MAGIC-TDAS 04-xx
04mmdd

New Image Cleaning Method for the MAGIC Telescope

M. Shayduk, Th. Hengstebeck, O. Kalekin, N. A. Pavel, Th. Schweizer,

April 01, 2005

Abstract

The MAGIC telescope reflector has a parabolic shape, which conserves the time structure of the cosmic showers. Together with the fast FADC data acquisition system it allows to optimize the integration range for the signal extraction, even on an *event-by-event* basis. A new Image Cleaning procedure, based on the time coincidence of neighbour pixels is described.

Contents

1 Introduction

The MAGIC telescope has an isochronous reflector with an intrinsic time spread of *400ps*, with the effect of mirrors staggering included. Considering the DAQ system the telescope has a time resolution of ∼*600ps*. Such a precision is sufficient to resolve the time structure of the cosmic showers. The FADC system samples the shape of the shower signal over the time window of *50ns*, while the typical time spread of the cosmic showers in the telescope camera plane is only *5-10ns*. Thus, exploiting the fast FADC system allows to optimize the signal to noise ratio by reducing the integration ranges to the relevant limits around the shower signal, decreasing the NSB contribution, even for each individual event.

We would like to emphasize that the main aim of the following Image Cleaning procedure is to improve the separation between the NSB noise and the shower signal. In this paper we will only make a slight touch of the further outgoing improvements in the g/h-separation. One should note, that extracting more signal does not significantly change the Hillas Parameters of the high energy events, but becomes extremely important for the signals near the telescope threshold, where the image consists of *4-10* pixels only. In other words, all significant gain from the new Image Cleaning procedure expected to be obtained in the low-energy region.

The method described below was tested on MC and real data. The exploit of the method on the real data concerns the delicate tasks of data calibration and signal extraction. Therefore we shortly discuss these questions in the first two sections.

2 Calibration

At present time several well-known methods of the absolute light calibration of the MAGIC telescope exist: the *"PIN diode method"*, where the LED light flux is directly measured at the camera plane by a calibrated PIN diode, the *"Blind pixel method"*, where the LED flux is measured by a PMT in the camera, working in the counting mode, and the so-called *"F-factor method"*, where the number of photoelectrons, induced by the LED flashes is calculated from Poisson fluctuation. All these methods are described in details in [1].

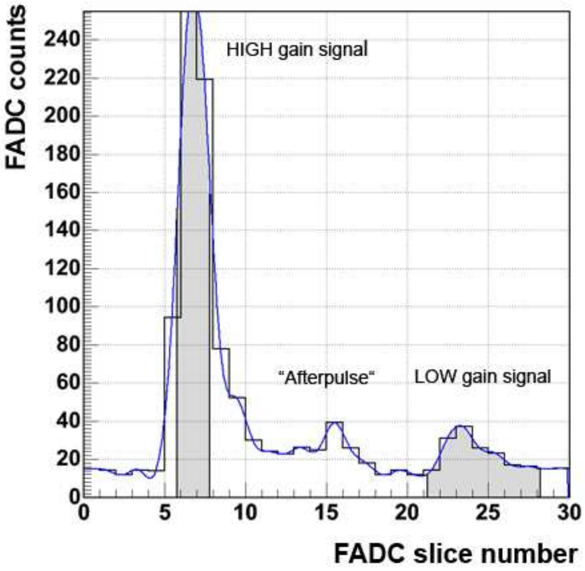
A good crosscheck for all these procedures is an alternative method for the absolute calibration with the muon rings [2]. With this method the whole signal processing chain *reflector-camera-DAQ* can be calibrated.

Hardware works on *PIN diode* and *Blind Pixel* methods are still in progress. In the present paper all data was calibrated using the *F-factor* method.

2.1 Calibration of low gain

The methods mentioned above can be directly applied to calibrate the high gain branch of the DAQ system (see the description of the DAQ system in [3]), using only one fixed LED intensity during the calibration run.

In order to calibrate the low gain the signal should appear in the low gain branch or better in both branches of the DAQ (in order to make a "smooth" high/low gain calibration). The high/low gain switches are set approximately to a signal amplitude of *100* FADC counts for all pixels. But due to the different PMT gain the low gain signal does not appear in some of the pixels. For the same reason the high gain saturation limits also differ from pixel to pixel. These facts make some difficulties in choosing the proper LED intensity and, in principle, demand a special run with a high LED intensity, devoted to the low gain calibration, which makes the whole calibration process rather sophisticated.



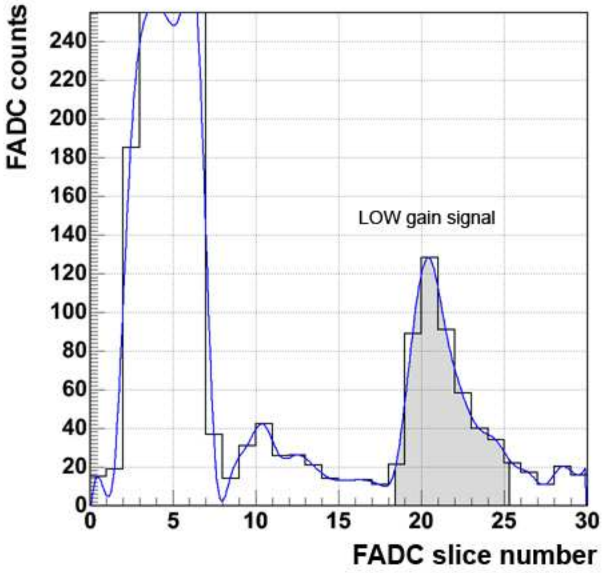


Figure 1: Example of the real signal with a high and low gain integration limits. Note that the "afterpulse" peak located in the first slice of the low gain range, has an amplitude which is a bit higher, as compared to the low gain signal.

Alternatively, the calibration of low gain can be done using the natural light from the cosmic showers with a big flux variability, providing the signal in both branches. Having the high gain signal already calibrated, one can determine the absolute calibration coefficients of the low gain, restoring the high gain amplification factors:

$$CONV_{high/low} = \frac{CHARGE_{highgain}}{CHARGE_{lowgain}}$$

In order to do that, the high gain signal in the amplitude range of *160* to *254* FADC counts was compared to the corresponding low gain signal as it is shown on Fig 1. One can clearly see a peak, which comes ~ 8 FADC slices later than the high gain pulse. This peak is probably caused by the effect of the so-called PMT afterpulses. Depending on the high gain signal position, that peak can be located even in the low gain range (as defined by a software).

Since the DAQ low gain branch has a known hardware time delay ($\sim 55ns$, because of the DAQ hardware), the peak search window for this branch was reduced to the size of *5* FADC slices around the expected position of the signal in order to avoid jumping to the "afterpulse" peak mentioned above (see again Fig.1).

The precision of the high/low ratio measurements depends dramatically on the signal extraction procedure. Different methods of the signal reconstruction were described in the recent TDAS note [4]. As it was shown, the reconstructed charge variance can be described as a square sum of different independent contributions:

$$RMS^2_{charge} = RMS^2_{nsb} + RMS^2_{el} + RMS^2_{extr}$$

where RMS^2_{nsb} denotes the NSB contribution, RMS^2_{el} the electronic noise and the last term RMS^2_{extr} describes the intrinsic signal extractor error. The absolute value of the NSB contribution remains constant with the increasing of the signal amplitude, depending on the integration window size only,

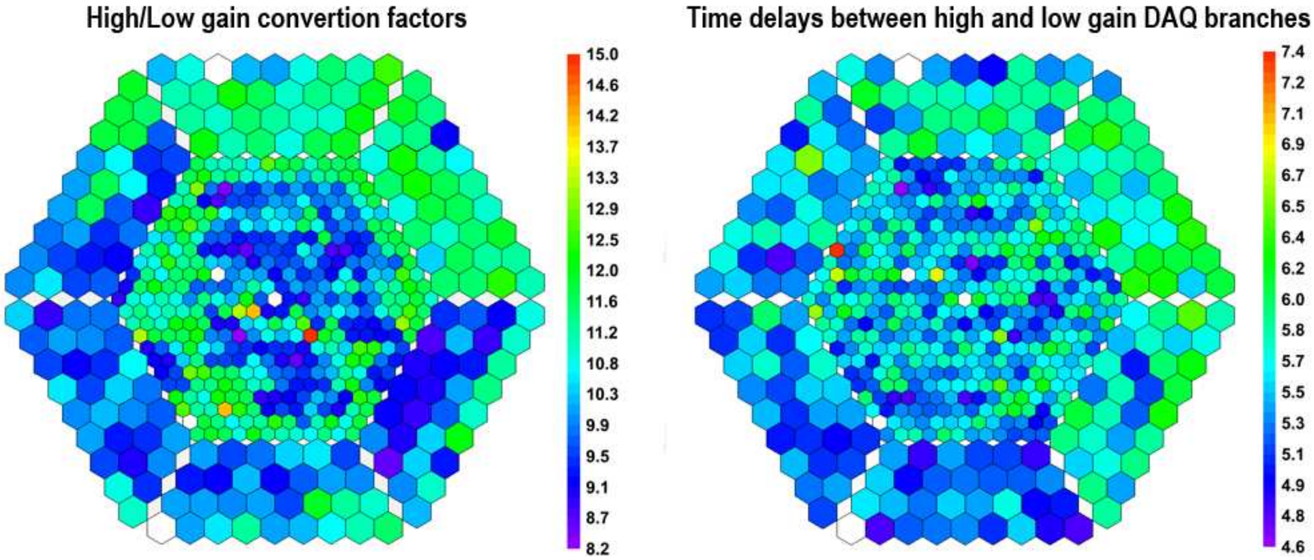


Figure 2: ***Left:*** Restored effective amplification factors, which are given by the ratio of high and low gain signals. The high and low gain signals was extracted with *2* and *7* slices integration window respectively. The fancy pattern is probably caused by the grouping of the readout channels in the receiver rack, where the high gain amplification is performed. Bunches of *8* pixels, wired to the one receiver board are quite noticeable. ***Right:*** Hardware time delays between the maximums of the high and low gain signals (a constant shift of *50ns* is substracted).

while the intrinsic extractor error varies linearly with the signal amplitude. The electronic noise is usually negligibly low.

Thus, for the small and medium charges (up to *50-100PHes*) one should minimize the NSB contribution to the charge variance to get a better resolution. Therefore the integration window should be small. Accordingly, the high gain charge was extracted using integration of spline over the window of *2* FADC slices around signal maximum.

For the range of relatively big charges, where the intrinsic extractor error exceeds the NSB contribution to the charge variance, the integration window should be larger in order to decrease that error (see [4, The Charge var...]). In addition, due to the analogue delay line the low gain branch has by default a broader signal (see [4, 2 Pulse shape reconstruction]). The other reason for the increase of the integration window for low gain is that for very big charges the intrinsic time properties of the shower might contribute to the resulting signal shape. Taking into account these facts, the low gain signal was extracted with the larger integration window of *7* FADC slices. The size of the window was tuned manually, while minimizing the error of the high/low conversion coefficients.

In principle, each DAQ channel has an individual time delay due to the hardware features. It is known that the signal in the outer pixels comes systematically later, mostly because of the bigger electron travel time in the outer PMT. These time delays can be easily restored, while calibrating the high gain DAQ branch. The artificial low gain time delays, in principle, should also be calibrated. It can be done on the way with the low gain calibration.

Some kind of a display to see the "smoothness" of the calibration is a cosmic events with a big *SIZE* parameter and huge time spread. Such events were already observed in the real data and considered as a candidates for the atmospheric fluorescence detection, but we found images of that type also in the MC data, where the atmospheric fluorescence was not simulated (see Fig.3).

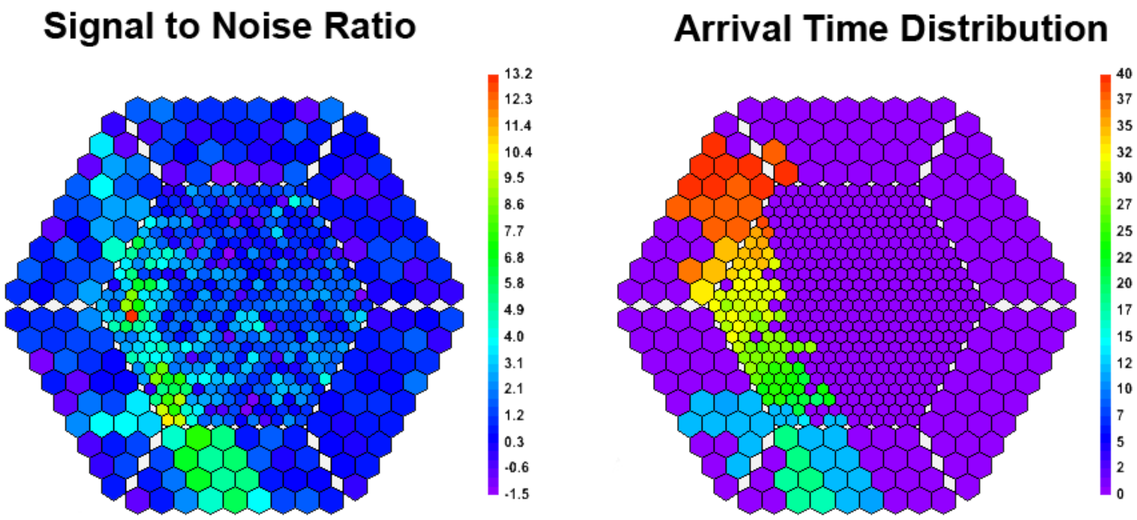


Figure 3: Image of the MC proton shower with an extremely big time spread and visible muon arc. Atmospheric fluorescence was not simulated.

The only disadvantage of the low gain calibration method described above is that the cosmic events usually cover only small part of the camera. For the pixels outside of the trigger area the probability to have a signal in the used amplitude range (*160-254* FADC counts) is even lower. Due to these reasons, one needs quite big amount of the data to be processed: to provide *50-100* entries for each outer pixel a set of about *400000* events is needed. Assuming the average data processing speed of *300event/sec* (see [4, Chapter 9. CPU requirements]), the low gain calibration procedure will take about *23* minutes.

3 Signal Reconstruction

3.1 Signal extraction methods

As explained in [4, Chapter 11. Conclusions], there are two most relevant signal extraction methods:

- Spline Interpolation over *2* slices
- Digital Filter

Actually, there is no significant difference in the charge resolution between these two signal extraction methods. The *Digital Filter* extractor demands the production of the constant signal shape for the high and low gain slices. The other disadvantage of this extractor is that it demands the obligatory substraction of the pedestal before the signal extraction, what involves looping over the data to calculate the pedestal RMS. Despite these disadvantages the Digital filter method has a big potential and the software works on it's careful implementation are almost finished.

For simplicity reasons we decided to use the *Spline Interpolation* method for the signal extraction, although the *Digital Filter* extractor is faster.

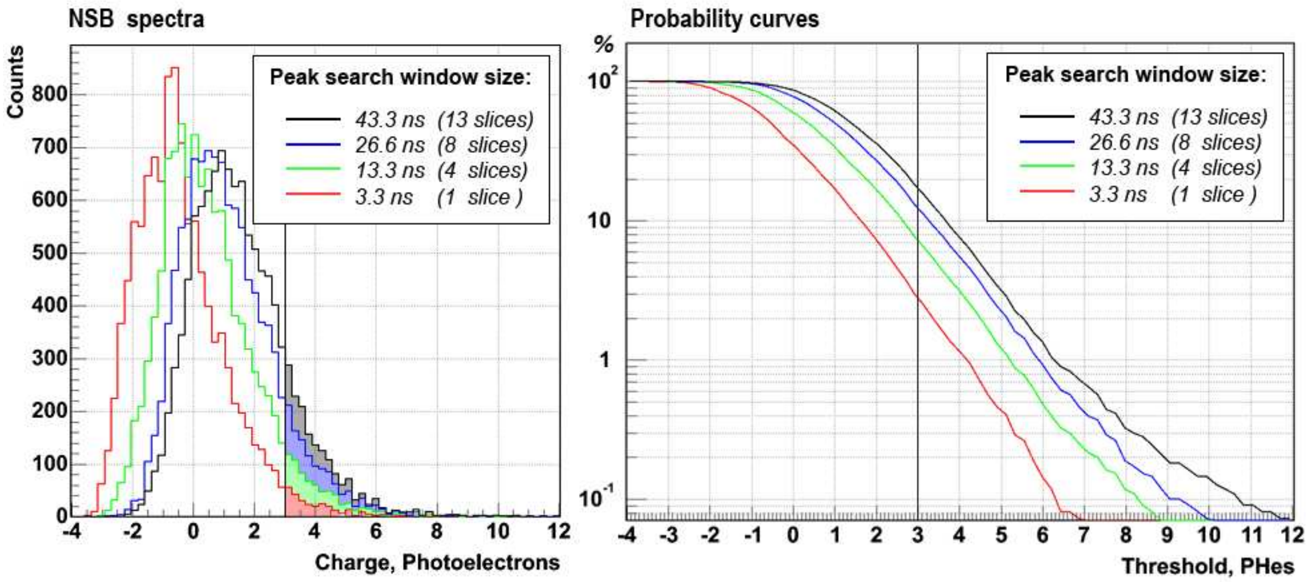


Figure 4: ***Left:*** NSB charge spectra, "seen" by the *Spline Interpolation Extractor* for different peak search windows. ***Right:*** Cumulative probability curves for different peak search windows, obtained from the integration of the NSB charge spectra, presented on the left picture.

3.2 Signal Extractors applied to pedestal run

In order to define the cleaning thresholds for the Image Cleaning procedure, it is necessary to know the confidence levels for the reconstructed charge in the pixel being a signal from the shower. These cleaning thresholds mainly depend on the NSB level. The pedestal RMS is a good indicator of the NSB level, but the confidence levels for the signal can not be calculated by assuming a gaussian *PDF* with parameters: *(0, RMS)*, because the pedestal NSB charge spectrum has an asymmetric non-gaussian shape.

The other effect, which also affects the gaussian-based calculation of the confidence levels, comes from the signal extraction procedure, where the signal peak with a $FWHM \approx 6ns$ should be found within the *50ns* FADC window. This procedure of the peak search produces an additional systematic shift in the probability levels, the so-called "signal extractor bias" (see [4, Section 4.1 Bias and Mean-squared Error]). This means that the probability curves are a function of the Signal Extractor properties, i.e. the charge resolution, which basically reflects the NSB level, and the bias. Thus, the whole Image Cleaning procedure is strongly dependent on the signal extraction method.

Taking into account these remarks, the proper experimental way to get the correct probability curves is to apply the signal extractor on pedestal data, pretending to find signals. An example of the resulting NSB "charge" histogram after such procedure is shown on the left plot of Fig 4.

3.3 Probability curves to catch the NSB signal

As it was already discussed, the procedure of searching for a peak shifts the mean of the pedestal charge histogram to the higher values, but makes the RMS smaller. The interplay of this two effects determines the final probability levels to identify noise as a signal. It is obvious, that the knowledge of the exact position of the signal reduces the signal extractor bias to zero values. Hence, it is extremely important to reduce the peak search ranges as much as possible.

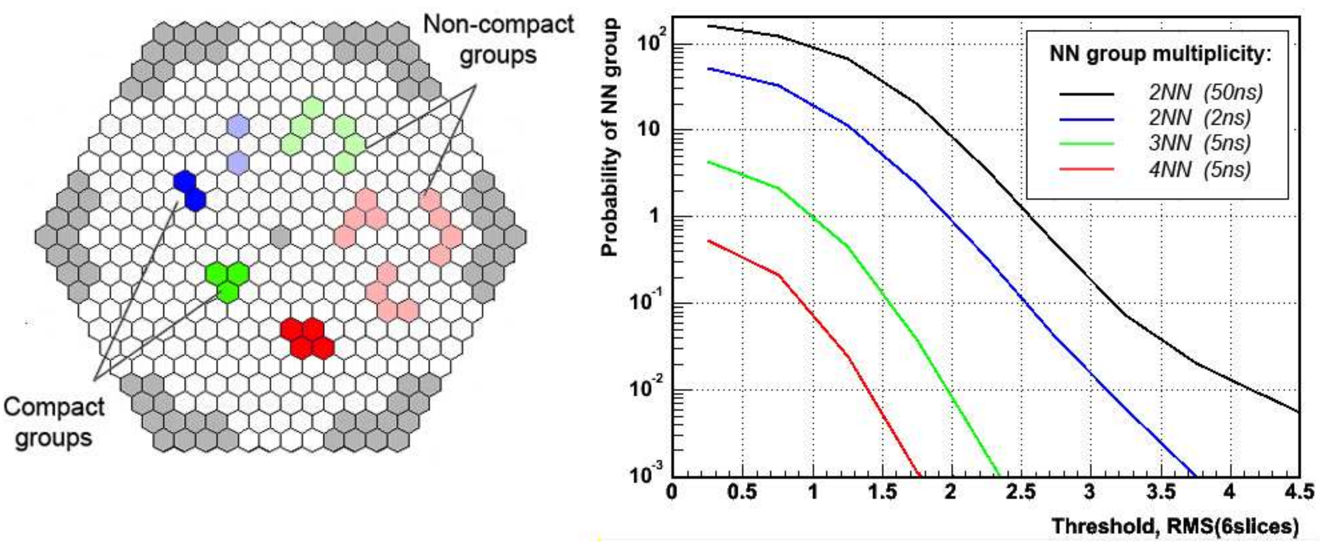


Figure 5: ***Left:*** Examples of the compact and non-compact *next-neighbor* group. ***Right:*** Probability curves to find the *2NN*, *3NN*, and the *4NN* compact group in the pedestal event. The threshold expressed in units of standard deviation of pedestal charge, integrated over *6* FADC slices. For the CRAB Nebula NSB level: $RMS(6slices) \approx 2.72PHes$.

The probability curves for the different peak search windows for the Crab Nebula NSB level are shown on the right plot of Fig 4. One can see a quite significant drop of the Image Cleaning Levels, mostly due to the decrease of the signal extractor bias, when reducing the peak search window from *45ns* to the values of just *5-3ns*.

4 Method

4.1 Probability of NN compact group

Since in the experimental data the position of the shower signal is not fixed and no information about the pixels, which caused the event trigger, is available, the extraction of the image core demands artificially high thresholds in the Image Cleaning procedure, due to the problem of the peak search discussed above. In order to reduce the peak search range one can try to restore the signal position, exploiting the topological and temporal properties of the image. It involves the calculation of the topological probabilities to find the compact *next-neighbour* groups, denoted here as *NN* groups (see Fig. 5).

For practical use this probabilities can be obtained by measuring a frequency distribution of the *NN* groups in the pedestal data, taking into account also the time coincidence. The right plot on Fig 5. shows the measured probabilities to find *2NN*, *3NN* and *4NN* group with a time coincidence, specified in the legend. These probability curves strongly depend on the star field in the telescope camera, which can dramatically increase the noise in some of the channels, locally raising the probability to find the compact *NN* group.

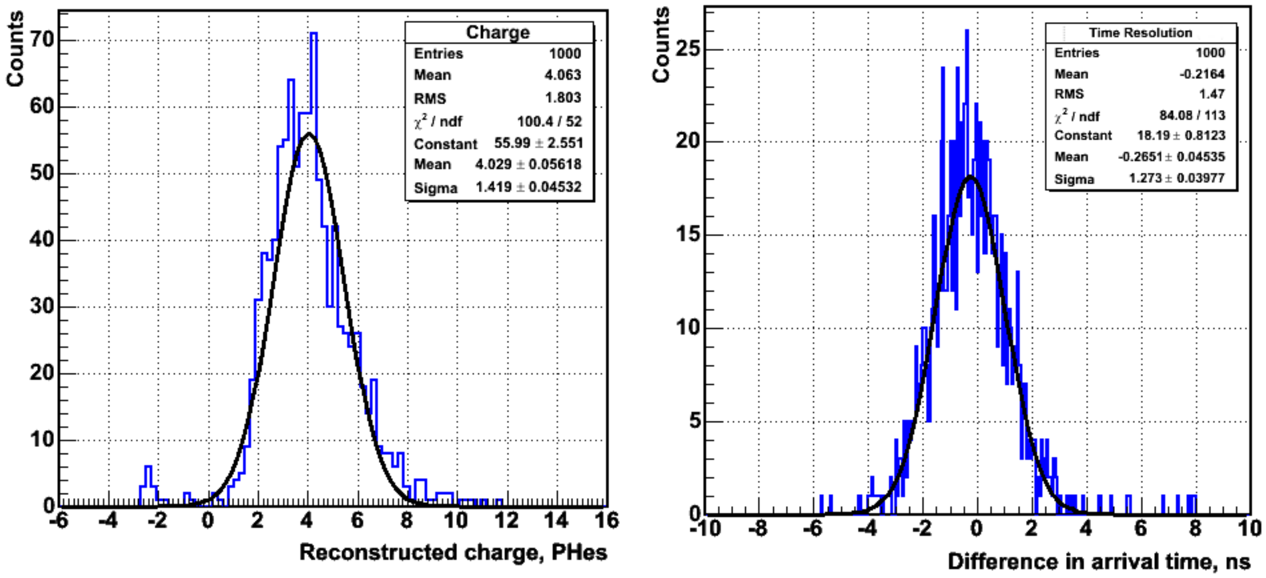


Figure 6: Charge and time resolution for *Spline Interpolation* signal extractor for the pulpo charge, equivalent to $4PHes$. The equivalent pulpo charge in reference pixel (pixel 315) was about $52PHEes$. Thus, the effective time difference variation for two signals of $4PHes$ each, can be estimated by multiplying the absolute time resolution (1.27ns) by a factor of $\sqrt{2}$.

4.2 Software NN Trigger

The so-called standard Image Cleaning procedure demands a *2NN* group above *8.2PHes* for the Crab Nebula NSB in order to accept the image, without checking the time coincidence (what corresponds to 3σ level for 6 slices integration window. Here and below σ denotes this RMS). That level is artificially high, because the data purity is provided only by the topology of the image. In this case, all images with *2NN* group below the *8.2PHes* threshold will be lost, while the individual pixel threshold, "allowed" by the hardware NSB trigger rate can be lower, namely about *4PHes*. If the signal is low, the probability to catch a high amplitude (but not time-correlated) *2NN* group caused by the NSB can be even higher that to catch a real signal. So, the image will be organized around NSB "core" pixels.

Let us call the decision of accepting the image a *Software NN Trigger*. In order to avoid using erroneously NSB for the organization of the image with low image cleaning levels, one should change the *Software NN Trigger* condition, taking into account the arrival time information.

The real hardware trigger has a *4NN* multiplicity and a coincidence window of *6ns* (after the last changes of the trigger condition). The trigger threshold is not yet known precisely, but as can be seen from the right plot on Fig 5. the probability to get a *4NN* topological group from the NSB drops already above the threshold of *3.5-4.5PHes* (1.6σ). These levels correspond approximately to 1-0.5% probability to catch a *4NN* group, induced by the NSB. So, one can set similar conditions for the *Software NN Trigger*.

The real signal in some of the pixels, composing the hardware *4NN* trigger group, can be lost during the Signal Extraction procedure, because it can fail the amplitude threshold or the time coincidence condition due to the poor charge or time resolution respectively (for the Crab Nebula NSB level the relative time resolution for two signals of *4PHes* is about *2ns*. The charge resolution is 35% only, see Fig.6).

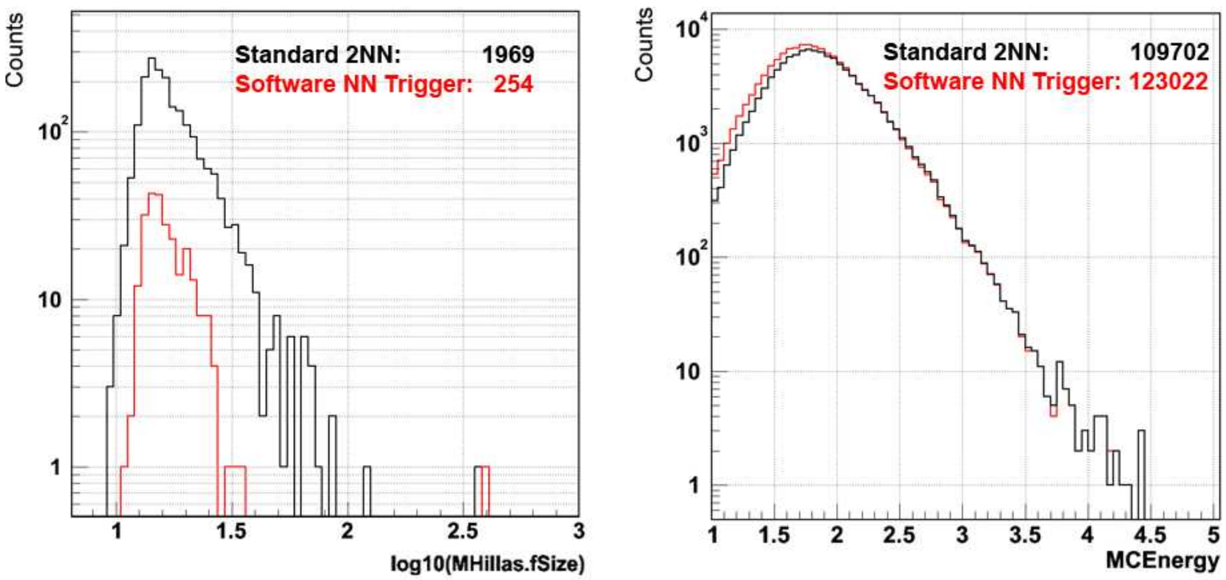


Figure 7: ***Left:*** *SIZE* distribution for pedestal events. *Standard trigger* and the new *Software NN trigger* procedures applied to the same pedestal data (total number of events - 25000). One can clearly see that new procedure gives ~ 8 times less triggers. ***Right:*** Energy distribution for MC gammas. Gain in number of events for small energies is clearly seen (total gain -12%).

Thus, in order to trigger the event in such cases additional conditions can be set: for a *3NN* group, at about the same probability level of 1%, corresponding threshold should be $\approx$ *5.4PHes* or 2σ(see again Fig.5). This additional condition will not provide trigger in all cases, but will surely increase the number of triggered events.

Summarizing all concepts, the *Software NN Trigger* should find at least one *4NN* group *or* at least one *3NN* group *or*, completing the concept, at least one *2NN* group in order to accept the event. The probability level to catch NSB for all these types of groups could be kept about the same, in order to avoid the unnecessary complexity of the procedure. Having chosen the probability level one can either require a certain amplitude threshold and reduce accordingly the time coincidence window or viceversa. As it was mentioned in previous section, the probability to find the compact *NN* group depends on the star field in the camera. Thus, the above specified thresholds are quite rough estimators of this probabilities.

The relevant comparison of the "standard" trigger and the new *NN Trigger* can be done applying both procedures to the pedestal runs. The criteria of better method is the smaller number of triggered events in the pedestal data and bigger or same number of the accepted events in the data run. The method can be optimized by minimizing the number of trigger events in pedestal runs and at same time maximizing the number of accepted cosmic events. Here the fraction of triggered NSB events was chosen to be 1%. This results in an increase of the number of true events being found with the new method in the low energy range as shown on Fig.7 (right plot).

4.3 Core pixels search

The next step in forming the image is the search of the so-called *core pixels*. One can define such pixels as a pixels with a signals at high confidence level. In the standard Image Cleaning procedure such high confidence was provided only by the image topology and, accordingly, quite high threshold

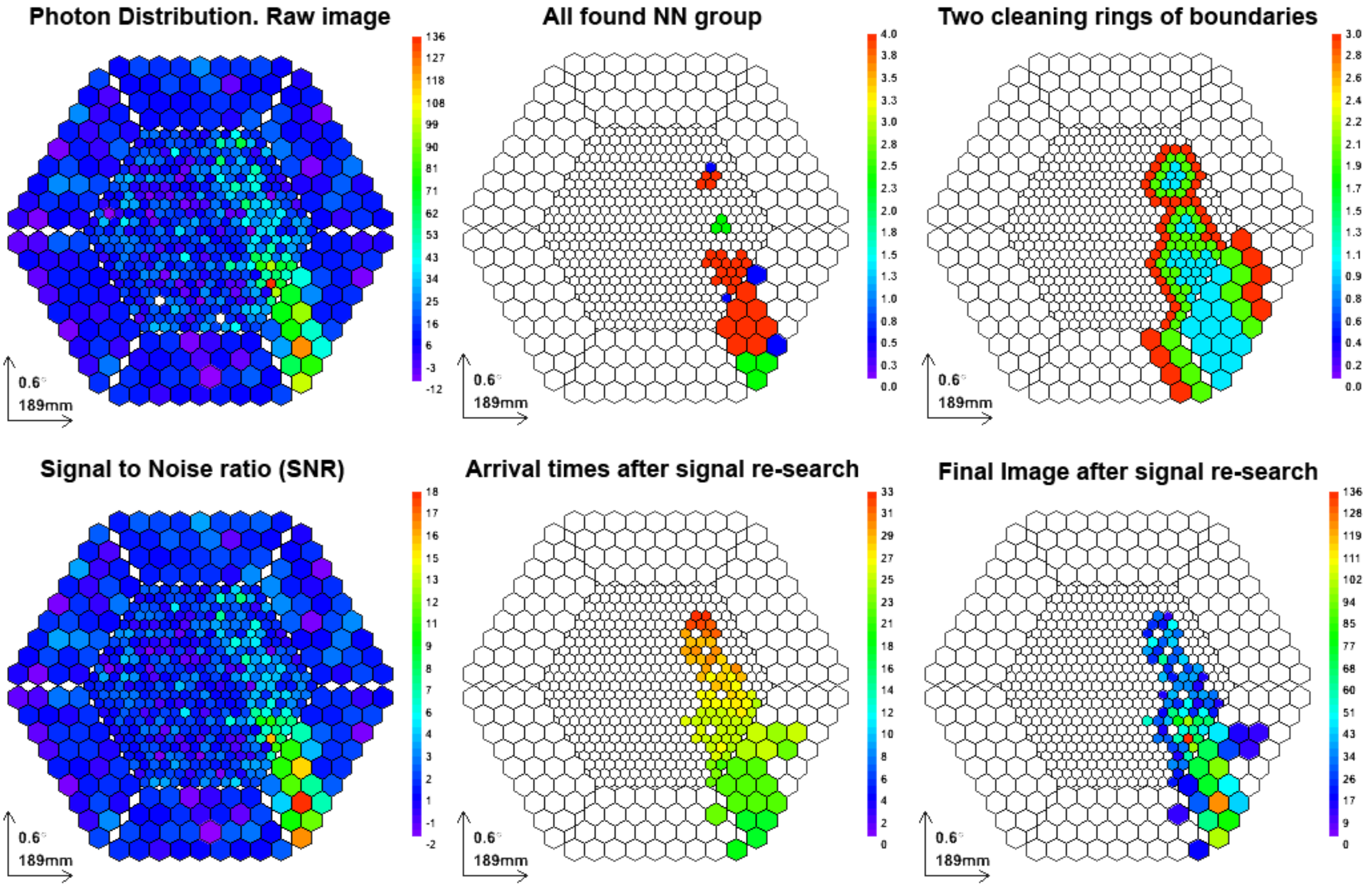


Figure 8: Image Cleaning stages. On the second figure in the first row all found compact next-neighbour groups are shown - *core pixels* (red color indicates all *4NN* groups, green - additional pixels, found by *3NN* group search, blue - additional pixels, found by *2NN* group search). The last figure in the first row shows two cleaning rings of the the *core* pixels neighbours, where the signal will be re-extracted.

to accept the signal.

As discussed above, by exploiting the additional time information, the cleaning levels can be decreased quite significantly. The natural way to do that is to use again the concepts of the "Software NN trigger" algorithm, described in the previous section. The *core* of the image has to be composed from compact time-correlated groups: *all 4NN* groups, *all 3NN* groups, *all 2NN* groups, which can be found by the "Software NN Trigger" procedure. It is obvious that such groups will overlap within the main island region, but all significant islands also will be found. All pixels participating in those groups can be set as high confidence *core pixels*. During such a procedure the image will be formed step by step, requiring a small difference in arrival time for neighboured pixels. The total time spread of the image, however may be any. Also, parts of the image with different local light intensities will be found: low concentration *4NN* parts and high concentration *2NN* parts.

4.4 Boundary pixels search

The input event data for the "standard" Image Cleaning Procedure is the signal and RMS of each individual pixel. All signals are already extracted from the raw FADC slices and already affected by the peak search procedure. The real low charge signals from the shower might be lost because a high fluctuation of NSB has been selected instead of the low signal in another time slice. This concerns mostly the boundary pixels, located in the periphery of the image core i.e. in the first, second and further cleaning rings. The first cleaning ring defined as the group of all pixels with at least one neighbour from the image *core*, the second cleaning ring comprises all pixels with at least one neighbour from the first cleaning ring and so on (see Fig.8). The possible way to improve the signal reconstruction for such pixels, is to "unify" the Signal Extraction and the Image Cleaning procedures.

When the *core* of the image is formed, the arrival time of the triggered signal is known. Thus, one can, finally, reduce the peak search window for the boundary pixels in the following way:

- Calculate for each boundary pixel the "expected" position of the signal as the average arrival time of neighbouring CORE pixels .
- Reduce the peak search window to the small window around the calculated signal positions.
- ***Loop again*** over raw FADC time slices, looking for the signal within the defined above limits. Set the very low threshold to accept the signal.

In case of a misidentification of the low signal peak with the higher NSB peak, while searching over all high gain slices, the second loop could help to find the signal peak around the proper position.

The unification of the Calibration, Signal Extraction and the Image Cleaning procedures has a disadvantage, that the reprocessing of the data is required since the signal thresholds for the Image Cleaning may have to be varied. The possible partial solution is to preprocess the data with a loose cleaning thresholds for boundaries and saving also more cleaning rings.

All Image Cleaning levels and the algorithm steps are summarized in the following table:

Trigger condition (OR)	**Signal level**	**Time Window**
one *4NN* group	*4.4PHes* (1.6 σ)	*5ns*
one *3NN* group	*5.4PHes* (2.0 σ)	*5ns*
one *2NN* group	*8.2PHes* (3.0 σ)	*2ns*
Image CORE		
all *4NN* group	*4.4PHes* (1.6 σ)	*5ns*
all *3NN* group	*5.4PHes* (2.0 σ)	*5ns*
all *2NN* group	*8.2PHes* (3.0 σ)	*2ns*
Boundary Pixels		
Re-search in 2 cleaning rings	*3PHes* (1.1 σ)	*3ns*

The numbers in the round brackets are the Image Cleaning thresholds, expressed in terms of standard deviations of the pedestal charge, calculated for the integration window of 6 FADC slices ("standard display" of the NSB level).

5 Results

The method described above was applied to the MC data together with the standard Image Cleaning procedure in order to make a relevant comparison of the algorithms. All MC data was simulated with

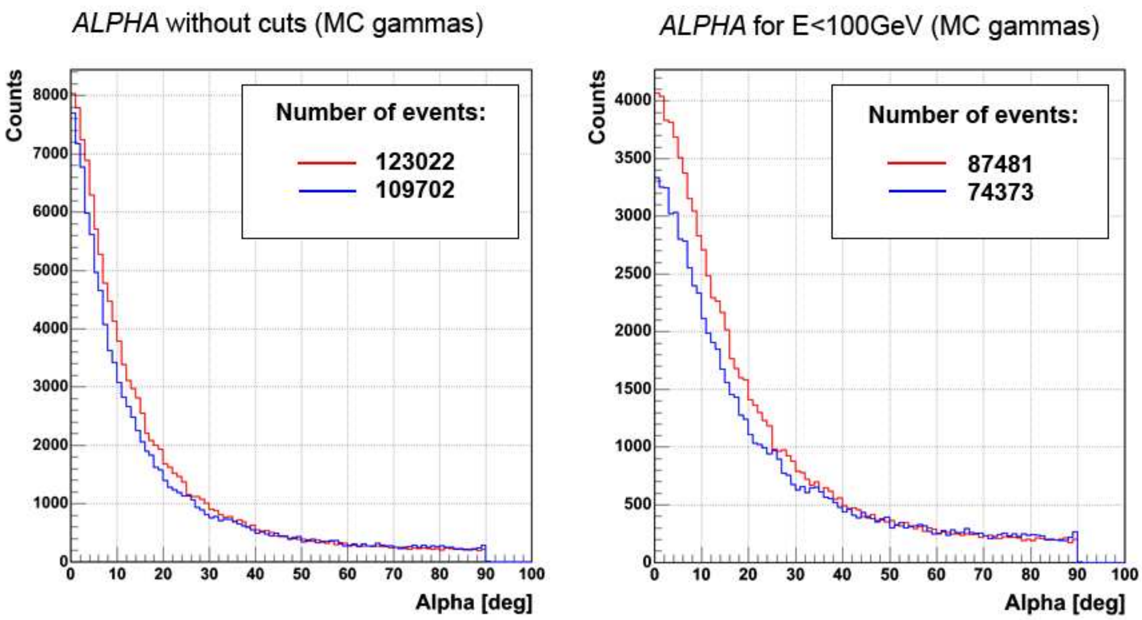


Figure 9: ***Left:*** *ALPHA* distribution for MC gammas without any cuts (0.4 phe/ns NSB noise is added). ***Right:*** *ALPHA* distribution for MC gammas after cut in energy $E < 100GeV$. Gain of 21% in number of events is clearly seen.

the NSB noise of $0.4PHes/ns$. The comparison of *ALPHA* distributions for MC gammas is shown on Fig.9. A clear gain of 12% in number of events is seen, which reaches the values of 21% for the energies below $100GeV$.

The distributions of other Hillas parameters *WIDTH*, *LENGTH* and *SIZE* are shown on Fig.10. All plots, presented in this section, are produced with a following precuts:

$$0.3^0 < DIST < 1.4^0$$
$$0.01^0 < WIDTH < \infty$$
$$Leakage1 < 0.1$$

The most significant effect of the new method is a reduced number of the so-called "pixelized" events, i.e. event in which the number of pixels in the image is so small that the discrete structure of the camera can already be seen. The new method with low Image Cleaning thresholds in such cases allows the additional pixels containing a signal to be found. Thus a more complete image can be recovered. The *WIDTH* and *LENGTH* distributions for such events are smoothed an do not show discontinuities or peaks, which are produced artificially.

The other quite remarkable effect of the new method can be seen in *SIZE* distribution plot and the plot MCEnergy vs *SIZE* in Fig.10. As can bee seen from these plots, the *SIZE* distribution, obtained with the new method, reflects the incident gamma spectrum in the low energy region more precise, because a large fraction of the true shower signals can be extracted from the image. As a result, for the new method the MCEnergy vs *SIZE* plot has a better linearity between the shower energy and *SIZE* of the image, as compared to the standard image cleaning procedure.

6 Conclusions

Resuming all the results, we would like to list the basic outcomes of the New Image Cleaning Method:

- Less NSB triggers. Higher purity of data
- 21% Gain in the number of events in the energy region below 100GeV (important for the pulsar studies).
- More informative Image Parameters, especially in the low energy region.
- Possibility to lower the analysis threshold.
- Sensitive to islands.

Further overall improvements in the g/h-separation and the energy estimation, caused by the higher purity of the images and extraction of low signals, should be studied, for example with using the *Random Forest* procedure [5]. In addition, a new possibility of the g/h-separation might be found due to the high sensitivity of the method to the islands.

References

[1] T. Schweizer et.al., *The Optical Calibration of the MAGIC Telescope Camera*, Nucl. Sci. 49 (2002) 2497.

[2] K. Mase, *Calibration of the MAGIC Telescope Camera with Muon Rings*, TDAS, Nov. 6 2001.

[3] J. Cortina et al. (MAGIC Collab.), *The data acquisition of the MAGIC telescope, Prepared for International Symposium: The Universe Viewed in Gamma Rays, Kashiwa, Chiba,*, Japan, 2528 Sep 2002.

[4] H. Bartko et al., *Comparison of Signal Reconstruction Algorithms for the MAGIC Telescope*, TDAS Feb. 21 2005.

[5] Th. Hengstebeck, *The Random Forest*, TDAS, to be published

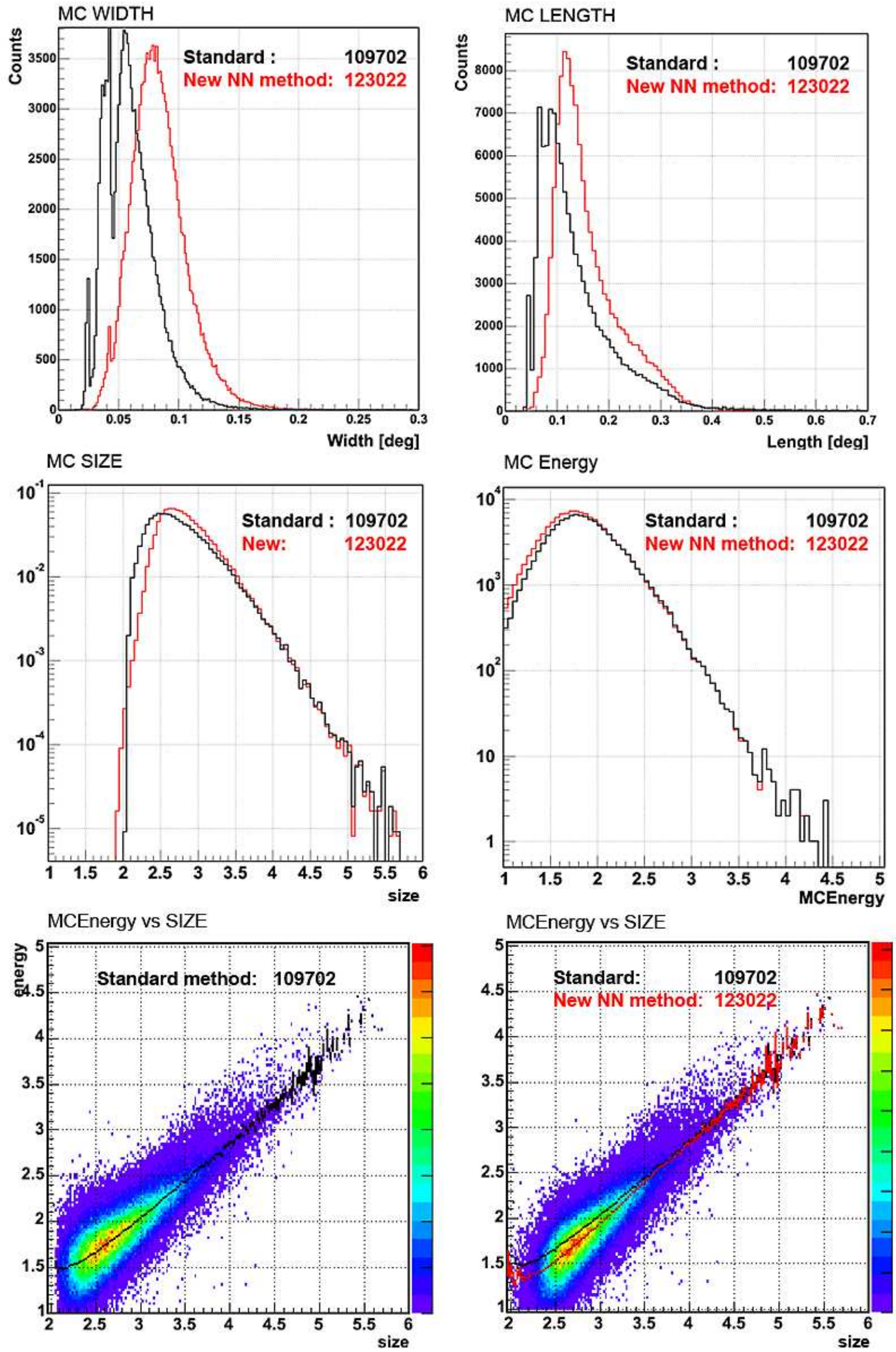


Figure 10: Comparison results for the standard Image Cleaning procedure and the new method. The camera pixelization effect in *WIDTH* and *LENGTH* distributions is dramatically reduced. MCEnergy vs *SIZE* plot for the new method has a better linearity (black and red curves indicate the mean energy for each *SIZE* bin for standard and new image cleaning method respectively).