EAS data classification into light and heavy mass groups by MAKET installation (Investigation of the influence of detector response)

Ararat A. Vardanyan, Gagik G. Hovsepyan

1 The MAKET Experiment

The MAKET installation 1 consists of 92 plastic scintillation detectors with thickness of 5 cm. Twenty four of them with the area 0.09 m^2 are arranged in the central part and the others with the area 1 m^2 in the center and on the periphery. The central part consists of 73 scintillation detectors and is arranged in a rectangle of 85×65 m^2. Two peripheral points of a distance of $95m$ and 65 m from the center of the installation consists of 15 and 4 scintillation detectors respectively.

In order to estimate the zenith and azimuthal angles 19 detectors (1 m^2) are equipped with timing readout measuring the EAS front appearance with an accuracy of ca. 5 ns.

The photomultipliers (PM-49) are placed in light-tight iron boxes. Logarithmic amplitude-digital converters (ADC) and constant fraction discriminators (CFD) are combined with the PM. The dynamic range of the registered particle number is $\sim 5 \times 10^3$, corresponding to the code of the logarithmic ADCs from 1 to 100.

Three types of triggers are used:

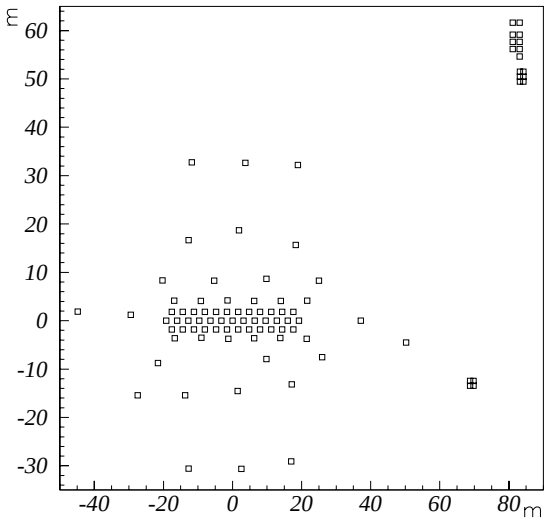

Figure 1: *Schematic view of the MAKET detector array.*

1. A hardware trigger by 11 preselected detectors with the conditions that at least 7 of them are firing with more than 3 particles.

2. A timing trigger by 19 timing detectors with the requirement that at least 9 (arranged symmetrically relatively to the center) are fired. The time delays are measured with respect to the central detector.

3. A software trigger derived from two groups of 8 detectors symmetrically arranged, with signals of at least one particle.

If the first two conditions are fulfilled in time window of 20 μs the event is stored. If the third condition is fulfilled the event is transferred to the data base. The whole system of the trigger elaboration and data readout is done in the *CAMAC* standard.

Monte-Carlo calculations have shown that this trigger corresponds to the threshold of registering EAS with $N_e \geq 5 \times 10^4$ when the shower core lies within the rectangle of $40 \times 12m^2$ from the geometrical center of the installation.

The uncertainties of the reconstruction of EAS parameters is: for the shower size $\sigma(N_e) \sim 10\%$, the shape (age) parameter - $\sigma(s) \sim 0.06$. The accuracies of EAS angles determination is: $\sigma(\theta) \leq 1.6^\circ$ and $\sigma(\phi) \leq 5^\circ$. The angles of incidence are determined by a novel method described in [1].

In the period of 1998 - 2002 approximately 7,788,000 EASs were registered with effective registration time of about 24,000 hours. From these showers $\sim$ 963,000 events were selected for the analysis. The selection criteria was to have more than 90% efficiency of registration, the following cuts were applied for the events selection: $N_e \geq 10^5$, $0.3 \leq s \leq 1.7$, $-24m \leq X_0 \leq 24m$, $-12m \leq Y_0 \leq 12m$, $\Delta(x, y) \leq 1.5m$, $\Delta s \leq 0.15$, $\theta \leq 45^\circ$

2 The Detector Response

The influence of the detector response function for the MAKET-ANI installation is investigated using CORSIKA [2] simulations. The EAS simulations were done for the Aragats level (3200m a.s.l.) using QGS [3] strong interaction model and the NKG approximation.

The main aim of this study is the EAS data classification into light and heavy groups of nuclei.

The simulated samples generated by the CORSIKA code were compared with those of including the MAKET-ANI detector response. For such comparison different methods were used, including comparison of the single EAS parameter distributions, one-dimensional statistical tests, correlation analysis and misclassification matrices obtained by neural and Bayesian classification.

Proton and iron simulated EAS events were used for these study. The N_e, S - shower age, and θ parameters are used for classification. The results of different kind of comparisons are presented bellow.

2.1 One dimensional tests for proton and iron samples with and without detector response

If we compare single EAS parameter distributions in terms of mean values and variances, we'll see that no large differences are introduced by the response function (Table 1). The only parameter somehow disturbed is the shower age parameter, which gets smaller difference for proton and iron samples when taking into account response function. This difference is 0.18 for pure CORSIKA generated samples and becomes 0.16 after including

the response function. these findings are demonstrated also by the one-dimensional statistical tests presented in Table 2. All three statistical tests (for the detailed description of these tests see [4, 5]) give practically the same P-values for all the EAS parameters when comparing proton and iron samples. The larger P-value demonstrates large difference in two compared samples. The only parameter which shows less differences for these two samples, when the detector response is included, is the shower age parameter.

Table 1: Mean values and variances

	without detector response				with detector response			
	μ		σ		μ		σ	
	H	Fe	H	Fe	H	Fe	H	Fe
lnE_0	13.35	14.35	1.21	1.15	13.34	14.35	1.19	1.15
lnN_e	12.08	12.36	1.26	1.32	12.15	12.41	1.20	1.29
S	0.97	1.15	0.10	0.10	0.94	1.10	0.15	0.14
X_0	-0.52	-0.61	18.70	19.2	-0.68	-0.69	18.64	19.24
Y_0	0.13	0.18	10.0	10.0	0.21	0.07	10.06	10.27
$cos\theta$	0.87	0.87	0.09	0.09	0.87	0.87	0.09	0.09
φ	3.13	3.14	1.82	1.82	3.14	3.12	1.82	1.82

Table 2: One-dimensional tests (t- Student, D- Kolmogorov-Smirnov, U- Mann-Whitney

	without detector response			with detector response		
	t	D	U	t	D	U
lnN_e	20.87	9.88	22.97	20.49	8.97	21.41
S	158.94	56.69	125.03	104.63	41.31	95.37
X_0	0.43	1.35	0.29	0.06	1.38	0.31
Y_0	0.44	0.89	0.42	1.27	1.43	1.00
$cos\theta$	5.32	2.55	5.36	5.97	2.85	6.06
φ	0.29	0.89	0.29	1.16	1.11	1.17

2.2 Correlation matrices

More important and informative are the results from correlation analysis. The Tables 3, 5 and 4, 6 demonstrate the correlation matrixes without and with detector response for proton and iron samples respectively. From these tables one can see crucial differences in pairwise correlations of primary energy and N_e parameters with the shower age parameter. For instance, for the proton events the correlation of N_e and shower age parameters is rather significant in sample without detector response and it becomes nearly two times smaller when the response function is taken into account. For the iron events these correlation is very high in CORSIKA simulated sample and decreases dramatically for the sample with detector response. Correlation analysis shows no significant changes of differences for the rest of parameters.

Table 3: Correlation matrix for proton sample without detector response

	lnE_0	lnN_e	S	X_0	Y_0	$cos\theta$	φ
lnE_0	1.00	0.91	-0.16	-0.00	-0.01	-0.23	0.00
lnN_e	0.91	1.00	-0.47	-0.00	-0.01	0.04	0.00
S	-0.16	-0.47	1.00	-0.01	0.01	-0.40	0.02
X_0	-0.00	-0.00	-0.01	1.00	0.02	0.00	-0.01
Y_0	-0.01	-0.01	0.01	0.02	1.00	0.01	0.01
$cos\theta$	-0.23	0.04	-0.40	0.00	0.01	1.00	-0.01
φ	0.00	0.00	0.02	0.01	0.01	0.00	1.00

Table 4: Correlation matrix for iron sample without detector response

	lnE_0	lnN_e	S	X_0	Y_0	$cos\theta$	φ
lnE_0	1.00	0.93	-0.70	-0.01	-0.02	-0.32	0.00
lnN_e	0.93	1.00	-0.83	-0.01	-0.01	0.03	0.00
S	-0.70	-0.83	1.00	0.02	0.01	-0.19	0.03
X_0	-0.01	-0.01	0.02	1.00	0.00	0.00	0.01
Y_0	-0.01	-0.01	0.01	0.00	1.00	0.01	0.00
$cos\theta$	-0.32	0.03	-0.19	0.00	0.01	1.00	-0.02
φ	-0.00	0.00	0.03	0.01	0.00	-0.02	1.00

2.3 Comparison of one and two dimensional distributions of EAS features for the MC data with and without detector response

For the visualization of the comparison results the one and two dimensional distributions of the used EAS parameters is presented. Figures 2, 3 demonstrate the N_e spectra of

Table 5: Correlation matrix for proton sample with detector response

	lnE_0	lnN_e	S	X_0	Y_0	$cos\theta$	φ
lnE_0	1.00	0.90	-0.06	-0.01	-0.01	-0.24	0.00
lnN_e	0.90	1.00	-0.28	-0.02	-0.01	0.05	0.01
S	-0.06	-0.28	1.00	-0.04	-0.04	-0.29	0.02
X_0	-0.01	-0.02	-0.04	1.00	0.01	0.01	0.00
Y_0	-0.01	-0.01	-0.04	0.01	1.00	0.00	0.00
$cos\theta$	-0.24	0.04	-0.29	0.01	0.00	1.00	-0.01
φ	0.00	0.00	0.02	0.00	0.00	-0.01	1.00

Table 6: Correlation matrix for iron sample with detector response

	lnE_0	lnN_e	S	X_0	Y_0	$cos\theta$	φ
lnE_0	1.00	0.92	-0.38	0.01	-0.02	-0.32	-0.00
lnN_e	0.92	1.00	-0.45	-0.01	-0.03	0.03	0.00
S	-0.38	-0.45	1.00	0.05	0.02	-0.17	0.01
X_0	0.01	-0.01	0.05	1.00	-0.01	0.01	0.00
Y_0	-0.02	-0.02	0.01	-0.01	1.00	0.00	0.01
$cos\theta$	-0.32	0.03	-0.17	0.01	0.00	1.00	-0.01
φ	-0.00	0.00	0.01	0.00	0.01	-0.01	1.00

proton and iron simulated events without and with detector response respectively. One does not observe very much differences, which is clearly demonstrated in Table 1, which compares the mean values and variances of the single EAS parameters distributions. The results of one dimensional tests (Table 2) also confirm this finding.

As seen from the Figures 4, 5 and shown by the results of one-dimensional tests and distributions shape comparisons (Tables 2 and 1) the shower age parameter is altered by the detector response. The mean values of age distributions for proton and iron samples became slightly closer and variances larger, which has led to more overlapping of these distributions. From the two dimensional plots of $S - N_e$ dependence (Figures 7, **??**) one can easily see the influence of the response function. Due to the significantly decreased correlations of these two EAS features in proton and iron samples when the detector response is taken into account, (see Tables 3, 5 and 4, 6) the distributions became smeared and highly overlapping. Thus, the separation between EAS events induced by proton and iron primaries is obviously more difficult when the detector response is taken into account. This is illustrated by the results of neural and Bayesian classification of simulated EAS events as induced by the light and heavy groups of nuclei.

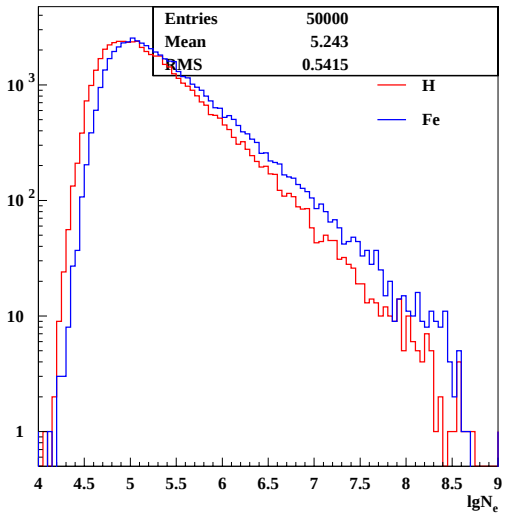

Figure 2: *N_e spectra of proton and iron for the MC data without detector response*

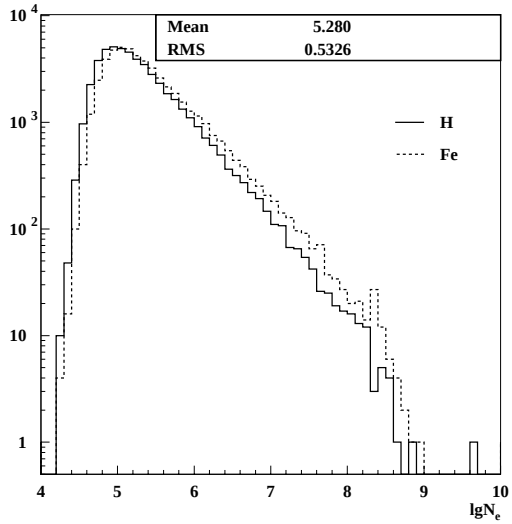

Figure 3: *N_e spectra of proton and iron for the MC data including detector response*

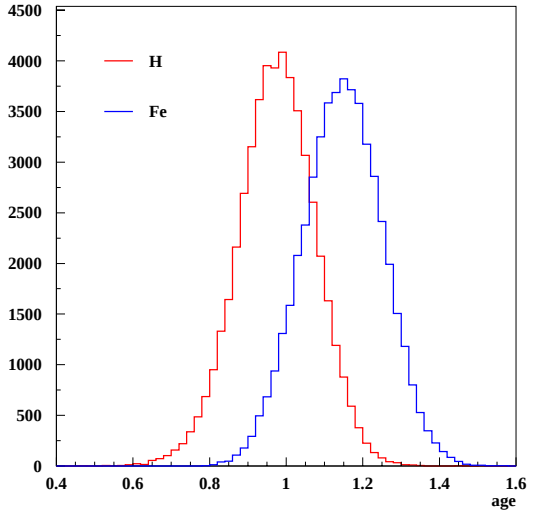

Figure 4: *The shower age parameter of proton and iron for the MC data without detector response*

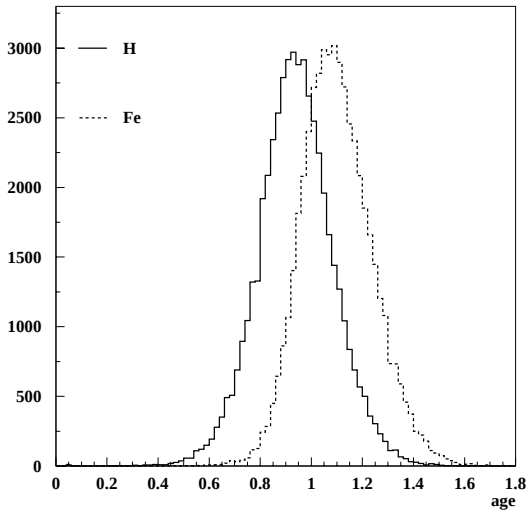

Figure 5: *The shower age parameter of proton and iron for the MC data including detector response*

2.4 Neural and Bayesian classification of EAS samples with and without detector response into light and heavy groups

The neural and Bayesian classification results (Tables 7, 8) show that CORSIKA generated light and heavy events can be distinguished with very high accuracy. The percent of correct classifications is somewhat 95% for both classes. This value decreases down to $\approx$ 76% for the samples with detector response included. So, the loss of the important

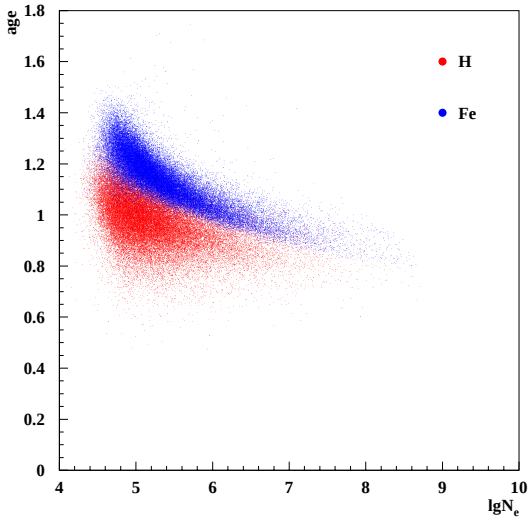

Figure 6: *The shower age versus shower size f proton and iron for the MC data without detector response*

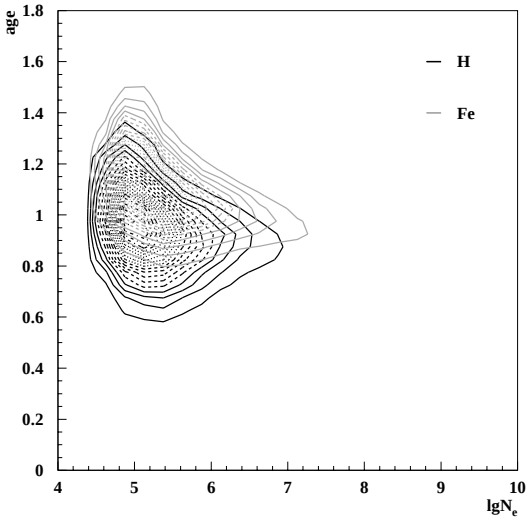

Figure 7: *The shower age versus shower size f proton and iron for the MC data with detector response*

Table 7: Neural classification into two classes using H+He and Si+Fe events without and with detector response

	without response		with response	
	light	heavy	light	heavy
light	0.925	0.075	0.720	0.280
heavy	0.045	0.955	0.240	0.760

Table 8: Bayes classification into two classes using H+He and Si+Fe events without and with detector response

	without response		with response	
	light	heavy	light	heavy
light	0.938	0.072	0.712	0.288
heavy	0.043	0.957	0.237	0.763

correlation information of the parameters used for classification leads to the significantly poor classification power in case of realistic simulations with full detector response.

Of course, taking into account the detector response and making EAS simulations more realistic introduces more fluctuations to the EAS parameters and affects some of these parameters, but it does not introduce significant systematic distortions as applied to the simulated events for MAKET-ANI installation and as shown by the results of different kind of comparisons. Such realistic simulations are the basis for comparative study of

MC generated and experimentally measured EAS features and allow to perform reliable analysis of the EAS data and to make physical inference. It is important to mention that more than 75% of correct classifications having realistic simulations is rather high accuracy for the two-way classification of EAS data.

3 Data classification into light and heavy groups of nuclei, purification of selected groups of nuclei

A Feed-Forward Neural Network (FFNN) is used to determine the mass and energy of individual primary cosmic rays (PCR) measured by MAKET-ANI installation in the knee region. The basics of the Neural Networks can be found in [6]. The general procedures for the application of neural network methods for EAS data analysis are given in [7].

For the classification and estimation tasks the N_e, s and θ parameters measured by MAKET installation are used. Each EAS event is classified as induced by light or heavy group of nuclei.

The efficiency and accuracy of two way classification are presented in Tables 9 and 12 As one can see from these tables the percent of correct classification for light (H, He) and heavy (Si, Fe) nuclei is rather high, the intermediate nuclei are classified into these two classes by nearly the same fraction - 0.45 and 0.55 for the first and second class respectively. The purity of the light class is significantly higher than that of the heavy group in the particular case of assumed primary composition of the control sample, which consists of: 30% H, 24% He, 17% Ox, 17.5% Si and 11.5% Fe.

Table 9: The efficiency of Neural classification of EAS data into two mass groups

	light	heavy
H	0.720	0.280
He	0.691	0.309
O	0.453	0.547
Si	0.352	0.648
Fe	0.240	0.760

Table 10: The efficiency of Neural classification of EAS data into two mass groups for the I purity interval

	light	heavy
H	0.672	0.160
He	0.589	0.216
O	0.348	0.430
Si	0.256	0.530
Fe	0.153	0.688

Table 11: The efficiency of Neural classification of EAS data into two mass groups for the II purity interval

	light	heavy
H	0.567	0.095
He	0.475	0.135
O	0.252	0.303
Si	0.176	0.393
Fe	0.099	0.561

The purification of selected light and heavy samples is done in the following way: the neural net analyses perform a nonlinear mapping of multidimensional characteristics of the EAS to the real number interval $[0, 1]$ (Fig. 8). Particular class assignments for the two way classification are the subintervals $[0., 0.5)$ and $[0.5, 1.]$ for the light and heavy class respectively. This classification procedure results in a purity[1] of 85% for the light class, and of 60% for the heavy class. The neural information technique [5] allows to decrease the contamination of misclassified events in each class of nuclei. Of course, the efficiency[2] of the classification is reduced at the same time.

If the neural net (NN) is satisfactorily trained to have generalization capabilities, the NN output distributions for the different classes are overlapping at the subinterval boundaries. Therefore, by shrinking the subintervals, one can remove a large proportion of misclassified events. But, simultaneously one loose parts of the true classified events. Figure 8 illustrates this procedure of purification.

Figure **??** plots the purity versus the efficiency for two classes. The purity of light nuclei can be reached more than 95% and for the heavy nuclei nearly 80%, while the efficiency is still remaining above 50%. The purity and efficiencies are obtained by classifying ~ 35000 light (H,He) and ~ 17000 heavy (Si,Fe) control events, which are not used for the training of the neural net. The high purity for both classes is achieved, since the intermediate nuclei (simulated oxygens) are not considered here.

[1] purity: fraction of true classified events in an actual number of events assigned to a given class

[2] efficiency: fraction of true classified events in total number of events of a given class

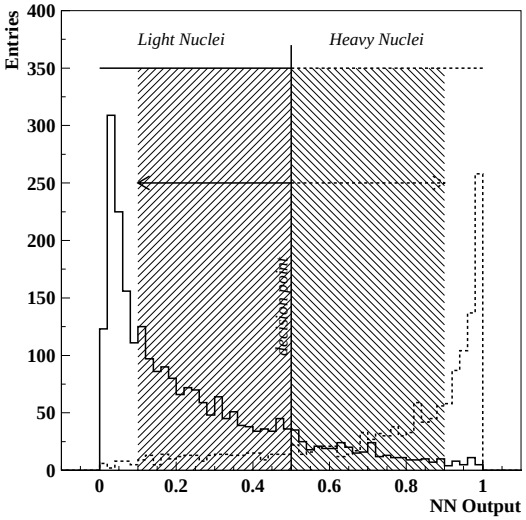

Figure 8: *An example of NN output distribution for the two way classification*

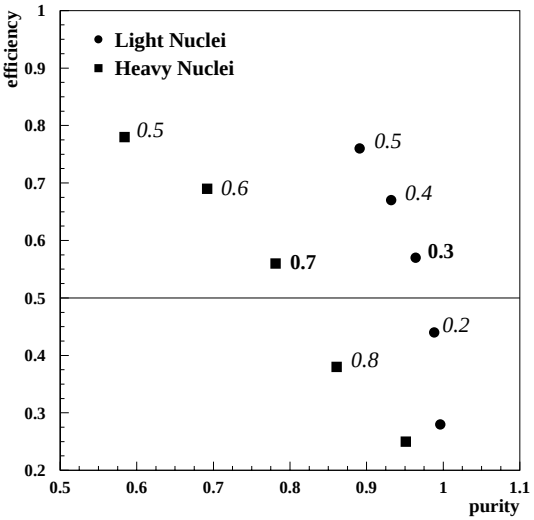

Figure 9: *Purity versus efficiency of classification into two mass groups*

Table 12: The contamination of different nuclei in light and heavy groups

	H	He	O	Si	Fe
light	0.407	0.298	0.137	0.111	0.047
heavy	0.162	0.167	0.208	0.255	0.208

Table 13: The contamination of different nuclei in light and heavy groups corresponding to the I purification interval

	H	He	O	Si	Fe
light	0.432	0.305	0.127	0.097	0.039
heavy	0.138	0.150	0.210	0.267	0.235

Table 14: The contamination of different nuclei in light and heavy groups corresponding to the II purification interval

	H	He	O	Si	Fe
light	0.459	0.310	0.115	0.084	0.032
heavy	0.115	0.131	0.207	0.278	0.268

The consideration of the intermediate nuclei spoils the situation with purity efficiency dependence. It turns out that purifying both, light and heavy classes is not so easy and

effective, nevertheless less there is a slight improvement in eliminating misclassified events from both classes. The purity of the light group increases from 70 to 76% and for the heavy ones from 46 to 55% (Tables 10, 11, 13, 14). As it is easy to see from Table 14 still the contamination of intermediate nuclei in the heavy sample is $\sim$ 20%, the other 25% are misclassified light nuclei. The purity of the light class is significantly higher since the contamination of intermediate and heavy nuclei is some 22%.

4 Estimation of the primary energy of different groups of nuclei

Primary energy of each EAS event is estimated separately for the light and heavy nuclei induced events. The accuracy of the energy estimation, displayed in Figures 10 and 11 demonstrate the high reliability of the procedure. The energy estimation bias does not exceed the 5% level for the light group of nuclei in the whole energy range except the lowest energies. For the heavy group of nuclei again the estimation bias in the energy range of $10^{15} - 10^{16}$ eV is not larger than 5%, nevertheless, one can observe some underestimation of the primary energy for the middle energies and overestimation for low and high energy regions. The energy resolution for heavy group of nuclei is significantly better as compared to the one of light group of nuclei due to the smaller fluctuations of EAS features induced by heavy nuclei.

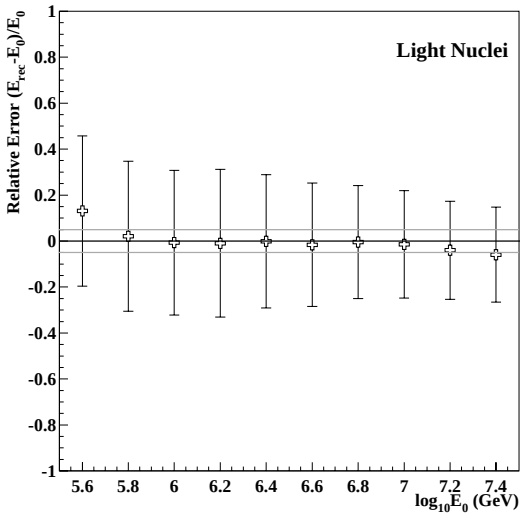

Figure 10: *Primary energy estimation of light nuclei (H,He)*

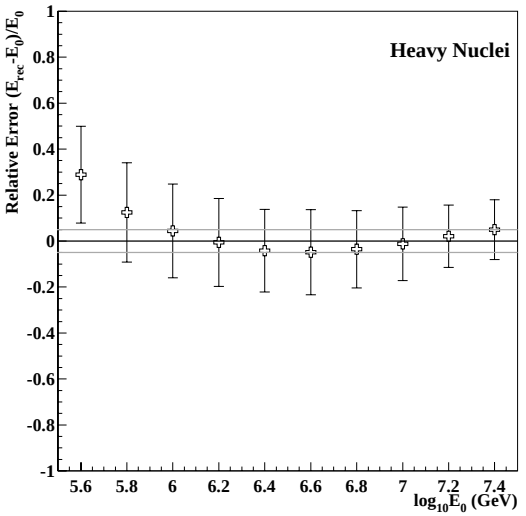

Figure 11: *Primary energy estimation of heavy nuclei (Si,Fe)*

5 Results

After performing a primary energy estimation with rather high accuracy and classifying primary particles into light and heavy groups of nuclei the primary energy spectra of these two groups of nuclei and their relative abundances in the energy region of $10^{15} - 10^{16}$ eV is constructed. In Figure 12 these spectra and the resulting all-particle spectra are shown. As one can see from this figure the spectrum of the light group of nuclei shows a knee around the 4×10^6 GeV energy. The knee feature is not observed for the spectrum of the heavy component, it shows a constant spectral index in the investigated energy region. The abundances of light and heavy nuclei at $\sim 10^{15}$ eV is approximately equal and the abundance of heavy nuclei is obviously larger after the knee in the energy spectrum of the light nuclei. This could be caused by the fact that the so-called light and heavy groups of nuclei have some contamination of the intermediate nuclei. As one can see from the Tables 9 and 12 slightly more intermediate nuclei are assigned to the heavy class than to the light one. As the assumed primary composition is light dominant, the mixture of intermediate nuclei is two times larger in the heavy class as compared to the mixture of those in the light class. So it is interesting to apply the purification procedure described in the previous section to these light and heavy samples in order to eliminate mainly the oxygen contamination in both classes as much as possible and to compare the primary energy spectra for pure light and pure heavy nuclei with ones in the samples with oxygen mixtures.

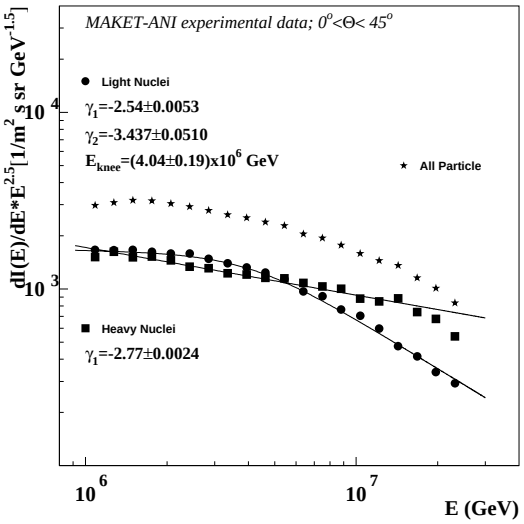

Figure 12: *Primary energy spectra of the light and heavy groups of nuclei obtained by NN classification and estimation*

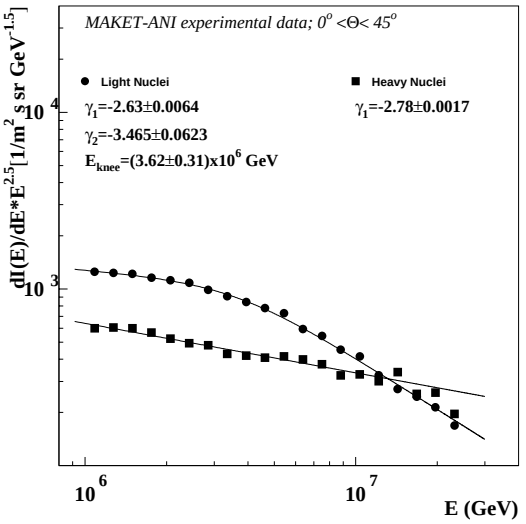

Figure 13: *Primary energy spectra of the purified light and heavy groups of nuclei obtained by NN classification and estimation of MAKET-ANI data.*

The resulting spectra of the "pure" light and heavy groups of nuclei are presented in Figure 13. The comparison of these spectra with initial ones demonstrates the following: obviously the flux intensities for both classes are shifted down, because of removed oxygen

events. More events are removed from the heavy group of nuclei than from the light one, this finding accepts the fact that intermediate nuclei have larger mixture in the sample classified as heavy. Comparison of spectral indexes, knee position and the shape of spectra shows that the changes in the purified samples are negligible, which means that events are removed from the whole primary energy interval nearly uniformly, and thus, the purification procedure does not introduce any significant systematic distortions.

The Figure 14 demonstrates the dependence of the relative abundances of purified light and heavy nuclei on primary energy. At a lower part of investigated energy region the primary composition is light dominant with a trend to change to a heavier composition at higher energies.

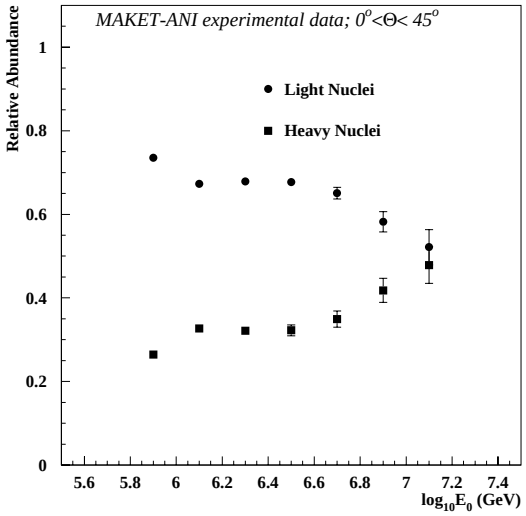

Figure 14: *Relative abundances of the purified light and heavy groups of nuclei versus primary energy*

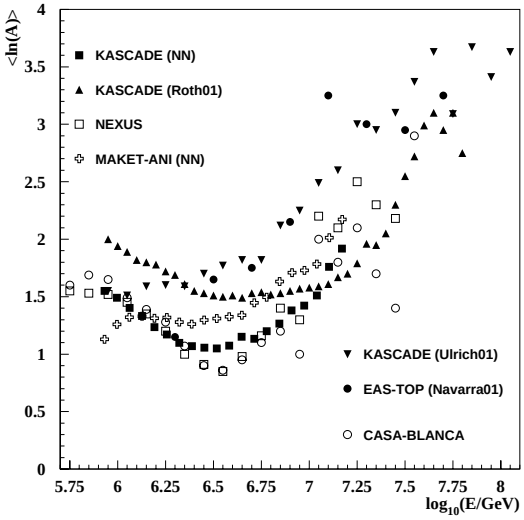

Figure 15: *Mean logarithmic mass as obtained by different experiments using different methods*

After these findings the obtained results were compared with the results reported by different EAS experiments. In Figure 15 the mean logarithmic mass versus the primary energy is plotted for different experiments. One observes some discrepancy of the results in terms of the values of mean mass as well as in terms of the behavior of it. Nevertheless, the behavior and the absolute values of the mean logarithmic mass of the KASCADE and MAKET-ANI experiments agree rather well. It is necessary to mention that the same NN technique of energy estimation and mass classification was used in both cases, although KASCADE results are obtained after three way classification in contrast to the two way classification performed in the current analysis.

Figures 16 and 17 compare the all-particle and light component primary energy spectra respectively. These spectra are obtained by different experiments using different data analysis methods. The agreement of results is rather well for the all-particle spectra. Energy spectra of the light group of nuclei obtained from KASCADE and MAKET-ANI data using the same analysis method is very good in terms of intensities, shape of the

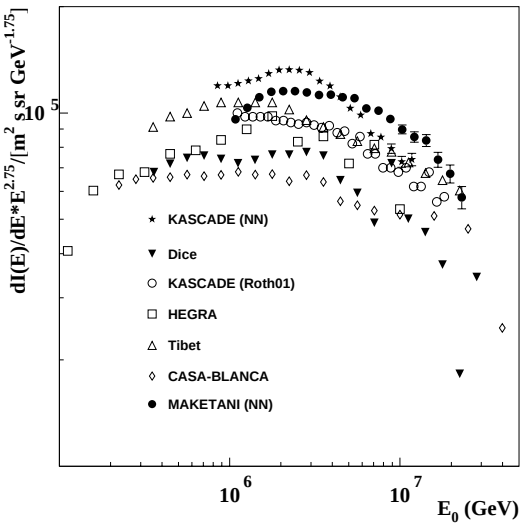

Figure 16: *Comparison of all-particle primary energy spectra of different experiments.*

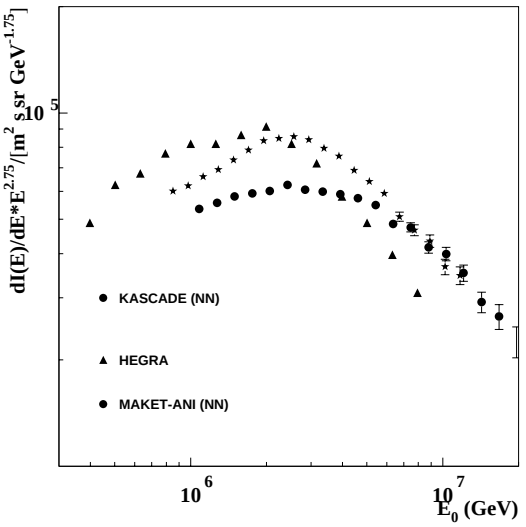

Figure 17: *Comparison of the primary energy spectra of the light group of nuclei for different experiments*

spectra and spectral indexes as well. The spectrum obtained by the data of the HEGRA experiment shows the same intensities and the shape like the other ones.

6 Conclusions

- It is possible to each EAS event registered by MAKET-ANI installation as induced by light or heavy group of nuclei and to estimate its primary energy with rather high accuracy.
- the estimated energy spectrum of the light group of nuclei shows a sharp knee at $\sim 3.5 \times 10^{15} eV$ energy.
- the energy spectrum of the heavy component of CR shows no knee in the investigated interval of primary energies.
- the relative abundance of light group of nuclei decreases with increasing energy while that of the heavy group of nuclei shows an opposite behavior.
- thus, the mean logarithmic mass is increasing with primary energy.
- the obtained spectra of light and heavy components coincide well with the spectra estimated by the KASCADE experiment using the same NN technique.
- The resulting all-particle energy spectrum agrees with the results of different experiments obtained using different methods of data analysis.

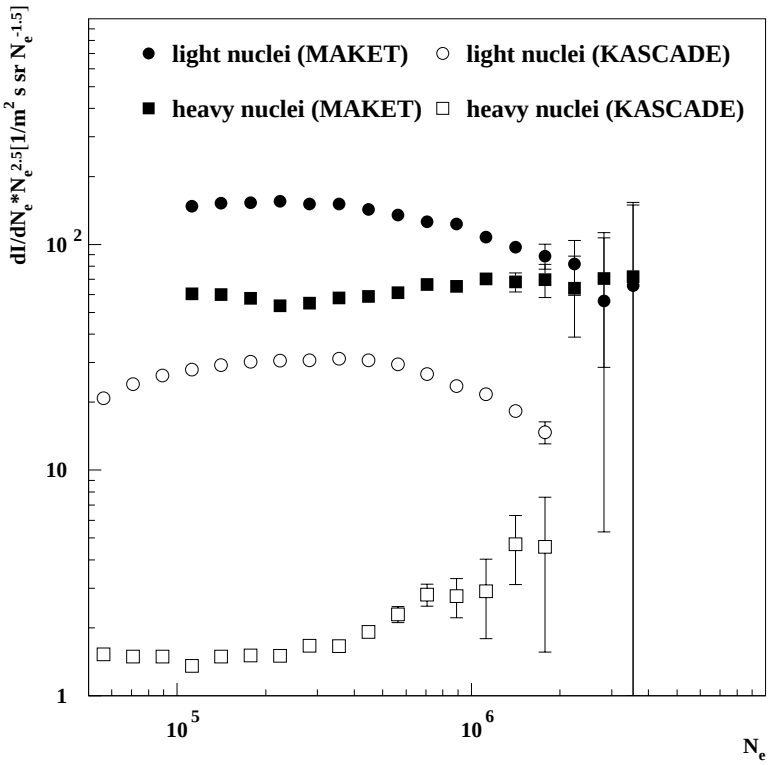

Figure 18: *N_e spectra of light and heavy groups of nuclei obtained by NN classification of MAKET data. (Compared to the KASCADE data from our previous analysis)*

References

[1] Chilingarian A.A., Gharagyozyan G.V., Martirosyan H.S. and Hovsepyan G.G.: Determination of angular accuracy of the MAKET-ANI surface array. In *Proc. of the 27th ICRC*, p. 590. Hamburg, 2001

[2] Heck D., et al.: The Karlsruhe Exensive Air Shower Simulation Code CORSIKA. FZKA-Report 6019, Forschungszentrum Karlsruhe, Forshungszentrum Karlsruhe, Germany, 1998

[3] Kalmykov N. N., Ostapchenko S. S., Pavlov A. I.: Quark-Gluon-String model and EAS simulation problem at ultra-high energies. *Nuclear Phys. B*, **52B**; 17, 1997

[4] Fukunaga K.: *Introduction to Statistical Pattern Recognition.* Academic Press, Harcourt Brace Jovanovich Publishers, 1990

[5] Chilingarian A. A.: Analysis and Nonparametric Inference in High Energy Physics and Astroparticle Physics, 1998. Program Package ANI, (Users Manual, unpublished) http://crdlx5.yerphi.am/proj/ani

[6] Bishop C. M.: *Neural Networks for Pattern Recognition.* Oxford Univ. Press, New-York, 1995

[7] Chilingaryan A. A.: Neural classification technique for background rejection in high energy physics experiments. *Neorocomputing*, **6**; 497, 1994