Romanian Reports in Physics, Vol. 62, Nos. 1, P. 27–36, 2010

FUZZY CLUSTERING ALGORITHM FOR GAMMA RAY TRACKING IN SEGMENTED DETECTORS

G. SULIMAN, D. BUCURESCU

"Horia Hulubei" National Institute for Physics and Nuclear Engineering(IFIN-HH), Magurele, Romania

(Received February 10, 2009)

Abstract. A tracking algorithm for reconstruction of the gamma-ray trajectories in a segmented detector system was developed as part of our collaboration in the AGATA project. The main characteristics and performances of the algorithm are presented.

Key words: gamma-ray; segmented germanium detectors; gamma-ray tracking.

1. INTRODUCTION

The need for better gamma-ray detector systems for nuclear physics made the scientific comunity to start building segmented HPGe detectors that, instead of rejecting the photons that do not have all their interaction points in the same detector, try to get the full information they are carrying through the reconstruction of their trajectory by using all individual interaction points in the detector. This principle allows for a gain of at least a factor of two in the solid angle that can be covered by the Ge detectors, since the BGO shields used in current systems can be removed.

In order that such a detector works, a few conditions must be fulfilled. The first one is to be able to find the "exact" interaction point and the amount of energy deposited in each such point. This condition is accomplished through the use of highly segmented detectors and pulse shape analysis. A second condition is to be able to use these interaction points to reconstruct the real trajectories of the photons in the detector.

This approach was tested to work for one segmented detector [1], and one large detector array based on this principle is already under construction in Europe: the Advanced GAmma Tracking Array (AGATA) [2, 3, 4, 5]. This detector array will be built out of 180 tapered large volume 36-fold segmented Ge crystals packed in 60 triple-clusters. It covers 78.26% of the solid angle and reproduces as closely as possible an empty germanium shell with inner radius of 23 cm and outer radius

of 32 cm. On-line pulse shape analysis will allow the determination of the interaction points of the photons to an accuracy of 4–6 mm [7]. Tracking algorithms reconstruct the trajectory of the photon inside the detector, leading to high efficiencies and peak-to-total ratios.

The purpose of this article is to present a tracking algorithm appropriate for systems of segmented detectors. First, we present the algorithm, with an emphasis on the method used to get an estimate of the number of photons that entered the detector, then procedures used to validate the obtained clusters of interaction points, and finally we show some performances of the algorithm.

2. ALGORITHM DESCRIPTION

A tracking algorithm for reconstructing the trajectory in a nearly 4π Ge ball such as the AGATA detector will have to be able to deal with events of high multiplicity (up to 30 or maybe even higher). It must be able to reconstruct photons with energy from as low as a few tens of keV and up to 10 MeV. In this energy range, the main effects that account for the photon interaction in the detector are the photoelectric absorbtion, the Compton scattering and, at higher energies, the pair generation.

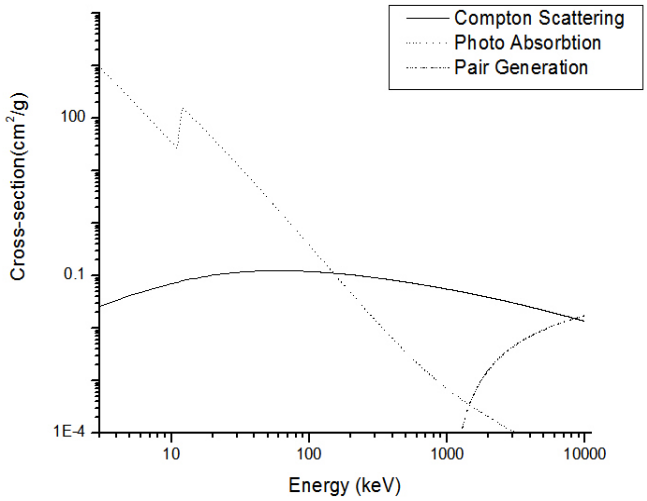

Fig. 1 – Cross section of the processes of the interaction of a γ-ray with a germanium detector.

When a 1 MeV photon leaves its entire energy in a germanium detector it has an average of 4 interactions. Assuming that all the interaction points are found with absolute precision, reconstructing the trajectory of the photon is mainly a problem of choosing one ordered combination of points from all the possible combinations. In a brute force approach, this would lead to trying out 4! combinations. The problem

becomes difficult when a higher number of photons hits the detector at the same time. Taking the example of an event in which 30 photons with an energy around 1 MeV each hit the detector, the number of such combinations rises to $(30 \times 4)!$. As a consequence any realistic tracking algorithm must incorporate ways to reduce the number of combinations that need to be verified.

The algorithm we have developed has two separate sections. The first section deals with the partitioning of the initial set of interaction points in smaller groups (that we call "clusters") while the second one tries to reconstruct a photon which passes through all the points inside each cluster, by using a forward tracking procedure.

The algorithm provides a list of possible photons, obtained from these clusters, and their associated "quality coefficients" that give us an image of how well the photon trajectory is reconstructed, allowing the rejection of badly reconstructed photons.

For the first part of the algorithm, we have chosen to use the Fuzzy C-Means algorithm [6]. This is an objective function based numerical algorithm suitable for identifying well separated groups of points in a multidimensional space. Because the length of the photon path in Germanium decreases with the decrease of energy we can assume that all the interaction points belonging to a certain photon will be grouped together in the 3D space.

The Fuzzy Clustering C-Means algorithm has a long standing history in the image recognition field [6]. It determines the best partition of points such that a certain function describing the partition is minimized. The ingredients of this algorithm are the points to be divided in groups, the number of groups, and a metric to be used as a distance function. The fuzziness of the algorithm lies in that each of the initial interaction points belongs to each of the groups to a different degree, called "membership degree". The algorithm is iterative, at each step a new set of positions for the group centroids and the new set of membership degrees being computed. The objective function of the algorithm is generally based on the least-squares formula, but including as weighting factor the degree of membership of the points:

$$J(f) = \sum_{x \in X} \sum_{k \in K} f^m(x)(k) \cdot d^2(x,k), \tag{1}$$

where by X we denote the points to be clusterized, by K the groups (clusters) that are to be constructed, $f(x)(k)$ is the membership degree of point x to group k (a number varying between 0 – non-membership and 1 – full membership), $d(x, k)$ is the distance from point x to cluster k and m a constant of the algorithm called "fuzzyfier" which modifies the involvement of the weighting membership degree in the final objective function. In our implementation of the algorithm the fuzzyfier was given the value $m = 2$, commonly used when there is no information about how separate the clusters are in 3D space.

It was demonstrated that such an iterative algorithm is convergent [6]. Also, in order to have a minimum objective function, the membership degree of a point to a certain group has to be:

$$f(x)(k) = \begin{cases} \dfrac{1}{\sum_{j \in K} \left(\dfrac{d^2(x,k)}{d^2(x,j)} \right)^{\frac{1}{m-1}}} & \text{if} \quad I_x = \emptyset \\ \sum_{i \in I_x} f(x)(k) = 1 & \text{if} \quad I_x \neq \emptyset, k \in I_x \\ 0 & \text{if} \quad I_x \neq \emptyset, k \notin I_x \end{cases}, \tag{2}$$

where $I_x = \{j \in K | d(x, j) = 0\}$.

The position of the clusters (groups) is computed using the formula

$$k_i = \frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\sum_{j=1}^{n} u_{i,j}^m}, \tag{3}$$

where $f(x_j)(k_i) = u_{i,j}$ denotes the membership degree of point x to cluster k, and k_i and x_j are the cartesian coordinates of the clusters and points, respectively.

The algorithm that we employed to obtain the groups of points from the initial pool of points is the following:

1. find the number of clusters;
2. initialize the cluster positions and the degree of membership of each point to each cluster;
3. determine the new positions for the cluster centers using the formula (3);
4. determine the new membership degree of each point to each cluster using formula (2). The distance function we used is the geometrical distance between the interaction point and the center of the cluster;
5. repeat from step (3) until a certain criterion is fulfilled. The criterion we used is to stop after a certain number of iterations.

After running the algorithm, we obtain the optimal positions of the centers of the clusters and the degree of membership of each point to each of the clusters. The next step in the code is the *defuzzyfication*, which makes the move to crisp clusters (one point belongs only to one cluster), which are then checked against the photon scattering formula. In the following sections we will make a short description of the parts of the algorithm that are not covered by this general description.

2.1. Finding the number of clusters

One basic caveat of the Fuzzy C-Means algorithm is that it must know from the beginning the number of clusters it needs to find. Put into the language of our track reconstruction problem, this means that we should supply the code with the number of photons interacting with the detector at each event. Because this number cannot be known *apriori*, we have built an independent procedure to get an estimate of the number of photons, starting only from the known interaction points.

The procedure to get an estimate of the number of clusters starts from the first interaction point in the list and marks all the interaction points that are closer to it than a certain distance $\sqrt{DMAX}$. As soon as all the points are checked against the first point in the list, the algorithm moves to marking all the points in the list that are close enough to the next unchecked point in the list, and the procedure goes on until all the interaction points are marked. This algorithm does not need a reordering of the list of interaction points in order to work.

The external parameter *DMAX* is one of the "free" parameters of the algorithm and its value was chosen such as to get optimal results. The optimization was caried out with events generated by using an AGATA event simulation code [10]. The number of clusters found by our algorithm for each event was checked against the known number of photons that interacted with the detector within that event. The results are shown in Fig. 2.

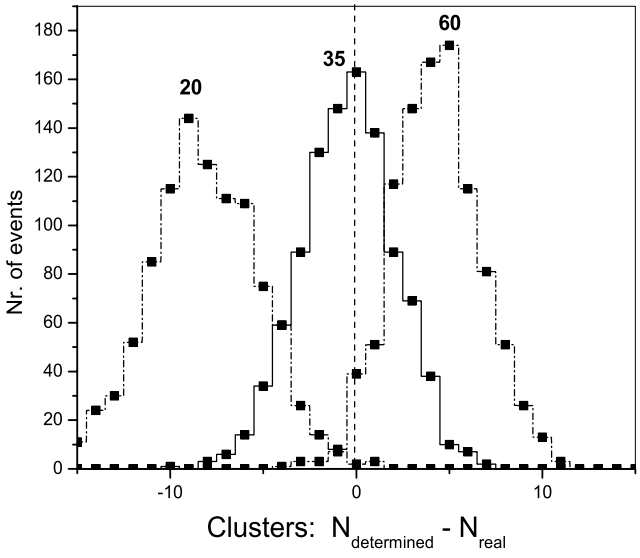

Fig. 2 – The difference between the real number of clusters and the estimated number of clusters for different values of the *DMAX* parameter (measured in cm^2), as indicated.

It can be seen (Fig. 2) that we can find a value of the distance parameter DMAX that makes the number of found clusters to be similar to the number of photons that interact with the detector. The *DMAX* value we used in calculations is 35 or slightly smaller (in order to account for the photons that interact with the detector in the same region of the detector, creating a single group of points from two interacting photons). This parameter value is characteristic for the given material of the detector (Germanium).

2.2. Defuzzyfication

We call defuzzyfication the process that makes the transition from a description of the problem in which each interaction point belongs to each cluster to a number of well defined clusters each having a well defined number of points. This part is necessary because we want that in the end each interaction point belongs to only one photon. In order to achieve this we apply a cut on the membership degree of each cluster. The value of the cut is chosen on the basis of the final membership degree of each point to the cluster. The average membership degree for each cluster is computed and all the interaction points with membership degrees to that cluster more than a certain fraction of this value are included in that crisp cluster. The advantage of this approach over a simple threshold comparison procedure is that it allows the clusters made of spread out interactions (with lower membership degree) to form a cluster.

2.3. Validation procedures

At this point in the algorithm we have reduced the problem of finding the right sequences of interaction points at a level that is feasible with the existing computing power. The collection of interaction points is divided among a number of clusters, each of them representing all the possible interaction points of one photon. While this already makes it possible to obtain basic information about the energies of the photons involved, there is still one step to be done towards revealing the true power of the segmented detector approach: finding the order in which the interactions occurred. This step is called validation of the clusters because in the process of finding the direction, some of the clusters are found not to contain a valid scattering sequence, and are therefore discarded, thus improving the peak-to-total ratio, although at the expense of some decrease of the efficiency.

The purpose of the validation procedure is to reconstruct the correct path of the photon inside the detector. By identifying the first interaction point (together with the knowledge about where the photon originated) the direction from which the photon came can be extracted, allowing in real experiments a very good Doppler correction, while knowing the first two interaction points opens the way

towards studying the polarization of the photon. At the heart of the validation procedure is a function that gives a numerical value to the degree of correctness of a certain scattering sequence. The Compton scattering formula allows the calculation of the scattering angle when the photon energy and the first energy deposition are known. By comparing this value to the angle between the points being considered, a quality factor can be computed that gives an estimate of the correctness of the scattering sequence.

We have developed two validation procedures, both aiming at the recovery of the trajectory of the photon among the interaction points in a cluster. Forward tracking is used as a tracking philosophy. This requires that the center of the detector is used as a starting point for the photons, improving the results of the algorithm [8], but at the same time it limits its use to experiments where the photon emission takes place in the center of the detector.

The first validation procedure is based on computing a quality factor for each possible sequence in the cluster in a similar way to [9] and then by deciding whether the cluster is valid by comparing this value to a predefined threshold.

The second validation procedure chooses as the first two interaction points of the photon the points that best comply with the compton scattering formula by using the center of the detector array and the two points, and then adds to the path the point that best continues the sequence. The quality value of the chain that is obtained in this way is the given by the worst fitting sequence in the chain, in a way similar to the idea that the strength of a chain is given by the strength of the weakest link.

Both procedures are completed with a test concerning the last found point in the sequence, which has to be a photoelectric absorbtion. The probability that this effect occurs during the last part of the interaction sequence in competition with all the other possible processes is taken into account.

3. RESULTS

The test ground for the performance of the code were data obtained from Monte-Carlo simulations of the AGATA detector [10]. The parameters that describe the performance of the code are those that will be used for the detector characterization: the peak-to-total ratio and the absolute efficiency in the case of single energy gammas, and the absolute efficiency in the case of multiple gamma events, such as, e.g., the decay of a rotation-like band.

The obtained results are highly dependent on the multiplicity of the event and on the precision with which the photon interaction positions are computed [12]. Two main factors influence this precision: the first one is the imprecision in the determination of the position of interaction through pulse shape analysis, so that the interactions will be in reality displaced from the real interaction points

(smearing); the second one is that it will be impossible to separate interactions that take place within a certain distance from each other. The reproduced interaction point will have the energy sum of the interactions and be placed in a position between the real interactions (packing). In our estimate we have only used a multiplicity of 30 (in each event 30 photons are shot at the same time into the detector) and both the packing and smearing distances were considered to be 5 mm, in line with the latest measurements on the AGATA prototype detectors [7].

Figures 3, 4 and 5 show the results obtained with the algorithm. Figures 3 and 4 demonstrate the performance in the almost theoretical case that only single energy gammas are entering the detector. This event configuration is the only one in which the peak-to-total ratio is unambiguously defined. The decrease in efficiency with the increasing energy is a result of both the decreasing intrinsic efficiency of the array with increasing energy, as well as the increasing importance of the pair generation effect, that is not included in the present code.

Table 1

Comparison of performances of the present code with those of other existing codes, for the detection of one 1.33 MeV photon

Algorithm	Efficiency (%)	Peak-to-Total (%)
MGT[9]	28	49
OFT[8]	24	55
Fuzzy C-Means	27	46
Deterministic Annealing[11]	26	48

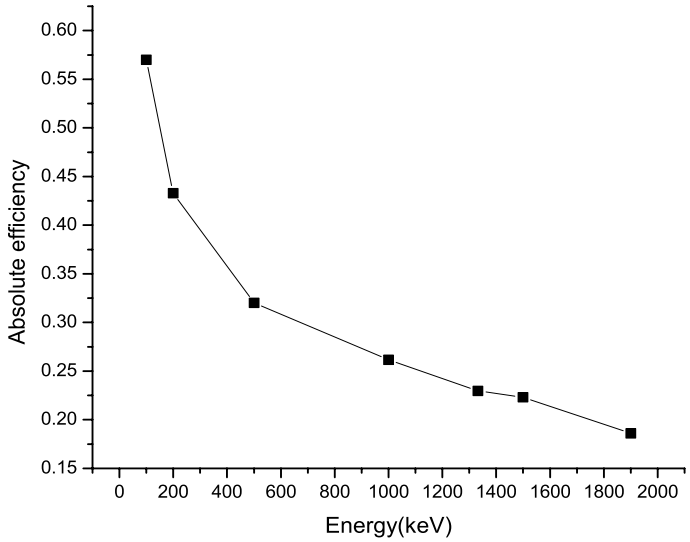

Fig. 3 – Absolute efficiency after tracking multiplicity 30 single energy (1.33 MeV) gamma-ray events. The packing and smearing distances for the simulated data are 5 mm.

A comparison with other tracking codes developed for the AGATA detector can be found in Table 1, concerning the absolute efficiency and peak-to-total ratio. It can be seen that the other existing codes give similar results for simulated data that were treated in the same way (using a 5 mm packing and smearing distance).

Figure 5 shows a comparison of the results obtained with our code and those obtained with the mgt code [9]. The algorithm was applied to an event consisting of 30 photons emitted in a rotational band. The first gamma-ray transition of the band was chosen at 80 keV, and the energy gap between two successive transitions was chosen to be 90 keV. This allows also the investigation of the sum peaks that appear in the spectrum, as they will differ in energy from the peaks in the band.

From the presented results it is clear that a segmented array brings an improvement of an order of magnitude to the absolute detection efficiencies compared to the detector arrays that are currently working. This improvement comes with a considerable increase of the complexity of operation and construction costs. As a result, the construction of the AGATA detector array will take place in stages, the first one, the AGATA demonstrator (consisting of 15 detectors, or 5 triple clusters) being scheduled to be fully operational in 2009. While all the software required for operating AGATA (pulse shape analysis codes, tracking codes, etc.) are now ready for use, tests of experimental data will be soon done on the AGATA demonstrator. As such, because all the codes were tested on simulated data, there is no definite conclusion on whether a tracking code is superior to the others, as the experimental data may differ from the simulated ones in subtle, unpredictable ways.

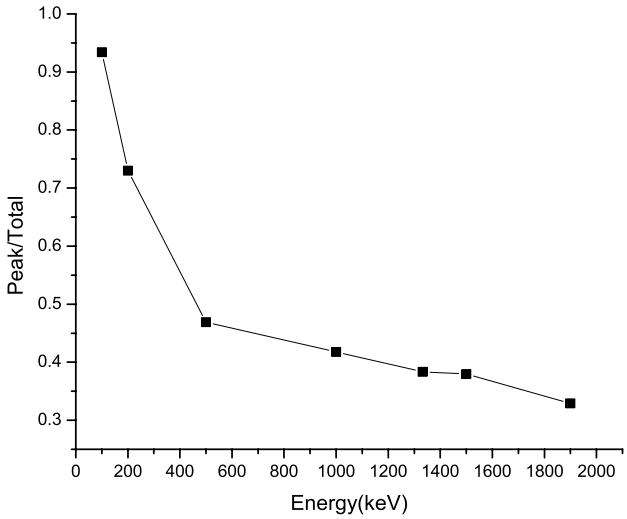

Fig. 4 – Peak-to-total ratio after tracking in the case of multiplicity 30 single energy (1.33 MeV) gamma-ray events. The packing and smearing distances are 5 mm.

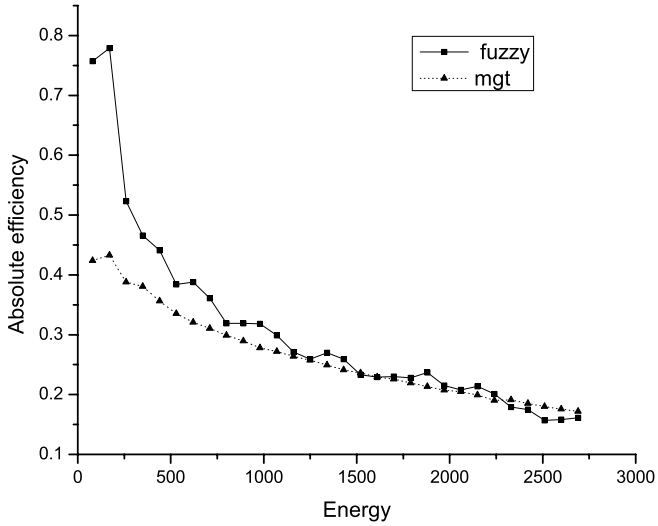

Fig. 5. – Absolute efficiency after tracking in the case of multiplicity 30 rotation-like band gamma-ray events. The packing and smearing distances are 5 mm. Comparison between the code and the mgt code of Padova University [9].

Another important aspect of the characterization of the code performance comes from the fact that the validation of the clusters embeds an important trade-off: by increasing the required quality for accepting the clusters the peak-to-total ratio is increased at the expense of decreasing efficiency. This opens the way for physics cases in which a better detection efficiency makes up for the decrease in peak-to-total ratio, and as such brings a new dimension to the way nuclear physics experiments are planned and carried out.

REFERENCES

1. Th. Kröll *et al*., Nucl. Instr. Meth., **A586**, 421 (2008).
2. J. Gerl and W. Korten (eds.), AGATA technical Proposal, 2001; wwww2k.gsi.de/agata and http://www.agata.org/reports/AGATA-Technical-Proposal.pdf.
3. J. Simpson and R. Krücken, Nuclear Physics, News, **13**, *4*, 15 (2003).
4. J. Simpson, J. Phys. G: Nuclear Physics, **31**, S1801 (2005).
5. J. Eberth and J. Simpson, Prog. Part. Nucl. Phys., **60**, 283 (2008).
6. F. Hoppner, *Fuzzy cluster analysis*, John Wiley & Sons, 2000.
7. F. Recchia, Acta Phys. Pol., **B38**, 1297 (2007).
8. A. Lopez-Martens *et al*., Nucl. Instr. Meth., **A533**, 454 (2004).
9. D. Bazzacco, LNL-INFN(REP), 2000.
10. E. Farnea, *et al*. LNL-INFN(REP) 202/2004, 158 (2004).
11. François Didierjean, 7th AGATA Week, Uppsala, 2008 (http://www.agata.org/sites/default/ files/ AW7/Talks/aw7-s03-didierjean.ppt).
12. J. van der Marel, B. Cedervall, Nucl. Instr. Meth., **A477**, 391 (2002).