The estimation of EAS core position by means of non parametric Parzen kernel probability density

Varlen Grabski
Universidad Nacional Autónoma de México, México

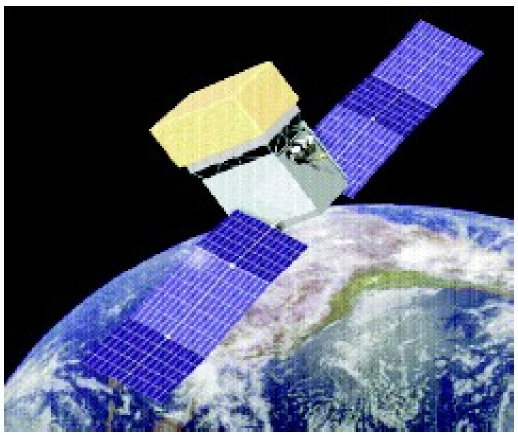
- Briefly about HAWC Observatory
- Simulation and detector response
- Algorithms for the core position
- Results
- Conclusions
- Acknowledgments

HAWC gamma observatory

HAWC Observatory

Comparison of Gamma-Ray Detectors

Low Energy Threshold
EGRET/Fermi



Space-based (Small Area)

"Background Free"

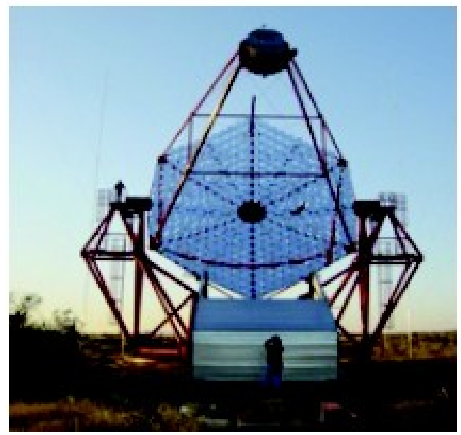
Large Duty Cycle/Large Aperture

Unique potentials at $\lesssim$ 200 GeV:

extragal. sources (AGNs and GRBs), mqso, pulsars

Dark Matter

Best Angular and Energy resolutions
HESS, MAGIC, VERITAS, CTA/AGIS (2018?)



Large Effective Area

Excellent Background Rejection

Low Duty Cycle/Small Aperture

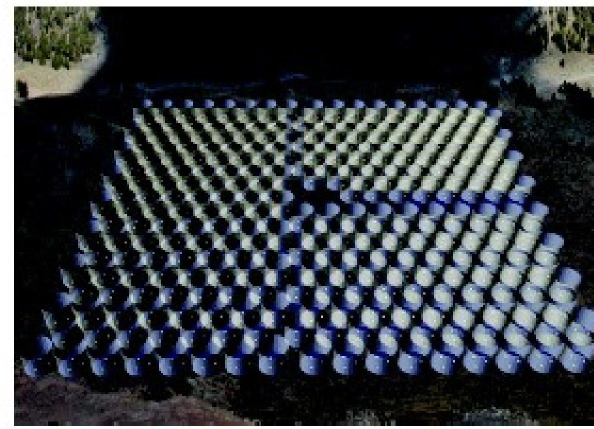
Morphology of TeV emissions (SNRs, PWN)

High Resolution Energy Spectra up to ~20 TeV

Studies of known sources (galactic and nearby galaxies)

Surveys of limited regions of sky

Large Aperture/High Duty Cycle
Milagro, Tibet, ARGO, future: HAWC,...



Moderate Area

Good/Excellent Background Rejection

Large Duty Cycle/Large Aperture

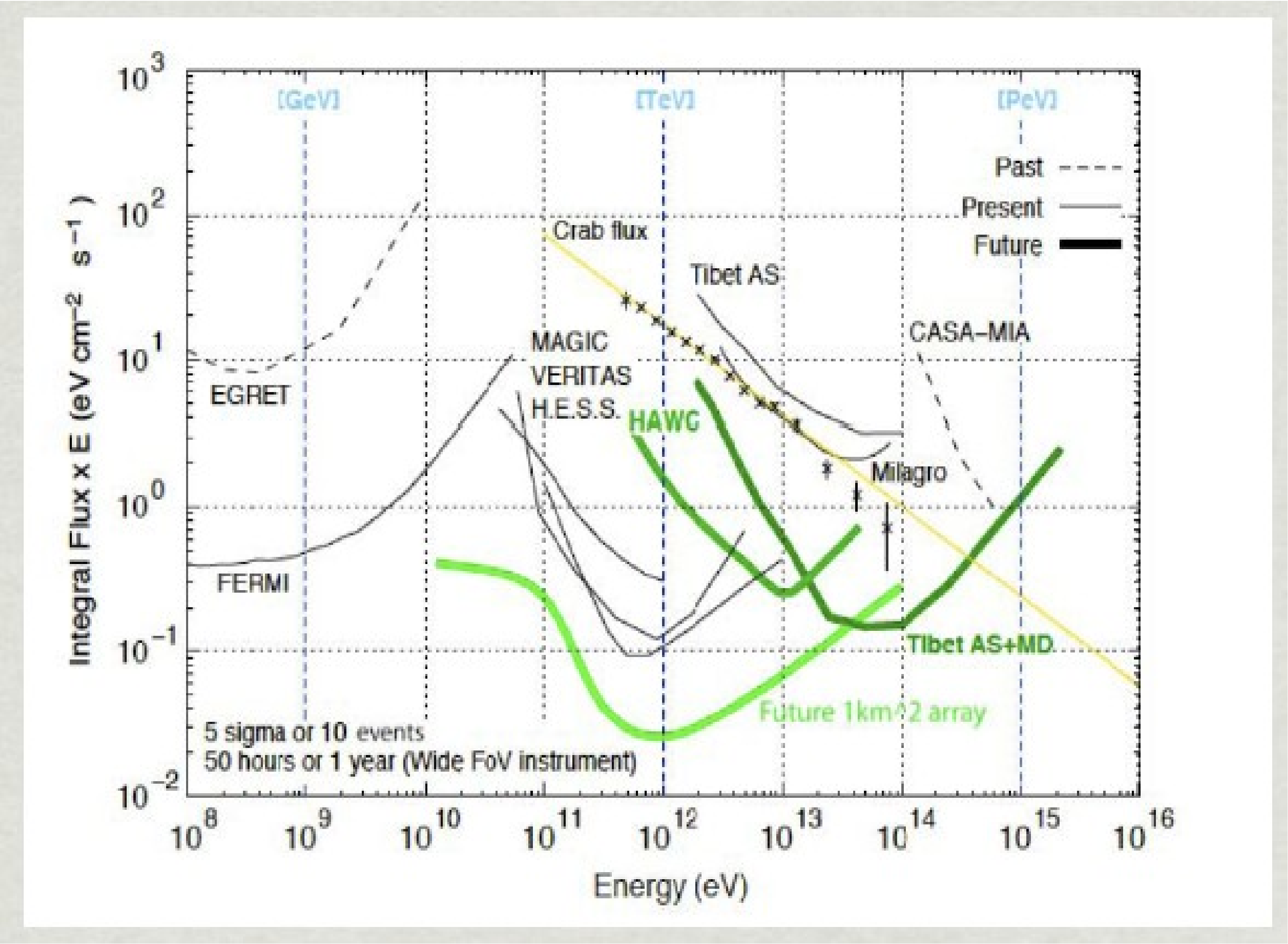
Unbiased Sky Survey

Extended sources and Cosmic Ray PeVatrons

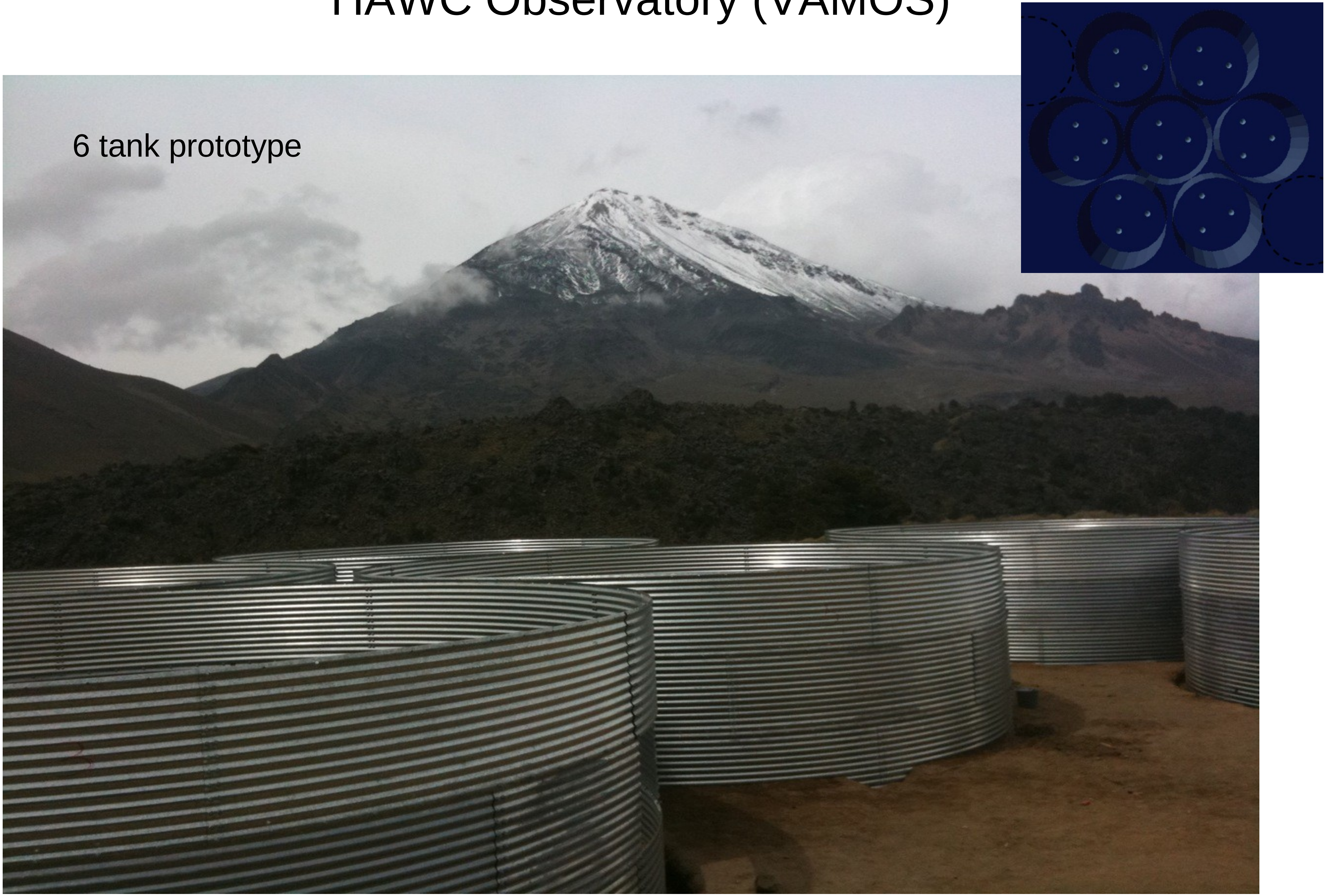
Transients (GRB's) > 30 GeV

Energies up to about 100 TeV

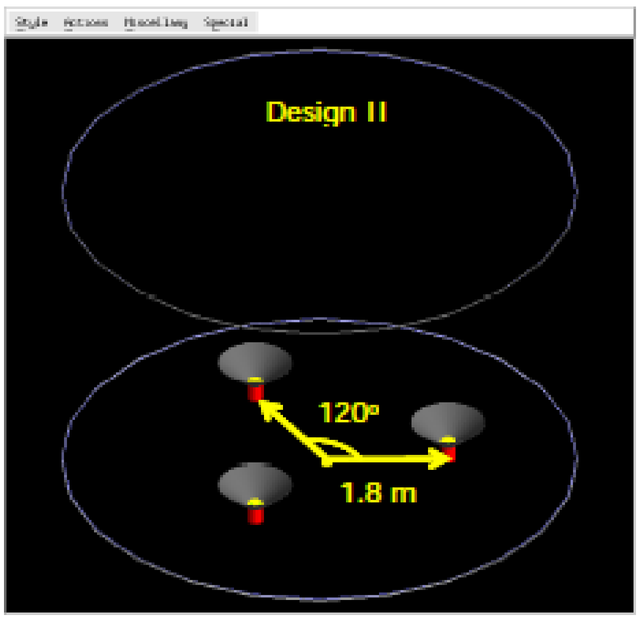
HAWC Observatory



HAWC Observatory (VAMOS)

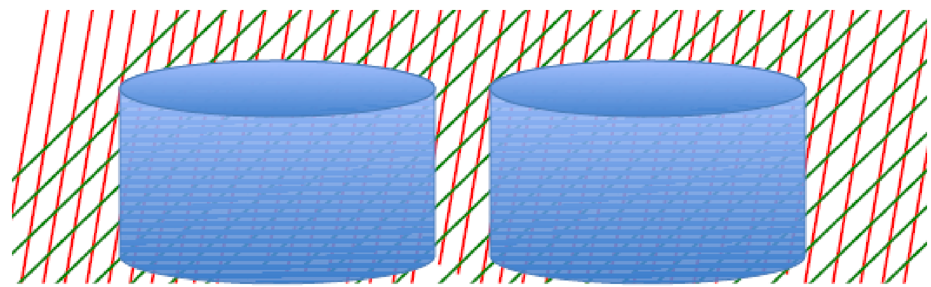


HAWC Tank design



7.3m in diameter

8m between the centers

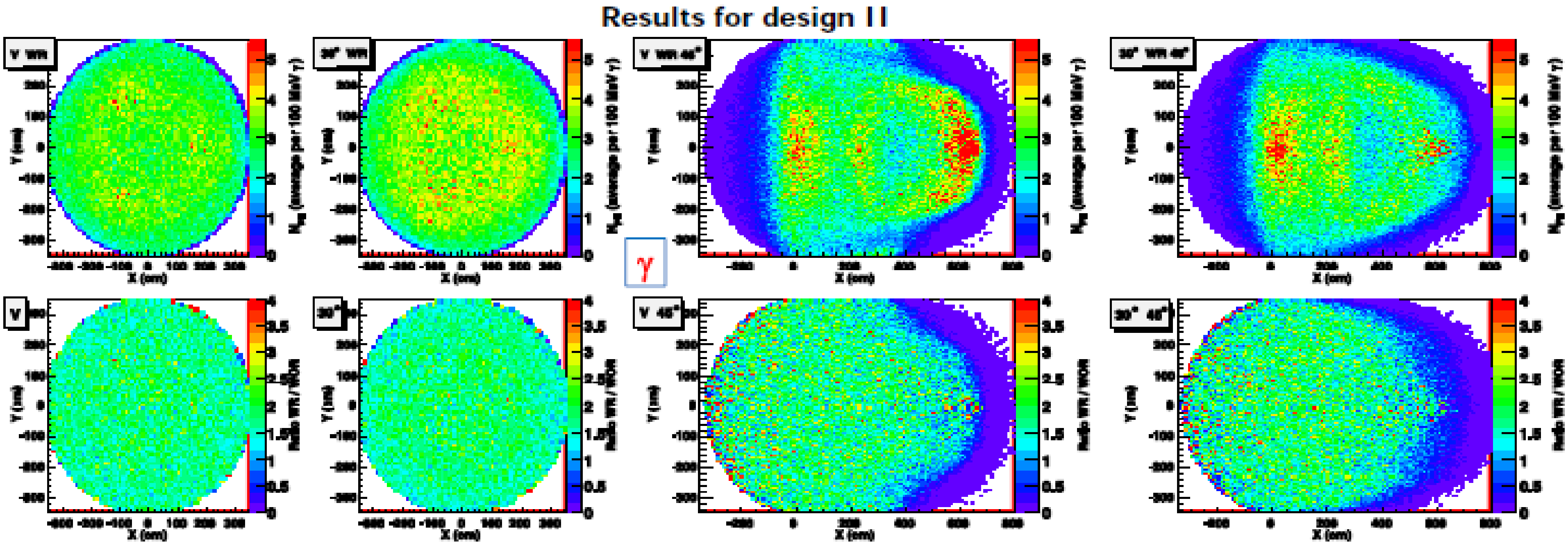


The dense replacement of tanks is to make the energy threshold as low as possible, though the effective area will reduced. The accent is on the gamma astronomy and not on the astroparticle physics.

Simulation Study to develop a new algorithm for the core position estimation (why?)

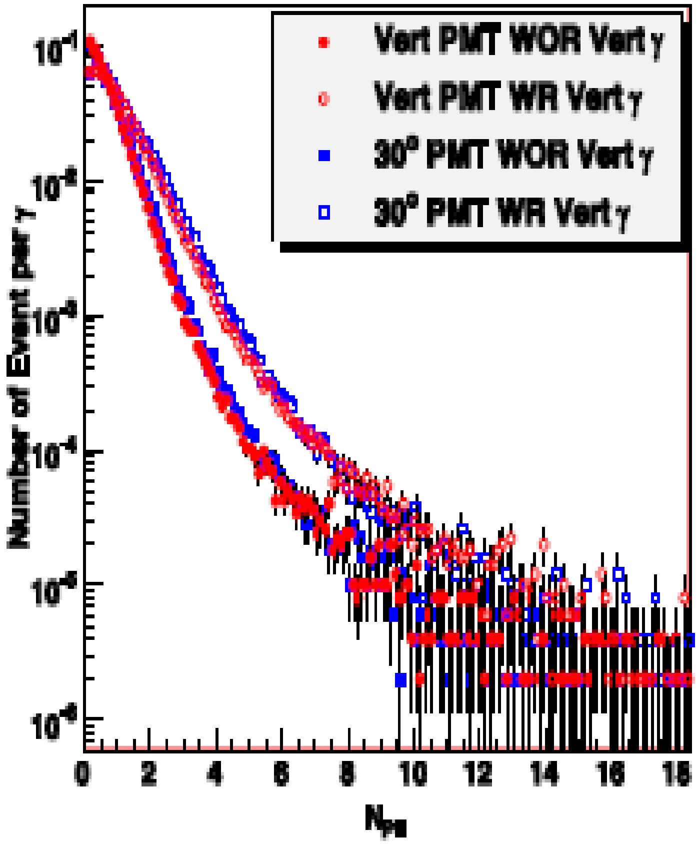
- Gamma and proton showers are simulated by CORSIKA in the energy region 50-100000GeV and for zenith angles 0-30 degree.
- The core position is drown uniformly in the area 500x500m2 having the same center as the detector, but this area is more than one order of magnitude larger than the detector area.
- Spectral index for gammas and for protons is 2.5 and 2.7 respectively
- The detector response is simulated using the parametrization of response distributions obtained from GEANT4 simulation results for the several fixed energy and angle values.
- In the data analysis the sum signal from 3 PMTs is used. The summation is performed over 3 PMTs if the signal is larger than the cut value (0.25PE).
- The data are considered useful if the number of fired tanks are larger than some values (30 is the minimum value that can be used for HAWC observatory). The limitation comes from the trigger rate (large trigger rates make DAQ expensive and difficult for a small collaboration like HAWC).

Tank response for 100MeV gammas ICRC 2009

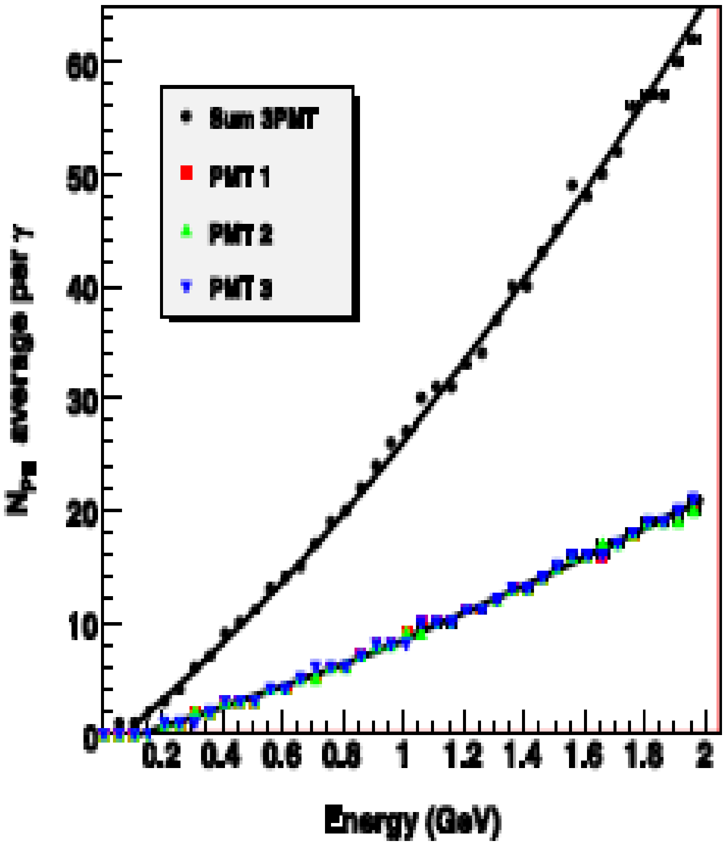


Vertical and 45 degree gamma incident
Uniform response only for the vertical gammas.
The effective area of the tank for non vertical gammas is larger than for the vertical ones

PMT response for 100MeV gammas

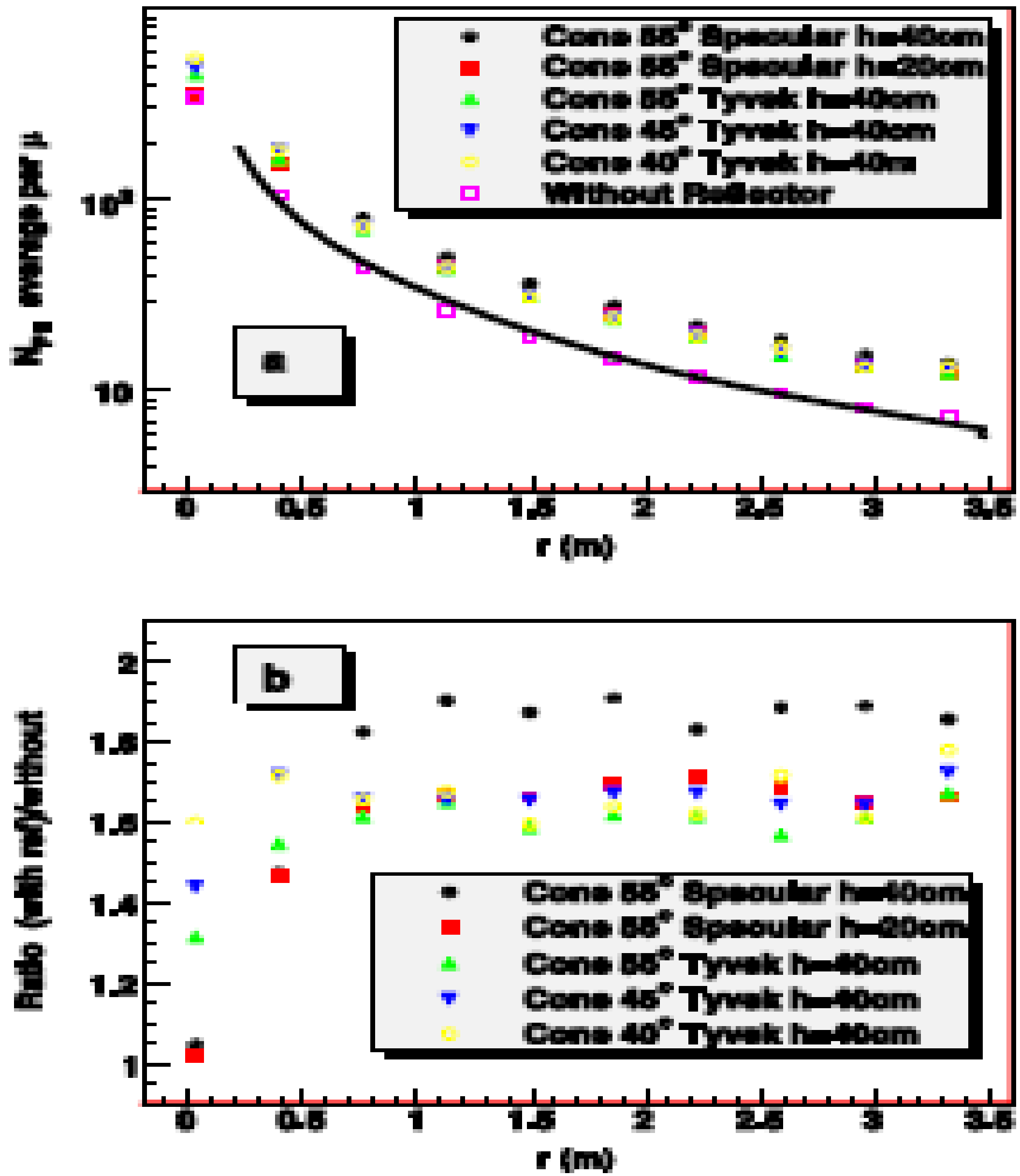


N_{PE} normalized distributions of single PMT for 100MeV γ.

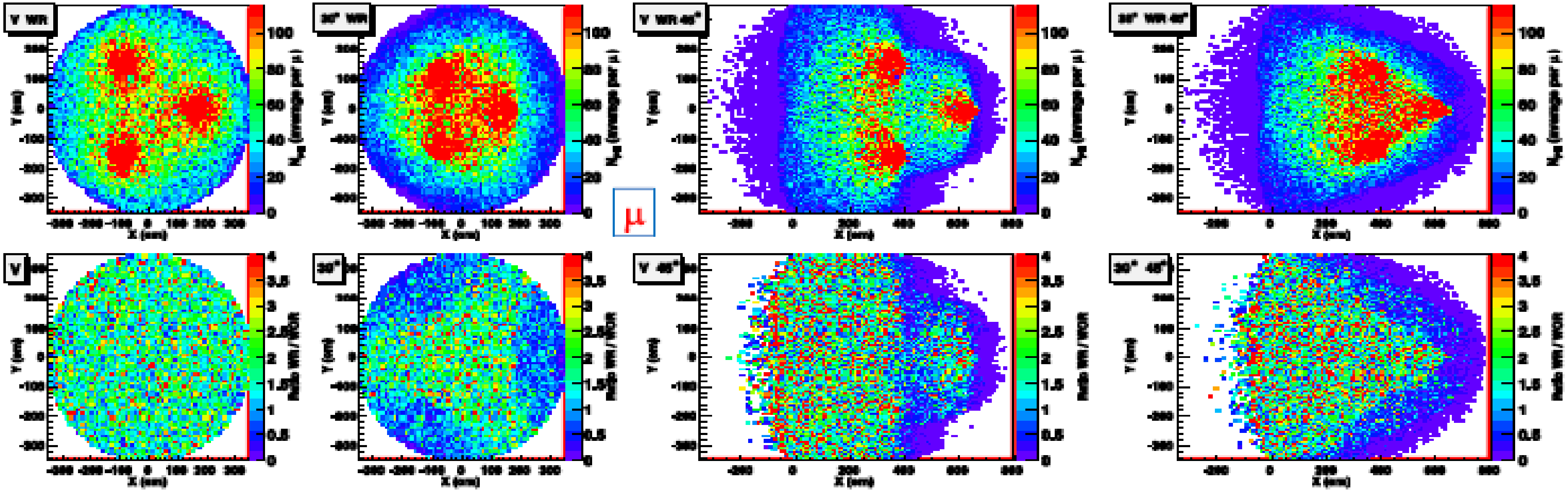


N_{PE} dependent on the energy of γ. Lines are polynomial fit results for single PMT and the sum signal of 3 PMTs

2 GeV vertical muon response: GEANT4 simulation results. Almost 1/r dependence from the PMT center



Coordinate dependence of vertical and 45 degree muons



Vertical and 45 degree muon incident
The effective area for non vertical muons is larger than for vertical ones

Algorithms for the core position estimation

- Centre mass (more simple and less acurate than the others.)
- Gaussian Fit (it is strange but seems it is usable: milagro)
- Tree (ARGO ICRC 2007) uses the shower space topology (the individual distance between fired detectors is also important)
- NKG Fit (usual for energies > 10 TeV and has a physical base). NKG parametrization is for the charge particle density. The usage of this parametrization for the water Cherenkov detectors is not obvious.
- Response smoothing by some function and then the use of the maximum as the core position (ARGO ICRC 2007)
- Non parametric pdf estimation (new) and core position estimation by the maximum of pdf. This algorithm has not been tried yet and it is expected to be successful because of similarity to the smoothing algorithm, but it has a strict mathematical base.

Non parametric pdf estimation

This method developed in this study is based on the usage of the non parametric probability density estimation using the measured lateral profile. For low energies (< 1TeV) due to large fluctuations we consider the usage Parzen probability density estimator, which provides better performance than the KNN.

For two dimensional case the Parzen pdf can be written:

$P(\mathbf{x}) = |S|^{-0.5}\Sigma_k w_k \exp(-d_k/2h^2)/(2\pi h)$,

where h is the Parzen kernel width, d_k is the distance in Mahalanolobis metric and defined as:

$d(\mathbf{x_1},\mathbf{x_2}) = \sqrt{((\mathbf{x_1}-\mathbf{x_2})^T S^{-1}(\mathbf{x_1}-\mathbf{x_2}))}$,

S is the covariance matrix and for non zero weights defined as:

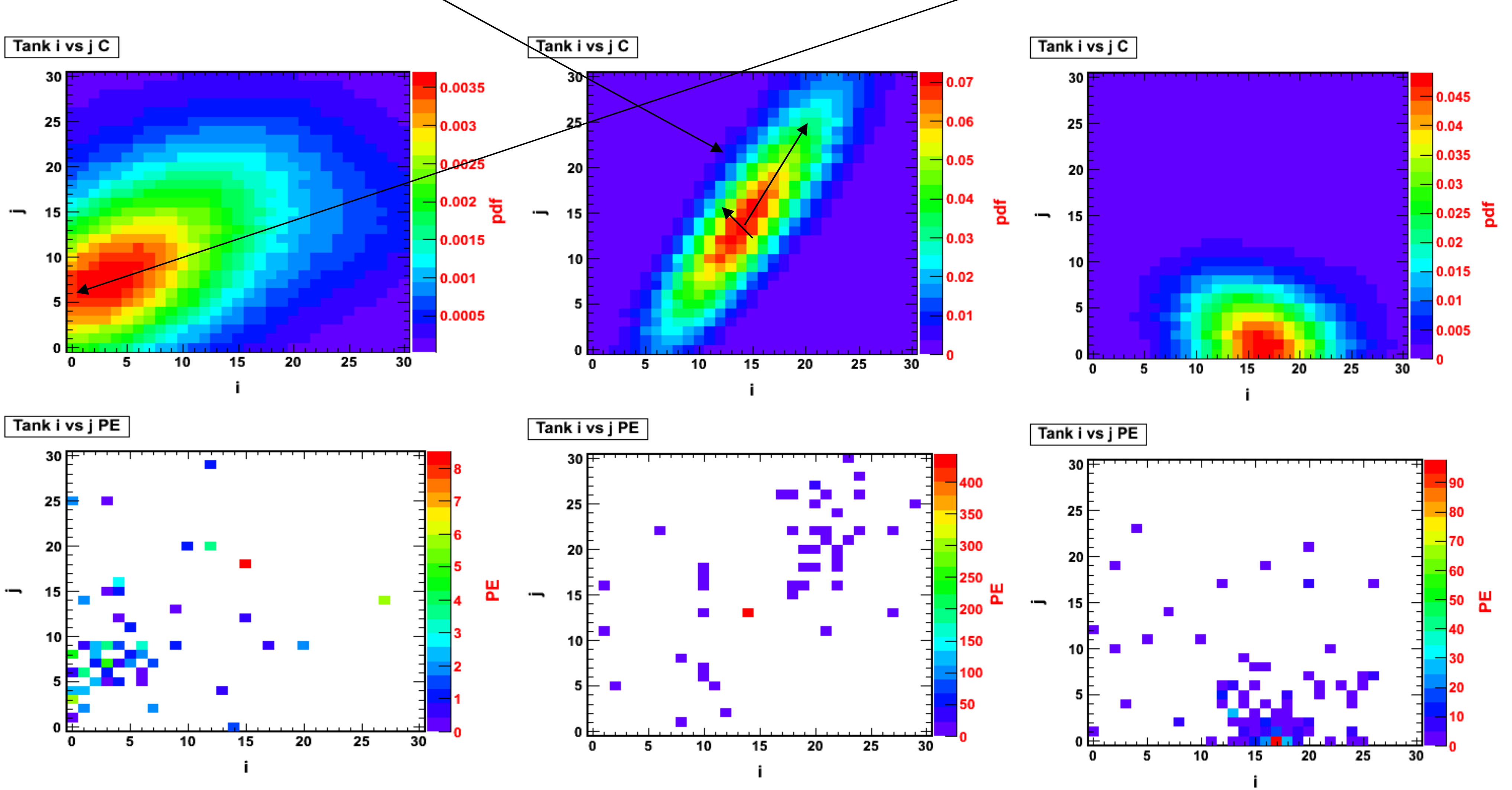
$S_{ij} = \Sigma_k(w_k(x_{ki}-\langle x_i\rangle)(x_{jk}-\langle x_j\rangle))/(1-\Sigma_k(w_k{}^2))$,

where $\langle x_i\rangle$ is the average value of x_i and w_k is the weights equal to- $PE_k/\Sigma_\iota PE_i$; PE_k is the tank response; k =1...N(number of fired tanks)

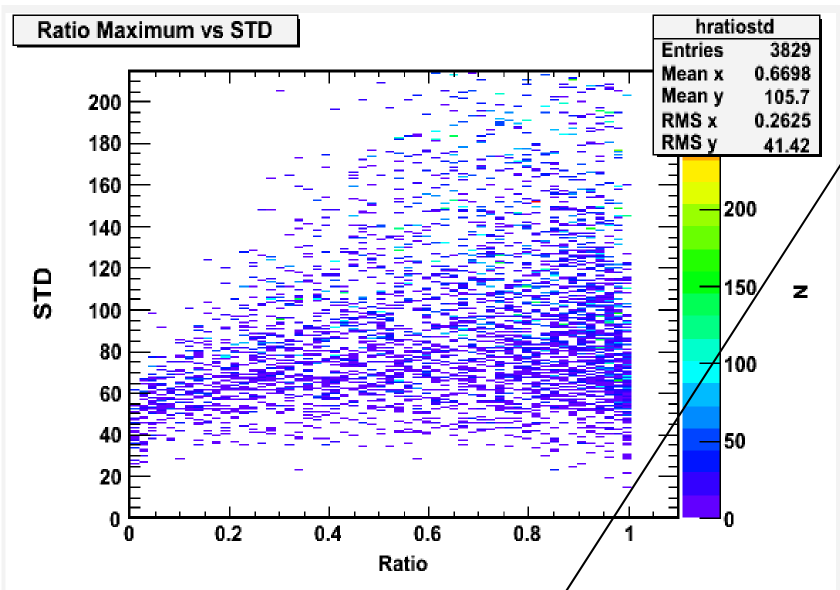
As in previous case this method includes a single parameter h to be optimized using simulation data. After reconstruction the pdf it is possible to construct an algorithm to separate showers having core position inside and outside the detector area using pdf's characteristics.

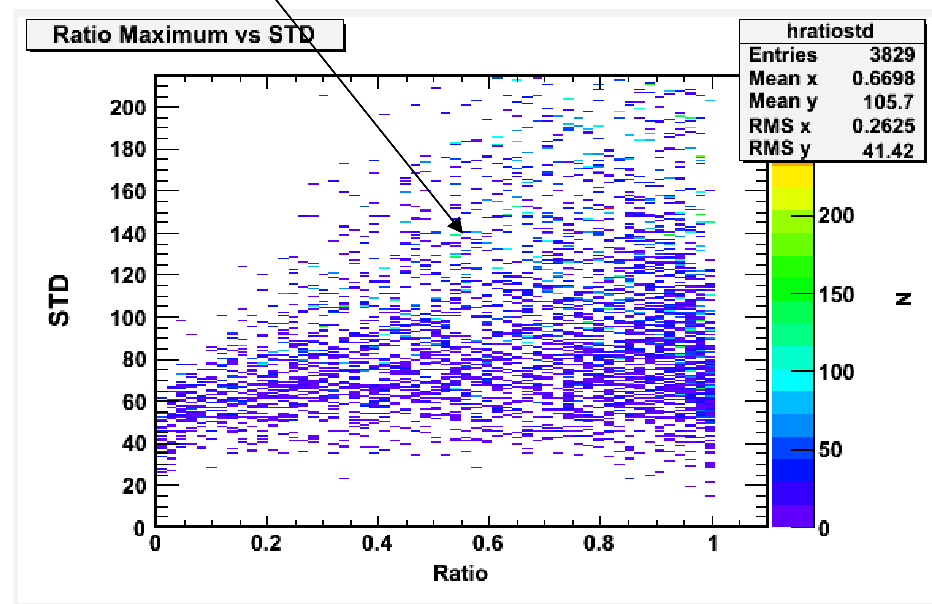
Detector response and corresponding pdf
two parameters to characterize the localization inside the detector

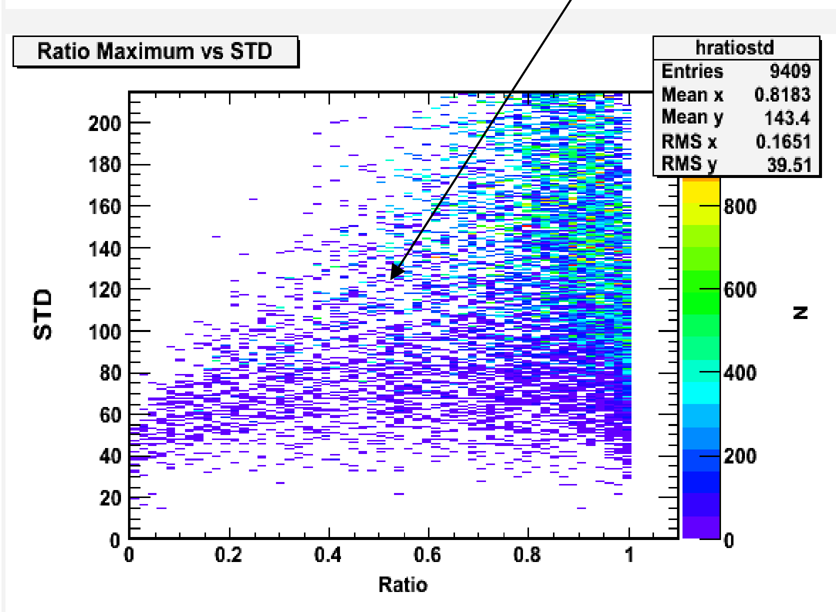
- The ratio of the maximum value at the edge to global maximum of pdf. (Ratio)
- Character size of the shower as the quadratic sum of ellipse semi-major and semi-minor axis's (STD)

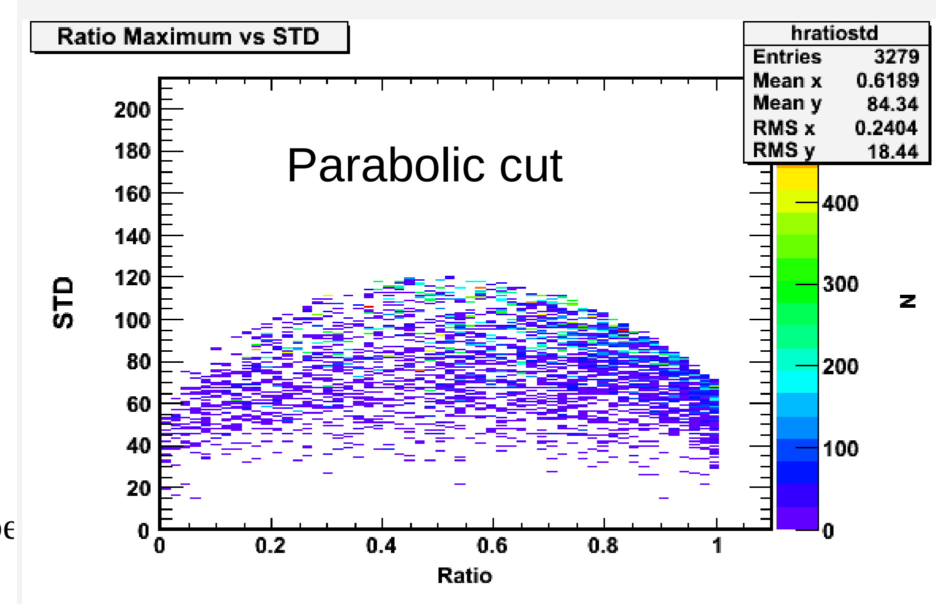


Two dimensional plots for the events inside the detector and inside an area 500x500m2

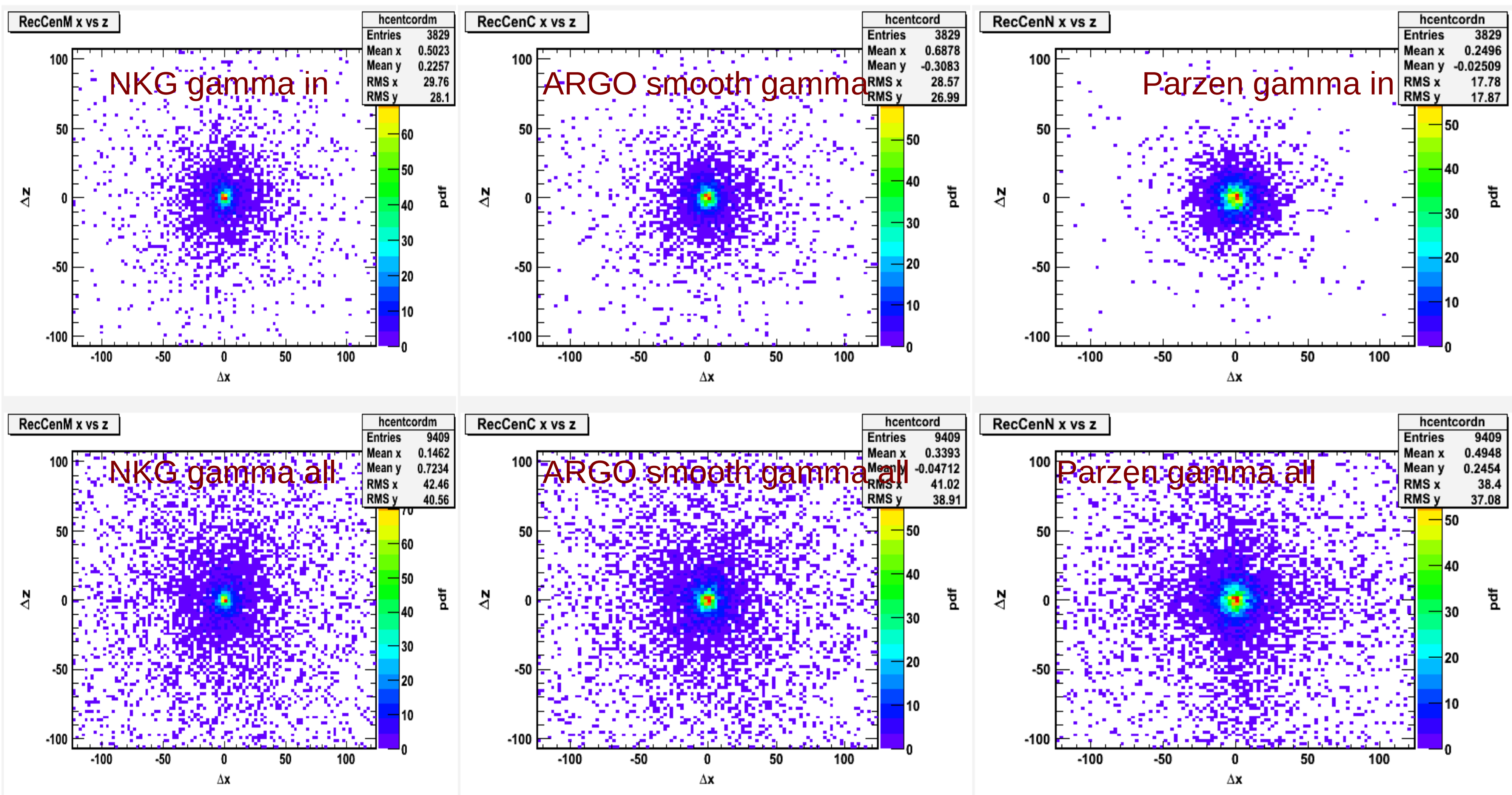




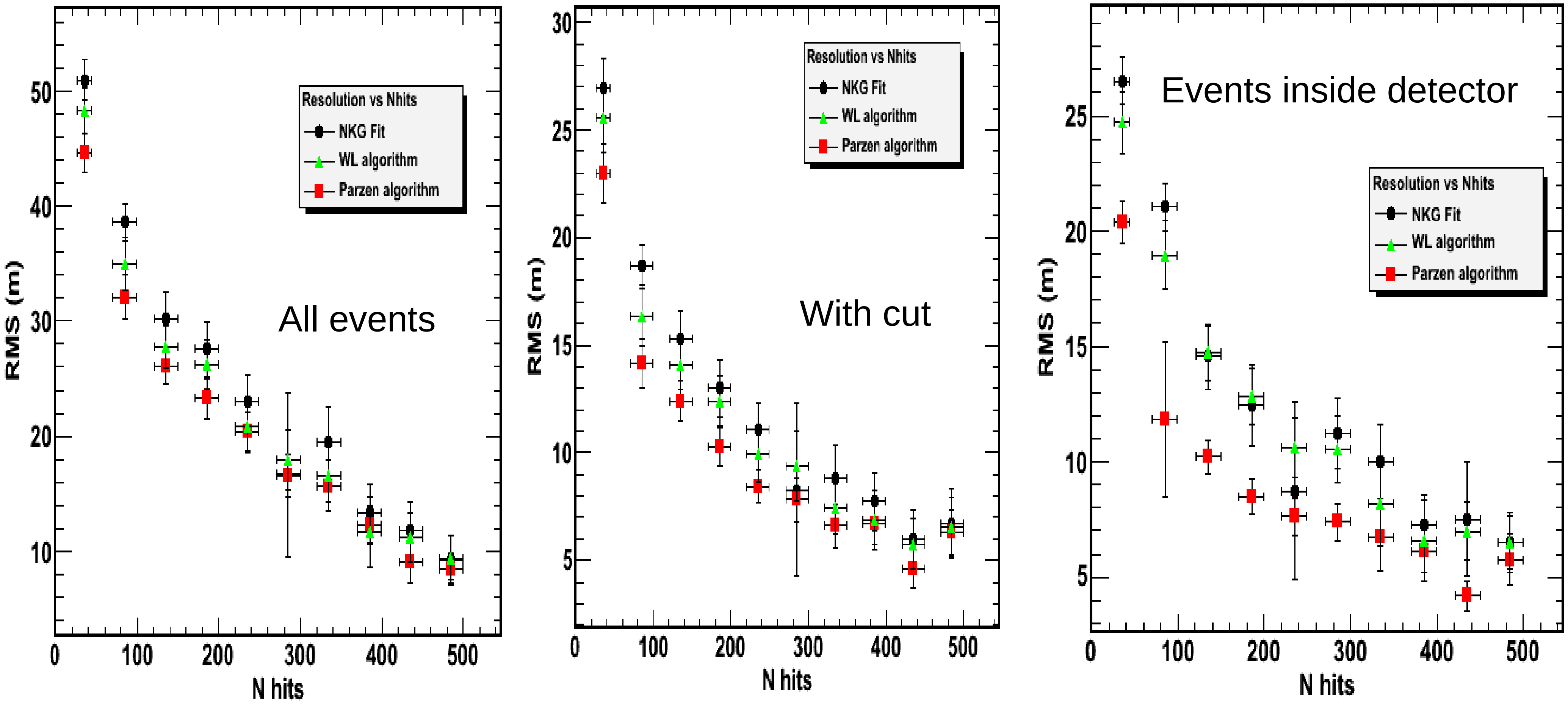




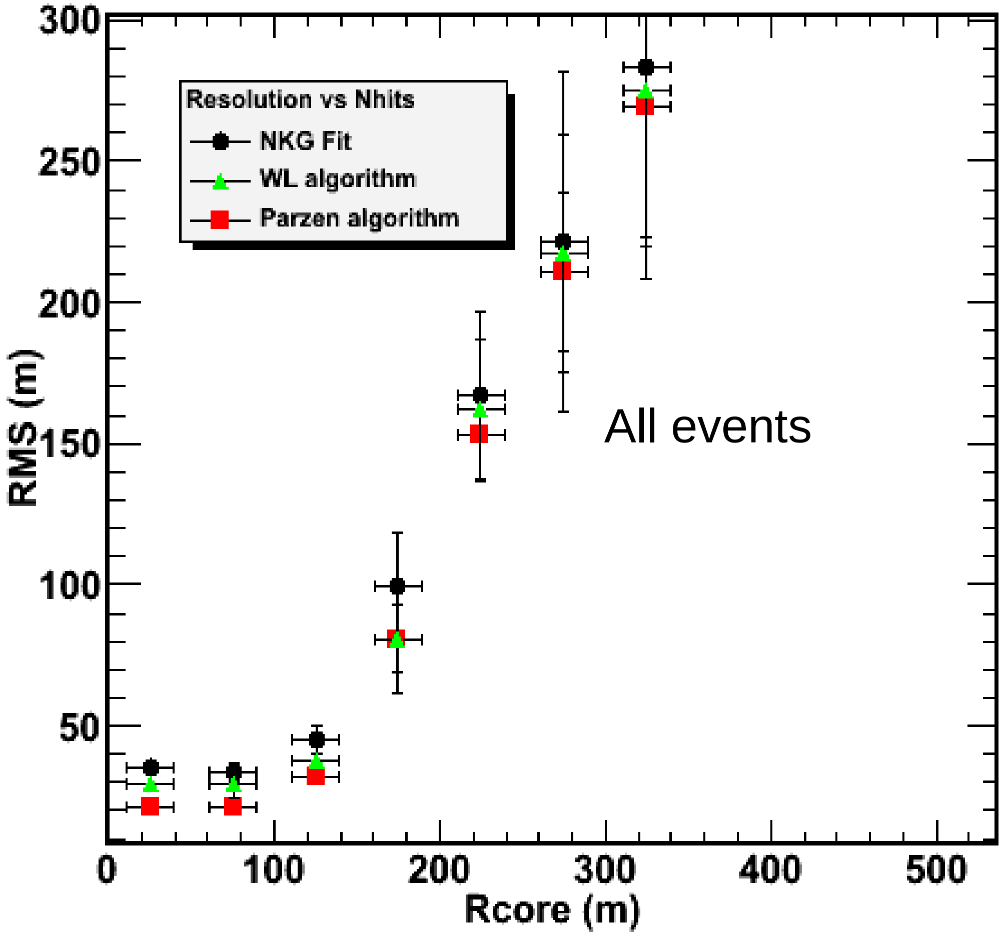
Two dimensional plots of differences of core position (original and reconstructed)

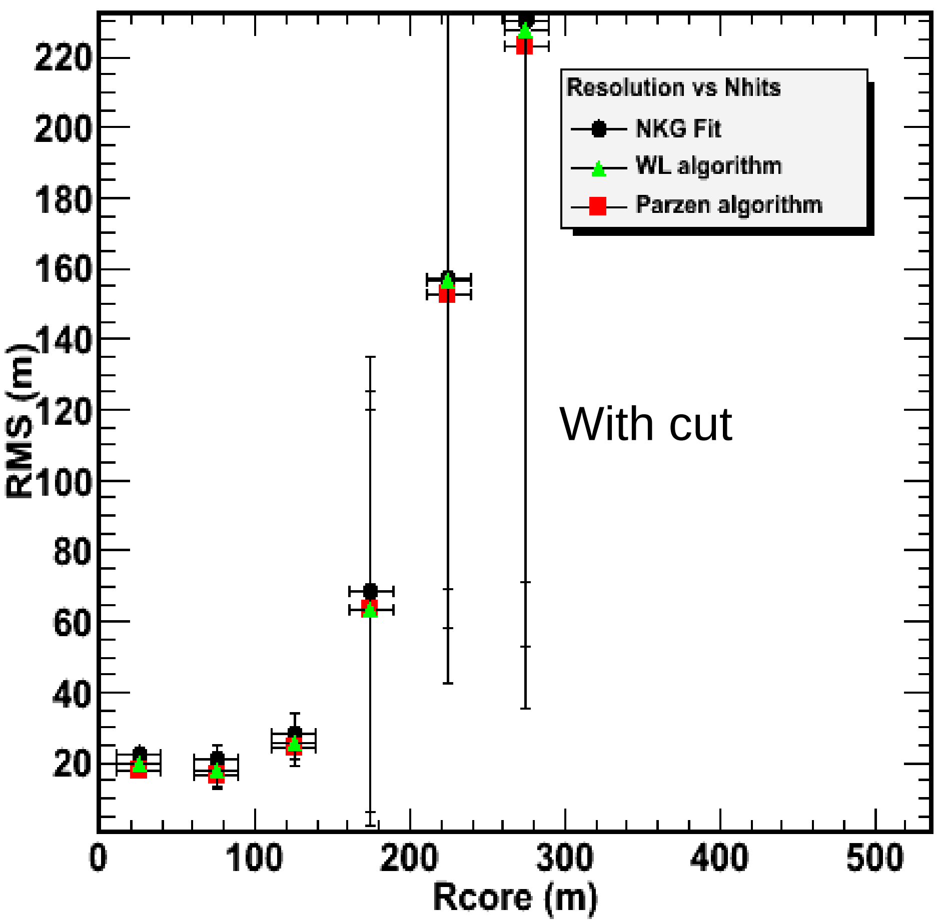


Core position resolution vs hit multiplicity



Core position resolution vs core position for ntrigger > 30





Conclusions

The new approach based on non parametric pdf, for the core position estimation is better than the existing algorithms

This method allows to develop an algorithm to separate core positions inside and outside the detector.

It has strict mathematical base and is good enough unless something like NKG will be found for the shower electromagnetic energy with large fluctuations (when shower tail is detected)

Acknowledgments

The author is grateful to Yerevan physics institute to spent a sabbatical year in YerPhi and PASPA DGAPA of UNAM for a partial financial support.