

CLUSTER ANALYSIS OF HZE PARTICLE TRACKS AS APPLIED TO SPACE RADIOBIOLOGY PROBLEMS

M. Batmunkh^{a,b}, *L. Bayarchimeg*^{a,c}, *O. Lkhagva*^b, *O. Belov*^a

^a Joint Institute for Nuclear Research, Dubna

^b National University of Mongolia, Ulaanbaatar

^c Institute of Physics and Technology, Mongolian Academy of Sciences, Ulaanbaatar

A cluster analysis is performed of ionizations in tracks produced by the most abundant nuclei in the charge and energy spectra of the galactic cosmic rays. The frequency distribution of clusters is estimated for cluster sizes comparable to the DNA molecule at different packaging levels. For this purpose, an improved K-mean-based algorithm is suggested. This technique allows processing particle tracks containing a large number of ionization events without setting the number of clusters as an input parameter. Using this method, the ionization distribution pattern is analyzed depending on the cluster size and particle's linear energy transfer.

Выполнен кластерный анализ ионизаций в треках заряженных частиц, характеризующихся наибольшей частотой в зарядовом и энергетическом спектрах галактических космических лучей. Рассчитана частота распределения кластеров с размерами, соответствующими масштабам молекулы ДНК на различных уровнях компактизации. Для этой цели предложен усовершенствованный алгоритм, основанный на методе К-средних. Данный алгоритм позволяет обрабатывать треки, содержащие большое количество ионизаций, без предварительного ввода параметра, характеризующего общее число кластеров. С использованием предложенного метода характер распределения ионизаций исследован в зависимости от размера кластера и линейной передачи энергии частицы.

PACS: 87.53.Bn; 87.55.kd

INTRODUCTION

Clarifying the biological effect of high-charge, high-energy (HZE) atomic nuclei of space origin has been one of the main problems of space radiobiology in recent years. The intense study of the biological consequences of exposure to this space radiation component is mainly determined by tasks connected with planning long-term manned space missions outside the Earth's magnetosphere. The measurements of space radiation spectra and numerous Earth-based radiobiological experiments showed that HZE particles are a major health risk factor limiting the maximum mission length. The charge spectrum of the galactic cosmic rays (GCR) is usually broken down into three main groups, namely, the carbon group ($Z = 5-8$), magnesium group ($Z = 10-15$), and iron group ($Z = 25-30$). Protons and alpha particles ($Z = 1$ and $Z = 2$, respectively) represent a special GCR fraction, which is usually considered separately from the HZE nuclei. The fraction constituted by the nuclei with Z higher than two is relatively small. On the average, GCR consist of 85% protons, 14% alpha particles, and 1% heavier nuclei [1]. The most probable energy of GCR particles is 0.5–1.0 GeV/nucleon. Despite the small portion of HZE particles, their penetration ability,

as well as biological effectiveness, is quite significant. Due to their specific pattern of energy deposition, heavy ions cause a unique kind of DNA lesion called clustered damage. These lesions can lead to potentially harmful effects such as cell transformation, induction of chromosome aberrations, apoptosis, etc. The abundant induction of clustered DNA lesions makes HZE nuclei qualitatively different from X or gamma rays.

The severity of HZE exposure strongly depends on particles' ability to form so-called ionization clusters in biological targets. From the radiobiological point of view, the most important clusters have sizes comparable to the DNA molecule at different packaging levels. These clusters are of major interest for investigation due to their ability to induce clustered DNA lesions, which are the hardest to repair.

The comparative analysis of charged particles' ability to form ionization clusters requires using special techniques that can estimate the order of the developing clusters. Here, «order» means the number of ionizations within the area occupied by a cluster. One of the widely used techniques for clustering data processing is the K-means method [2, 3]. The application of this method for ionization density analysis in clusters comparable to the DNA molecule size is well demonstrated in [4] for low- Z particles. However, the implementation of K-means for HZE atomic nuclei of GCR requires modification of this algorithm. A standard representation of this technique implies finding a user-specified number of clusters (k) represented by their centroids. Due to the necessity of setting the number of clusters each time, this approach is not well suitable for the comparative analysis of large data sets representing ionization coordinates for different HZE particles. Another peculiarity which should be taken into account is the necessity to fix the size of the clusters according to a chosen biological target. In this paper, we present an improved K-means clustering algorithm which meets the requirements mentioned above. This method is applied to several most abundant GCR nuclei.

1. METHOD

1.1. Track Structure Simulation. Cluster analysis implies two main subtasks. Except the development of a cluster analysis algorithm itself, it also includes the simulation of the particles' track structure. To obtain a track structure of the most abundant GCR particles, we used Monte Carlo-based simulation technique implemented in the Geant4-DNA code, which was developed for modeling early biological lesions induced by ionizing radiation at the DNA scale [5, 6]. The physics of the atomic interaction taken into account includes ionization, excitation, and charge transfer for protons and alpha particles, and only ionization for ions heavier than helium [7, 8]. Here, a $1 \mu\text{m}^3$ cubic volume of liquid water is used for the simulation of all particles' tracks. The length of each track segment is set at $1 \mu\text{m}$.

1.2. Cluster Analysis Technique. We improved the standard K-means methods suggested by Späth [9] and Hartigan [10], which are widely used for the analysis of clustering. These methods imply a strong dependence of the final cluster position on the initial coordinates of the cluster center. The initial coordinates of the cluster centers for the large data sets are chosen randomly or by standard initial partitioning [9]. Such calculations lead to a locally poor solution and become time-consuming for large data sets. In our study, we generate initial cluster centers and find the global minimum. This is a better approach as it yields more precise final positions of clusters. It also does not require the number of clusters as an input parameter for calculations.

The other steps of the algorithm are identical to the ones in [4]. The virtual center of a cluster is calculated as

$$\bar{x}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} x_i. \tag{1}$$

It is represented as a three-dimensional coordinate vector. Here, m_j is the number of ionizations in the C_j cluster, and x_i is the Cartesian coordinate of the i th ionization. To cut down the process of cluster growing, we introduce the cluster size p and the following condition to be satisfied:

$$|x_m - x_n| \leq p. \tag{2}$$

In Eq. (2), x_m and x_n are the coordinates of any two ionizations belonging to the same cluster C_j . In this model approach, each cluster could be enfolded by a sphere of the diameter p . The value of this parameter is defined according to the size of a chosen biological target.

When all track ionizations are divided into spherical clusters of a given size p and the coordinates of these ionizations are defined, we can estimate the frequency distribution of the clusters. This distribution represents the number of clusters containing j ionizations. Here, we consider j as the cluster order. The frequency distribution depends on the cluster size p and the mean free path λ_i of a charged particle in the medium. Figure 1 shows the scheme of the improved clustering algorithm compared to the one developed in [4]. An example of the improved algorithm applied to a 1 GeV/nucleon iron ion track in liquid water is shown in Fig. 2.

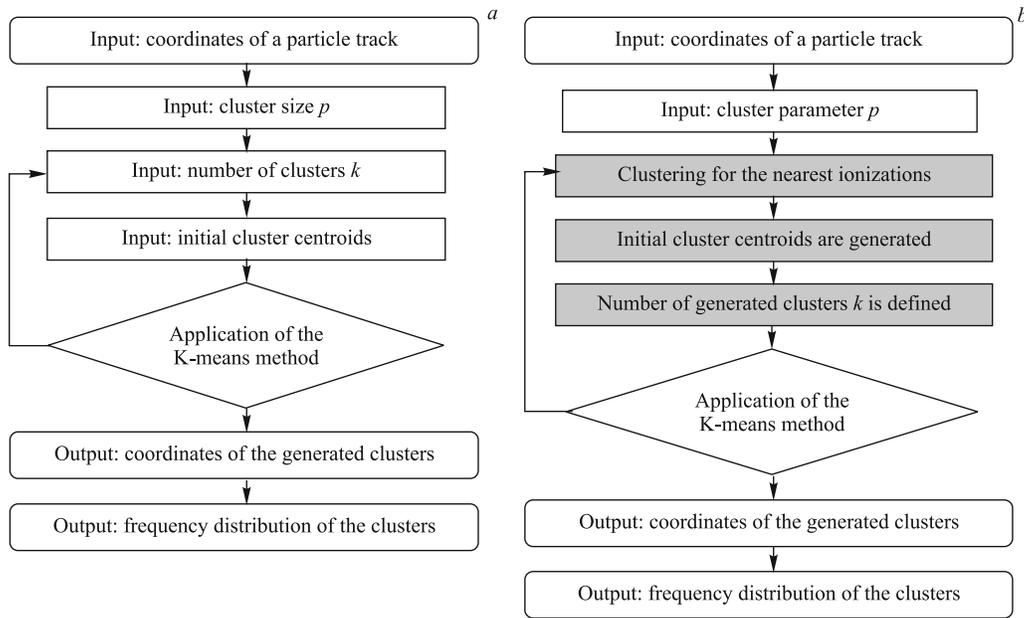


Fig. 1. The scheme of the initial clustering algorithm proposed in [4] (a) and the improved one (b)

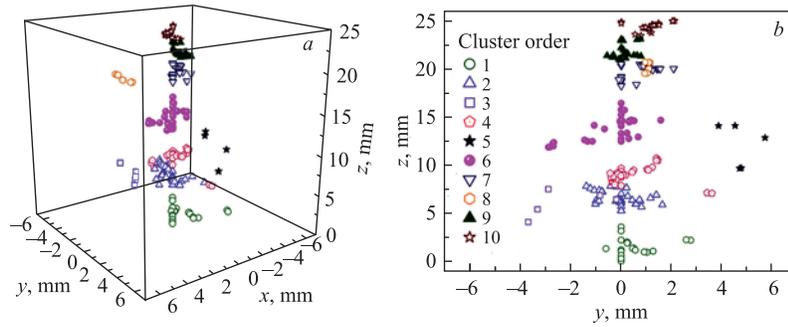


Fig. 2. An example of three (a) and two (b) dimensional representation of ionization clusters formed in a 1 GeV/nucleon ^{56}Fe ion track in liquid water. The total number of ionizations is 280; the number of the produced clusters is 10; the cluster size p is 6 nm

2. RESULTS

2.1. Particle Track Structure Simulation. To estimate the ability of space radiation to form ionization clusters in a biological medium, we simulated tracks of several most abundant GCR particles. We chose protons, alpha particles, and carbon and iron ions with energies of

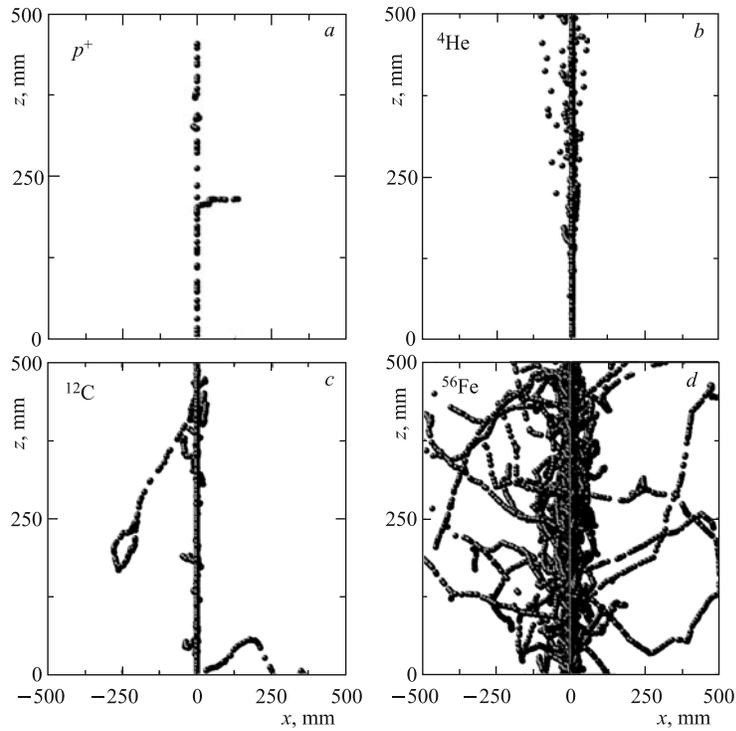


Fig. 3. Two-dimensional projections of particle tracks in liquid water simulated by the Geant4-DNA code. The energy of protons is 0.1 GeV (a); the energy of ^4He (b), ^{12}C (c), and ^{56}Fe (d) is 0.3 GeV/nucleon

0.1–1 GeV/nucleon. Examples of two-dimensional projections of tracks of these particles are shown in Fig. 3. The simulation results are represented as the Cartesian coordinates of the produced ionizations. The coordinates of each ionization correspond to an interaction with a nonzero energy deposition. In our calculations, the initial energy of ^4He , ^{12}C , and ^{56}Fe ranged from 0.1 to 1 GeV/nucleon. This energy corresponds to the maximum of the GCR differential flux for the chosen particles. As Geant4-DNA code has limitations to simulate high-energy proton tracks, we set the maximal possible value for protons at 0.1 GeV. We also introduced a 100 eV energy production cut-off for elastic electron scattering.

2.2. Frequency Distribution Estimation. The mean frequency distribution of clusters is computed for protons, alpha particles, and carbon and iron ions using at least 100 tracks per each particle. Figure 4 shows the calculated frequency distribution for the considered particles of an energy of 0.1 GeV/nucleon. The data is obtained for the cluster sizes of 2, 4, and 6 nm. In fact, these results represent the probability of the appearance of clusters containing certain number of ionizations. Thus, for particles of the same energy we can conclude that increasing the projectile's charge reduces the probability of low-order cluster induction. This effect becomes stronger when the cluster size increases.

In our study, the cluster sizes were set according to the DNA molecule size at different packaging levels. The 2 nm clusters correspond to the diameter of the native double-stranded DNA. The 4 nm clusters are close to the effective diameter of DNA taking into account the

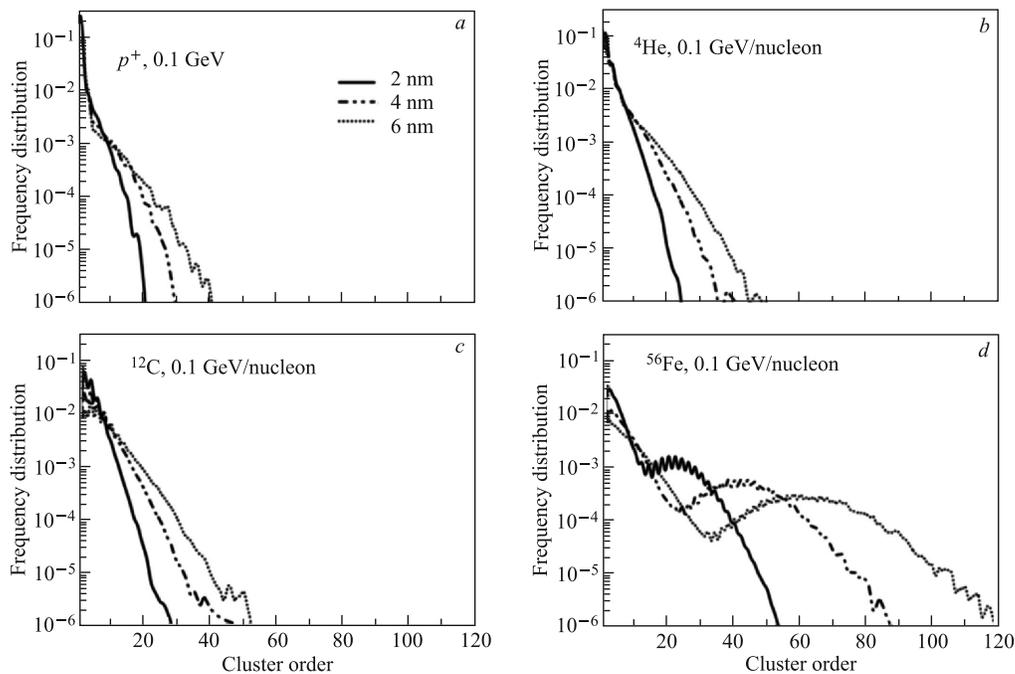


Fig. 4. The frequency distributions of clusters calculated for protons (a), ^4He (b), ^{12}C (c), and ^{56}Fe (d) with an energy of 0.1 GeV/nucleon. The LET values of the particles are, respectively, 0.73, 2.9, 26.2, and 489.68 keV/ μm . The calculations are performed for the cluster sizes of 2, 4, and 6 nm

area of free radical production. The size of 6 nm is comparable to the height of two turns of a DNA filament when it winds around the histone protein core.

The calculations performed for other energies revealed the dependence of the frequency distribution on linear energy transfer (LET) (Fig. 5). The shape of the distribution curves becomes more complicated as LET increases. In particular, a bend appears in the distribution curve (Figs. 4, *d* and 5, *c*). According to our calculations, the formation of this bend starts from the LET values close to 230 keV/ μm . It also strongly depends on the cluster size. For larger clusters, the bend is observed at lower LET. For example, the curve for 0.3 GeV/nucleon ^{56}Fe ions (LET = 237 keV/ μm) has an appreciable bend when the cluster size is 6 nm, while 2 nm clusters show no such changes, and 4 nm clusters have just a weak bend. The higher LET values yield a significant bend for 2 nm clusters as well (Fig. 4, *d*).

As the calculated frequency distribution reflects ionization density in particle tracks, the obtained results show how this density changes depending on the ion charge, energy, and LET. It is shown that high-LET particles demonstrate a completely different pattern of ionization density. A specific fraction of clusters is observed, which has intermediate values of the cluster order. This fraction corresponds to a curve dip obtained for all the considered cluster sizes at high LET values (Fig. 4, *d*). The location of this dip in the curve and the second peak for higher cluster orders depends on the cluster size. We defined the ranges of the cluster order values corresponding to the curve dips. As the results for 0.1 GeV/nucleon ^{56}Fe ions

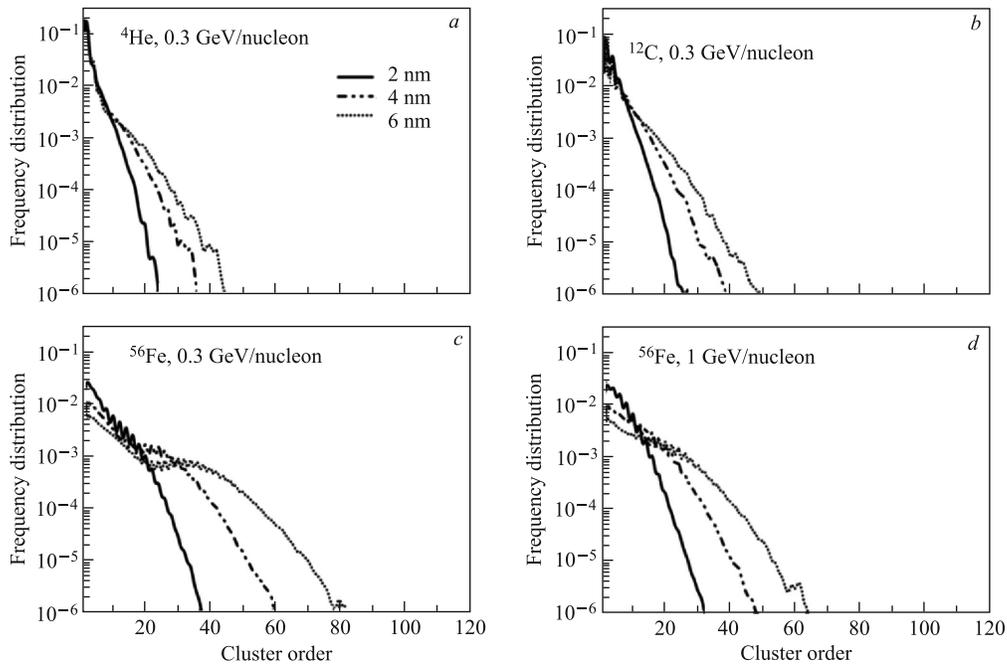


Fig. 5. The frequency distributions of clusters calculated for 0.3 GeV/nucleon ^4He (*a*), ^{12}C (*b*), ^{56}Fe (*c*) and 1 GeV/nucleon ^{56}Fe (*d*). The LET values of the 0.3 GeV/nucleon particles are, respectively, 1.41, 12.65, and 237.45 keV/ μm . For 1 GeV/nucleon ^{56}Fe , LET is 145.24 keV/ μm . The calculations are performed for the cluster sizes of 2, 4, and 6 nm

are the most representative, we show estimation for this case only. For $p = 2$ nm, the cluster order of the intermediate fraction is ranged from 10 to 20 ionizations per cluster; if $p = 4$ nm, this fraction includes clusters of the 16–40th order; the 6 nm clusters correspond to the cluster order of 22–57.

CONCLUSION

Studying the ability of charged particles to form ionization clusters in a medium plays a central role in the estimation of the damaging capacity of different radiations. The clustering of ionization events in biological targets is responsible for the induction of clustered DNA lesions of different complexity, which are the most harmful and the hardest to repair.

Many problems of space radiobiology require clarifying the specifics of GCR particle interactions with biological targets. This paper presents a comparative analysis of clustering patterns demonstrated by several most abundant GCR nuclei. It is shown that the frequency distribution of clusters differs depending on particles' physical characteristics and the cluster size. Much attention is paid to the analysis of the distribution at different LET. It is demonstrated that the distribution curve pattern changes significantly when LET reaches the value of about 230 keV/ μ m. The curve shape alterations become more visible with further increasing LET. This indicates that the distribution of clusters by their ionization density is not even for high LET. For each considered cluster size, a special portion of clusters is identified with intermediate ionization density. The ranges of cluster order values for these portions are defined for different cluster sizes. These intervals also depend on LET and the cluster size.

In this paper, we offer an approach to analyzing ionization density taking into account the specifics of energy deposition in sensitive biological structures. This method is focused exclusively on ionization events occurring at the early stages of the radiation damage process. Here, we are not concerned with the free radical species that form later as a consequence of water radiolysis and play an important role in radiation damage to DNA. Nevertheless, we suppose that the free radical species distribution should strongly depend on the ionization event distribution — at least at the early stages of track time evolution.

Acknowledgements. The study was conducted as part of the research project between the Laboratory of Radiation Biology of the Joint Institute for Nuclear Research (JINR) and the National University of Mongolia. The authors would like to thank Dr.Sc. O. Chuluunbaatar (the Laboratory of Information Technologies, JINR) for his suggestions on writing the program code.

REFERENCES

1. *Wiebel B.* Chemical Composition in High Energy Cosmic Rays. Univ. of Wuppertal. Report WUB 94-08. 1994.
2. *Steinhaus H.* Sur la division des corps materiels en parties // Bull. Acad. Polon. Sci. 1956. V. 4, C1 III. P. 801–804.
3. *Lloyd S.* Least Square Quantization in PCM's // IEEE Trans. Inf. Theory. 1982. V. IT-28. P. 129–137.
4. *Michalik V.* Particle Track Structure and Its Correlation with Radiobiological Endpoint // Phys. Med. Biol. 1991. V. 36. P. 1001–1012.

5. *Agostinelli S. et al.* Geant4 — a Simulation Toolkit // Nucl. Instr. Meth. A. 2003. V.506. P.250–303.
6. *Allison J. et al.* Geant4 Developments and Applications // IEEE Trans. Nucl. Sci. 2006. V.53. P.270–278.
7. *Incerti S.* Comparison of Geant4 Very Low Energy Cross Section Models with Experimental Data in Water // Med. Phys. 2010. V.37. P.4692–4708.
8. *Francis Z.* Stopping Power and Ranges of Electrons, Protons and Alpha Particles in Liquid Water Using the Geant4-DNA Package // Nucl. Instr. Meth. B. 2011. V.269. P.2307–2311.
9. *Späth H.* Cluster Analysis Algorithms for Data Reduction and Classification of Objects: Engl. transl. New York: Halsted Press, 1980. 226 p.
10. *Hartigan J.A., Wong M.A.* A K-means Clustering Algorithm // Applied Statistics. 1979. V.28. P.100–108.

Received on March 29, 2013.