv}\left(\frac{|1/\lambda|}{\sigma}\sqrt{\frac{2}{\pi}}\right) + \lambda\sigma^2. \quad (12)$$

Here $\operatorname{erfcxinv}(x)$ is the inverse of the scaled complementary error function which is given by:

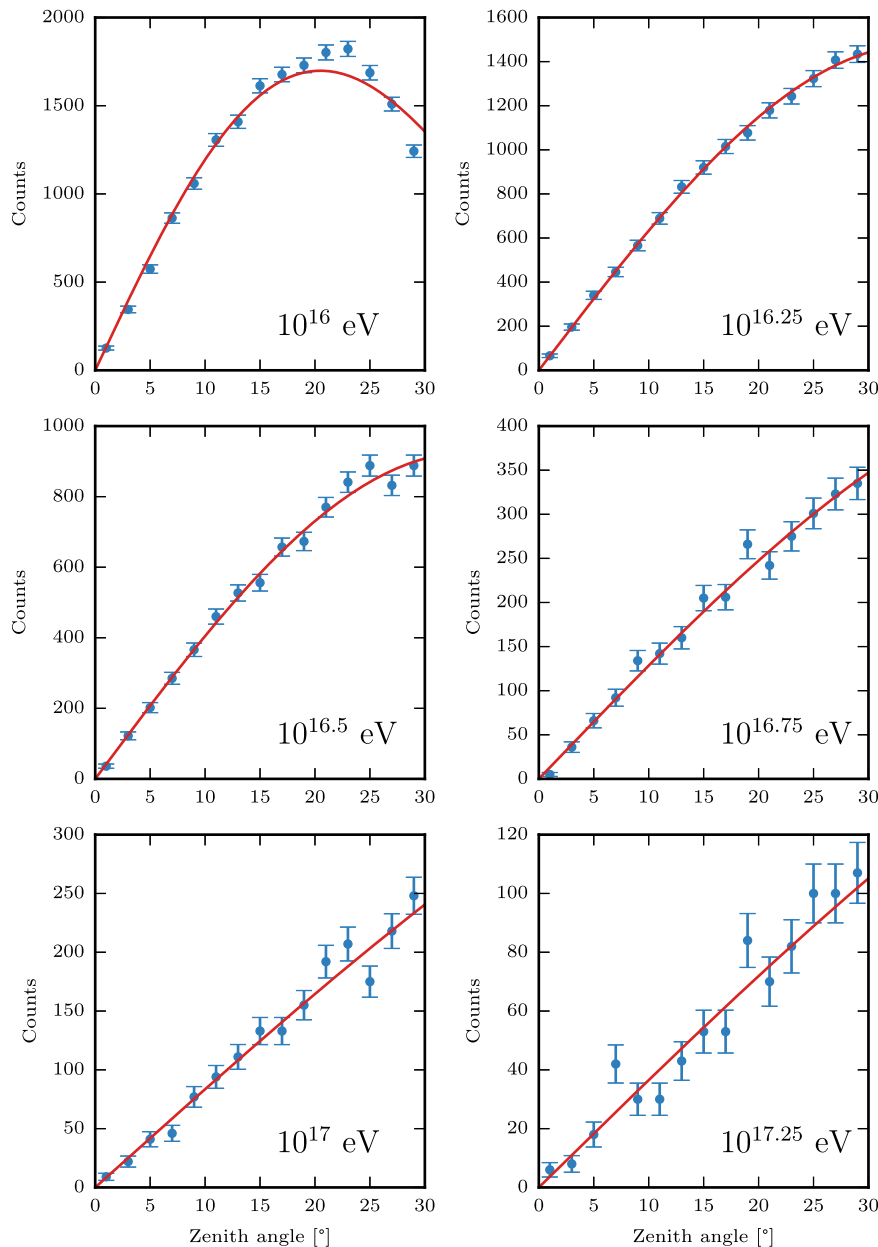$$\operatorname{erfcx}(x) = e^{x^2} \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \quad (13)$$

Fig. 7. Zenith angle distributions for six EAS energies. The uncertainty is given by the square root of the number of events in each bin. The red lines show the fits of Eq. (7). Because of the small solid angles subtended by bins (rings) at small zenith angles, the number of events in these bins is limited.

At $10^{17.25}$ eV (bottom right plot) the number of events is too low for an adequate fit.

With the detection efficiency as function of core distance, the effective surface area can be calculated by integrating the circumference at r:

$$A = \int_0^\infty 2\pi r \, \frac{g(r)}{g(0)} \, dr \tag{14}$$

Similar to Eq. (8), the detection efficiency is set to 1 at $r = 0$ m. Fig. 11 shows the effective surface areas as function of EAS energy. As the radius of the shower footprint is typically much larger than the distance between the four stations in the selection criterion (501, 502, 503 and 514), the effective surface area of the array is approximately the effective surface area of the shower. The shower radius increases linearly with the logarithm of the shower energy. Because of this, the effective surface area increases quadratically with the logarithm of the energy (πr^2 with $r \propto \log(E)$). The red line shows a quadratic function:

$$A(\log(E)) = a(\log(E) - \log(E_0))^2 \tag{15}$$

fitted to the data points excluding the point at 10^{16} eV. Here A is the effective surface area, a a fit parameter and E and E_0 the energy in eV and a fit parameter resp. At 10^{16} eV the quadratic parametrization is not valid since the shower footprint is not much larger than the distance between the four stations. The effective surface area below $10^{16.25}$ eV is described linearly (straight line versus dotted quadratic line).

2.8. Energy spectrum

The cosmic ray energy spectrum can now be obtained from the reconstructed energies (Fig. 6), effective solid angle, effective surface area and the detection time period ($t = 22$ months):

$$F = \frac{N_{\text{events}}}{A \cdot \Omega \cdot t \cdot \Delta E} \tag{16}$$

with ΔE the width of the energy bin. Fig. 12 shows the HiSPARC cosmic ray energy spectrum obtained with the Science Park cluster (values in Table 2). There is good agreement between HiSPARC measurements
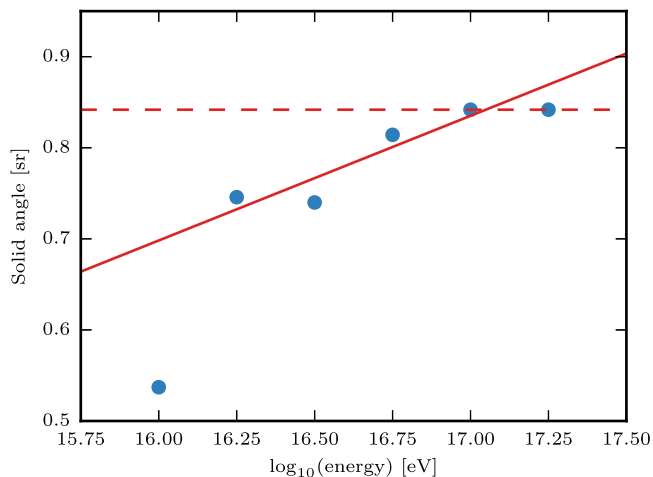
Fig. 8. Effective solid angle as a function of EAS energy. The effective solid angles are obtained by integrating the detection efficiency (Eq. (6), fits in Fig. 7) over the circular field of view subtended by a rotated zenith angle. The dashed line indicates the maximum solid angle at 30°. Above 10^{17} eV the efficiency is 100%. The red line shows a linear fit excluding the point at 10^{16} eV at which the efficiency fit was inadequate (Fig. 7).

Table 2
Cosmic ray flux values.

Energy [eV]	Flux [$m^{-2}\,s^{-1}\,sr^{-1}\,GeV^{-1}$]
$10^{16.0}$	$(4.19 \pm 0.05) \cdot 10^{-15}$
$10^{16.1}$	$(2.04 \pm 0.02) \cdot 10^{-15}$
$10^{16.2}$	$(1.04 \pm 0.01) \cdot 10^{-15}$
$10^{16.3}$	$(4.79 \pm 0.07) \cdot 10^{-16}$
$10^{16.4}$	$(2.04 \pm 0.03) \cdot 10^{-16}$
$10^{16.5}$	$(9.78 \pm 0.18) \cdot 10^{-17}$
$10^{16.6}$	$(4.76 \pm 0.10) \cdot 10^{-17}$
$10^{16.7}$	$(2.24 \pm 0.05) \cdot 10^{-17}$
$10^{16.8}$	$(1.09 \pm 0.03) \cdot 10^{-17}$
$10^{16.9}$	$(5.20 \pm 0.17) \cdot 10^{-18}$
$10^{17.0}$	$(2.62 \pm 0.10) \cdot 10^{-18}$
$10^{17.1}$	$(1.38 \pm 0.06) \cdot 10^{-18}$
$10^{17.2}$	$(6.60 \pm 0.33) \cdot 10^{-19}$
$10^{17.3}$	$(3.00 \pm 0.19) \cdot 10^{-19}$
$10^{17.4}$	$(1.64 \pm 0.12) \cdot 10^{-19}$
$10^{17.5}$	$(7.15 \pm 0.64) \cdot 10^{-20}$
$10^{17.6}$	$(3.89 \pm 0.40) \cdot 10^{-20}$
$10^{17.7}$	$(1.35 \pm 0.20) \cdot 10^{-20}$
$10^{17.8}$	$(7.57 \pm 1.26) \cdot 10^{-21}$
$10^{17.9}$	$(4.25 \pm 0.80) \cdot 10^{-21}$
$10^{18.0}$	$(1.87 \pm 0.45) \cdot 10^{-21}$
$10^{18.1}$	$(1.44 \pm 0.34) \cdot 10^{-21}$
$10^{18.2}$	$(4.69 \pm 1.66) \cdot 10^{-22}$
$10^{18.3}$	$(3.01 \pm 1.14) \cdot 10^{-22}$
$10^{18.4}$	$(2.53 \pm 0.89) \cdot 10^{-22}$
$10^{18.5}$	$(6.99 \pm 4.04) \cdot 10^{-23}$
$10^{18.6}$	$(3.45 \pm 2.44) \cdot 10^{-23}$
$10^{18.7}$	$(3.83 \pm 2.21) \cdot 10^{-23}$

and data from other experiments (open circles) [8,9, and references therein]. The black line shows a power law fitted to the data in log space with a slope of -3.08. This is consistent with the reference value of -3.1 [1]. A change of the slope near the ankle ($\sim 10^{18.5}$ eV) cannot be identified because of the limited number of events in that energy region.

3. Flux estimation with a two-detector station

In the previous section the cosmic ray energy spectrum was determined by combining data from multiple HiSPARC stations. However, as the majority of high schools employs a two-detector station which is often too far away from neighbouring stations to sample the same

EAS, this method is not applicable for many HiSPARC stations. In this section it will be shown that also a single two-detector station can be used to probe the cosmic ray flux, albeit with limited precision.

3.1. Pulse integral distribution

The value of the PMT pulse integral is directly proportional to the number of scintillation photons that reach the PMT, and thus to the number of particles traversing the detector (MIPs). The energy loss of a MIP follows Landau's theory [10]. The Landau distribution has a peak at the most probable energy loss with a pronounced tail towards higher energies. If multiple particles simultaneously traverse a detector, the energy loss is described by a sum of Landau distributions. The shape of the pulse integral distribution (PID), i.e. a histogram of the pulse integral values, is thus determined by the number of MIPs that traverse the detector. This number is determined by the EAS particle densities in the footprint at the ground, which relates to the energy of the primary cosmic ray. The small particle densities at low energy EASs ($\lesssim 10^{13}$ eV) generate a steeply falling PID, i.e. there is a relatively high probability to detect just 1 or 2 simultaneous particles whereas higher multiplicities at this energy become rare. For higher energy EASs it is also quite likely to detect, for example, 10 simultaneous particles. This means that the PID provides an indirect measure for the energy of the primary cosmic ray. Since it is impossible to select single energy EASs, a PID obtained by a station can be thought of as a sum of single energy PIDs.

3.2. EAS simulations

A large sample of EASs for primary cosmic rays with an energy of 10^{12}, 10^{13}, 10^{14}, 10^{15} and 10^{16} eV is generated with CORSIKA [11]. Only proton initiated showers are considered. The showers have been generated with zenith angles ranging from 0° to 60° in steps of 3.75°. For high energy hadronic interactions QGSJET-II [12] was selected. Hadronic interactions below 80 GeV are simulated using GHEISHA [13] and electromagnetic interactions are incorporated with EGS4 [14]. The full particle shower was simulated (no 'thinning' [15]). The location of the EAS shower cores were randomly chosen within a circle with a radius of 150 m. The two-detector station is at the centre of the circle. The arrival direction (azimuth and zenith) was chosen isotropically. The response of the scintillator and light-guide to particles traversing the detector was simulated with GEANT4 [16]. A parameterized PMT response was used [3]. Finally, the HiSPARC trigger conditions were applied.

Fig. 13 shows the simulated PIDs for the five different energies normalized to the number of events. At higher energies the relative abundance of large pulses increases considerably. By fitting a linear combination of the simulated distributions to the experimentally observed spectrum, the cosmic ray flux at a fixed energy interval can be estimated.

3.3. Effective surface area and solid angle

EASs at large zenith angles result in lower particle densities at ground level because of increased absorption due to the larger path length through the atmosphere. Moreover, the particle density observed in the detection station strongly depends on the distance to the shower core. These two effects need to be quantified in order to calculate the cosmic ray flux. For this, an effective surface area and effective solid angle are introduced. The fraction of events that results in a trigger is defined as the 'EAS detection efficiency' (ε). The left plot in Fig. 14 shows ε for 10^{15} eV air showers as function of the zenith angle and distance to the shower core. The maximum efficiency occurs at small r and small zenith angle θ.

Integrating over the solid angle yields the effective solid angle as a function of core distance (upper right plot):

$$\Omega(r) = \int_0^{2\pi} \int_0^{\pi/2} \varepsilon(\theta, r) \sin\theta \, d\theta \, d\phi \qquad (17)$$
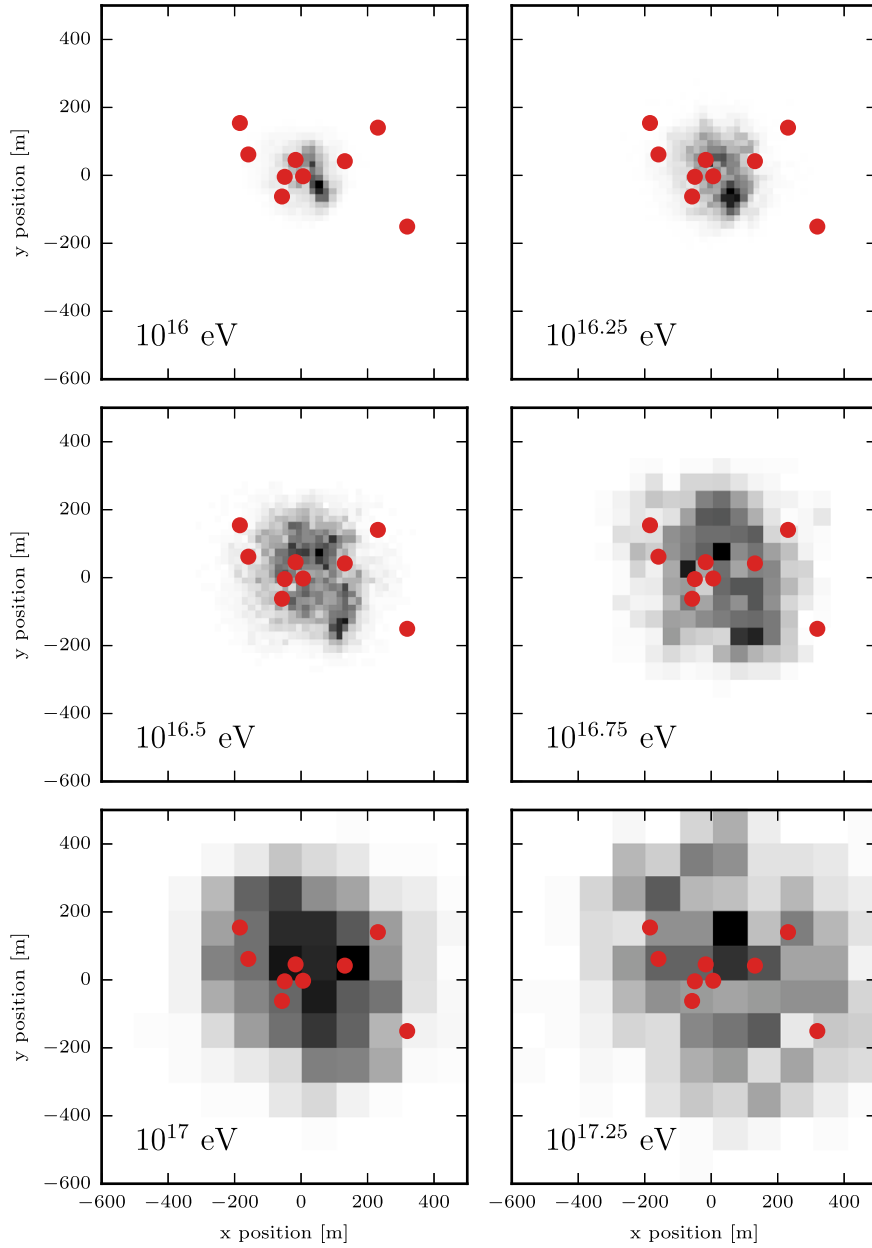
Fig. 9. Two dimensional distribution of core positions for six EAS energies. The red dots indicate the station positions. The maximum distance of the detected EAS shower cores to the central stations increases with energy.

The effective solid angle can accurately be parameterized using the same exponentially modified Gaussian distribution earlier used in Eq. (10) (reprinted here):

$$\Omega(r; \alpha, \mu, \sigma, \lambda) = \alpha \exp\left[\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2r)\right]$$
$$\times \operatorname{erfc}\left(\frac{\mu + \lambda\sigma^2 - r}{\sqrt{2}\sigma}\right) \tag{18}$$

with α a scaling parameter, μ and σ the mean and standard deviation of the Gaussian part of the distribution and λ the rate of the exponential part. The complementary error function, $\operatorname{erfc}(x)$, is given by Eq. (11). Integrating the detection efficiency over the surface area (polar coordinates r and ζ) yields the effective surface area as a function of zenith angle (bottom right plot in Fig. 14):

$$A(\theta) = \int_0^{2\pi}\int_0^\infty \varepsilon(\theta, r) r \, dr \, d\zeta \tag{19}$$

The effective surface area can be parameterized using the same formula used in Eq. (6) (reprinted here):

$$A(\theta) = a \cdot \exp[-b \cdot (\sec\theta - 1)] \tag{20}$$

with a and b fit parameters. Since, in order to obtain the flux, the number of events needs to be divided by both the effective solid angle and the effective surface area; there is no need to evaluate them separately. Instead, the two are combined:

$$A\Omega = \int_0^{2\pi}\int_0^\infty \Omega(r) r \, dr \, d\zeta \tag{21}$$

$$= \int_0^{2\pi}\int_0^{\pi/2} A(\theta)\sin\theta \, d\theta \, d\phi \tag{22}$$

Table 3 shows the fit parameters that describe the effective solid angle and surface area for the simulated energies (for 10^{15} eV proton showers $A\Omega$ is $8.58 \cdot 10^3$ m² sr). The $A\Omega$ values can be calculated from Eqs. (21) or (22), or by direct summation of the two-dimensional
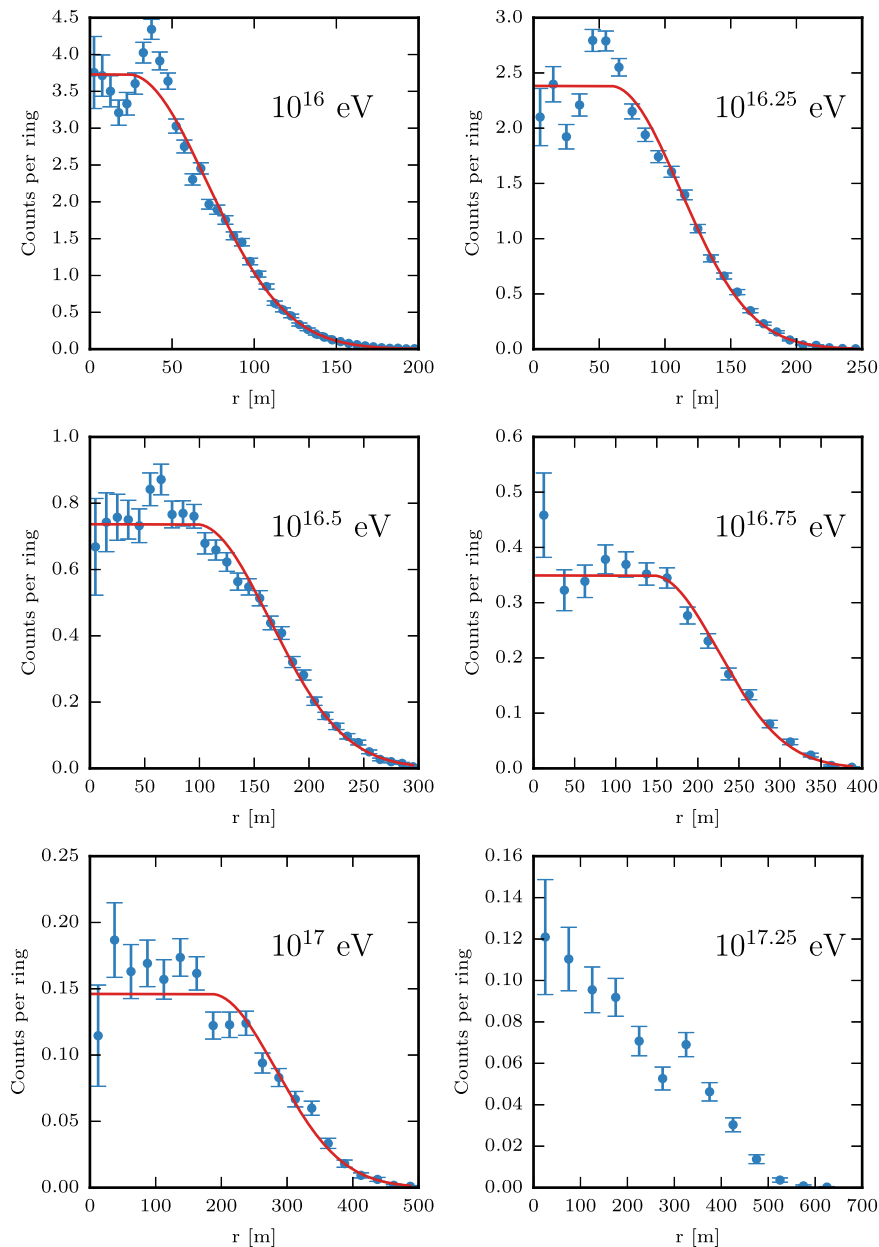
Fig. 10. Distribution of the radial distances of the reconstructed core positions to the centre of mass of the core positions (Fig. 9) divided by the surface area of each radial bin (rings of $2\pi r$) for six energies. The uncertainty is given by the square root of the number of events in each bin. The red lines show the fits of Eq. (9) to the distributions. At $10^{17.25}$ eV the number of events is too low for an adequate fit of Eq. (9).

histogram in Fig. 14. The last column in Table 3 lists the averages and standard deviations of these three methods.

Fig. 15 shows the five AΩ values. The interpolation is defined by the following equation:

$$\log_{10}(A\Omega) = ax^2 + bx + c \qquad (23)$$

with $x = \log_{10}(E)$, $a = -0.239$, $b = 8.31$, $c = -66.8$. Extrapolation leads to an estimate at 10^{11} eV of AΩ = $4.2 \cdot 10^{-5}$ (red circle). A precise direct estimate at this energy from simulation is difficult because it is computationally expensive to collect a sufficiently large data set; the number of EASs that satisfy the selection criteria is extremely small. The number of triggers at an EAS energy of 10^{11} eV is approximately 400 times smaller than at 10^{12} eV. This is consistent with the estimate obtained by extrapolation of Eq. (23).

3.4. Fitting single energy PIDs to experimental data

About 8×10^5 events collected by a two-detector station (station #4) between the 5th of March and the 2nd of April 2018 were used to obtain a PID. The contribution to the PID from random coincidences was subtracted. A Bayesian method was used to fit a combination of simulated single energy PIDs in the range 10^{12}–10^{16} eV to the experimental data. Only pulse integrals smaller than 35 Vns are considered to limit systematic differences between the detectors [4]. Signals are affected by gamma rays, and Cherenkov light generated in the light-guide [3]. Especially at lower multiplicities this contribution becomes apparent. Pulse integrals below 6 Vns (2 MIPs) are therefore discarded as well.

As shown in Fig. 15, the probability that a 10^{11} eV shower results in a footprint that triggers the station is negligible. The slope of PIDs with

Table 3

The parameters in Eq. (18) (effective solid angle $\Omega(r)$) and Eq. (20) (effective surface area $A(\theta)$) are listed as a function of energy. The last column gives the value of the combination $A\Omega$ (Eq. (21) or (22)).

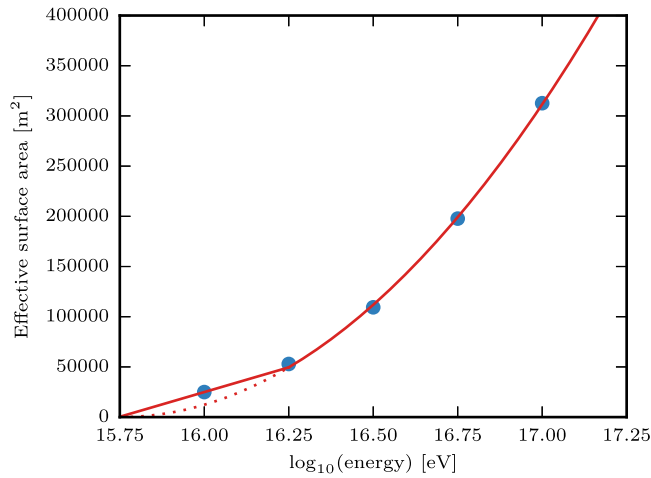| Energy [eV]	$\Omega(r)$				$A(\theta)$		$A\Omega$ [m^2 sr]
	α	μ	σ	λ	a	b	
10^{12}	$1.73 \cdot 10^{-5}$	-2.3	1.0	$8.00 \cdot 10^{-2}$	$4.86 \cdot 10^{-2}$	9.66	$(2.75 \pm 0.09) \cdot 10^{-2}$
10^{13}	$8.71 \cdot 10^{-3}$	-3.1	3.6	$1.22 \cdot 10^{-1}$	7.85	6.89	5.50 ± 0.14
10^{14}	$5.44 \cdot 10^{-1}$	-10.4	8.1	$8.78 \cdot 10^{-2}$	$6.30 \cdot 10^{2}$	7.90	$(4.23 \pm 0.21) \cdot 10^{2}$
10^{15}	4.40	-14.2	24.5	$7.25 \cdot 10^{-2}$	$9.21 \cdot 10^{3}$	4.76	$(8.58 \pm 0.28) \cdot 10^{3}$
10^{16}	6.92	-50.7	63.3	$2.71 \cdot 10^{-2}$	$6.33 \cdot 10^{4}$	3.62	$(7.33 \pm 0.13) \cdot 10^{4}$

Fig. 11. Effective surface areas as function of EAS energy. The effective surface areas are obtained by integrating the detection efficiency (Eq. (14), fits in Fig. 10) over the surface area of a disk with radius of 1000 m. The red line shows a quadratic function (Eq. (15)) which is used to parametrize the effective surface area as function of energy. Below $10^{16.25}$ eV this parametrization is not valid and the effective surface area is described linearly (straight line versus dotted quadratic line).
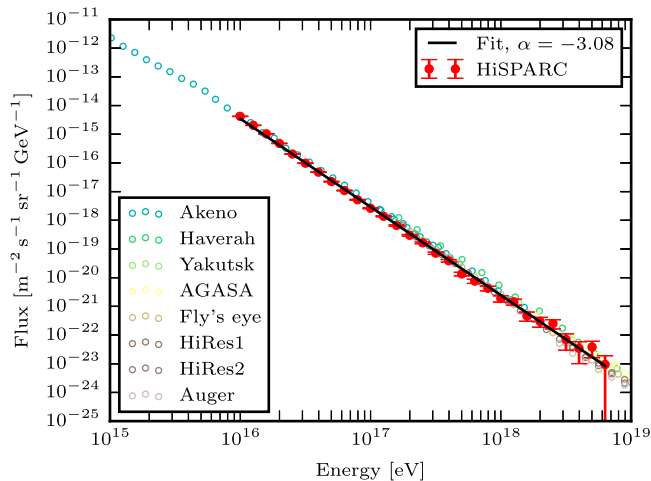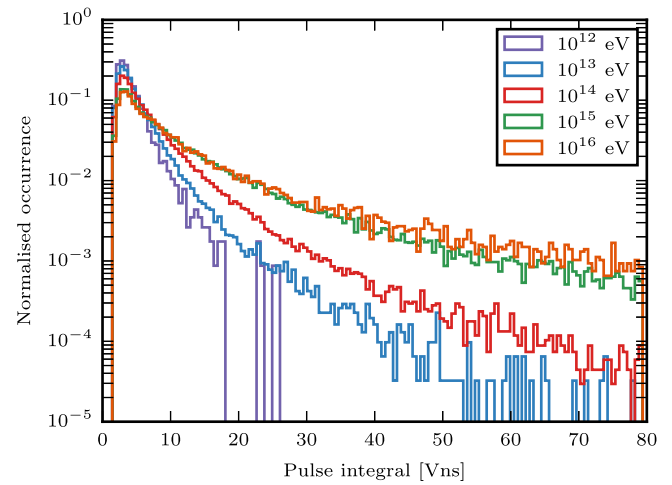

Fig. 13. Simulated PIDs at five different energies. The PID depends on the energy of the EASs. A full detector simulation is carried out with isotropically selected EASs and uniformly chosen core distances. Only proton induced EASs are considered. The difference becomes evident at large pulse integrals.

$$+ \; x_3 \cdot \mathrm{PID}_{14}(n_i) + x_4 \cdot \mathrm{PID}_{15}(n_i)$$
$$+ \; x_5 \cdot \mathrm{PID}_{16}(n_i) \tag{24}$$

Here $\vec{x} \equiv (x_1, x_2, x_3, x_4, x_5)$, n_i is the bin value, $z(n_i)$ is the expected number of events in each bin and each $\mathrm{PID}(n_i)$ is a single energy model (Fig. 13). The uncertainty in the number of entries in the bins of the experimentally observed PID is described by a Poisson distribution. The probability to obtain a number of counts y_i in bin n_i given $\vec{x}$ is given by:

$$p(y_i, n_i | \vec{x}) = P(y_i, z(n_i, \vec{x})) \tag{25}$$

with the Poisson distribution:

$$P(y, z) = \frac{z^y}{y!} e^{-z} \tag{26}$$

Bayes' rule can be exploited to define a probability to obtain the fit parameters $\vec{x}$.

$$p(\vec{x} | N, Y) = \frac{p(N, Y | \vec{x}) p(\vec{x})}{p(N, Y)} \tag{27}$$

Here N and Y are the combined collection of elements n_i and y_i. The denominator $p(N, Y)$ is a normalization constant. The $p(\vec{x})$ is known as the prior probability distribution. The prior states that the parameters x_1, x_2, x_3, x_4 and x_5 cannot become negative. The function $p(\vec{x} | N, Y)$ is the posterior probability distribution. The $p(N, Y | \vec{x})$ function is the likelihood (the product of conditional probabilities):

Fig. 12. The HiSPARC cosmic ray energy spectrum obtained with the Science Park cluster under the assumption of a pure proton composition. The black line shows a power law fitted to the data in log space with a slope of -3.08. There is good agreement between HiSPARC measurements and data from other experiments (open circles, see Fig. 18 to distinguish the different experiments).

energies 10^{17} eV and beyond becomes rather similar to that of 10^{16} eV in the pulse integral range between 6 to 35 Vns whereas the flux at those energies rapidly decreases. The experimentally observed PID will therefore be restricted to the sum of five single energy contributions ($10^{12}-10^{16}$ eV):

$$z(n_i, \vec{x}) = x_1 \cdot \mathrm{PID}_{12}(n_i) + x_2 \cdot \mathrm{PID}_{13}(n_i)$$

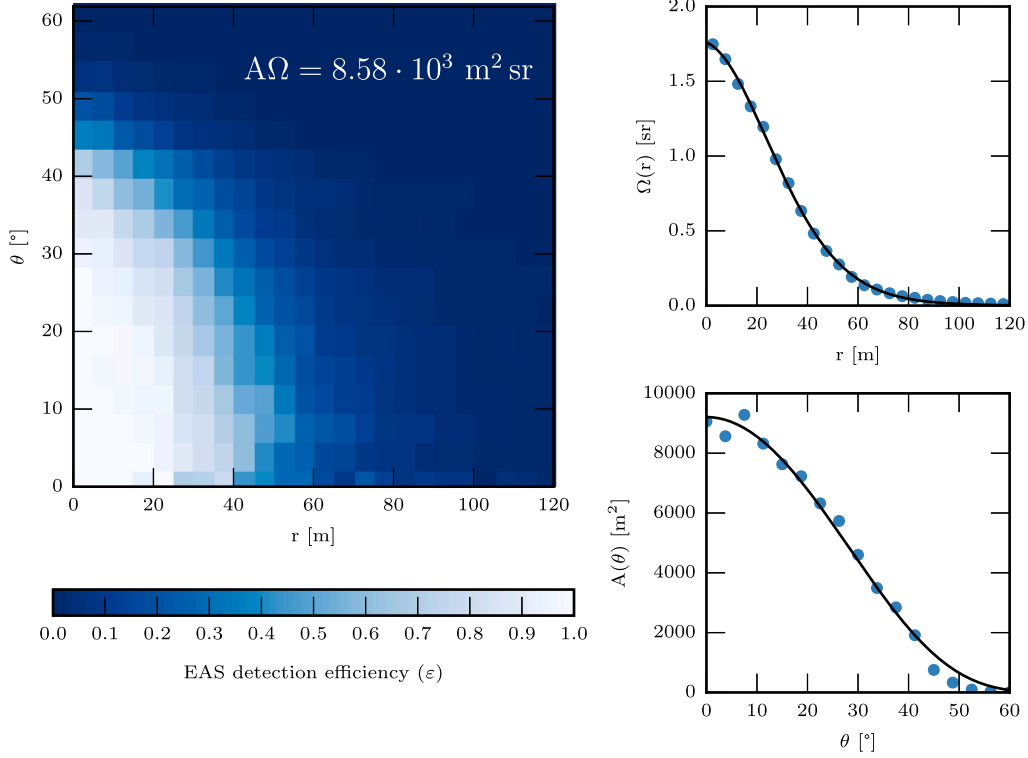$$p(N, Y | \vec{x}) = \prod_{i=1}^{M} p(y_i, n_i | \vec{x}) = \mathcal{L} \tag{28}$$

Fig. 14. Left: The detection efficiency (ε) for 10^{15} eV proton induced showers as function of core distance (r) and zenith angle (θ). Top right: the detection efficiency integrated over the zenith angle yields the effective solid angle as function of core distance ($\Omega(r)$, blue points) which can accurately be described using the parametrization in Eq. (18) (black line). Bottom right: the detection efficiency integrated over the core distance yields the effective surface area as function of zenith angle (A(θ), blue points) can be described by the parametrization in Eq. (20) (black line).
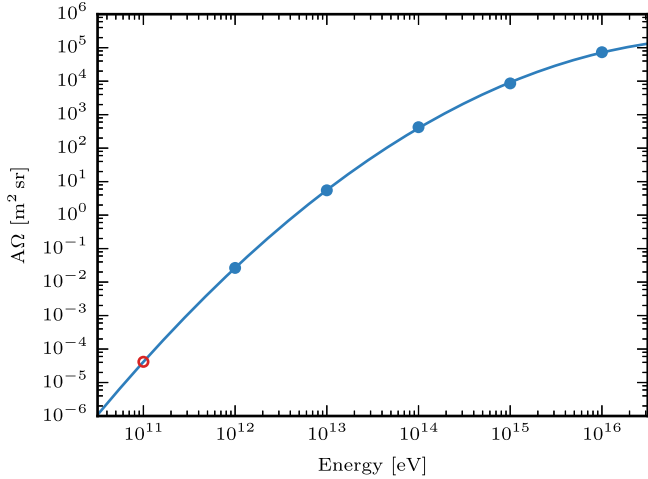

Fig. 15. The blue curve (Eq. (23)) shows the interpolation of the AΩ values (blue dots) listed in Table 3. AΩ at 10^{11} eV (red circle, A$\Omega = 4.2 \cdot 10^{-5}$ m^2 sr) is obtained by extrapolation.

Since the slope differences between the PIDs occur at higher multiplicities, a gradually larger weight is assigned with increasing multiplicity. This leads to a weighted likelihood:

$$\widehat{\mathcal{L}} = \prod_{i=1}^{M} p(y_i, n_i | \vec{x})^{w(n_i)} \tag{29}$$

with $w(n_i)$ the weight function:

$$w(n_i) = 1.161^{n_i} \tag{30}$$

which corrects for the number of entries in the lower ($1.7 \cdot 10^4$ at 6 Vns) and higher bins (231 at 35 Vns). By taking the logarithm of this weighted likelihood, the product converts into:

$$\log(\widehat{\mathcal{L}}) = \sum_{i=1}^{M} w(n_i) \cdot \log \left(\frac{z_i^{y_i}}{y_i!} e^{-z_i} \right) \tag{31}$$

$$= \sum_{i=1}^{M} w(n_i) \cdot (y_i \log(z_i) - z_i - \log(y_i!)) \tag{32}$$

$$= \sum_{i=1}^{M} w(n_i) \cdot (y_i \log(z_i) - z_i - C) \tag{33}$$

with $z_i = z(n_i, \vec{x})$ and C is a constant.

Instead of directly maximizing the posterior probability distribution, a range of parameters ($\vec{x}$) are explored. This is done using a Markov Chain Monte Carlo (MCMC) algorithm [17]. Fig. 16 shows (lower dimensional) subsets of the sampled (five dimensional) posterior probability distribution. The upper (on the diagonal) Gaussian shaped histograms show the sample selection in the dimension of the fit parameters. The median of the histogram (red lines) is taken as the best fit. The uncertainty is shown by the standard deviation (green lines). The best fit values are $(3.01 \pm 0.35) \cdot 10^4$, $(2.32 \pm 0.05) \cdot 10^5$, $(1.09 \pm 0.03) \cdot 10^5$, $(5.63 \pm 0.20) \cdot 10^4$, $(2.63 \pm 0.16) \cdot 10^4$ for 10^{12} to 10^{16} eV resp. The other subplots display the relation between two fit parameters. There is some interdependence between the fit parameters of neighbouring energies. This is especially pronounced at fit parameters x_4 and x_5 (fourth sub-plot in bottom row) due to the relatively small slope differences between the 10^{15} and 10^{16} eV PIDs (Fig. 13). The resulting fit to the experimental data is shown in Fig. 17. Also here it can be seen that the 10^{15} and 10^{16} eV distributions (green and orange lines resp.) are very similar. The cosmic ray flux at 10^{15} eV is approximately 1000 times higher than at 10^{16} eV. Taking the AΩ factor into account gives a ~100 times larger number of events at 10^{15} eV. The figure thus suggests that a ~1% sample of 10^{16} eV EASs can be distinguished using the

Table 4

Best fit for the number of events with their uncertainties.

Energy [eV]	Number of events
10^{12}	$(4.09 \pm 1.76) \cdot 10^4$
10^{13}	$(2.04 \pm 0.23) \cdot 10^5$
10^{14}	$(1.25 \pm 0.20) \cdot 10^5$
10^{15}	$(5.65 \pm 1.32) \cdot 10^4$
10^{16}	$(2.43 \pm 1.04) \cdot 10^4$

Table 5

Cosmic ray flux values obtained using a two-detector HiSPARC station.

Energy [eV]	Flux [$\text{m}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}\,\text{GeV}^{-1}$]
10^{12}	$(2.08 \pm 0.93) \cdot 10^{-4}$
10^{13}	$(5.19 \pm 0.60) \cdot 10^{-7}$
10^{14}	$(4.14 \pm 0.70) \cdot 10^{-10}$
10^{15}	$(9.22 \pm 2.20) \cdot 10^{-13}$
10^{16}	$(4.64 \pm 2.01) \cdot 10^{-15}$

slight difference between the two PIDs. Even though the obtained best fit results, and the x_4 versus x_5 subplot in Fig. 16 not being entirely linear (i.e. there is some possibility to differentiate between both PIDs), seem to indicate this, it could be argued that it is better to interpret the data point at 10^{15} eV as the number of particles with an energy of 10^{15} eV and higher, and disregard the 10^{16} eV point. However, for the remainder of this article the data point is kept as a separate measurement.

In addition to the statistical uncertainty in the experimental data, the single energy models also have an intrinsic statistical uncertainty. Creating larger data sets for the single energy PIDs is currently limited by the generation of simulated EASs which is computationally expensive. The uncertainty has been estimated by resampling the single energy PIDs. The number of events in each bin was randomly resampled following Poisson statistics. A combination of these new PIDs was fitted to the experimental data as well. This procedure was carried out multiple times. The mean and standard deviation of the best fit parameters are listed in Table 4.

3.5. Cosmic ray flux

The estimated number of events per energy decade derived from the fit can be used in combination with the effective surface area and solid angle to obtain the cosmic ray flux as a function of energy. The fluxes are calculated using:

$$F = \frac{N_{\text{events}}}{A\Omega \cdot t \cdot \Delta E} \tag{34}$$

with t the duration of the experiment ($t = 29$ days $= 2.506 \cdot 10^6$ s) and ΔE the width of the energy bin (e.g. $10^{15.5} - 10^{14.5} \approx 2.85 \cdot 10^{15}$ eV). The flux values are listed in Table 5. Fig. 18 shows the cosmic ray energy spectrum (circles) from several experiments [8,9, and references therein] together with the flux values derived from HiSPARC data (red dots). The uncertainties are smaller than the dot size. The grey line represents the function in Eq. (1). The HiSPARC values agree well with the other measurements. Note that the reference flux values below 10^{14} eV are measured using spacecraft (e.g. Proton satellite [18]).

4. Discussion and conclusion

HiSPARC is a widespread, irregular EAS detector array with the majority of the stations located at high schools. For some clusters of stations, the stations are located close enough to each other such that they can sample the same EAS. An example of such a cluster is the Amsterdam Science Park array consisting of 12 four-detector stations. The majority of high schools employs a two-detector station which is
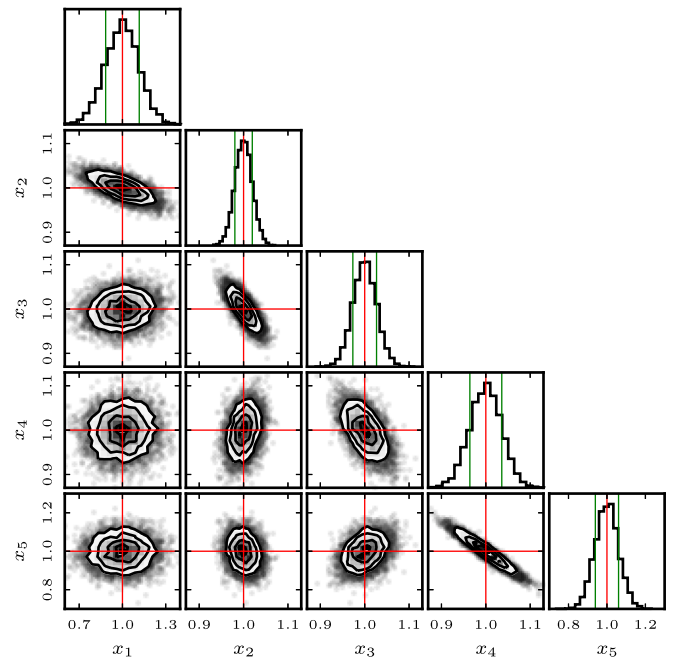

Fig. 16. Subsets of the sampled posterior probability distributions. The upper Gaussian shaped histograms show the sample selection in the dimension of the fit parameters ($\bar{x}$). The median of the histogram (red lines) is taken as the best fit. The uncertainty in the fit is shown by the standard deviation (green lines). The other subplots display the relation between two fit parameters. There is some interdependence between the fit parameters of neighbouring energies. This is more pronounced for x_4 and x_5 (fourth sub-plot in bottom row). See also Fig. 13.
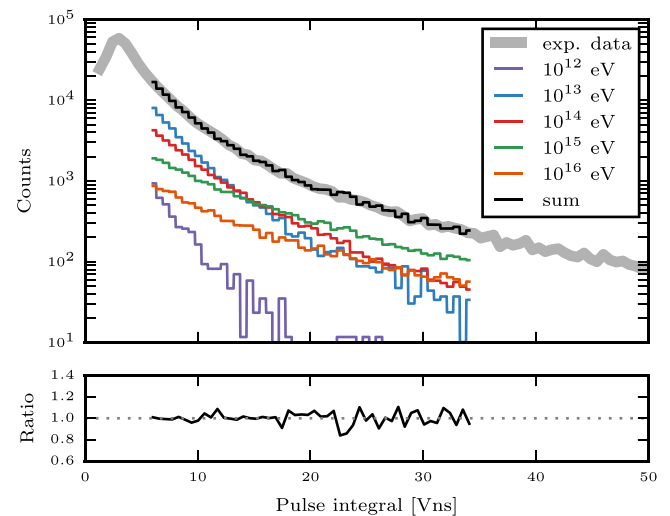

Fig. 17. A linear combination of five simulated PIDs (coloured lines, see Fig. 13) fitted to the experimentally obtained PID (thick grey line). The black line shows the best fit. The bottom plot gives the ratio of the experimental data and the fit.

often too far away from neighbouring stations. In this article two methods are presented to reconstruct the cosmic ray energy spectrum with HiSPARC: one for clusters of stations and one for a single two-detector station.

Data from 9 stations in the Science Park cluster have been combined to determine the cosmic ray energy spectrum. The direction of the EASs was reconstructed from the detector trigger times by planar approximation of the shower front. The uncertainty in the arrival direction roughly agrees to that of a single station. The position of the shower core and energy of the EASs were reconstructed by fitting the LDF
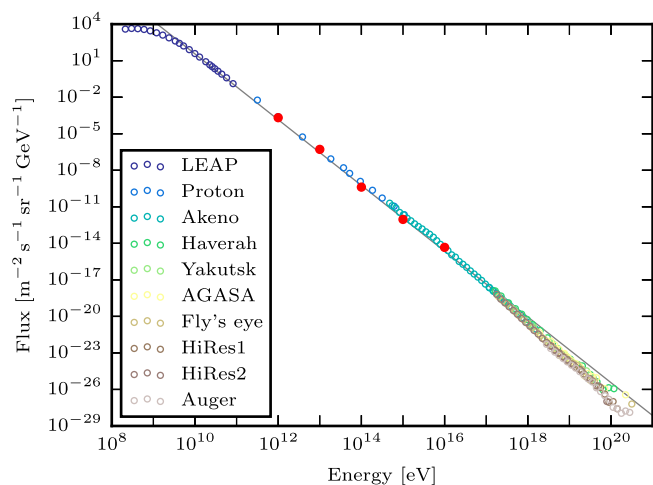
Fig. 18. Cosmic ray fluxes at $10^{12}-10^{16}$ eV from HiSPARC data (red dots) compared to results from other experiments [8,9, and references therein] (circles). The uncertainties are smaller than the dot size. The fluxes obtained by HiSPARC agree well with the other measurements. The grey line represents Eq. (1).

Fig. 19. Cosmic ray energy spectrum obtained with a two-detector station (blue dots) and the Science Park cluster (red dots). There is good agreement between HiSPARC measurements and data from other experiments (open circles). Two power laws have been fitted to the separate data sets. The blue line shows the best fit to the two-detector measurements. Its slope of -2.71 ± 0.07 agrees well with the reference value of -2.7 below the knee. The red line shows the best fit to the flux values obtained by the Science Park cluster. The slope of -3.08 ± 0.02 agrees well with the reference value of -3.1 at energies beyond the knee.

to the sampled particle densities projected on the arrival plane. The effective solid angle and effective surface area were determined from the experimental data. EASs up to an energy of $\sim 10^{19}$ eV were detected within a 22 month data taking period. Below $\sim 10^{16}$ eV the particle densities in the shower footprints are typically too low to be detected by multiple stations. The cosmic ray flux values obtained by the Science Park cluster agree well with data from other experiments. The red dots in Fig. 19 show the flux values. A power law has been fitted to the data in log-space. The slope of -3.08 ± 0.02 agrees well with the literature value of -3.1 [1].

As a two-detector station only provides two samples of an air shower footprint, it cannot be used to reconstruct single EASs. However, with a two-detector HiSPARC station it is possible to probe the cosmic ray energy spectrum directly. A new method has been presented which relies on deriving the energy dependent particle multiplicity distributions in a single scintillation detector. If a detector detects a large number of shower particles simultaneously traversing the scintillator (i.e. a large pulse integral), they could be part of a relatively low energy EAS with its shower core close to the detector, or originate from a high energy EAS with its core further removed from the detector. The probability for both possibilities differs. By accurately modelling the pulse integral distribution as a function of energy, these probability characteristics are exploited to determine the cosmic ray flux. One month of data is sufficient to obtain flux values in the range $10^{12}-10^{16}$ eV (five energy decades) that are in good agreement with the results from dedicated (space-based) experiments. The blue dots in Fig. 19 show the flux values obtained with a two-detector station. The open circles denote the flux values obtained by other experiments. The blue line shows a power law fitted to the data point in log-space. The slope of -2.71 ± 0.07 agrees with the literature value of -2.7 [1].

For the Science Park Array method, the agreement between the HiSPARC spectrum and the data from other experiments shows that the discussed energy and flux reconstruction methods are correct. However, more research is needed to investigate the uncertainty in the energy estimation. A more advanced LDF that is specifically tailored to the Science Park array could also further improve the energy reconstruction. Additionally, new reconstruction algorithms based on artificial intelligence or Bayesian methods could be developed.

Concerning the two-detector station method, there are many assumptions that greatly limit the scientific value of the obtained flux measurements. By integrating over all zenith angles in data and simulations and by assuming that cosmic rays consist of protons only,
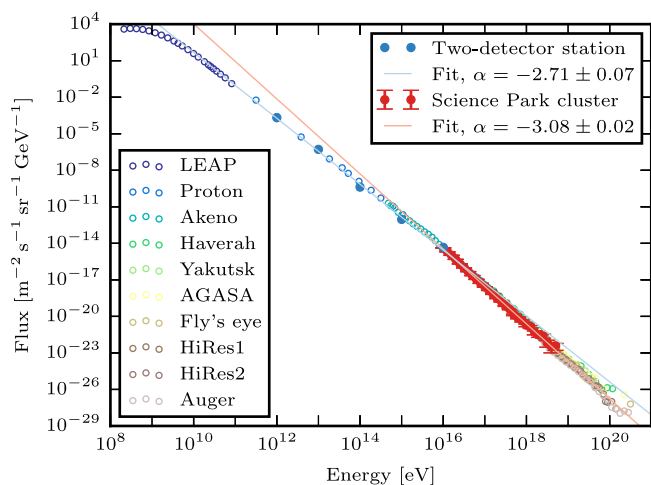
the large dependence of the acceptance on the primary mass and on the interaction model employed in the EAS simulations are ignored. Furthermore, as the method relies on approximate cancellation of the decreasing flux with energy on one hand and the increasing acceptance with energy on the other hand, uncertainty in these acceptances play a large role. Nevertheless, the gross features of the cosmic ray energy spectrum could be retrieved using a simple two-detector station which adds to the educational value of the equipment. Unfortunately, because of the mathematical complexity and need for Monte Carlo simulations similar studies cannot be replicated by high school students. The results and methods, however, can be shown and discussed.

In conclusion, it can be shown that reconstruction of the cosmic ray energy spectrum is certainly possible, albeit with limited precision compared to dedicated scientific arrays. This augments HiSPARC's primary goal, which is to engage high school students with modern physics and let them contribute to real scientific research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] T.K. Gaisser, R. Engel, E. Resconi, Cosmic Rays and Particle Physics, Cambridge University Press, 2016.

[2] J. Beringer, J. Arguin, R. Barnett, K. Copic, O. Dahl, D. Groom, C. Lin, J. Lys, H. Murayama, C. Wohl, et al., Review of particle physics, Phys. Rev. D 86 (1) (2012) 010001.

[3] K. van Dam, B. van Eijk, D. Fokkema, et al., The HiSPARC experiment, Nucl. Instrum. Methods Phys. Res. A 959 (2020) 163577.

[4] K. van Dam, HiSPARC: A Scintillating Experiment (Ph.D. thesis), University of Twente, ISBN: 978-90-365-5227-1, 2021.

[5] K. van Dam, Increased radioactivity during precipitation measured by the HiSPARC experiment, Phys. Scr. 95 (2020) 074011.

[6] J.M.C. Montanus, The Observability of Jets in Cosmic Air Showers, (Ph.D. thesis), University of Amsterdam, ISBN: 978-94-028-0549-9, 2017.

[7] G. De Mauro, Tasting the Ultra-High-Energy Universe (Ph.D. thesis), Radboud University Nijmegen, ISBN: 978-94-028-2094-2, 2020.

[8] Cosmic Ray Spectrum website by Dr. William Hanlon, https://www.physics.utah.edu/~whanlon/spectrum.html.

[9] J.W. Cronin, T.K. Gaisser, S.P. Swordy, Cosmic rays at the energy frontier, Sci. Am. 276 (1997) 44–49.

[10] L. Landau, On the energy loss of fast particles by ionization, J. Phys.(USSR) 8 (1944) 201–205.

[11] D. Heck, J. Knapp, J.N. Capdevielle, et al., CORSIKA: A Monte Carlo Code To Simulate Extensive Air Showers, Forschungszentrum Karlsruhe GmbH, Karlsruhe (Germany), 1998.

[12] S. Ostapchenko, QGSJET-II: Towards reliable description of very high energy hadronic interactions, Nuclear Phys. B - Proc. Suppl. 151 (2006) 143–146.

[13] H. Fesefeldt, GHEISHA The Simulation of Hadronic Showers, Tech. Rep., RWTH Aachen, 1985, PITHA-85/02.

[14] W.R. Nelson, D.W.O. Rogers, H. Hirayama, The EGS4 Code System, Tech. Rep., Stanford Linear Accelerator Center, Stanford, California, 1985.

[15] A.M. Hillas, Proceedings International Cosmic Ray Conference, Vol. 1, 1981, p. 193.

[16] J. Allison, K. Amako, J. Apostolakis, et al., Recent developments in Geant4, Nucl. Instrum. Methods Phys. Res. A 835 (2016) 186–225.

[17] D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, Emcee: the MCMC hammer, Publ. Astron. Soc. Pac. 125 (2013) 306.

[18] N. Grigorov, Y.V. Gubin, I. Rapoport, I. Savenko, B. Yakovlev, V. Akimov, V. Nesterov, Energy spectrum of primary cosmic rays in the 10 11-10 15 ev energy range according to the data of proton-4 measurements, in: 12th International Conference on Cosmic Rays, 16-25th August 1971, Hobart, Tasmania, Australia, 1971.