A PATTERN RECOGNITION APPROACH TO LEARNING TRACKS OF HEAVY-ION PARTICLES IN TIMEPIX DETECTORS

A Dissertation

Presented to

the Faculty of the Department of Computer Science

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Son Hoang

December, 2013

A PATTERN RECOGNITION APPROACH TO LEARNING TRACKS OF HEAVY-ION PARTICLES IN TIMEPIX DETECTORS

Son Hoang

APPROVED:

Dr. Ricardo Vilalta, Chairman
Department of Computer Science, UH

Dr. Lawrence Pinsky
Department of Physics, UH

Dr. Shishir Shah
Department of Computer Science, UH

Dr. Lennart Johnsson
Department of Computer Science, UH

Dr. Nikolaos Tsekos
Department of Computer Science, UH

Dean, College of Natural Sciences and Mathematics

ACKNOWLEDGEMENTS

I would like to thank my wife Xuan Anh Ma for her dedication and patience at all times. I also want to thank my parents and sisters who have given me their constant support and encouragement in all my endeavors.

I would like to express my special gratefulness to my principal advisor Professor Ricardo Vilalta for the help, guidance, and support he provided for the completion of this dissertation. I am extremely fortunate to have Professor Lawrence Pinsky as my second advisor. His knowledge and expertise in the field has helped me successfully achieve the goal for each phase of the project.

Many thanks to the contributions of Medipix collaboration, HIMAC, NASA Johnson Space Center; and the following people in particular: M. Kroupa, N. Stoffle, J. Idarraga, J. Jakubek, A. Empl, S. Pospisil, Z. Vykydal, D. Turecek, E. Semones, K. Lee, N. Zapp, A. Bahadori, H. Kitamura, N. Yasuda, Y. Uchihori for their substantial contribution in this research project.

I would also like to acknowledge the financial support from VEF and NASA; the academic and technical support of Department of Computer Science, University of Houston, the facility support of Brookhaven National Laboratory.

Last but not least, sincere appreciation goes to fellow graduate students in Dr. Vilalta's group for helpful discussions and great friendship; my colleagues and friends in Vietnam, United States and elsewhere for their support and encouragement throughout.

A PATTERN RECOGNITION APPROACH TO LEARNING TRACKS OF HEAVY-ION PARTICLES IN TIMEPIX DETECTORS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Computer Science

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Son Hoang

December, 2013

ABSTRACT

The rapid development in semiconductor detector technology at CERN has provided the capability to develop an active personal dosimeter for use in space radiation environments. The work reported here is based on the Timepix chip, which when coupled with an Si sensor, can function as an active nuclear emulsion, allowing the visualization of the individual tracks created as the different incident particles traverse the detector. The Timepix chip provides the capability of measuring the energy deposited by each incident particle that traverses the sensor layer. Together with the capability for online readout, this detector opens the door to a completely new generation of active Space Radiation Dosimeters.

Although recent advances in hardware technology promise a major step forward in the development of such active portable space radiation dosimeters, little effort has been devoted toward software tools for analysis and classification of sources of radiation. Coupling radiation dosimeter hardware with pattern recognition techniques and machine learning tools has the potential to greatly improve current applications on space dosimeter projects. Our focus is not only to measure dosimetric endpoints directly such as dose-equivalent, but also to determine the physical nature of the radiation field with sufficient precision to allow characterization of the radiation composition and energy spectrum.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

A PATTERN-RECOGNITION APPROACH TO LEARNING TRACKS OF HEAVY-ION PARTICLES IN TIMEPIX DETECTORS

1. Introduction

The future of human space missions is constrained due to long-term space radiation effects. The attempt to acquire an accurate knowledge of radiation fields, and a fast-warning mechanism of radiation hazards is the major motivation in developing a Space Radiation Dosimeter. Current standard devices used for radiation monitoring are not portable. They provide insufficient information about the radiation composition and energy spectrum. Our approach tries to take advantage of the rapid development in semiconductor detector technology at CERN. We are using a state-of-the-art pixel detector known as the Timepix, which can function as an active nuclear emulsion. The device allows the visualization of the individual tracks left as charged particles traverse the detector. In addition, this detector provides the capability to measure the energy deposited per pixel by these particles. Together with a capability of online readout, this detector opens doors to a completely new generation of Space Radiation Dosimeters.

A fundamental challenge in the effort to develop such dosimeters based on the current Timepix chips from the Medipix2 Collaboration is how dosimetric endpoints such as dose-equivalent along with the detailed characteristics of penetrating heavy-ion particles can be determined from the raw Timepix outputs. The measurement of Linear Energy

Transfer (LET), source and velocity of incident ionizing radiation are of initial interest because they provide the capability to calculate the Dose-equivalent and characterize the radiation field for future radiation protection purposes. In this section, we briefly introduce the hardware technology of Timepix detectors. We then describe a pattern recognition system that plays an important role in developing such a dosimeter. We also discuss specific applications of the pattern recognition system applied to data output from Timepix and other related work.

1.1 Timepix device

Timepix is a version of the hybrid pixel detector technology developed by the Medipix2 Collaboration [1]. The individual pixel circuitry of the Timepix chip is composed of a charge-sensitive preamplifier whose output is viewed by a digital discriminator whose threshold can be externally applied (Figure 1.1). The input charge-sensitive preamplifier is discharged by a constant current source producing an approximately triangular-shaped output voltage vs. time curve. The constant current source is also globally settable over a range of values. There is also a separate 4-bit Digital to Analog Converter (DAC) Threshold Offset that may be set individually for each pixel to establish a uniform discriminator noise response across the entire device. Figure 1.2 displays a standard Medipix 2 functional block diagram of the circuitry within each pixel[1].

A Si sensor detector layer is typically coupled to the Timepix for charged-particle dosimetry applications. For thermal neutron detection, a neutron conversion layer such as ^{6}LiF is supplied on top of the Si in order to convert the neutron kinetic energy into that

of a charged particle. In principle, one can distinguish a charged particle generated from captured neutron with the one of genuine penetrating charged particle by an analysis of the distribution of charge collected by the pixels.

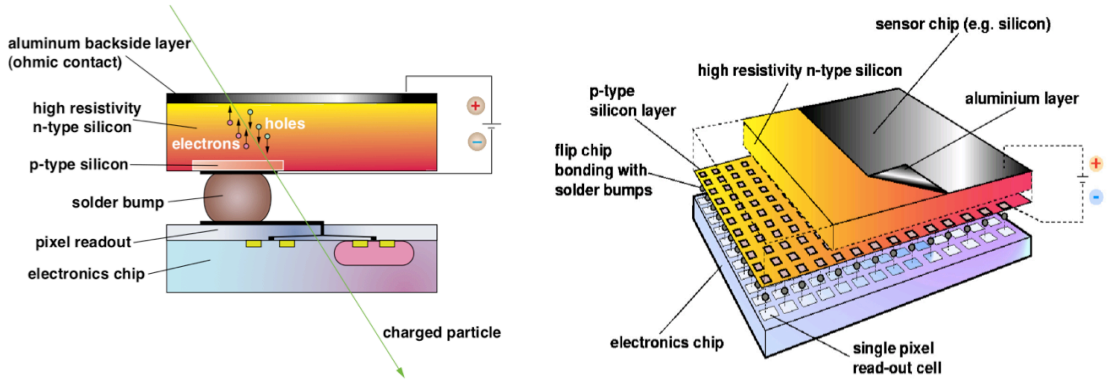


Figure 1.1. Medipix 2 configuration

The output of the discriminator is input to a logic-block that is also gated by an external global shutter signal. Additionally, the Logic-Block has access to a common externally applied counter-clock signal that can be as set as high as 100 MHz. The output of the Logic-Block is input to a 14-bit pseudo-random shift-register counter with an overflow bit. The shift-registers are read out in 256 pixel columns. The Logic Block also contains a 2-bit mode selection capability that may be used to individually program the function of the Logic-Block for each pixel separately. There are 4 possible modes for the Logic-Block to function in: Medipix - counting mode, Time-Over-Threshold (TOT) – energy sensitive mode, TimePix - time of arrival mode, and One-Hit - logical mode[2].

[1,2] These sections are excerpted from [28]

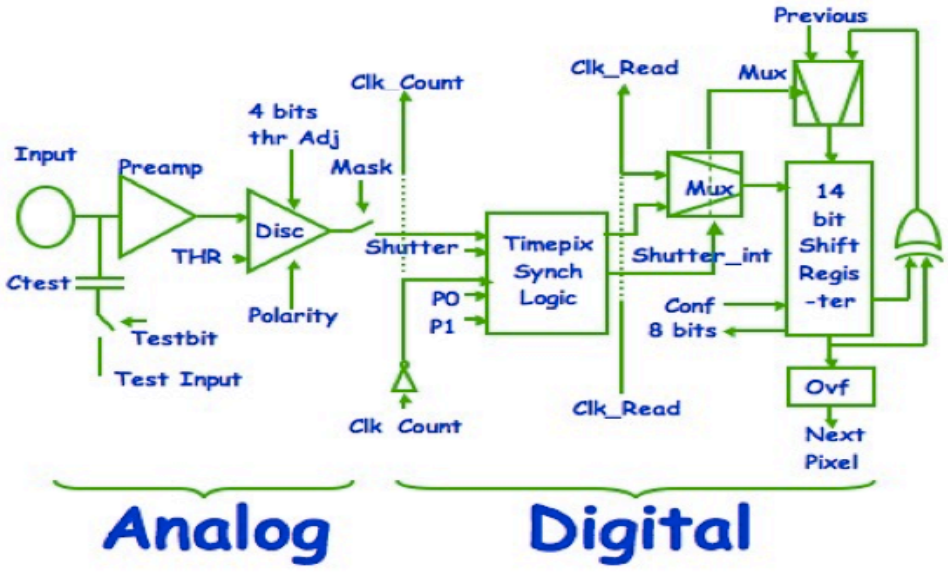


Figure 1.2. Standard Medipix 2 pixel electronic circuitry

The mode we used in this work is the TOT mode. In this mode, the Logic-Block functions act as a Wilkinson-type ADC by counting the number of Counter-Clock pulses that occur while the Discriminator input is above the threshold. TOT collected at each pixel can be converted into energy deposited by a calibration method [2].

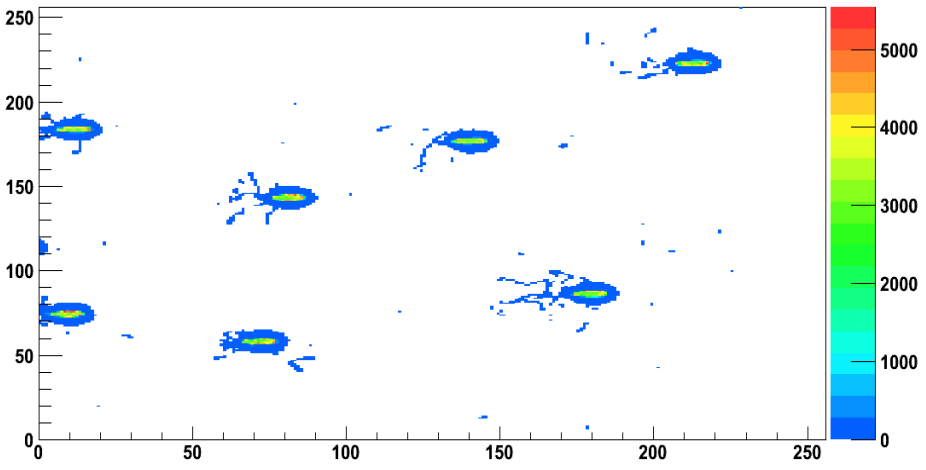


Figure 1.3. An output of Timepix detector exposed to 60degree iron particles

The Timepix chip that is attached to an overlying Si sensor layer using the bump-bonding

technique generates the raw output (Figure 1.3). The resulting detector is a hybrid semiconductor CMOS-based pixel detector made of 256 x 256 pixels, each 55µm x 55µm with the readout electronics for each pixel embedded within the footprint of that pixel. This device is able to survive and perform for extended periods in strong radiation fields.

When an incident charged particle penetrates the Si layer, it ionizes atoms along its path, producing a core of charge carriers that reflect the detail of the ionization along the track structures. These charges drift along and diffuse perpendicular to an applied bias-voltage field, which in turn causes the charge to be collected by the underlying pixels (Figure 1.4). Pixel information is read out in a frame that corresponds to a 256 x 256 pixel image containing the digitized charged information collected during the acquisition time window.

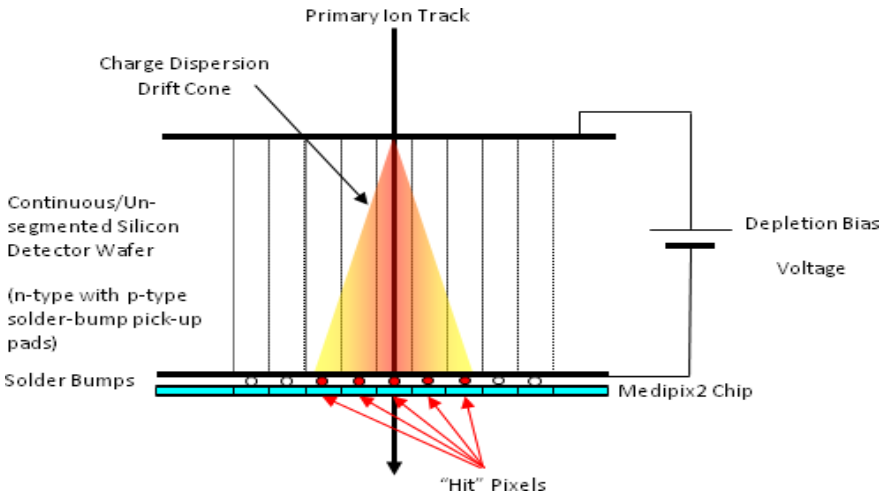


Figure 1.4. Charge drifting of a vertically incident ionizing particle

In this dissertation, pattern-recognition techniques and machine-learning tools are used for the automated characterization and classification of tracks of ionizing radiation during

space missions that are measured by the Timepix pixel detector system. Our analysis makes use of data taken in beams of heavy-ions at HIMAC (Heavy Ion Medical Accelerator Center at the National Institute for Radiological Sciences) in Chiba, Japan and NSRL (NASA Space Radiation Laboratory) at the Brookhaven National Laboratory in New York. The analysis techniques have been tested with both accelerator-based data and data coming from the ISS (NASA's-International Space Station). The classification problem consists of determining the charge and velocity (or energy) of the each individual incident particle by analyzing the track structures collected as pixel images. The main challenge remains extracting relevant features from the pixel images that can facilitate the calculations of the Dose-Equivalent, and determination of the particle types and energies in the incident radiation.

1.2 Pattern recognition and machine learning

"*The world we live in is not completely disordered, although it may sometimes seem so*" [3]. This judgment refers to a concept of pattern that plays an important role in real life. All objects, processes and phenomena in our life have their own patterns, implicit or explicit, that needs to be identified. Interpreting these patterns is crucial for us to establish the foundation for empathy, for critical and analytical thinking. It also provides the ability to explain the what, the how, the why, the right and the wrong.

Living organisms identify patterns to locate edible food, to classify friends versus enemies, and to seek a safe place for survival. Men learn how to recognize voices or distinguish handwritings of different people. All of this can be achieved through the

mechanism of discovering and learning patterns from experience. Examples include a function in cameras that automatically recognizes faces of people in a video, a detector that is capable of automatically identifying the nature of the source of incident upon it, and a car that can run on the road without being controlled by a human being. These machines are equipped with pattern recognition system that mimics the perceptual and cognitive ability of human beings to analyze information obtained from experimental data [4].

The concept of a pattern has been defined by pattern theory in which objects and processes in the real world can be expressed through mathematical constructs using combination of primitive structures [3]. Pattern recognition is a mechanism in which regularities in data -- called patterns -- are discovered to build models that can recognize objects of interest and take actions according to observed data [5]. In general, a pattern recognition system can be divided into two main tasks: feature extraction and classification as in Figure 1.5.

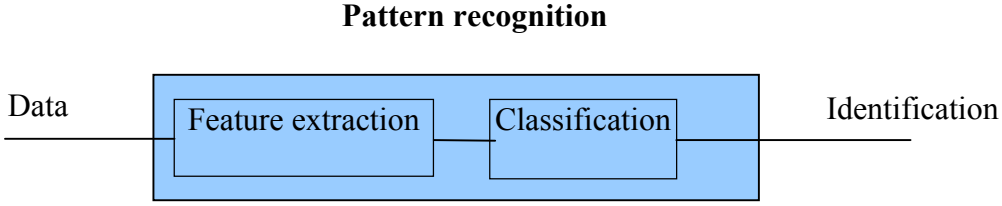


Figure 1.5. A general pattern recognition system

Human beings process data collected from environment through perceptual and cognitive abilities. In contrast, machines require algorithms designed for a pattern recognition system to process data obtained from sensors. Features will represent data in such a way

that pattern recognition system can understand and perform algorithms on it. Thus, feature extraction is the process of formulating data into a standard description, either a vector or a structure of features. Classification then uses these formulated features to build models that facilitate the identification process. The most effective techniques to build these models and develop classifiers are known as machine learning; it makes use of the capability of learning from data [6].

Machine learning is a *"Field of study that gives computers the ability to learn without being explicitly programmed"* (Arthur Samuel, 1959). Traditional applications are constructed by rules to solve a particular problem. This is done manually and requires programming skills to develop on a case-by-case basis. In contrast, machine learning uses data-driven techniques to generate a model with the capability of assigning instances of data into classes or groups. This model learns patterns extracted from data and its behavior changes significantly according to the input data or training set. As a consequence, if the distribution of data changes, behavior of models will adapt without explicitly being programmed.

In a pattern recognition system, feature extraction is an essential step, as it converts the raw data into relevant features appropriate for a particular domain and application. This representation has to be complete, powerful in describing patterns, and facilitating the classification process. Statistical pattern recognition [7] uses n-dimension vectors to represent n-feature instances. Another way of representing features is to use structural features together with a syntactic grammar to maintain a relationship between features. This technique is employed in structural pattern recognition [8].

A detailed pattern recognition system, shown in Figure 1.6, consists of preprocessing, feature extraction, model selection, training and testing:

- The preprocessing step extracts instances and objects from data. It removes noise, outliers, and includes processes of interpolating, smoothing data if necessary. For example, in image classification, preprocessing step may have to bifurcate the color image in order to segment the image into background and foreground object. Some logical operators such as dilation and erosion might be performed to remove small objects and fill holes inside objects. In Timepix data, one needs to use a blob-coloring algorithm to detect particles appearing in a frame based on connected components. Next steps are to remove fragmentation and crudely separate heavy-ion particles of interest (e.g., high energy particles) from those that are of no interest.

- Feature extraction represents instances output from the preprocessing step as a feature vector. Each dimension of a feature vector is a sub-pattern that can be a number or a nominal. A feature-selection step might be executed to perform dimensionality reduction and to remove potentially irrelevant features that are not helpful for the recognition process.

- Model selection tries to find an optimal model that is suitable for classifying the data of a particular domain. In fact, there exists no perfect model that can outperform others for solving all kinds of problems. In the state-of-art model, there are a variety of classifiers such as Naïve Bayes, decision tree, support vector machine, and boosting that can be used for instance recognition. Each model has

its own advantages and disadvantages. Variance and Bias need to be traded off in order to obtain a good model. One can use criteria such as classification accuracy, Receiver Operating Characteristic (ROC) curve to evaluate models. In addition, meta learning techniques are commonly used in this step [9].

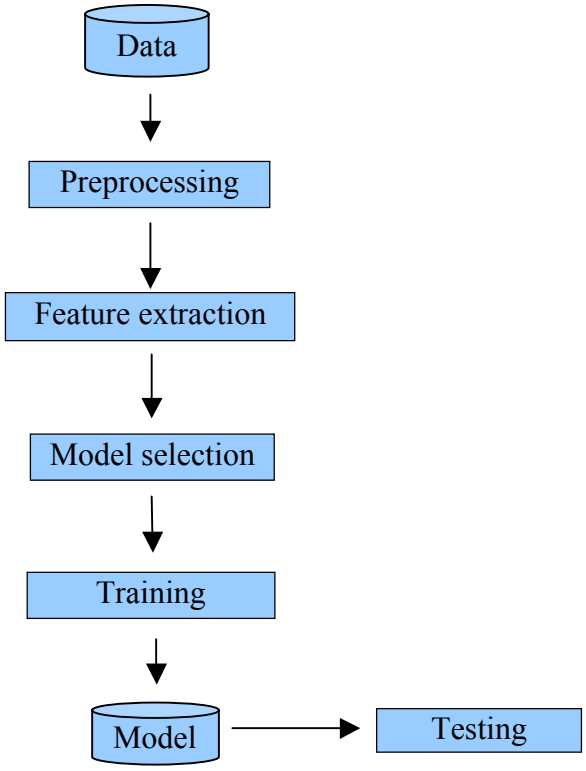


Figure 1.6. A detailed pattern recognition system

- Training and Testing: The aim of training step is to optimize parameters associated with the selected model in order to increase productivity. A cross-validation technique [10] can be used to avoid over-fitting issues in training by trading off between bias and variance. Testing evaluates the performance of the model with corresponding parameters.

1.3 Applications to Timepix data

1.3.1 Dose-equivalent calculation routine

Dose is defined as the energy deposited by a source of radiation per unit mass of traversed matter. Dose-equivalent is the most common form to express the biological effects of the dose caused by a particular type of particle for radiation protection purposes. To calculate the Dose-equivalent, each type of potentially incident radiation is given a Quality Factor (QF), also referred to as Relative Biological Effectiveness (RBE). As proposed by the National Council on Radiation Protection [NCRP 153 (2008)] regarding heavy-ions in the space radiation environment, the Quality Factor is a function of the LET of the traversing particle. LET is distinguishable from dE/dx (the kinetic energy lost by the primary ion per unit track length) in that it is a measure of the energy transferred to the local medium, which differs from the dE/dx due to the transport of the energy lost in a particular medium out of that medium.

LET_{∞} is defined in a theoretical infinite medium, and should ultimately be equal to dE/dx. In a practical sense, however, the actual local LET is what physical detectors measure, and what really matters to assess risk to tissue in human exposure. As discussed below, one can reasonably infer the amount of LET in tissue based on a measurement in another medium such as silicon. The Timepix chip device has a 300μm-thick Si sensor with a volume of 6×10^{-2} cm^3 (2 cm^2 x 0.003 cm). Si has a density of $d_{Si} = 2.3212$ g/cm^3, giving the Si detector layer a mass of ~0.14 g. Thus, one can determine the Dose in Si (D_{Si}), in Gray (Gy), by multiplying the total absorbed energy in a TimePix detector by a single factor:

$$D_{Si} = E_{TimePix} \text{ (in KeV) } \{1 / [\ 0.14 \text{ g x } 6.24 \text{ x } 1012 \text{ KeV/Gy g } \}$$

$$= E_{TimePix} \text{ (in KeV) } \{\ 1.153 \text{ x } 10^{-12} \ \} \text{ Gy} \qquad (1)$$

In order to determine the dose-equivalent, however, one needs to multiply the absorbed dose by a Quality Factor on a particle-by-particle basis. Since the Quality factor is a function of the LET (Figure 1.7), we need only determine the LET to achieve the value for the dose-equivalent:

$$H = Q(LET) \ . \ D_{si} \text{ (Sv)} \qquad (2)$$

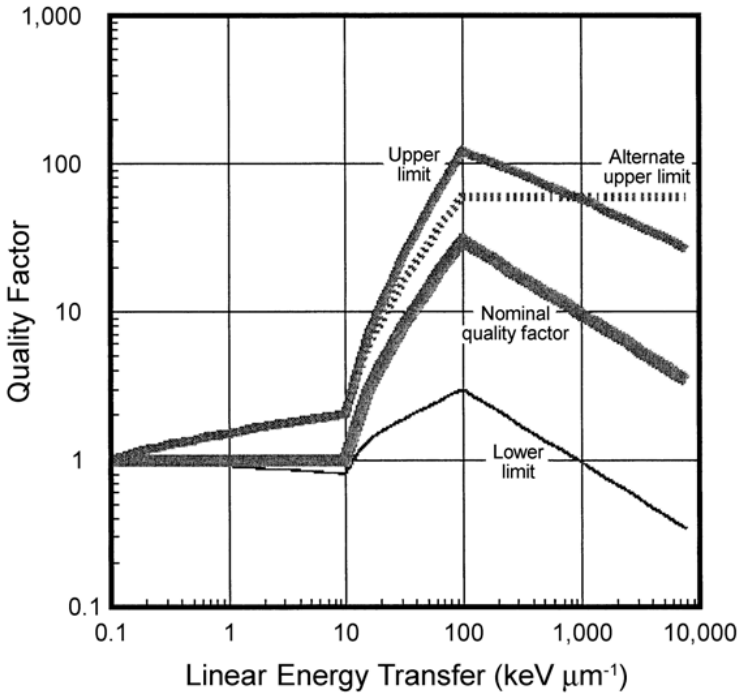


Figure 1.7. Quality Factor as a function of LET. NCRP 153 (2008)

Using the Timepix-based detector, LET can be measured from the total energy deposited (which is available from the calibrated Timepix chip) divided by the path length traversed

by the particle as it passes through the Si layer. The thickness of the Si layer, which is 300μm, can be used to calculate the total track length if it is possible to measure the track's angle of incidence, or its projected track length parallel to the surface of the detector. As discussed below, analyzing the angle patterns in cluster's structures can do this with a very good resolution.

In summary, with the availability of the calibration coefficients for each pixel, an accurate estimate of the energy deposited can be obtained, and with the availability of the frame times, the absorbed dose rates can be calculated directly. Finally, using pattern recognition, one can evaluate the LET for each track and use the NCRP 153 formula to calculate the QF-factor, track by track to yield the dose-equivalent.

1.3.2 Classification of type and energy of tracks

After exposure to sources of ionizing radiation, Timepix detectors produce tracks, which are in turn collected as footprints of the energy depositing events within an acquisition time. The source of energy deposited might be a penetrating particle of heavy-ion (e.g., H, He, C, Si, Fe…), a photon, a fragment from an interaction, or a product of neutron captured by a converter material. Stopping particles with low energy ranging in the detector layer might also create distinct tracks. Pattern-recognition techniques and machine-learning tools are expected to be able to identify the sources of the radiation and characterize their energy spectrum.

Identification of sources of tracks

It was experimentally determined that the short-range heavy nuclear fragments created by

the interactions of protons with sensor material can mimic genuine incident heavy-ions. Since protons are the most abundant type of incident particles in the space radiation, the nuclear interactions of incident protons require close investigation. Dose-equivalent can be efficiently assessed only when these interactions are discriminated from the tracks of genuine incident heavy-ions. The reason to exclude this energy is that it is due expressly to the interaction cross sections in Si. Our goal is to characterize the incident radiation field, so in principle we are only interested in knowing what the incident particle was, and not the secondary particles produced in any nuclear interactions in the Si.

The other consideration needed for an effective dosimeter is to segment the low energy particles that are not likely to stop in the sensor layer from those that most likely will range before exiting the sensor layer. The total track length would not be able to be determined accurately using projected track length or angle calculation if the particles did not penetrate the detector. The LET then would not be calculated correctly with this; there would be a miscalculation of total track length. Therefore, these stopping particles need to be identified to apply a special correction using Monte Carlo simulation method.

Photons (x-rays, γ-rays) and neutrons are electrically neutral. As such, they do not deposit any energy during their traversal of the sensor layer. However, neutrons can interact with the nuclei of the material, and photons can interact with both electrons and nuclei via the electromagnetic interaction by absorption and Compton scattering. In such cases, these interactions can lead to the production of energetic charged particles within the sensor volume. Neutrons and incident high-energy photons are treated differently from incident charged particles. Special correction factors for photons and for the detection of neutrons,

therefore, are necessary. This implies a need in the Timepix pattern recognition algorithms to be able to identify neutrons and photon interactions separately from penetrating charged particle tracks.

Classification of energy of tracks

For biological purposes, the dose in water or tissue is typically required, and a correction factor needs to be applied to convert the dose in Si into these other dose values. In practice, the conversion factor has a slight dependence on the velocity (or kinetic energy) of the particle depositing the dose. A plot in Figure 1.8 generated from SRIM simulation [26] can show this dependence. The conversion factor can be reasonably approximated by a single factor (~1.23) for higher energies (Figure 1.8). In addition, pattern recognition algorithm can help determine the energy of each individual particle from information contained in the combined pixel shape to obtain a good conversion ratio.

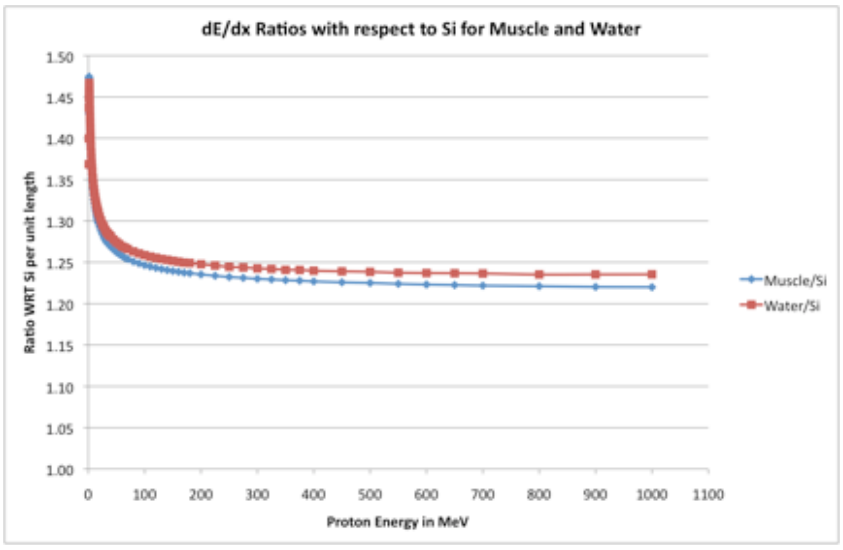


Figure 1.8. dE/dx ratios with respect to Si for Muscle and Water

More specifically, for velocities comparable to protons greater than 200 MeV, the ratio of 1.23 would be reasonably used as the conversion factor. For penetrating charged particle tracks with energies below 200 MeV/A, pattern recognition algorithms and machine learning tools can be used to bin the energy into several bins. The ranges of bins are sufficient for characterizing energy spectrum and obtaining good conversion factors.

In summary, pattern-recognition and machine-learning techniques are very important and essential for characterizing ionizing radiation in Timepix detectors, especially for Dose-equivalent calculation routines. The possibility to automatically classify different types of ionizing radiation and make use of the available energy deposition brings more benefit for dosimetry application as the equivalent dose can be calculated more accurately by incorporating weight factors for the different types of radiation.

The most important concern when developing such a pattern recognition system is the training set; it should be large enough to capture the underlying distribution of radiation data. For now, we mainly rely on data taken in HIMAC and NSRL, which provide marvelous facilities for understanding ionizing radiation of important heavy-ions. In the next step, we plan to develop a Monte Carlo simulation to generate a large amount of data for arbitrary heavy-ions penetrating Timepix. As such, we will have plenty of data for training, which will likely improve the performance of classifiers significantly.

1.4 Related work

Timepix detectors have a variety of applications in life sciences such as radiography with heavily charged particles, and neutron radiography [11]. A pattern recognition system

plays an important role in understanding and characterizing radiation fields based on Timepix detectors. A simple pattern recognition algorithm, together with a measurement for the efficiency of tracks induced by individual quanta of ionizing radiation in Medipix2 detector, has been developed [12][13]. This work described the results of distinguishing individual components of radiation field composed of proton, alpha, beta and gamma sources. However, it is a coarse pattern recognition system limited to a specific couple of radiation sources that cannot apply for heavier ion particles.

It is worth noting that the pattern recognition in [12] was developed for Medipix 2, which does not have the "time over threshold" mode for measuring energy deposited at each pixel. As a result, an exact angle of incident particles and the LET cannot be extracted. The proton with energy lower than 1MeV cannot be distinguished from photons and electrons due to a lack of energy spectrum at each pixel.

Generalized separation of particle tracks based on cluster patterns has been accomplished for low-energy and low-charge particles with Timepix detectors. Previous works have shown differences between electron, proton, and slow alpha tracks in silicon detectors [7][14]. However, visual inspection of heavy-ion tracks indicates that there is more information to extract from the cluster data beyond basic morphology. The detection, and the determination of heavy-ion charges, velocities and their applications to dosimetric characterization have been addressed in general in the past decade [15]. In [15], the work showed detailed structures of heavy-ion tracks in tissue. The structures of tracks are distinguished by two regions: core and penumbra. The core is a narrow central zone with a defined radius in tissue, while the penumbra is a peripheral zone enveloping the core.

This study also discusses the dissipation in tracks of high-energy heavy-ions in tissue, and shows a lateral spread depending on the energy of the primary particle. The relationship between the dissipation with the charge Z and the velocity of the particles has been investigated. The theoretical effects on LET and dose-equivalent are also discussed.

More recently, with the advances in hardware technology of the Timepix chip, one can now detect and visualize tracks of primary and secondary particles [16]. The capability of discrimination of penetrating heavy-ion charge and velocity with a TimePix-based Si detector has been demonstrated [17]. However, there is, to date, no other research implementing these ideas to extract detailed patterns to help characterize the composition and velocity of heavy-ion charged particles.

Pattern recognition for tracks of ionizing radiation and machine-learning approaches for identification of sources of heavy-ion charged particles have been developed [18][19]. However, the purpose of this study is to focus on extracting basic patterns and to test the capabilities of applying pattern recognition and machine learning for Timepix data. An advanced analysis of structures of tracks in term of LET, δ-rays and distributions of energy deposited should be taken into account. These are important factors for classification resolution.

2. Angle calculation method

In a Timepix-based detector, a penetrating energetic charged particle may be incident on the detector layer isotropically. While it is possible to determine the axis of the particle's trajectory through the sensor layer, there is an ambiguity in terms of the actual physical particle's direction of motion along the axis. By default, we assume that the particle direction is incident from above the sensor layer. There may be mechanisms, in some instances, to resolve the ambiguity, but we have not taken on that challenge yet. Calculation of the two angles --azimuth and polar angle-- is a critical step in determining the LET, Dose-equivalent and characterizing the type, velocity of the particles.

When an event occurs, electrons and holes liberated along the particle's path drift in the electric field of the depletion bias voltage, diffusing both laterally and horizontally with respect to the field. The diffused charge will generally spread out over multiple pixels creating a cluster of pixels associated with a single event. In cases where the quantity of charge is large, Coulomb repulsion between like charges has a significant effect as well. This spreading follows a characteristic pattern: the shape and energy intensity at pixels have precise information embedded within them about the azimuth direction and the angle of incidence with respect to the normal direction.

However, one cannot simply fit the shape with a certain model -- such as an ellipse -- and determine the angle resolution based on the elongation. The existence of energetic δ-rays, the effects of producing variations with different bias-voltage and source of charged particles [20][21], the heightened recombination of electrons and holes in regions with

very high levels of ionization (the so-called “plasma-effect”) [22], and any Coulombic effects, make the incident angle difficult to estimate accurately.

2.1 Azimuth direction determination

The shape of a cluster resembles a continent (geographic analogy) comprising a mountain in the center that is surrounded by peninsulas. Close study of the structure of clusters enabled us to identify three main components: a core part at the center of each cluster, the region immediately surrounding the core, that has been termed the “skirt”, and an δ-ray part which comes from recoil electrons liberated from the ionization process (Figure 2.1). The core contains most of the pixels with high energy, as it is the area where the detector collects multiple overlapping charges along the traversed path. The skirt contains low energy pixels, where charge densities are low. The core and the skirt parts constitute the primary track, or continent. The δ-ray part can be considered as peninsulas connecting to the continent. The δ-ray part also contains low energy levels that provide hints on the velocity or energy of the δ-rays in their length and total energy. The δ-rays begin deep within the core and emerge through the skirt, and they can be distinguished clearly if the energy of a particle is high enough. The maximum possible δ-ray energy is a function of the velocity of the primary ion.

The core part can be used to calculate the azimuth direction, or slope of the track. High-energy pixels provide more information about the direction than the low-energy pixels. Thus, instead of using a standard linear least square fitting, we propose a weighted least square linear fit to find an equation of a line that best describes x-y data pairs of pixels in the core part. The energy at each pixel will be utilized as the weighing factor for this

technique.

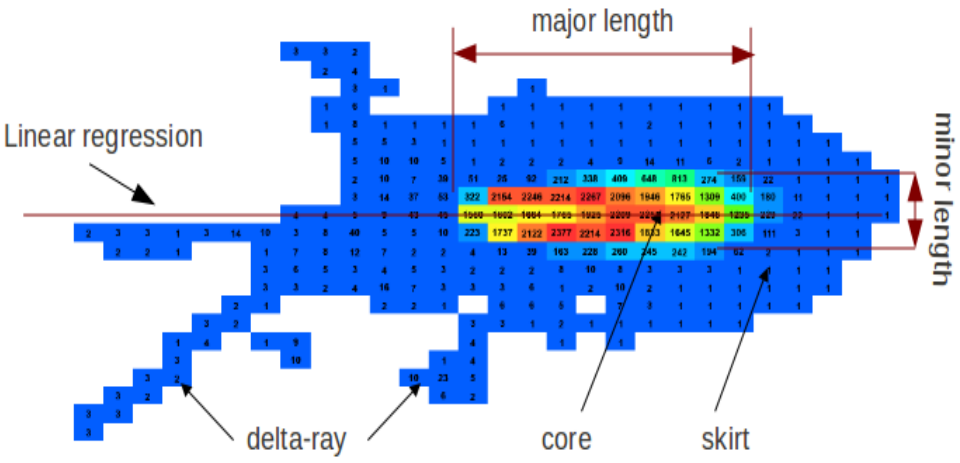


Figure 2.1. Cluster components

For computational reduction, we apply a threshold cut using a mean energy value of the cluster to get only pixels belonging to the core for model fitting. This threshold value will therefore exclude δ-ray and skirt pixels from data. In order to fit a line $y=\beta_0+\beta_1 x$ to pixels in the core of cluster, we minimized the function of weighted root mean square error as follows:

$$E = \sum e_i [y_i - (\beta_0 + \beta_1 x_i)]^2$$

where e_i is the energy of the pixel, β_0, β_1 are the cut and the slope of the regression line. x_i, y_i are the coordinates of corresponding pixels in the detector coordinate. E is the error function.

Taking the first derivative of E according to β_0 and setting it to 0, we have

$$\sum e_i(y_i - \beta_1 x_i) - \beta_0 \sum e_i = 0$$
$$\Leftrightarrow \beta_0 = \frac{\sum e_i(y_i - \beta_i x_i)}{\sum e_i}$$

Then, taking the first derivative of E according to β_1 and setting it to 0

$$\sum e_i x_i(y_i - \beta_0 - \beta_1 x_1) = 0$$

Replace β_0 , and we have:

$$\sum e_i x_i(y_i - \frac{\sum e_j y_j - \beta_1 \sum e_j x_j}{\sum e_j} - \beta_1 x_i) = 0$$
$$\Leftrightarrow \sum e_i x_i[y_i - \frac{\sum e_j y_j}{\sum e_j} + \beta_1(\frac{\sum e_j x_j}{\sum e_j} - x_i)] = 0$$

Finally, we can calculate the intercept and the slope of the regression line of the cluster as follows:

$$\beta_0 = \frac{\Sigma e_j x_j \cdot \Sigma x_i y_i e_i - \Sigma x^2{}_i e_i \cdot \Sigma e_j y_j}{\Sigma e_j x_j \cdot \Sigma e_i x_i - \Sigma x^2{}_i e_i \cdot \Sigma e_j}$$
$$\beta_1 = \frac{\Sigma e_j y_j \cdot \Sigma x_i e_i - \Sigma x_i y_i e_i \cdot \Sigma e_j}{\Sigma e_j x_j \cdot \Sigma e_i x_i - \Sigma x^2{}_i e_i \cdot \Sigma e_j} \quad (3)$$

Figure 2.2 show examples of clusters with azimuth directions of 0 degree (left) and -60 degree (right). A traditional least square linear fit without using energy weighting factors

will not capture the true direction of a small incident particle (less than 15 degree).

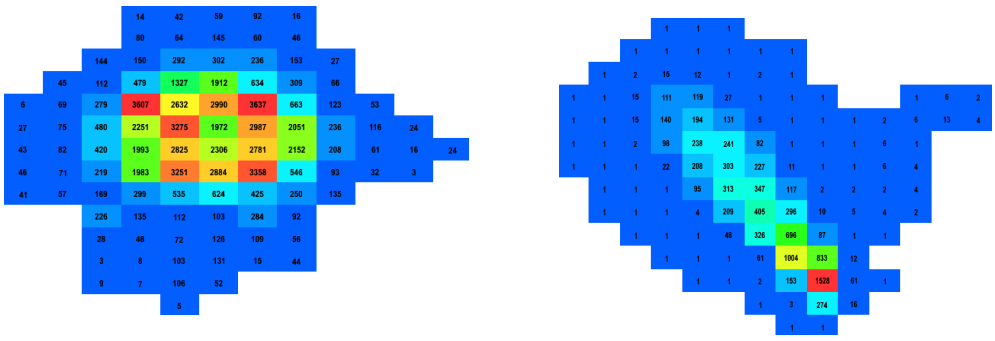


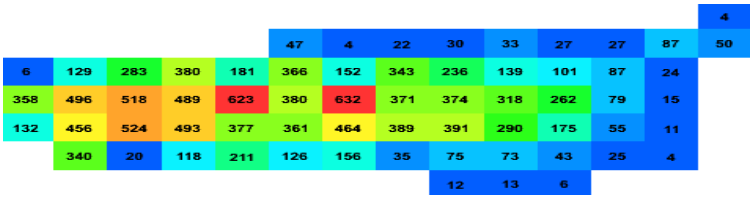
Figure 2.2 Azimuth direction of 0 degree (left) and -60 degree (right)

2.2 Angle calculation method

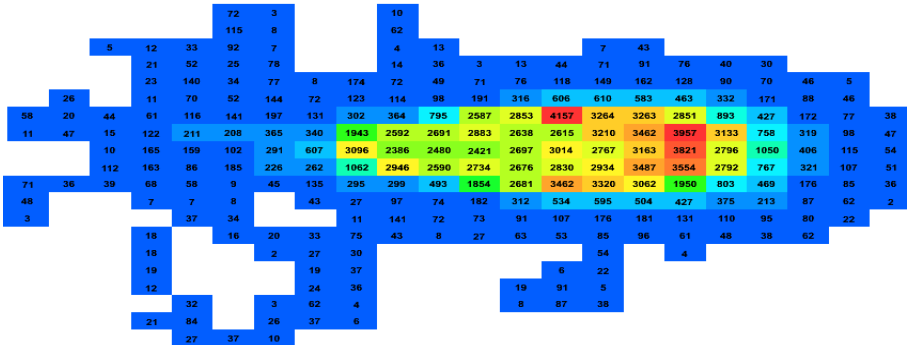
One of the biggest challenges in the determination of the uncertainty in the estimate of the projected track length is that the type of a charged particle at the point of penetrating Timepix chip is unknown. Each type creates a distinct shape and energy on the projected plane of the detector. We need to find a general pattern for all types of particles in order to come up with a method for accurately determining the angles of incidence.

Figure 2.3 shows the shape and deposited energy of the proton (a), carbon (b), Iron (c) at the same 60-degree incident angle. The major projected track length (major length) and the minor projected track length (minor length) are both proportional to the heavy or the atomic number of ions (Z). Incident angle resolution depends on these major and minor track lengths. Note that the δ-ray, skirt parts and bias-voltage variation also affect the resolution of these two track lengths. Also, the plasma-effect needs to be taken into account when it occurs with some heavy-ions (Figure 2.4).

a. Proton at 60 degree



b. Carbon at 60 degree



c. Iron at 60 degree

Figure 2.3. Variation of different source of charged particles at the same angle

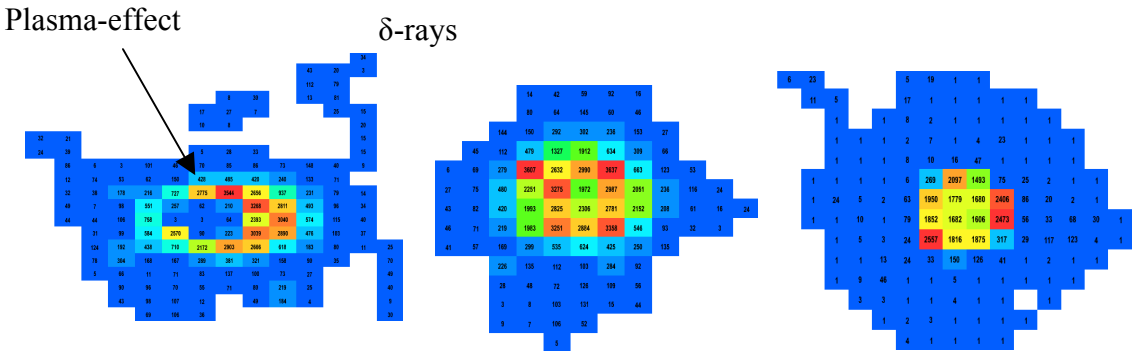


Figure 2.4. δ-ray, plasma-effect and bias voltage variation of 15V and 100V

In order to extract the major and minor track length, we need to rotate the clusters so that

the major direction of a rotated cluster is always parallel to x-axis. This can be accomplished by applying a rotation to each of the pixels once the slope of the track is obtained.

$$a = -\text{atan}(\beta_1)$$

$$X = \cos(a)*x - \sin(a)*y \qquad (4)$$

$$Y = \sin(a)*x + \cos(a)*y$$

Where β_1 is the slope obtained from (3), x y are original pixel coordinates, X Y are rotated pixel coordinates.

We then apply a formula that combines two track lengths to obtain the best estimate of the angle. Note that using only the major length for angular estimation is not appropriate, as this length will vary considerably according to different applied bias-voltage and different unknown sources of particles.

To extract the effective track length on each axis, we project and accumulate the energy of rotated pixels belonging to same bins of this axis. A profile of cluster is constructed that ameliorates the plasma-effect. The result is a histogram in which the horizontal axis shows the bins, and the vertical axis shows the accumulated energy for each bin. Since the energy collection at each pixel and bin is a stochastic process, which means that in each individual case there is a degree of randomness in the cutoff point that determines which bin belongs to the core part and which one belongs to the skirt part, as well as how much each bin contributes to the track length. For this reason, we applied fuzzy set theory [23] to assign a membership function for the concept “core part”. The membership value ranges from 0 to 1. For bins where the vertical value is above the mean of histogram, the

membership value is 1. Otherwise, the membership value is defined as the ratio of the vertical value of each bin divided by the mean value for the entire histogram. The sum of all membership values is the length of the track for this direction.

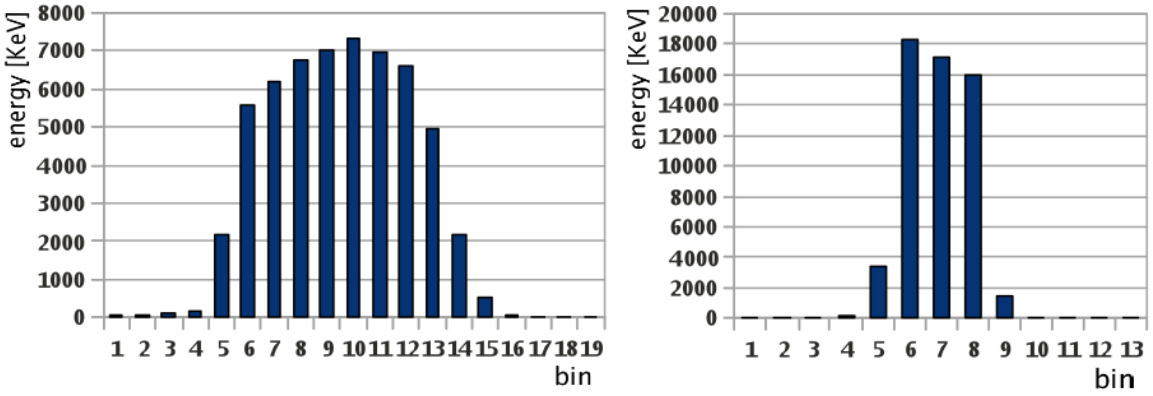


Figure 2.5. Energy histograms on the major direction (left) and the minor direction (right)

Figure 2.5 shows the accumulated energy histogram for the cluster in Figure 2.1 along the major direction and the minor direction. In this case, major direction has the slope of zero, which implies the horizontal direction. As we can see, the vertical values of bins on the left and right hand side (belong to the skirt part) are very small. Therefore, their membership values contribute almost nothing to the track-lengths, no matter how many bins are there. A significant contribution comes from bins in the middle and the ones that are fuzzy to be recognized between the core and the skirt part. Once the major and minor track-lengths have been calculated, we propose the following formula to compute the incident angle.

$$\tan(\alpha) = \frac{D}{T}\left(major - \frac{minor}{2major - \text{minor}} minor \right) \qquad (5)$$

In equation 5, T=300μm is the detector layer thickness, and D=55μm is the distance

between two pixels in the detector. The other constants are assigned by finding a reasonable fit to data for the angles and do not change as much for different configurations of the detector. Note that these assigned constants must satisfy the asymptotic principles in which $tan(\alpha)$ reaches ∞ as *major* reaches ∞ and $tan(\alpha)=0$ as *major*=0. An additional requirement for the vertical track formulation is $tan(\alpha)=0$ as *major=minor*. This formula ensures that the increasing or decreasing of both track lengths by bias-voltage variation effect does not have a big impact on the final angle resolution. Also, this formula can be applied for all types of heavy-ion particles that are unknown at this point. D and T parameters can be changed according to a specific design of the Timepix chip and the associated sensor. Experimental results will be shown in Section 7.

3. LET estimation

Using the Timepix-based detector, LET can be measured from the total energy deposited, divided by the path length traversed by the particle as it passes through the Si layer. The traversed path length can be calculated once the incident angle is determined in Section 2. As for the total energy, we need a calibration process to convert collected TOT to energy at each pixel. A further post-processing step to correct energy deposited needs to be done with heavy-ions suffering from volcano effect.

3.1 Energy calibration

We briefly describe the method developed by Jan Jakubek [2] to calibrate energy from the Timepix device to obtain the corresponding energy deposited on each pixel. By integrating the energy of all pixels belonging to a cluster, one can get the total energy deposited. The Time-Over-Threshold has been used for global operational mode in Timepix operation. In this mode, the integrated charge collected is measured by counting the number of Counter Clock pulses that occur while the output of the front-end charge sensitive preamp is above a default threshold.

As charge is collected at the input to the charge-sensitive preamp, the combined shaper circuit is set to produce a triangular-shaped pulse whose area and width are proportional to the charge collected. The leading edge of this pulse necessarily has an initial rise-time that introduces a simple nonlinearity for pulses very near the threshold value. For pulses that significantly exceed the threshold, the linearity of the effective ADC is excellent (as shown in Figure 3.1).

The shape of the curve in Figure 3.1 is caused by the rise time of the shaping circuit and the capacitance of the input stage. It is possible to fit this curve with a simple function:

$$TOT = a\,E + b - c/(E-t) \qquad (6)$$

where TOT is the recorded number of Counter-Clock pulses, E is the energy deposited in the detector that resulted in the charge collected by the pixel and a, b, c & t are the four parameters that need to be determined for each pixel. Equation 6 can be inverted to yield the Energy as a function of the TOT value.

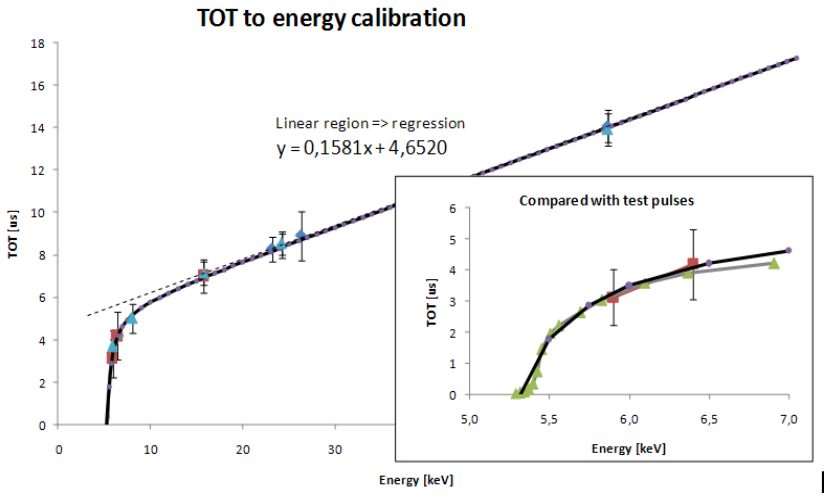


Figure 3.1. TOT to energy calibration

3.2 Energy compensation

Ionization along the primary particle track causes the creation of electrons and holes, which then drift under the influence of the applied bias voltage. As the carriers drift, they interact with opposing charge carriers and create the recombination of high densities of electrons and holes (the so-called “plasma-effect”) [22]. The estimate of the amount of energy deposited is reduced. This effect occurs to the greatest extent in the regions of

greatest ion density in the sensor, and where the bias voltage field direction is parallel to the track direction. Recombination may cause a reduced amount of energy in the measurement relative to what was deposited in the silicon by the incident ion. However, the major losses are introduced by the non-linear response of the in-pixel front-end pre-amplifier stage and the effect of the front-end over-voltage protection diodes. These effects will clearly affect the uncertainty in LET calculations. A reconstruction technique will be proposed to compensate for the energy loss due to these effects.

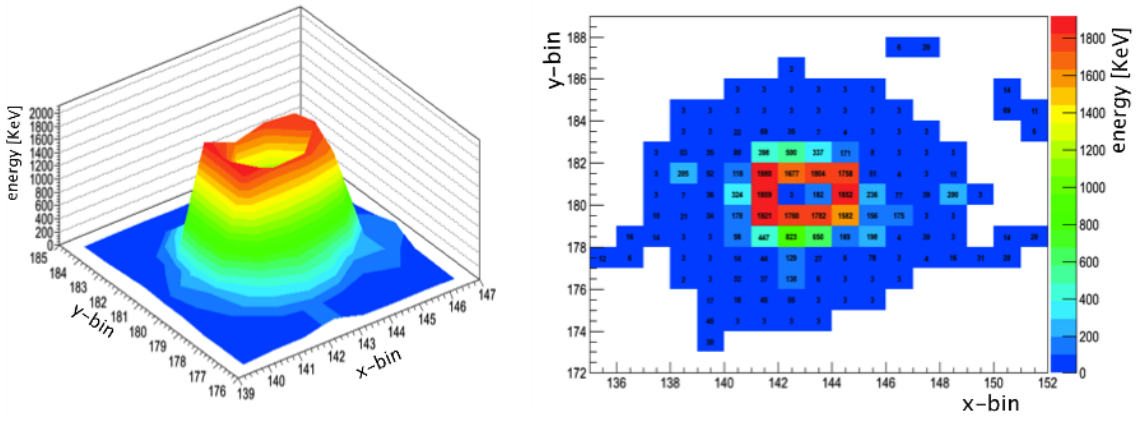


Figure 3.2. Volcano effect

As shown in Figure 3.2, the TOT values are suppressed for the pixels in the middle of the clusters where the density of the charges and the resulting input voltages are very high. If we assumed the normal cluster had the cone-shaped structure that was built by the more-or-less symmetrical accumulation of charges, the appearance of volcano would give a summit crater on the top of cluster. The slope of the summit crater varies depending on the density of the charges. This summit crater needs to be compensated first by flattening out the top of cluster as shown in Figure 3.3. Note that this step will not be necessary in the next generation of this technology, known as the Timepix 3 has a modified front end

pre-amplifier design so that the cluster will have a flat top response for large charges and voltages at the input.

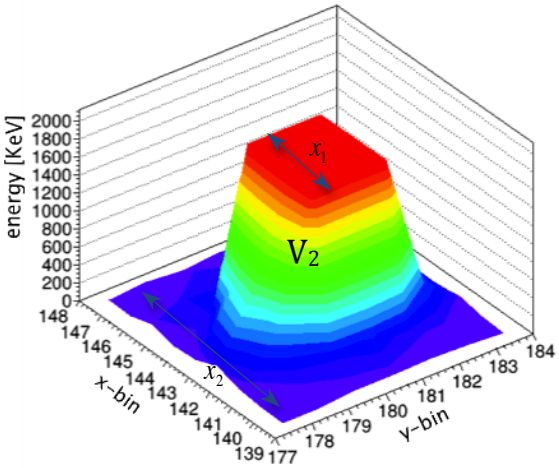


Figure 3.3. Energy compensation

We assume the total energy of the original cluster without suffering from volcano effect is V. After filling energy loss from the summit crater, we obtain total energy V_2. The remaining energy loss is the missing energy on the top of cluster, which is assumed as V_1. The purpose of the compensation process is to calculate V ($V=V_1+V_2$), given V_2 and the structure of cluster as in Figure 3.3. This structure allows calculations of minor projected track lengths x_1, x_2 at different the height of cluster according to different energy thresholds. We can calculate V following Equation 7.

In the formula below, y_1, y_2 are major projected track lengths at different energy thresholds. h_1, h_2 are the height of V_1 and V_2, and S_1, S_2 are the projected area of V_1 and V_2, respectively. $h=h_1+h_2$ is the height of the original cluster after compensation. Note that these formulas are only valid with the assumption that the cluster has a cone-shaped

structure and a projected shape of cluster has ellipse-shaped or symmetric quadrilateral. In fact, experimental results show that the assumption is reasonable to yield compensated LET accurately to within 10% at most. A disadvantage of this compensation method is that it is very sensitive to the variation of applied bias voltage, which is affecting the diffusion of charges and cluster structure. Experimental results will be shown in Section 7.

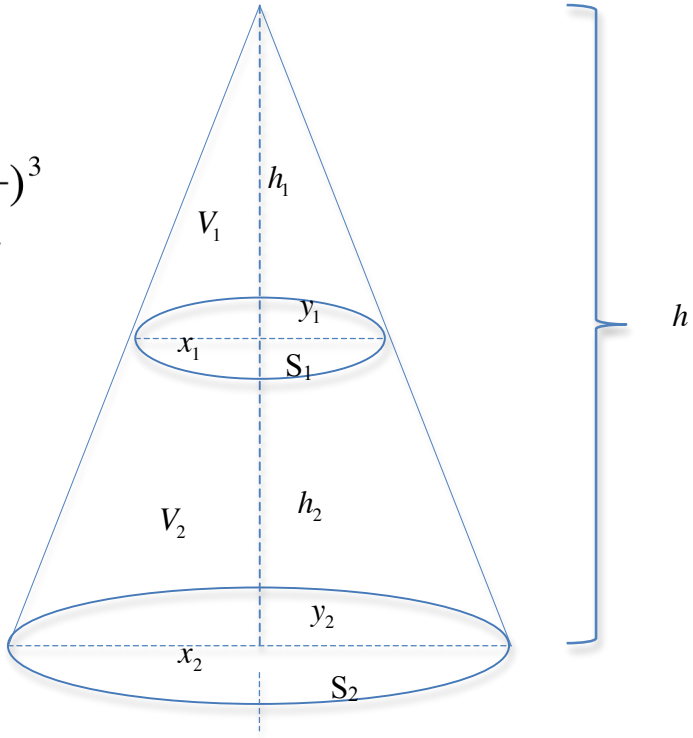
$$V = V_1 + V_2$$

$$\frac{V_1}{V} = \frac{S_1 h_1}{S_2 h} = \frac{x_1 y_1 h_1}{x_2 y_2 h} = \left(\frac{x_1}{x_2}\right)^3$$

$$\frac{V - V_1}{V} = \frac{x_2^3 - x_1^3}{x_2^3}$$

$$\frac{V_2}{V} = \frac{x_2^3 - x_1^3}{x_2^3}$$

$$V = \frac{V_2 x_2^3}{x_2^3 - x_1^3} \quad (7)$$



4. Delta-ray analysis

In this section, we explore δ-ray characteristics that are useful for identifying type and velocity of a particle. An energetic secondary electron (δ-ray) is produced through the net coulomb interaction of matter with a traversing charged particle. δ-ray electrons create a noticeable track of their own and can leave the immediate area of the primary particle's track when they have sufficient energy. Recent refinements on approaches of determining Relative Biological Effectiveness recognize that beyond just the LET, the velocity of the particle, which influences the maximum energy of the knock-on electrons (δ-rays) is also an important factor. Therefore, the extraction of δ-ray features is meaningful for obtaining the information of the velocity and type of a particle.

δ-rays always start in the core, emerge through the skirt and escape a significant distance away from the primary cluster as curly fragments when the energy is high enough. The range, energy and number of δ-rays play important roles in radial dose distributions near ion's path and far from the path. When coupling with LET, they provide a capability to identify the source of the primary incident particles. More details will be provided in Section 5.4.

In this dissertation, we extract four relevant features of noticeable δ-rays that are a function of energy and type of ions. They include a total number of δ-rays, a total energy of δ-rays, a maximum length of a δ-ray, a total length of δ-rays. In order to do that, we need to transform the cluster into a compact representation so that features are easy to derive. A skeleton representation is convenient for this purpose as it provides a way to

keep a compact structure of the whole track including δ-rays and primary cluster. Also, a distance transform algorithm is employed to distinguish which skeleton pixel belongs to the δ-rays and which one belongs to the primary track. They are described next.

4.1 Cluster skeletonization

Skeletonization is a popular technique used in digital image processing, computer vision and pattern recognition. This technique generates a skeleton that preserves the topology of an object and can be used as a representation of its shape. For our case, a skeleton of a cluster is a set of connected pixels representing the topology of the cluster as in Figure 4.1. Pixels in a skeleton are classified into three categories: end point which has exactly one neighbor, branch point which has exactly three neighbors, and skeleton point which has exactly two neighbors. End points are essential part of the topology that reflects a number of δ-rays of a cluster. Branch points and skeleton points can be used to compute the total length, maximum length and total energy of δ-rays. Some fragments of δ-rays are disconnected with the primary cluster in one or two pixels. One has to do a post-processing step to attach these fragments to the final skeleton.

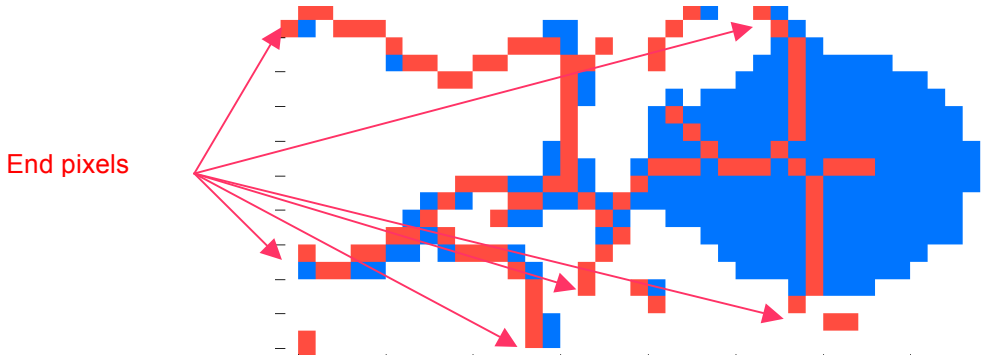
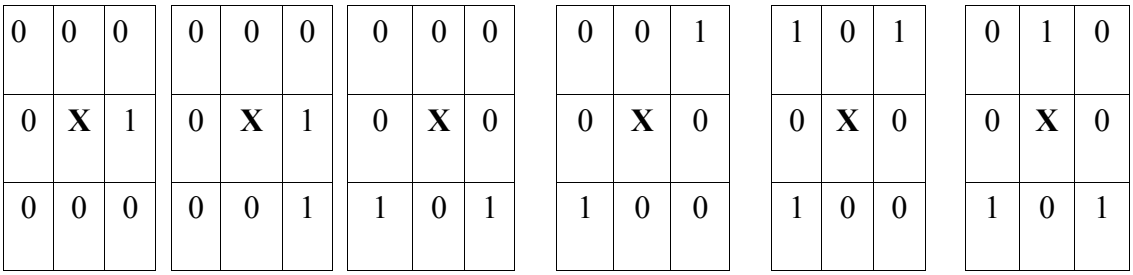


Figure 4.1. Cluster's skeleton



a. End point b. Skeleton point c. Branch point

Figure 4.2. Examples of categories of skeleton pixels

In general, a skeleton can be computed by either thinning the shape of an object or steadily eroding the boundary without breaking the topology. A survey on skeleton extraction in digital image processing can be found in [24]. In this project, we use an erosion technique to remove steadily boundary pixels that are not end points, branch points and skeleton points. We use 3x3 square block structure element to define the category of each pixel as in Figure 4.2. Each pixel has a corresponding signature specified by 8-connected neighborhood. Therefore, 256 different states are defined manually to obtain these signatures and corresponding categories. At each iteration of the erosion process, pixels are removed from a low energy pixel to higher energy pixels until no more pixels need to be removed. When a pixel is removed, it propagates its energy among the remaining neighbor pixels. At the end, a skeleton will contain the total energy of the cluster. A disadvantage of this method is that the formation of final skeleton is very sensitive to noise and fake ending points. The energy at each pixel should be incorporated to prune the skeleton in such cases.

4.2 Cluster distance transform

The generated skeleton of a cluster involves skeleton pixels of the δ-rays and primary cluster as well. Thus, it is necessary to identify which skeleton pixels belong to δ-rays and which to the primary cluster. As mentioned above, a δ-ray begins deep within the core of cluster, emerges through skirt part, and appears as a noticeable peninsula. In order to recognize the length and energy of a δ-ray, we need to determine how deep it goes inside a primary cluster. A distance transform algorithm can be helpful to capture this information.

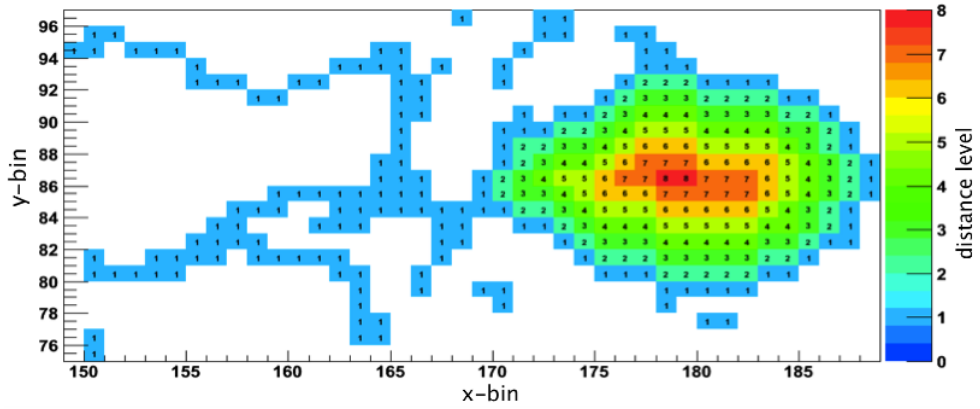


Figure 4.3. Distance transform of a cluster

Technically, a distance transform assigns each cluster pixel a value corresponding to a minimum distance from this pixel to the boundary. There are different common distance metrics that can be used such as Euclidean, city-block, and chessboard distance. A survey of distance transform and its applications is described in [25]. In general, Euclidean is often used in many applications but it is not easy to manipulate and not necessary for our

purpose. For the simplified computation, we use city-block distance metric with a cross-shaped element. Note that a primary cluster can be separated from it peninsulas (δ-rays) by utilizing the capability of distance transform. The majority of pixels in the noticeable δ-rays are assigned a distance level 1. For some cases, a δ-ray pixel is assigned a distance level 2 which can be corrected to distance level 1 by a simple post-processing step. Higher distance level will be assigned only for pixels inside the primary cluster.

The distance transform can also be used to detect the volcano effect. As the level of a pixel elevates, the higher energy should be expected. Otherwise a cluster is considered to suffer the volcano effect. Figure 4.3 shows the result of applying distance transform and a post-processing step for a cluster of the Iron particle. Using this transform, one can easily recognize δ-rays and select a deep level that δ-rays can go inside. Also, a maximum distance level of a cluster is also a good feature to identify the type of heavy-ions. For example, a proton cluster is always recognized with a straight shape for any incident angle that generates a small maximum distance transform level (less than 3).

Once skeleton pixels are combined with the distance level for each pixel, we can extract δ-ray features as follows. An estimate for the minimum total length of δ-rays can be computed by the total number of skeleton pixels belonging to δ-rays. Similarly, total energy of δ-rays is the sum of energy of δ-ray skeleton pixels. Number of δ-ray end points will be number of δ-rays. For calculating the maximum length of δ-ray, we find a maximum path length from end points to the other pixels corresponding to this δ-ray. These four features will be used to characterize whether a particle has a high or low energy, and what type of the particle generates such kind of δ-ray.

5. Ion type identification

5.1 Interaction recognition

As the charged particles traverse the sensor, they occasionally have nucleus interactions with the sensor material. These interactions are normally negligible because of the small relative size of the nucleus as compared to the size of the electron clouds. As is well known, protons are the most dominant particles in space. The interactions of incident protons with detector material can produce short-range heavily ionizing nuclear fragments that can mimic genuine incident primary heavy-ions that look like high LET particles. These misidentified heavy-ions will contribute significant errors to the Dose-equivalent estimates due to their potentially large Quality Factors.

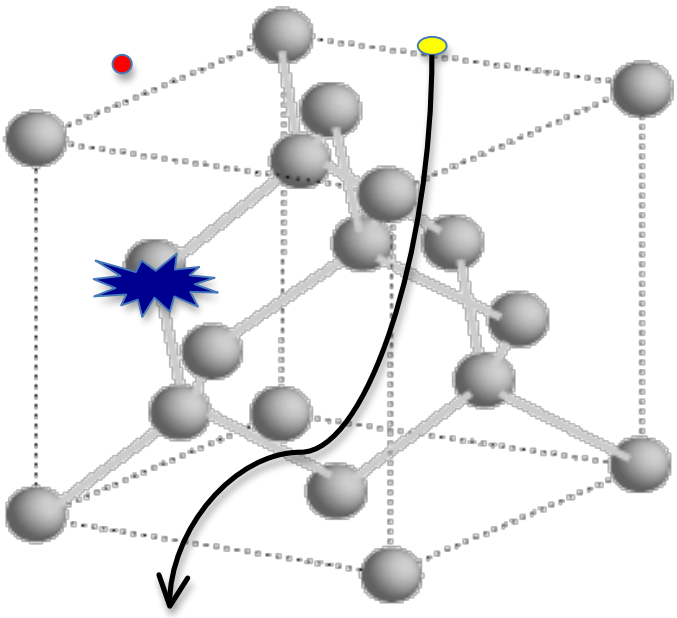


Figure 5.1. Nucleus interaction

As noted earlier, since our goal is the characterization of the incident radiation field and

such interactions are of interest only insofar as we need to know only what the primary incident particle was. Therefore, discrimination between these interactions and the tracks of genuine incident heavy-ions need to be addressed in order to exclude them from the total effective Dose-Equivalent assessment.

A proton can interacts with a nucleus via the strong interaction via either elastic or inelastic channels. In elastic channels, the proton retains its identity, maintains most of its original momentum and scatters off of the "target" nucleus at some angle, with the recoiling intact nucleus absorbing some of the proton's energy. The proton then continues to propagate depositing its energy by ionizing atoms as it had been doing prior to the collision. In inelastic collisions, the proton transfers most of its energy to break up the nucleus, and possibly to the production of new subatomic particles such as pions. The heavier nuclear fragments can mimic genuine incident primary heavy-ions (Figure 5.1).

In order to identify the variety of patterns of clusters created from such interactions, we conducted experiments with a large sample of proton energies. From that date we can separate high LET clusters that are definitely products of proton interactions. We have been able to identify different types of interactions associated with following patterns:

- **Interaction with two of more clear peaks**. A proton interacts with a Si nucleus producing two or more heavy-ions connecting via lower energy pixels. This pattern can be easily recognized by using a suitable energy threshold cut and calculating number of connecting components from the remaining pixels.

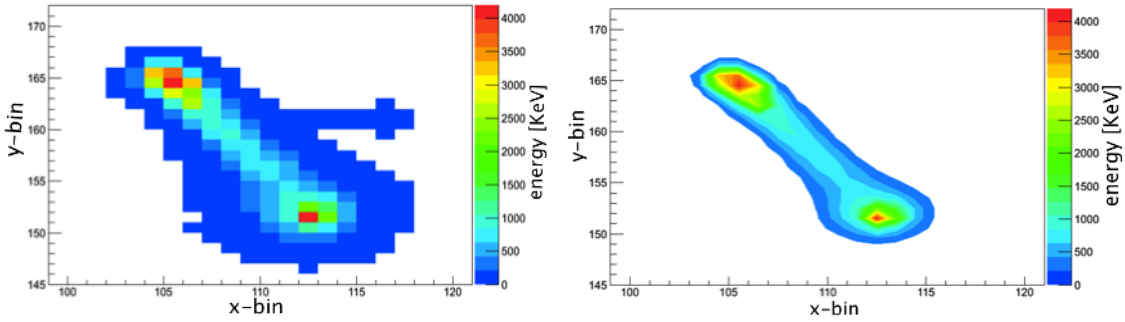


Figure 5.2. Interaction with 2 primary fragments

- **Interaction that creates a stopping particle**: This interaction transmutes the Si nucleus into a slow particle that ranges immediately or shortly at the vicinity of interaction position. As a result, the forming cluster contains characteristics similar to products of captured neutron (round shape, special morphology, etc.) or stopping particle (Bragg curve).

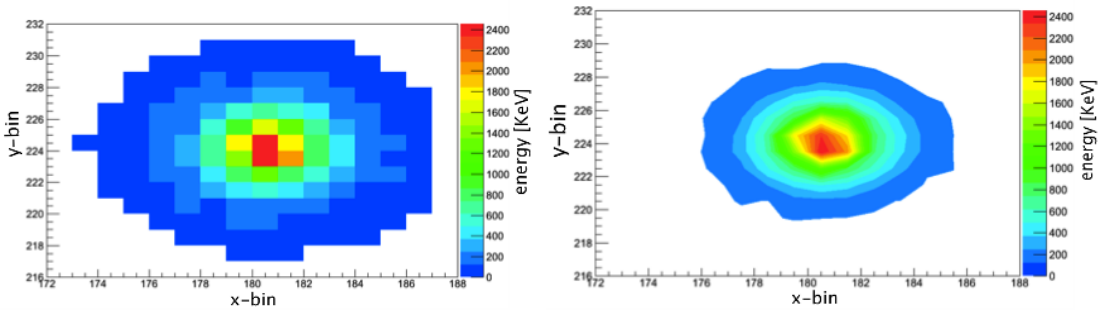


Figure 5.3. Interaction with stopping

- **Interaction with visible fragments surrounding primary particle**: Unlike the first case, which produces two or more primary fragments, this interaction creates one primary core of cluster and secondary fragments that are clearly different from the skirt of a genuine cluster. We use distance transform to determine levels of pixels and the core of the primary. Calculating the average distance between

pixels level 3 in the core and remaining pixels level 2 will help to recognize pixels of fragments. In a genuine cluster, pixels of levels 2 must be the immediate neighbors of pixels level 3 in the core, but this is not the case for these interacting clusters.

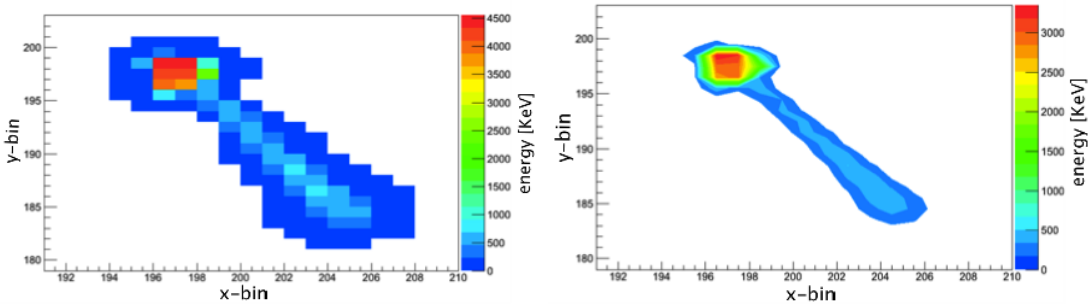


Figure 5.4. Interaction with one primary and one surrounding secondary

- **Interaction with fragments similar to delta rays**: Similar to the third case, the cluster consists of one primary fragment surrounded by many other fragments. However, the surrounded fragments that can mimic the δ-rays are not easy to recognize. We also use the distance transform to isolate the core and calculate the number of inner pixels in fragments. The number of inner pixels will help us determine whether secondary particles are δ-rays or interaction fragments.

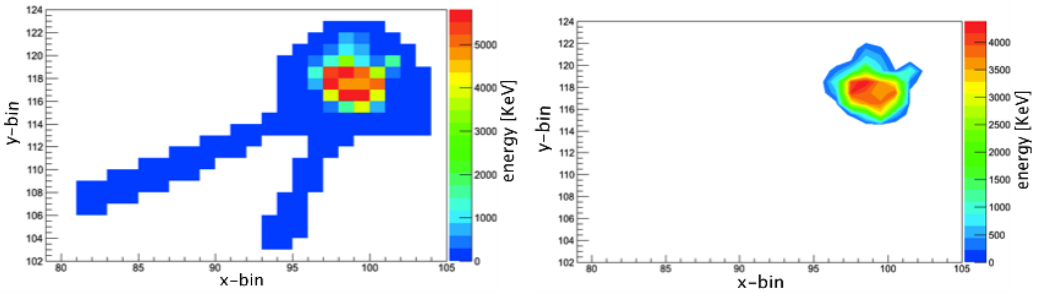


Figure 5.5 Interaction with one primary and fragments similar to delta-ray

5.2 Stopping recognition

For an effective dosimeter, there is a need to discriminate the low-energy particles that are likely to stop in the clothing or skin, from those that most likely will enter deeply into or penetrate through the body. The reason is that the calculation path length using the angle or projected path length as described above is no longer correct for particles that stop within the volume of the sensor layer. Thus, the estimation of LET is incorrect when dividing stopping energy by the miscalculation total track length. This phenomenon has been observed from ISS data with a bump roughly at 14 KeV/μm as shown in Figure 5.6.

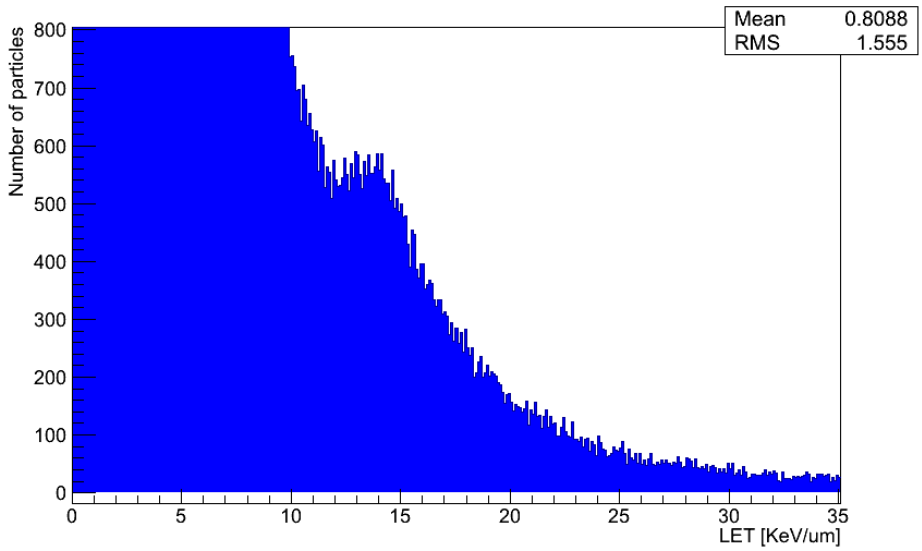


Figure 5.6. LET distribution of ISS data with an abnormal bump at 14KeV/um

Theoretically, there should not be such a bump in the LET distribution of ISS radiation data. This bump is the result of the miscalculation of LET with the assumption that all particles are penetrating the Si layer. Essentially, the pattern recognition system needs to provide a mechanism for recognizing these stopping particles. A correction can be made using Monte Carlo simulation.

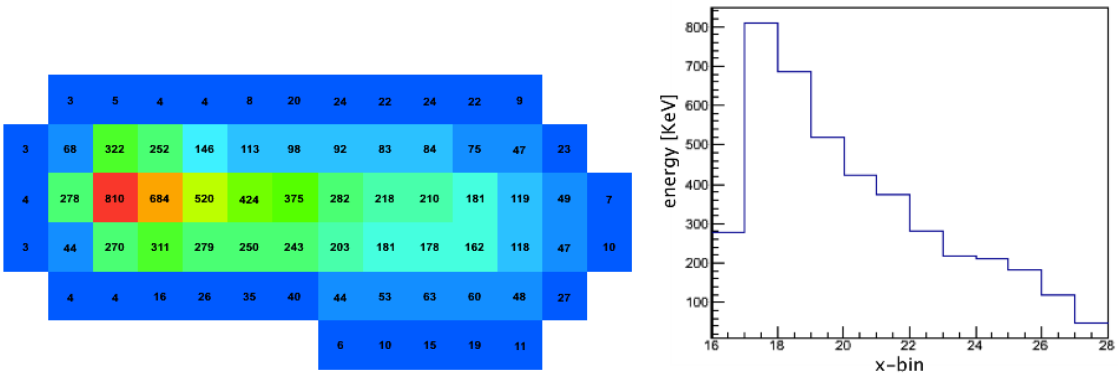
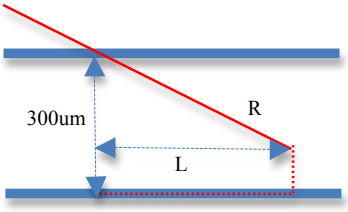


Figure 5.7. Stopping proton and its profile

To identify the stopping particles, we first use the method of proof by contradiction to narrow down the particles that are suspicious to stopping as follows:

- (1) Assume the particle is stopping
- (2) Calculate the total energy loss in Si layer
- (3) Calculate the range R of this particle with calculated energy loss in Si
- (4) Calculate the projected track length L using angle calculation routine
- (5) If L>R (adjacent > hypotenuse that is logical contradiction), then, conclude that the particle in question is not stopping
- (6) If L<=R, check the profile of particle to make a decision



The range R, a distance between a starting point where the particle enters and the ending point where the particle comes to rest, in step 3, is calculated by using Projected Range

Algorithm from SRIM [26] as follows:

$$R_p(E_0 + \Delta E_0) = R_p(E_0) + [\frac{4E^2 - (2E\mu S_n + \mu Q_n)R_p(E_0)}{4ES_4 - 2\mu Q_n}]\frac{\Delta E_0}{E}$$

where R_p is the projected range, E is the initial ion energy, $\mu=M_2/M_1$ with M_1, M_2 is the ion and target mass. S_n is the nuclear stopping power and S_t is the total stopping power. Q_n is the second moment of the nuclear energy loss.

One can use Bragg curve [27] as a profile in step 6 to check whether the particle is stopping or not. To do so, we identified the azimuth angle of this cluster by employing the weighted linear least square described in Section 2. The cluster is then rotated to be parallel with the X-axis. We divided each pixel of the cluster into two points in which each point contains half of the energy of this pixel. All points are projected onto X-axis and the maximum energy at each bin is chosen to build the histogram. If this profile has the Bragg curve structure, it should have the maximum-energy bin at the position of two or three. It also has to be monotonically decreasing along two sides as shown in Figure 5.7. A simpler approach is to look at the total number of local peaks in which the pixels at these peaks have energies higher than the energy of their neighbors. The ratio of this number to the size of cluster will help provide a stopping characteristic.

There is a disadvantage of using Bragg curve or local peaks for detection of stopping events, as these methods are not applicable in case of vertical stopping tracks. One has to perform further contour and δ-ray analysis in such cases.

5.3 Neutron recognition

While the interaction of photons (x-rays, γ-rays) with the sensor material typically results in the production electrons with isolated curly-shaped tracks that are easy to recognize, neutron detection is somewhat problematic. By themselves, neutrons do not deposit energy when they travel through the material. However, when they interact with nuclei in the material, they can create recoils that produce ionizing tracks that can deposit energy. In addition, slow neutrons can be captured when they strike nucleus in Si layer transmuting the Si nucleus into heavier isotope. Typically this isotope is not stable and commonly decays through β-decay, which produces an energetic electron depositing energy. It is also possible that the isotope may be in an excited state and subsequently de-excites by emitting γ-ray. The emitted γ-ray then may travel a long way before it deposits its energy, which may or may not be in the sensor material.

A method for detecting slow (thermal) neutrons is to use a thin layer of converter material external to Timepix sensor to capture neutrons. The next step is to identify the detectable products created from the capturing. For thermal neutrons, ^{6}LiF is a commonly employed absorber that can be used to capture neutrons. This process splits a ^{6}Li nucleus into a Triton ^{3}H and an energetic α particle that has energy >2 MeV/A. The α particle has enough energy to be able to move out of the thin layer of converter material and ranges in the immediate vicinity of entering the surface of the Timepix sensor. A pattern recognition system needs to be able to identify the patterns produced and that employs a classification algorithm to distinguish the α-particles created from captured neutron and those from genuine penetrating α-particles.

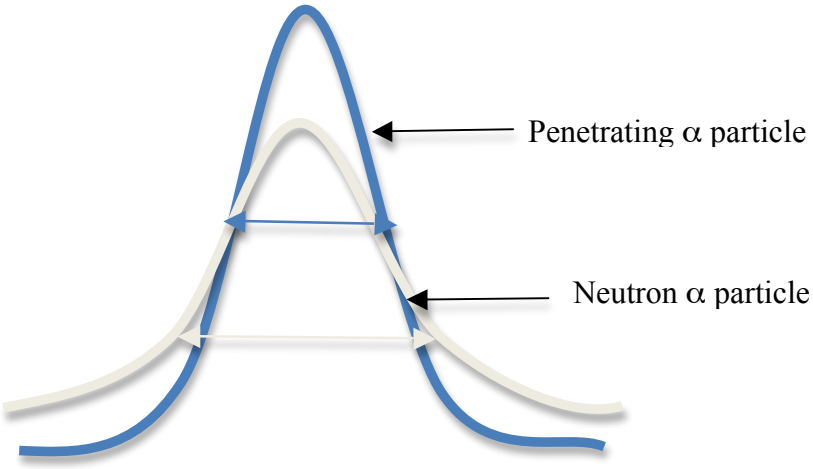


Figure 5.8. Neutron α-particle profile versus penetrating α-particle profile

As the α-particle associated with neutron particles stops immediately at the entry point of Timepix, the cluster created from this event should have the round shape with special morphology. The angle of this cluster is identified as 0 degree or so as a result. Its profile in term of energy distribution of cross section will also be different from the profile of the penetrating α-particle (Figure 5.8). We use the angle calculation to filter suspicious candidates for neutron products first. We then generate accumulated energy at different distance transform levels to characterize the morphology of clusters. Finally, we calculate the minor length of the cluster and combine previous morphology features to support the classification process. A binary classification with machine-learning approach is employed to fulfill this task. Details will be described in Section 6. A situation exists for energetic neutrons when the converter layer is rich in H, such as polyethylene. In that case, the neutron can elastically scatter from the protons at the nuclei of the H and the protons can recoil with a large fraction of the neutron's kinetic energy, leaving a thermal neutron and with the recoiling proton being detected as it enters the sensor layer.

5.4 Ion type classification

After identifying special events from interaction, stopping, and neutron particles, we need to classify tracks of ions into types corresponding to their atomic mass. Features extracted from previous analysis can be used to perform this classification process. Relevant features are extracted from angle calculation, LET estimation, delta-ray analysis, and cluster morphology. They include: LET, angle, size, minor projected track length, level extracted from distance transform, number of delta-ray, total length of δ-ray, maximum length of δ-ray, total energy of δ-ray and skirt, total energy of different distance transform levels. We are trying to classify unlabeled tracks into desired ion types: H, He, C, N, O, Ne, Si, Ar, Fe. For space-radiation dosimeters, H and He are more important and need to be recognized as well as possible.

One can build a multiclass classifier in machine learning to classify these classes simultaneously with the unknown distribution assumption. However, there are some potential cuts that can be used to simplify the classifier, to get higher performance and to be easier to explain the classification model. LET distribution plays an important role in determining these cuts.

LET is one of the most important features to identify the ion type of charged particles. The energy lost when a charged particle travels through Si layer is described by the Bethe-Block equation:

$$\frac{-dE}{dx} = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} [\frac{1}{2} \ln(\frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2}) - \beta^2 - \frac{\delta}{2}]$$

This equation relates the term dE/dx (the rate of energy loss of the particle with respect to the distance that it has traveled through the material) to the energy that the particle possessed upon entering the material. The energy is found in terms of β and γ

$$\beta = \frac{v}{c} \qquad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Here, v is the particle velocity, c is the speed of light, and Z, A are material atomic number, and mass respectively. T_{max} and I are the maximum and the mean excitation lost in collision with the electron. The term (δ), T_{max} and I can be considered as constant in this formula. Using SRIM [26], a package software that is able to calculate dE/dx versus energy and the range of the transport ions in matter, we plot the relationship between dE/dx, energy and type of particles (Z) as shown in Figure 5.9.

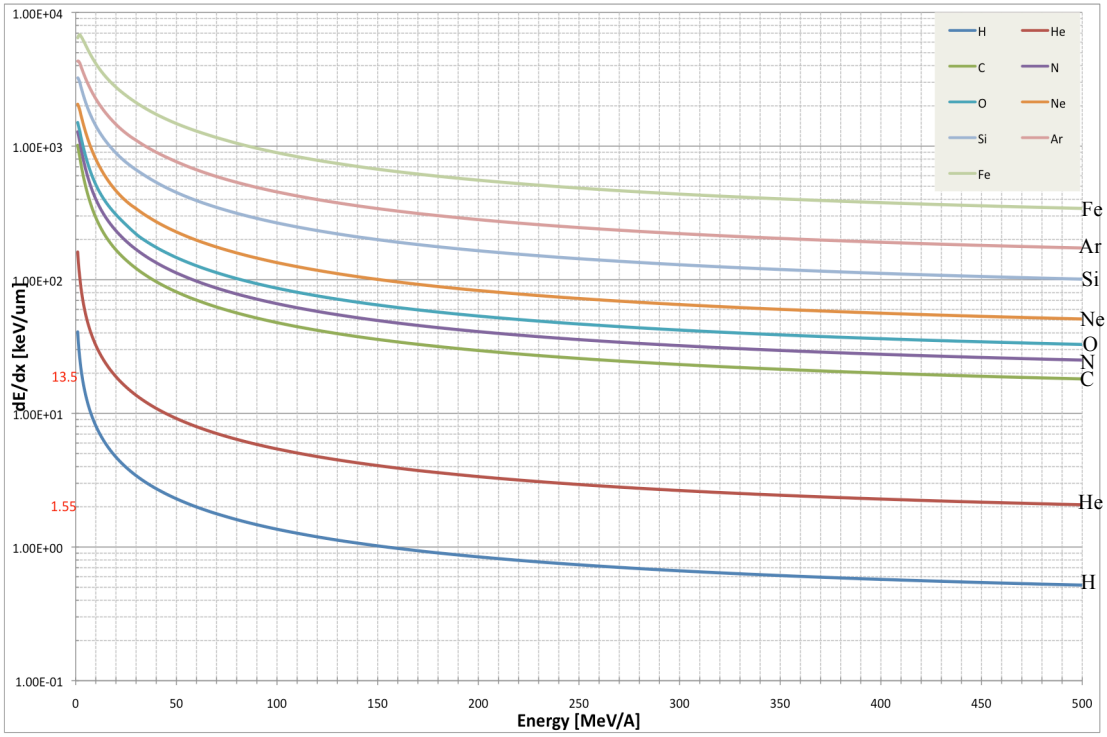


Figure 5.9. dE/dx version energy of heavy-ion particles

Figure 5.9 shows a big gap in LET between H, He and the rest of particles. A LET threshold at 13.5 can be used to isolate H and He from the rest. Note that there is a small range of LET from 13.5 KeV/um to 161 KeV/um (energy <30MeV) in which low energy He and high energy C cannot be distinguished by using solely LET cut. However, using the stopping identification in Section 5.2 and δ-ray analysis in Section 4 is sufficient to separate them apart. The reason is that H, He particles whose energy<30MeV range in the Si layer and produce no δ-rays, while same LET particles such as C, N, O have very high energy producing clearly detectable δ-rays. In the other words, LET feature and δ feature should be sufficient to separate classes into two groups: the first group includes H, He and the second group includes C, N, O, Ne, Si, Ar, Fe. Then, machine learning can be applied in individual group to reduce the complexity of the multiclass classification.

In the first group (H and He), a threshold at 1.55 KeV/um can be used to distinguish high energy H from He. For particles that have the LET in the range from 1.55 KeV/um and 13.5 KeV/um, other features such as distance transform level, minor length and size will be useful to facilitate the classification. A binary classification described in Section 6 will be demonstrated for such case. For the second group, LET cuts can still be used to classify particles if the LET values are estimated accurately. However, due to saturation effects in the current version of Timepix 2 and volcano effects, the stochastic level and the erroneous degree are increasing in according to the heavier of atomic mass and the lower of energy. Therefore, the underlying LET distribution of the data is hard to predict accurately for the time being. In this case, one has to combine all relevant features in addition to LET to support the classification. When the number of dimensions (number of

features) increases, the decision boundary is not easily interpretable by a human being. Machine learning approaches equipped with the capabilities of automatic learning form data should be utilized. Multiclass classification algorithms can be efficiently employed to solve this task. New features should be investigated such as the existence and the impact of volcano effects towards energy deposited.

Figure 5.10 shows a schema of all ion type classification described so far. However, as the training data are not enough to perform the multiclass classification problem on the right hand side of this schema (group 2), we only perform binary classification on the left hand side for H and He classification in this dissertation (group 1). Details will be described in Section 6.

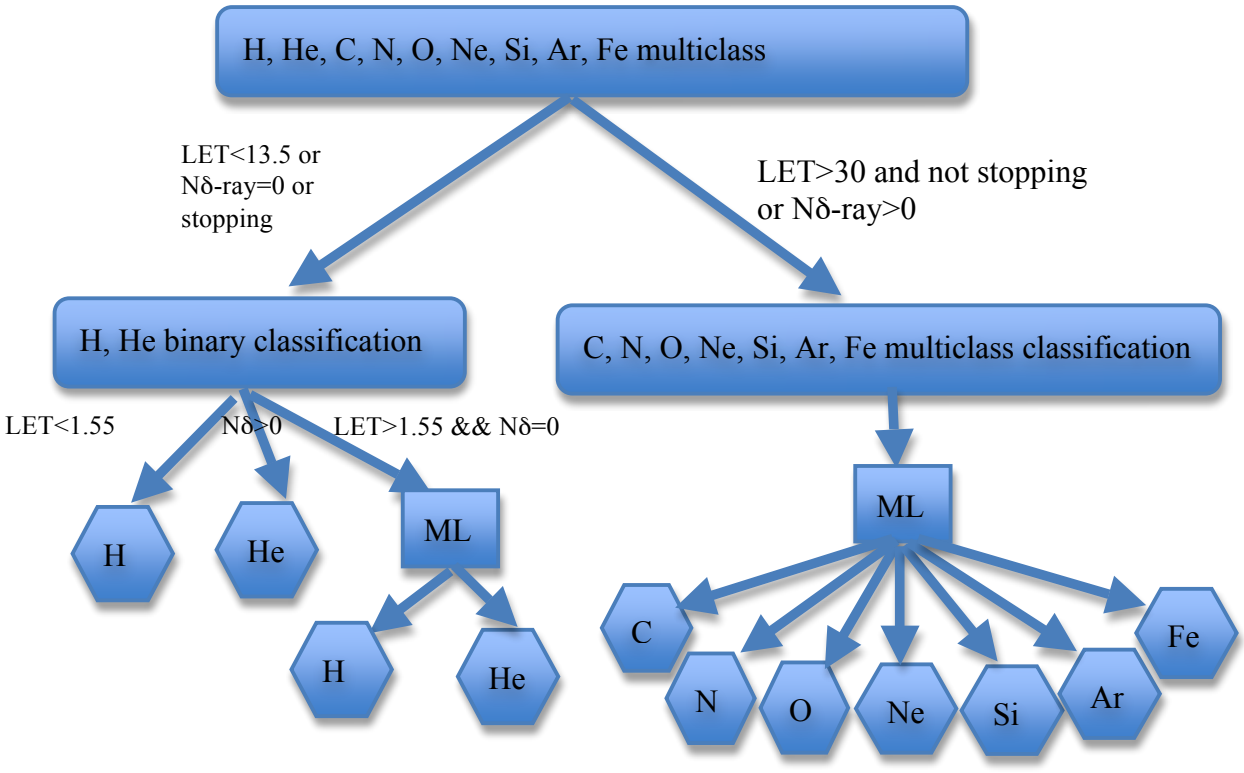


Figure 5.10. Ion type classification schema

6. Machine learning for Ion type classification

As described in Section 5, there are three tasks that need to be done using machine learning methods. They include two binary classification problems (slow H and He, Neutron and genuine He) and one multiclass classification problems (C, N, O, Ne, Si, Ar, Fe). These problems are not simply solved using one or two features. In addition to LET and δ-rays, more features need to be investigated. Discriminant models need to be built with a high dimension classifier. The distribution of the input space is complicated and needs to be automatically learned by invoking machine learning methods.

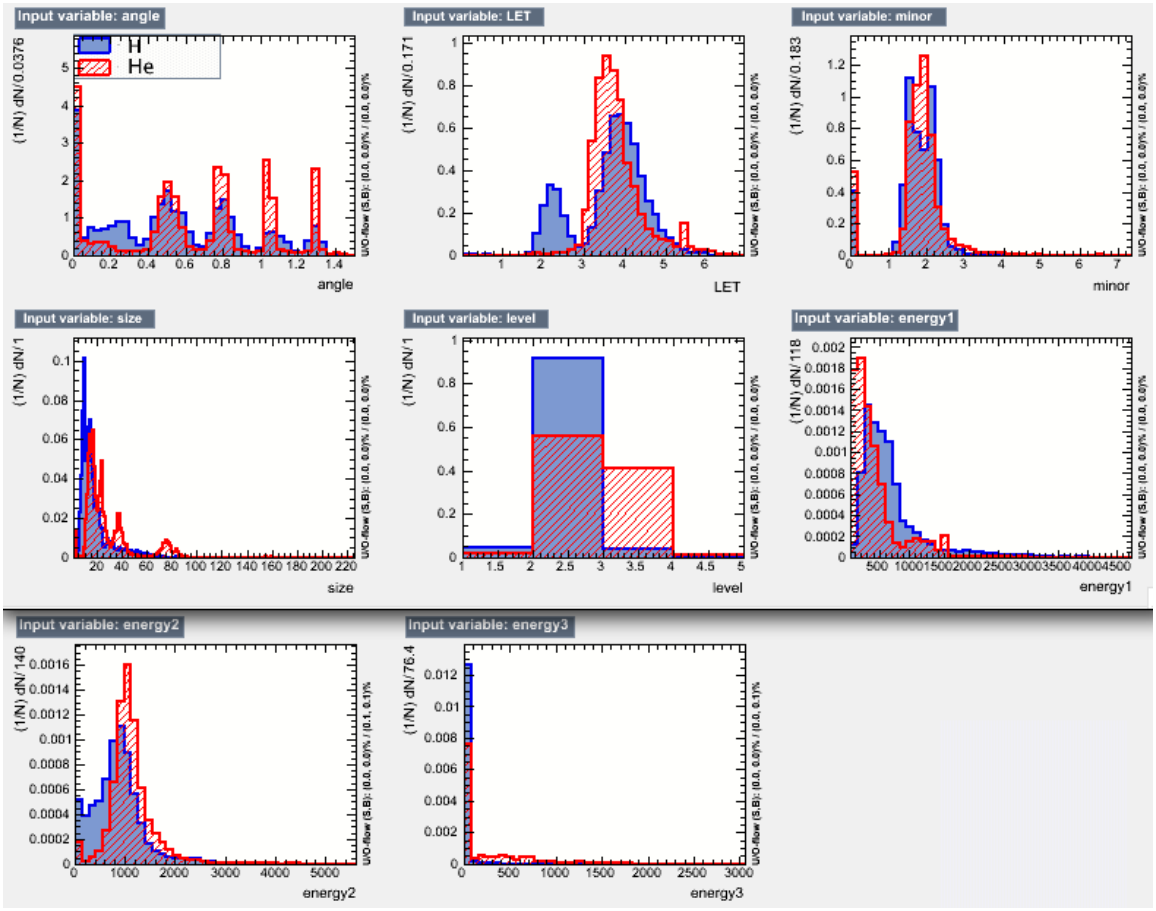


Figure 6.1. Feature distribution of training data of slow H and He

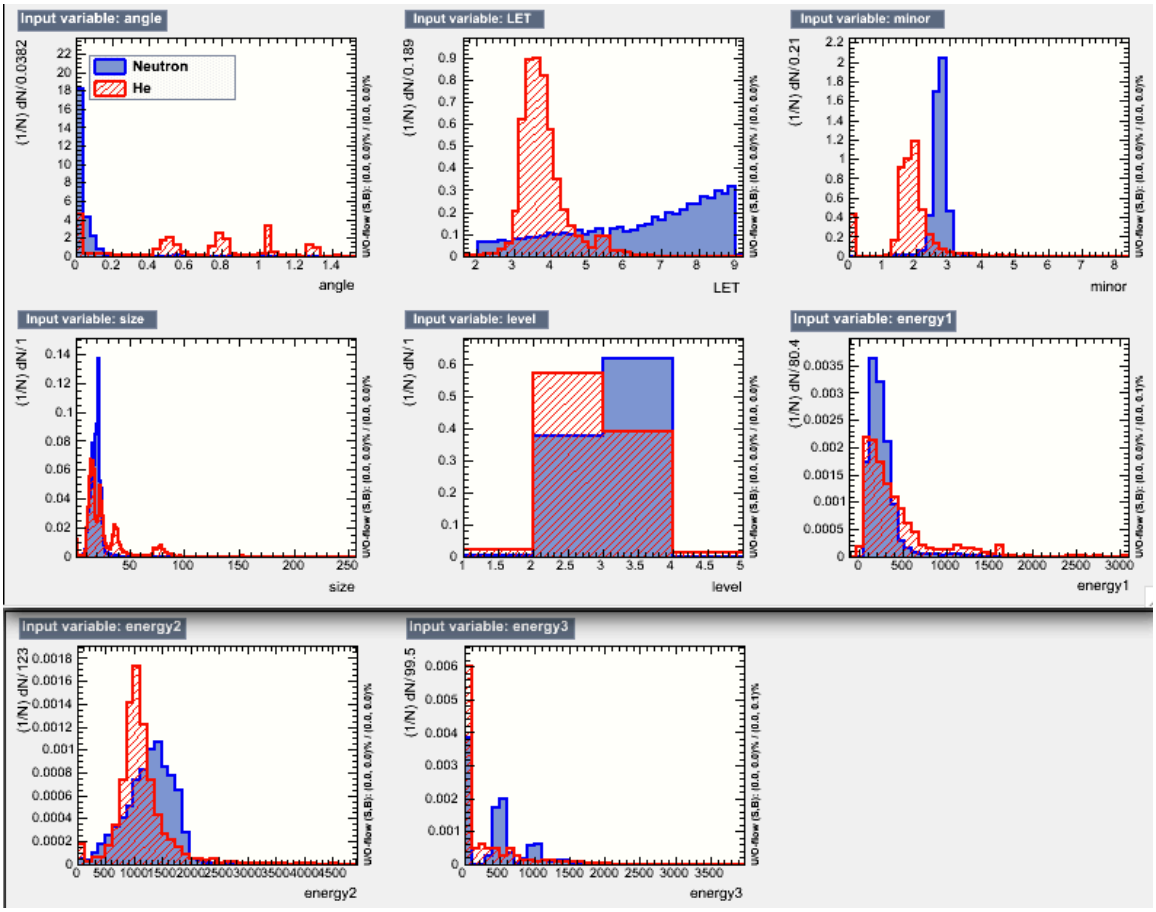


Figure 6.2. Feature distribution of training data of Neutron and He

We assume training data consisting of a set of training points T = {(x,y)}, where x=(a_1, a_2,…, a_n) is a n-feature vector. Each vector x ∈ X is attached a label y ∈ Y in which |Y| = k, a number of classes. In our case, k=2 for binary classification (Y = {slow H, He} or Y = {Neutron, He} representing negative and positive classes), k=7 for heavier ion multiclass classification (Y = {C, N, O, Si, Ne, Ar, Fe}). In training phase, a machine learning method will provide a classifier which is a function of mapping a cluster (represented by the n-feature vector) into one element of class, f: X→Y. This function f can be used to predict unseen clusters in testing phase and application. In this dissertation, we focus on two binary classification problems. Multiclass classification

problems will be investigated in future work. The features (a_1, a_2,…, a_n) will be used for these two binary classification as shown in Figure 6.1 and Figure 6.2

6.1 Data preprocessing

It's necessary to do data preprocessing prior to training and testing. The aim is to understand the correlation between features, learning the individual distribution of each dimension. This step might transform extracted features from input space into a transformed space in order to normalize features and to perform dimensionality reduction. A corresponding transformation needs to be taken that depends upon the specific method. We test methods with the identity transformation, variable normalization and principle component analysis to find the best preprocessing technique.

Variable normalization

The range of dimensions might be different from each other in units. For example, the LET is in KeV/μm while the angle is in radians or degrees. The former dominates the latter. To avoid this problem, the minimum and maximum of values in each dimension are calculated. A linear scale of the values is taken to get a range [-1, 1] for all dimensions.

Principle component analysis

Principal Component Analysis (PCA) is a mathematical technique used to identify the most important gradients. PCA rotates the cloud of data such that the maximum variability is visible. PCA describes observations from a data table in multiple inter-correlated quantitative dependent variables. It allows important information to be

extracted from the table, for the data to be represented as a set of new orthogonal variables called principal components, and for the similarity pattern of observations and of variables to be displayed as points in maps. One can use the PCA technique to reduce dimension of classification problem by selecting the first n-components that have highest variability among the variables.

6.2 Model selection

There are many classification models that can be used for our classification problem. In this dissertation, we explore parametric and non-parametric methods that have been known to give good results in high-energy physics. We select the best model based on several criteria described at the end of this Section. The following models will be explored:

Linear discriminant analysis

This is a simple method that provides data classification based on a linear model. A discriminant function $y(x) = x^T w + w_0$ with the parameter w and the bias term w_0 is built so that $y(x) \geq 0$ for positive class and $y(x) < 0$ for negative class. This method is used as a base line to check whether a simple linear classifier can be applied to our problem or not.

K Nearest Neighbor Classifier (k-NN)

The k-NN algorithm searches for k instances closest to a test instance based on a certain metric. The simplest metric is Euclidian distance. Depending upon the distribution of class label of these k closest instances, a decision will be taken based on a combination or

voting technique. The k-NN classifier is preferred when the decision boundary to separate two classes is complicated that is not approximated by parametric methods. A default value k = 20 has been used in this dissertation.

Artificial Neural Network

Artificial neural networks for classification are computational models that follow the principle of neural system operation. They are composed of units, and connections between units. They are divided into layers. Each node in the first layer (input layer) corresponds to one dimension of the input space. Nodes in the last layer (output layer) represent class labels. The internal layers connecting the input and output layers are called hidden layers. The network can be trained using back-propagation or feed-forward depending on the problem at hand. In this dissertation, we assume feed-forward architecture with one hidden layer. A Bayesian regulator is used to avoid over-fitting [32].

Support Vector Machines

Support Vector Machines use kernel technique to implicitly transform the original input space into a higher dimension space that can be separated by a hyper-plane. The position of the hyper-plane is determined by maximizing the margin between support vectors, a minimal subset of the training set closest to the decision boundary. This machine learning approach provides good performance for a wide variety of problems. It can compete well with Neural Networks and probability density estimator approaches. Moreover, its computation is simplified by using the kernel trick. In this dissertation, we assume Radial

Basis Function as the kernel with default parameters.

Boosted Decision Tree

A decision tree is a binary tree-structured classifier in which internal node represents a test operator on an attribute, each branch represents a sub decision tree or a leaf node representing a class label. Classification rules are defined by paths from root to leaf. A decision tree is built by using information gain or the entropy obtained from the training data set. It's a powerful technique but not stable, as a small change in the training set will make a large change in tree structure.

Boosting of decision trees can be used to overcome this shortcoming by building several trees forming a forest. Misclassified instances are reweighted so that the subsequent tree is built paying more attention to these instances. Finally, all generated trees are combined into a single classifier by a weighted average or a voting technique. The performance of the boosting is enhanced significantly as compared to a single-tree classifier. We employ Gradient Boosted Decision Tree in which the reweighting algorithm uses binomial log-likelihood loss instead of exponential loss to overcome the shortcoming of AdaBoost [33], a popular boosting algorithm, in presence of outliers or mislabeled data points.

The optimal method to be used depends on the problem at hand. There is no classification algorithm that outperforms the others in general. Model selection provides a way to select the best classifier for our problem. Several authors have proposed methods for model selection such as meta learning [9]. In this dissertation, we employ three criteria for

ranking and evaluating the performance of classification methods.

- The Receiver Operating Characteristics (ROC) curve in which true positive (the actual positive instance is predicted as positive) versus true negative (the actual negative instance is predicted as negative) is plotted at different threshold cuts. The larger the area under the curve, the better the performance.
- The separation, defined by the integral:

$$S^2 = \frac{1}{2}\int \frac{(y_P(x) - y_N(x))^2}{y_P(x) + y_N(x)}\,dx$$

  where $y_P(x)$, $y_N(x)$ is the probability density function of positive, negative class, respectively. If the separation =0, the distribution of two classes are completely overlapping. The closer to 1 the separation is, the better the performance.
- The discrimination significant, defined by the difference between classifier means divided by the quadratic sum of root mean square. The larger the significant, the better the performance.

7. Experimental results

We use data taken at the HIMAC (Heavy Ion Medical Accelerator facility in Chiba), Japan, NSRL (NASA Space Radiation Laboratory at Brookhaven) in USA and actual spaceflight data from the ISS (the International Space Station) for our experiments. The data frames at HIMAC and NSRL were taken at different angles and with different particle species. In particular, we show experimental results for H @ 50MeV, H @ 100MeV, He @ 100MeV, He @ 180MeV, C @ 300MeV, C @ 400MeV, N @ 290MeV, O @ 400MeV, Si @ 400MeV, Ar @ 290MeV, Fe @ 400MeV at 0, 30, 60, 75 degree. We typically analyzed several thousand frames, such that each frame contains several clusters associated with corresponding events, but not enough to create a significant likelihood of cluster overlap. Firstly, we compare our proposed angle and LET calculation methods with the expected values for HIMAC and NSRL data. Secondly, we compare dose, dose-equivalent results from our calculation with TEPC for ISS data. Finally, we show classification performance for ion identification.

7.1 Angle and LET evaluation

We evaluate the angles generated by our method and compare them to the true angles, which are generally known to a small fraction of a degree. Then we perform the analysis to determine the LET values for the tracks. The LETs generated from different angles are expected to be the same because the relatively small thickness of the sensor does not typically show a significant variation in the particles LET due to the reduction in the kinetic energy of the incident particle. These results demonstrate the resolution of our algorithm for determining the angle of incidence and LETs for these tracks.

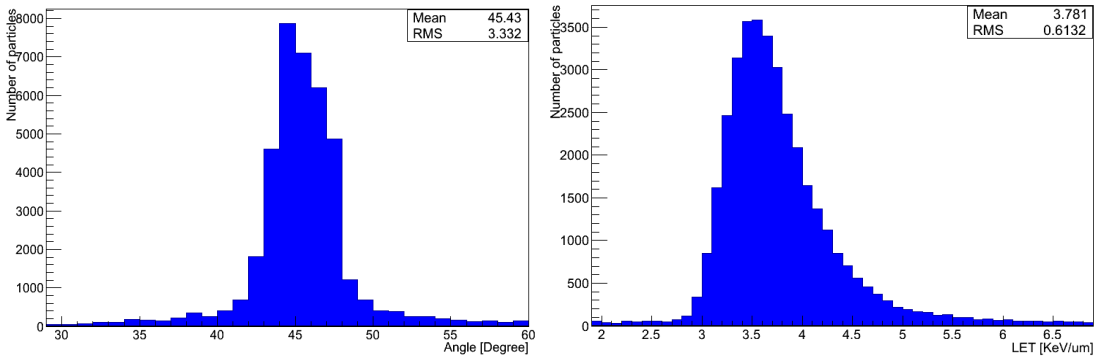


Figure 7.1 Angle and LET distribution at 45 degree of He-180MeV with 300μm sensor

Figure 7.1 is the angle and LET distributions of 45 degree of He-180MeV incidences, respectively. The results have the Gaussian distribution oscillating around a mean value with a standard deviation. The mean value is very close to the true angle and the standard deviation is small enough such that the error is acceptable. Table 7.1 shows the experimental angles compared to true ones for different angles and different ionizing particles. Cells in this table are left empty if data are not available.

In fact, if the LET is less than 10 then the resulting quality factor is approximately 1. The reason is that we define the LET by the total energy deposited (E) divided by the track length (L) of traversed matter, converted for tissue equivalent as described earlier. In our application, since L is always greater than 300μm (our sensor thickness), the quality factor will be approximately 1 if E is less than 3MeV. Therefore, we ignore clusters that have total energy deposited less than 3MeV and arbitrarily assign a quality factor of 1.

We can see from Table 7.1 that the errors in the angles are larger for vertical tracks. The primary reason is that the current generation of Timepix detectors has a pixel size of 55 μm and a 300 μm thick Si sensor layer, which makes it difficult to distinguish a 0 degree

incidence and a 15 degree incidence. We hope an improvement will be made with the next generation, which may have smaller-sized pixels.

However, in an isotropic flux of incident particles, the solid angle acceptance for vertical tracks is small, so the relative number will not be large. In addition, an estimate of the track length is relatively insensitive to the error in the angle of incidence near vertical, and is less than, or on the order, of the stochastic variation in the energy loss of the primary ions. So the errors on vertical track angle determination can be tolerated in determining the dose equivalent at this point.

We also took data at HIMAC with detector having a 1mm sensor thickness. The angle resolution is much better than the one with 300 μm sensor thickness as shown in Figure 7.2. For incident angles greater than 5 degree, the algorithm works perfectly with a very small degree of error. This can prove the efficiency of our approach for angle calculation with different settings of Timepix chip.

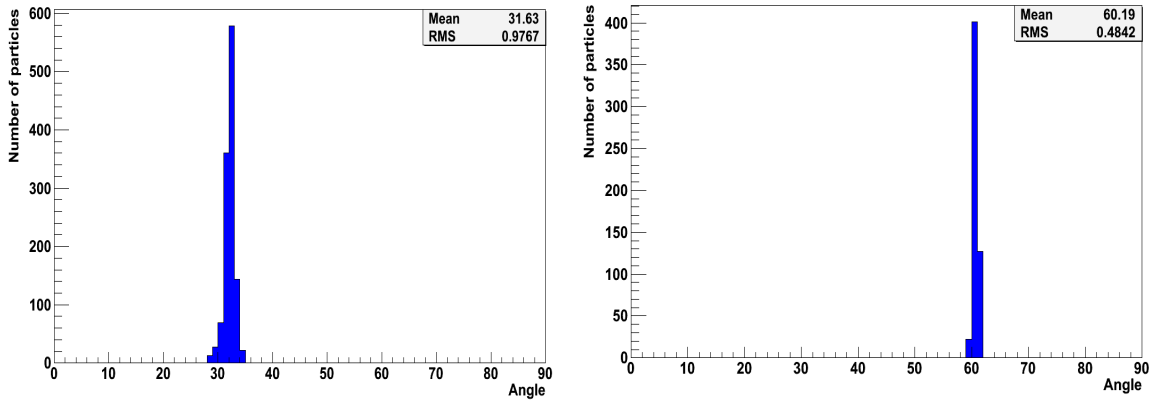


Figure 7.2 Angle distribution at 30 and 60 degree of He-230MeV

with 1mm sensor thickness

Table 7.1. Angle resolution

	0 degree		30 degree		60 degree		75 degree	
	Mean	RMS	**Mean**	RMS	**Mean**	RMS	**Mean**	RMS
H-50MeV	**8.4**	6.9	**28**	3.5	**59.7**	1.7	**74.2**	1.56
H-100MeV	**0.6**	1.3	**30.2**	3.8	**59.7**	1.2	**74.5**	1.3
He-100MeV	**2.1**	6.8	**30.2**	3.7	**60.1**	1.3	**74.7**	1.5
He-180MeV	**4.7**	7.2	**29.3**	3.5	**60.3**	1.4	**74.6**	1.04
C-300MeV			**29.7**	4.3	**60.9**	3.6		
C-400MeV	**2.1**	5.2			**60.4**	1.2		
O-430MeV	**4.3**	6.2	**29.7**	3.6	**60.1**	1.3	**74.7**	0.6
N-290MeV	**4.3**	6.6	**29.1**	3.6	**60.5**	1.1	**75.2**	0.6
Si-400MeV	**2.6**	3.5	**25.5**	4.6	**59.5**	1.8		
Ar-290MeV	**4.9**	7.6	**20.01**	6.38	**59.01**	1.57		
Fe-400MeV	**3.8**	4.5	**24.1**	5.2	**61.6**	1.1	**75.9**	1.5

Table 7.2 also shows the LET estimation using angular resolutions above, combined with total energy deposited from energy calibration method. In this table, the calculated LET values of the particles from H (50, 100MeV), He (100, 180MeV), C (300MeV), O (430MeV), N (290MeV) are very close to the ones calculated from SRIM simulation [26] even for different angles. Volcano effect does not occur with these particles. However, there is saturation effect affecting O (430MeV) and N (290MeV) that is not corrected completely. This effect will be removed when the next generation Timepix 3 is ready for use. In particular, this table shows very good results for LET values for H and He that are the most abundant particles that need to be carefully investigated.

Table 7.2. LET resolution

	SRIM	0 degree		30 degree		60 degree		75 degree	
	KeV/um	**Mean**	RMS	**Mean**	RMS	**Mean**	RMS	**Mean**	RMS
H-50	2.3	**2.26**	0.27	**2.27**	0.28	**2.27**	0.25	**2.35**	0.24
H-100	1.36	**1.4**	0.3	**1.3**	0.3	**1.26**	0.3	**1.3**	0.3
He-100	5.42	**5.5**	0.75	**5.67**	0.54	**5.7**	0.6	**6.0**	0.54
He-180	3.65	**3.65**	0.5	**3.8**	0.6	**3.7**	0.5	**3.76**	0.4
C-300	23.0			**25.4**	3.6	**26.6**	4.1		
C-400	20.0	**20.01**	2.8			**20.6**	2.2		
O-430	35.0	**32.2**	7.7	**38.8**	5.4	**39.6**	4.8	**41.0**	4.4
N-290	32.0	**31.0**	7.6	**29.1**	2.8	**31.9**	2.7	**33.0**	2.8
Si-400	110.0	**106.0** **57.2**	18.8 5.1	**113** **65.5**	24.2 6.1	**125.0** **74.9**	17.0 4.9		
Ar-290	225	**219.2** **88.15**	18.9 8.3	**237.5** **97.6**	19.5 6.6	**218.5** **102**	15.8 5.98		
Fe-400	377.0	**374.0** **126.3**	40.4 8.2	**380.0** **137**	49.0 9.5	**328.0** **134.7**	40.5 12.1	**340.0** **141.8**	23.8 11.2

However, due to the plasma-effect [22], which may come in part from the recombination of high densities of electrons and holes, the estimate of the amount of energy deposited is reduced for particles at the end of the table (Si-400MeV, Ar-290MeV, Fe-400MeV), although the angles are well-estimated in Table 7.1. This effect occurs to the greatest extent in the regions of greatest ion density in the sensor, and where the bias voltage field direction is parallel to the track direction.

The red numbers in this table represents raw calculated LET values without compensation for the plasma effect. The black numbers of the same cell show results after making a correction using the method described in Section 3. The corrected mean values of these are close to what is expected. The RMS representing the spread of statistical LET distribution is still big due to the saturation effect. The coming new version Timepix 3 will solve this problem by removing the saturation effect and by providing a flat response region on the top of the volcano.

7.2 Dose, Dose-Equivalent evaluation

We use the ISS data to evaluate Dose, Dose-Equivalent from our calculation with the calculation of The Tissue Equivalent Proportional Counter (TEPC) [31]. Tissue Equivalent Proportional Counter (TEPC) is an automatic gas-based microdosimetry system that is designed to characterize the space radiation environment. The mission of TEPC is to collect statistics of radiation environment of the International Space Station (ISS) to construct exposure records for the crew, to provide near real-time measurements during radiation events, and to perform shield verifications. Each TEPC consists of two main components, the spectrometer and the detector unit. The spectrometer allows real-time analysis of the data and calculations of total dose, total dose equivalent and incremental dose for penetrating radiation in space.

The comparison between TEPC and REM units is necessary to evaluate the performance of REM units and algorithms. We have had 5 REM units operating on ISS that transfer radiation data to ground for doing analysis on a daily basis. They are D03-W0094, E06-W0087, I03-W0094, I04-W0094, G03-W0094. In these detectors, G03-W0094 is half

covered by a thin ^{6}LiF layer in order to detect thermal neutrons. The other four units will be used to evaluate the performance of hardware and algorithms in terms of Dose and Dose-Equivalent estimates.

Figure 7.3 shows a comparison of dose rate in μGy/minute among 4 REM units and TEPC in GMT 296, 2012. A good match can be observed when looking at the trend of the five curves in this graph. The variance between them is not significant. The graph provides evidence to show the consistency and compatibly of out dose rate calculation methods with the existing NASA hardware.

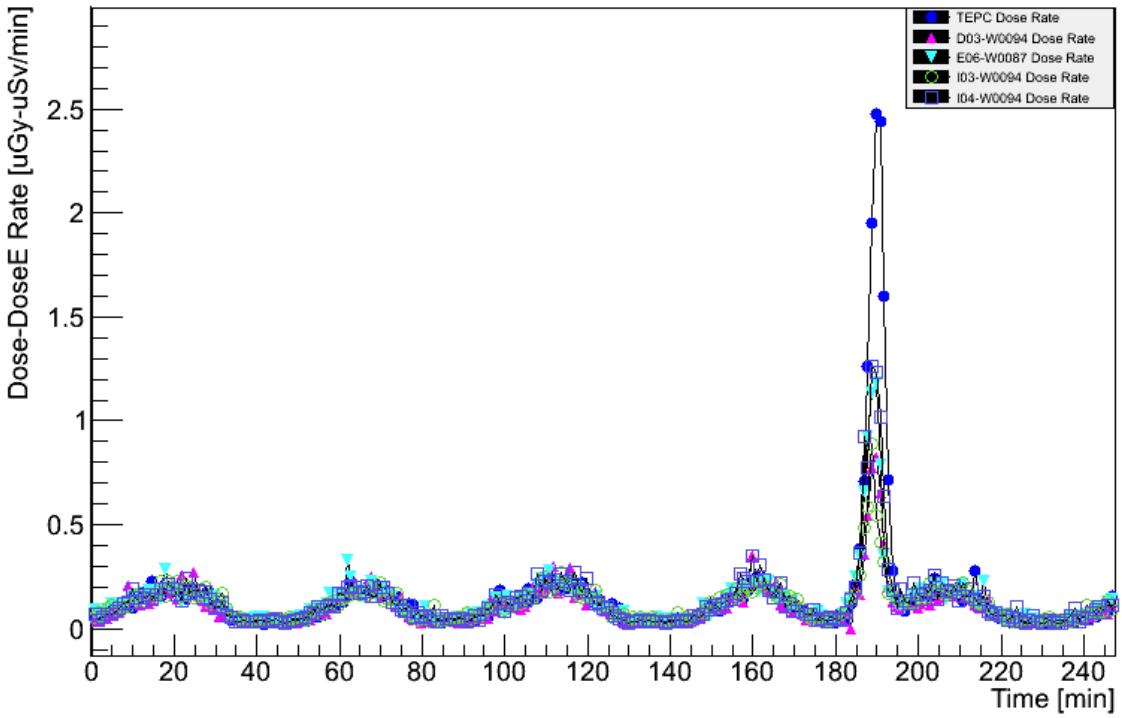


Figure 7.3. REM-TEPC Dose comparison

The comparison of dose equivalent rate in μSv/minute is shown in Figure 7.4. It should be noted that there are differences in the ways the TEPC calculates dose-equivalent from

ours. We calculate dose-equivalent by multiplying the dose with a quality factor that is a function of LET for each individual track. The TEPC analysis makes a global estimate of the average LET over a large number of hits. For the time being, our LET calculation for slow and heavy-ion particles (Si, Ar, Fe…) is as yet still not fully corrected for all of the effects mentioned earlier. Also, the statistical and systematical errors contribute significantly to the LET calculation. Therefore, the variance of curves in Figure 7.4 is larger than the one in Figure 7.3 although the general trend remains.

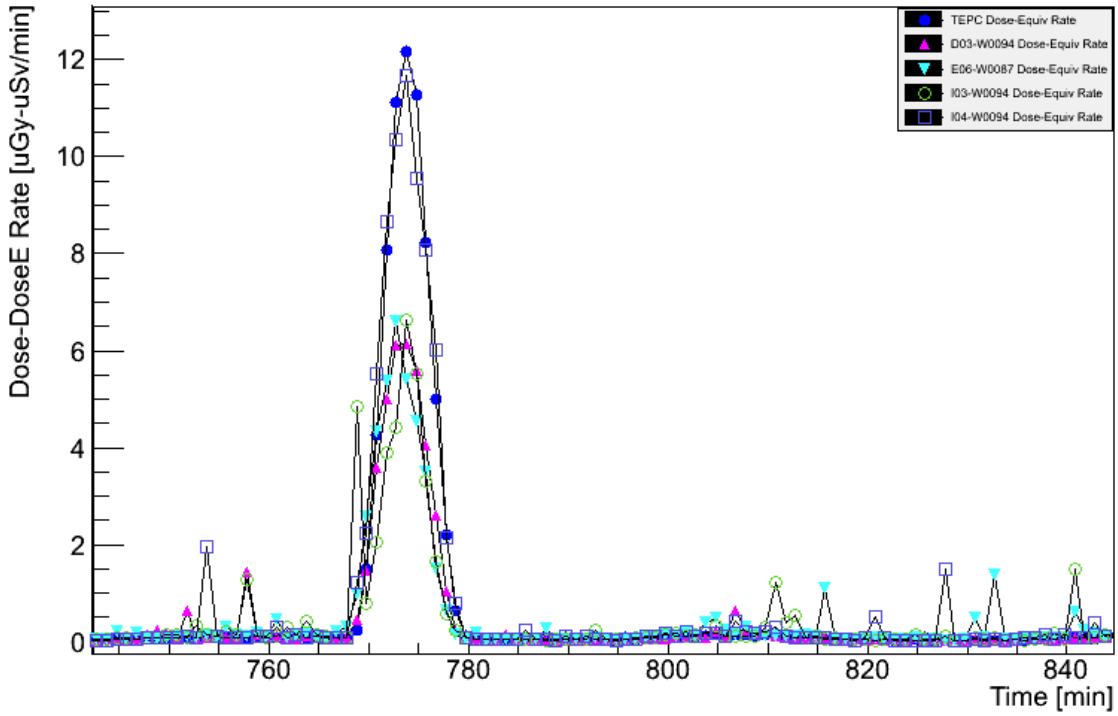


Figure 7.4. REM-TEPC Dose-Equivalent comparison

Figure 7.5 and Figure 7.6 show dose, dose-equivalent maps, respectively of ISS REM data versus ISS latitude and longitude. We calculate ISS orbit position of vehicle (longitude and latitude) based on the timestamp when taking the data frame and relied upon the SGP4 NORAD orbit propagator code. The calculation is automatically done in

the conversion process at the end of the day when ISS data are transferred from NASA servers to TLC2 server at University of Houston. The hot spots in these maps correspond to the higher dose, dose-equivalent rates during South Atlantic Anomaly (SAA) transit. The SAA is of particular interest in high-energy astrophysics because there is very dense area of trapped radiation above the South Atlantic Ocean.

As the Earth's magnetic field is not symmetric with the rotation axis and it is also offset away from the center of the Earth. This offset allows trapped particle trajectories to reach lower altitudes in this region than elsewhere. The highest radiation doses are typically seen during low altitude Earth orbits in this region. Our detectors were able to survive and provide dose rate and dose equivalent rates as expected during SAA passes.

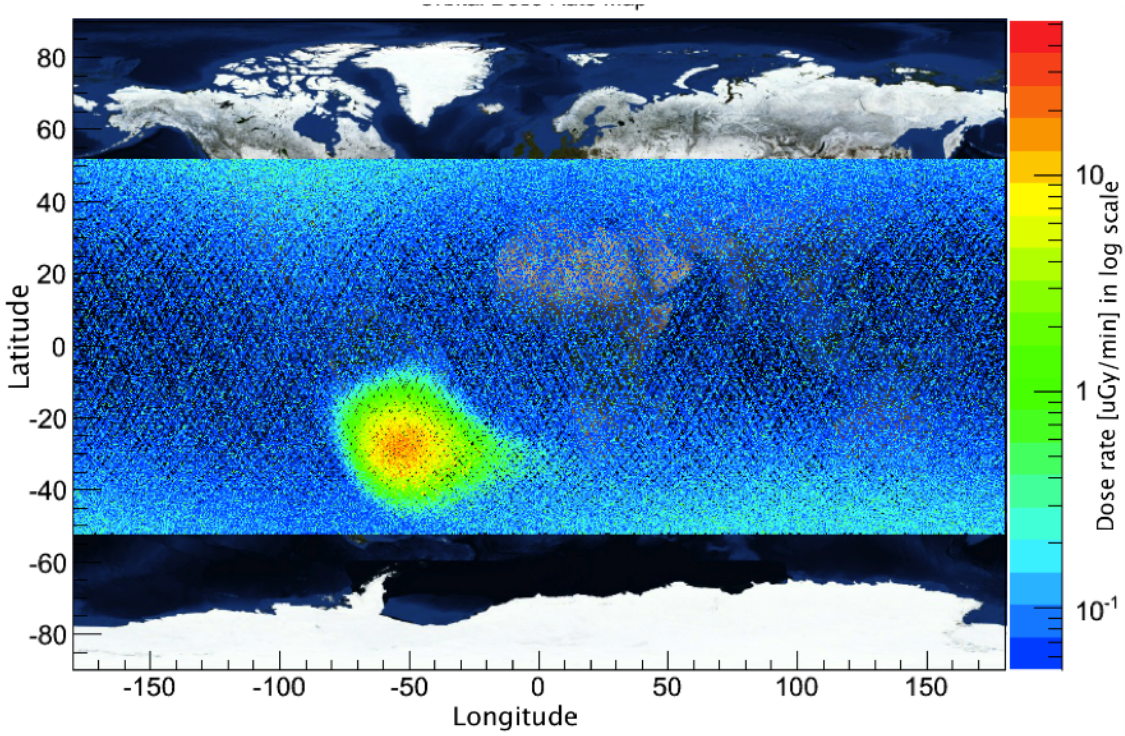


Figure 7.5. Orbital Dose Rate Map

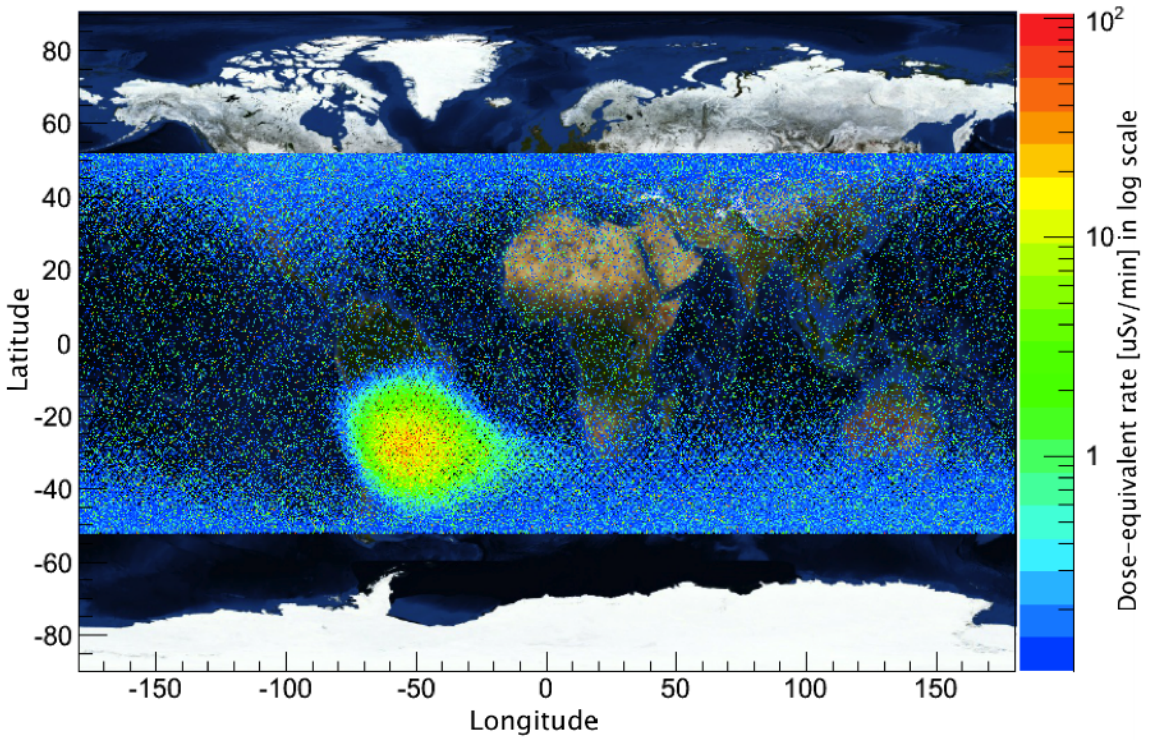


Figure 7.6. Orbital Dose-Equivalent Rate Map

7.3 Ion type classification evaluation

We investigated methods for two important binary classification problems: Slow H and He classification, the alpha particle classification from neutron and genuine He. Our analysis based on three criteria: ROC-integral, separation, and discriminant significant to select the best method for each problem. Five popular machine-learning methods were employed: Gradient Boosted Decision Tree (BDTG), Artificial Neuron Network with Bayesian Regulator (MLPBNN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Linear Discriminant (LD). The BDTG is considered to be the best method for multivariate data analysis in high-energy physics. The LD is used to provide a base line for classification performance. We use TMVA, toolkit for multivariate data analysis in doing experiments [29]. We extracted randomly 10000 events (5000 events for training

and 5000 event for testing). We run the experiments several times with different random seeds. Parameters of methods were chosen as default settings and not tuned up.

Table 7.3. Evaluation results for slow H and He classification

	ROC-integral	Separation	Significant
BDTG	**0.990**	0.867	3.565
MLPBNN	0.988	0.864	2.836
KNN	0.984	0.834	3.141
SVM	0.965	0.768	2.141
LD	0.929	0.636	1.214

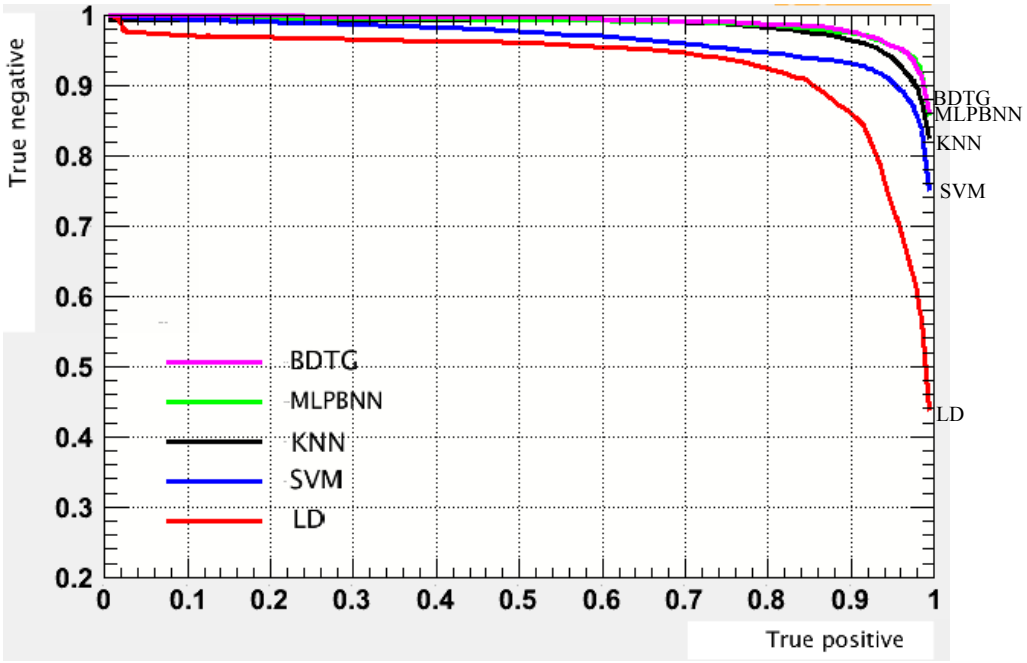


Figure 7.7. ROC curves for slow H and He classification

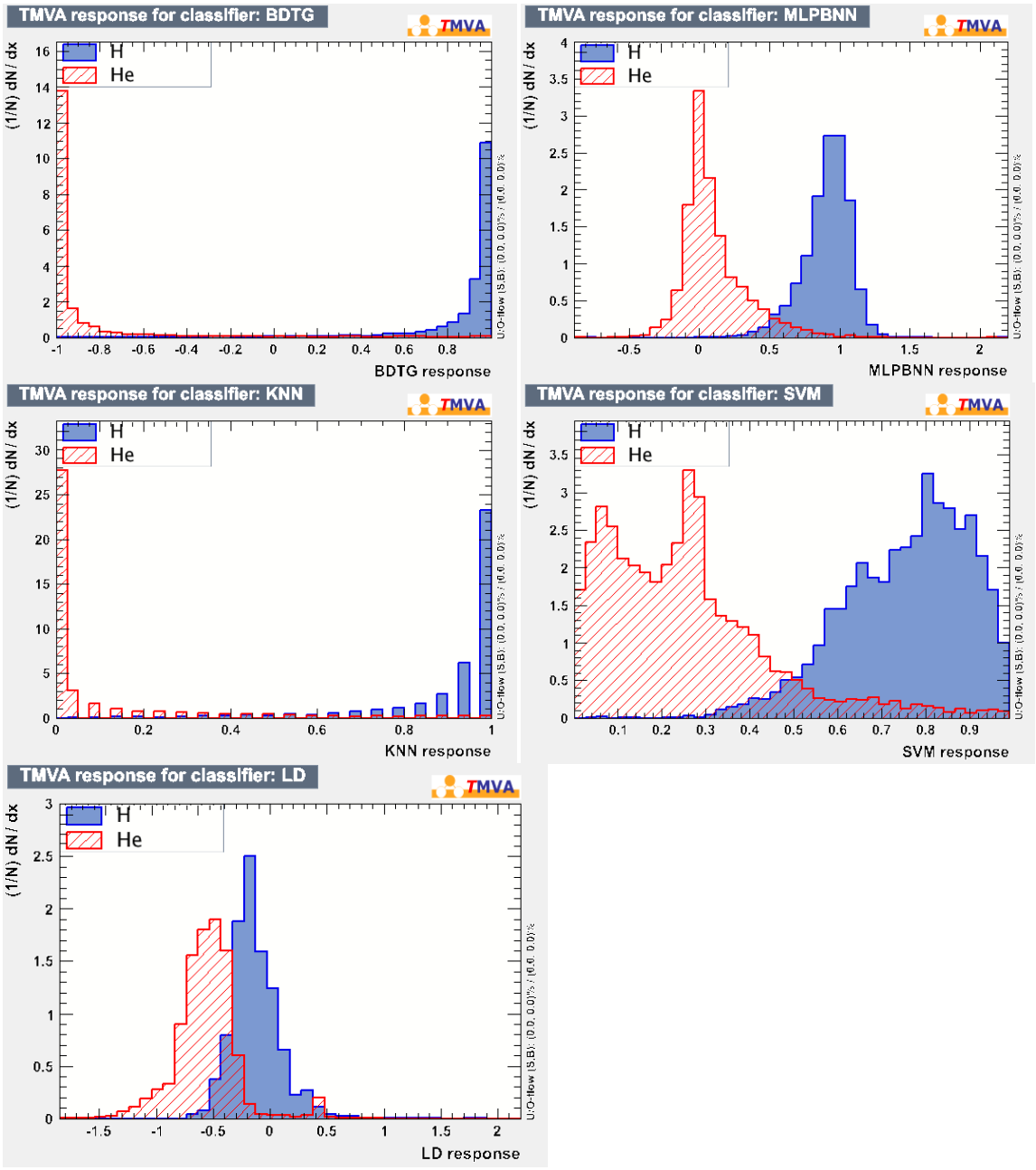


Figure 7.8. Method's responses for slow H and He classification

Table 7.3 shows evaluation results along with ranking methods according to the priority: ROC-integral, separation, significant. Figure 7.8 shows the distribution of responses for all employed methods. This distribution is used to determine a cut value for

distinguishing class labels. Based on these results, BDTG was chosen as the best candidate for classifying ion type of slow H and He. The separation of testing data using Linear Discriminant method is quite small (0.636), which means that a simple hyper-plane cannot separate well slow H data and He data using extracted features. BDTG is the best method with separation of 0.867, which good enough for our classification purpose.

Table 7.4. Evaluation results for Neutron and He classification

	ROC-integral	Separation	Significant
BDTG	**0.999**	0.964	7.608
KNN	0.999	0.958	7.018
SVM	0.995	0.935	3.996
MLPBNN	0.993	0.935	4.662
LD	0.988	0.873	2.656

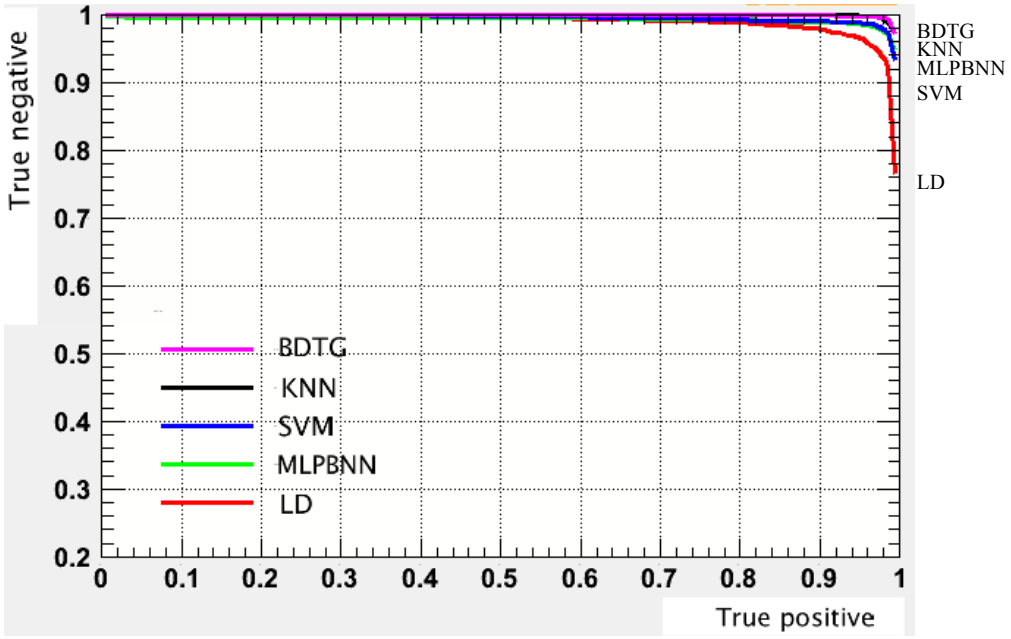


Figure 7.9. ROC curves Neutron and He classification

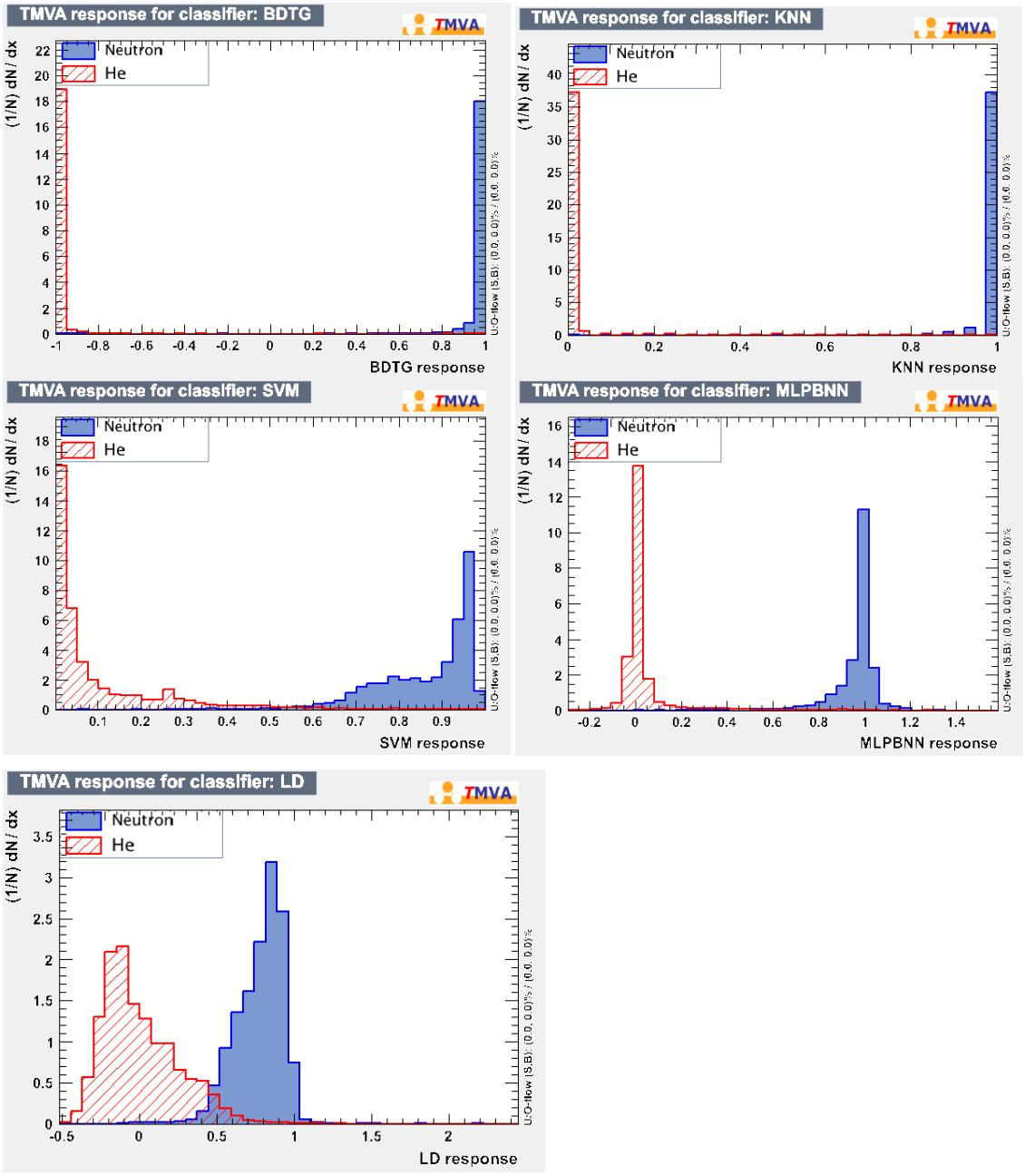


Figure 7.10. Method's responds for Neutron and He classifiaction

Table 7.4, Figure 7.9 and Figure 7.10 show results after invoking machine-learning methods for Neutron and He classification. The obtained performance is much higher as compared to the previous binary problem. Again, BDTG yields the best performance for

this classification problem with the separation 0.964. The separation of LD is 0.873, which means that even a simple hyper-plane can be reasonably used to separate neutron and He. This can be explained by looking at the feature distribution in Figure 6.1 and Figure 6.2. Feature distributions in Figure 6.1 between two classes are almost overlapping, causing more difficulty for the classifiers. The difference between profiles of a stopping particle and a penetrating particle (Figure 5.8) helps improving performance for neutron and He classifiers. The most important feature in this case is the minor projected track length as shown in Figure 6.2.

In summary, in two binary classification problems we are facing, BDTG is proved to be the best candidate and will be used as a final classifier in our application. The parameters for this method will be investigated in more detail in future work. More features need to be explored to improve the performance of the slow H and He classification.

8. Conclusion

In this dissertation, we provided methods and algorithms to not only measure dosimetric endpoints directly such as dose-equivalent, but also determine the physical nature of the radiation field. In term of dose-equivalent calculation, we proposed the energy compensation method for heavy-ions suffering from volcano effects. We also proposed the incident angle determination algorithm to facilitate LET estimation, quality factor and effective dose-equivalent measurement. Energy of penetrating particles is estimated as well to convert dose, dose-equivalent in Si into the tissue or water. The comparison of the results between the REM units and the TEPC on the ISS, and the world map display for dose and dose-equivalent show high accuracy of our methods.

In terms of determining the physical nature of source of the radiation field, we first provided pattern recognition techniques to extract features that characterize the nature of tracks of particles. Skeletonization and distance transform have been performed to derive features. We also recognized patterns that can be used to identify interaction, stopping, and neutron events. Machine learning algorithms were then applied to automatically learn underlying data distribution and to classify source of tracks with high accuracy.

Future work will extract more features and automate the detailed classification of the incident ionizing radiation for heavier ions (C, N, O, Si, Fe) using a multiclass classification approach. Also, we have to explore the core and the skirt of a cluster and extract more δ-ray information overlapping with these areas. A detailed list of future work as a continuation of the accomplishments reported in this dissertation is as follows:

- Apply energy compensation method to Timepix 3 when it is ready.

- Explore more relevant features that facilitate discriminating types of particles (skirt and core). Note that these features can be used for a machine learning approach and for a statistical purpose in physics as well.
- Do feature selection and model selection for multiclass classification.
- Optimize algorithms for an embedded system (DSP/ARM).
- Upgrade the software for data analysis and data taking to be stable for taking and processing ISS, HIMAC, and NSRL data.

References

[1] X. Llopart, R. Ballabriga, M. Campbell, L. Tlustos and W. Wong, Timepix, *a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements*. Nucl. Inst. And Meth. Phys. Res., A 581, 485-494, 2007

[2] J. Jakubek, Precise Energy Calibration of Pixel Detector Working in Time-Over-Threshold Mode. Nucl. Instr. and Meth. A 633 262-266, 2011

[3] U. Grenander, *Elements of Pattern Theory*. The John Hopkins University Press, 1996.

[4] K. Stephen. *Psychological Processes in Pattern Recognition*. Academic Press, New York, 1973

[5] C. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006

[6] V. Cherkassky, *Learning from Data*. John Wiley, 2007

[7] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, 1990

[8] K. Fu. Syntactic Pattern Recognition and Applications. Prentice-Hall, Englewood Cliffs, New Jersey, 1982

[9] P. Brazdil, C. Giraud-Carrier, C. Soares, R. Vilalta. *Metalearning: Applications to Data Mining*. Springer Verlag. ISBN: 978-3-540-73262-4, 2008

[10] R. Kohavi. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. *IJCAI*, vol. 14, no. 2, 1137-1145, 1995

[11] J. Jakubek, *Semiconductor Pixel detectors and their applications in life sciences*. Pixel 2008 International Workshop, Batavia, IL, USA, 2008

[12] T. Holy, E. Heijne, J. Jakubek, S. Pospisil, J. Uher and Z. Vykydal, *Pattern recognition of tracks induced by individual quanta of ionzining radiation in Medipix2 silicon detector*. Nucl. Instr. and

Meth. A 598 287-290, 2008

[13] J. Bouchami, A. Gutierrez, T. Holy, A. Houdayer, J. Jakubek, C. Lebel, C. Leroy, J. Macana, J.-P. Martin, S. Pospisil, S. Prak, P. Sabella, C. Teyssier, *Measurement of pattern recognition efficiency of tracks generated by ionizing radiation in a Medipix2 device*. Nucl. Inst. And Meth. Phys. Res., A 633, S187–S189, 2011

[14] C. Teyssier, J. Bouchami, F. Dallaire, J. Idarraga, C. Leroy, S. Pospisil, J. Solc, O. Scallon, Z. Vykydal, *Performance of the Medipix and Timepix devices for the recognition of electron-gamma radiation fields* . Nucl. Inst. And Meth. Phys. Res., A 650, 92–97, 2011

[15] A. Chatterjee and H. J. Schaefer, Rad. And Environm. *Mierodosimetrie Structure of Heavy Ion Tracks in Tissue Biophys*. 13, 215-227, 1976

[16] J. Jakubek, C. Granja, O. Jäkel, M. Martisikova, S. Pospíšil, *Detection and Track Visualization of Primary and Secondary Radiation in Hadron Therapy Beams with the Pixel Detector Timepix.* Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE, 1967-1969, 2010

[17] L. Pinsky, A. Empl, A. Gutierrez, J. Jakubek, H. Kitamura, J. Miller, C. Leroy, N. Stoffle, S. Pospisil, Y. Uchihori, N. Yasuda and C. Zeitlin *Penetrating heavy ion charge and velocity discrimination with a TimePix-based Si detector (for space radiation applications)*. Nucl. Instr. and Meth. A 633 190-193, 2011

[18] R. Vilalta, S. Kuchibhotla, R. Valerio, and L. Pinsky. *Development of pattern recognition software for tracks of ionizing radiation in medipix2-based (timepix) pixel detector device*. In 18th International Conference on Computing in High Energy and Nuclear Physics (CHEP-10), Taipei, Taiwan, 2010

[19] R. Vilalta, S. Kuchibhotla, F. Ocegueda Hernandez, S. Hoang, and L. Pinsky. *Machine learning for identification of sources of ionizing radiation during space missions*. In International Joint Conference on Artificial Intelligence, Workshop on AI in Space: Intelligence Beyond Planet Earth, 2011

[20] J. Bouchami, A. Gutierrez, A. Houdayer, J. Jakubek, C. Lebel, C. Leroy, J. Macana, P. Martin, M. Platkevic, S. Pospısil, C. Teyssier. *Study of the charge sharing in silicon pixel detector by means of heavy ionizing particles interacting with a Medipix2 device*. Nuclear instruments and Methods in Physics Research A 633 117-120, 2011.

[21] M. Kroupa, J. Jakubek, F. Krejci. *Charge Collection Characterization with Semiconductor Pixel Detector Timepix*. IEEE Nuclear Science Symposium Coferrence Record, 2008.

[22] R. Williams and E. Lawson. *The plasma effect in silicon semiconductor radiation detectors*. Nucl. Instr. and Meth. 120 261-268, 1974

[23] L. Zadeh Fuzzy sets. Information and Control 8 338–353, 1965

[24] J. Lakshmi and M. Valli. *A Survey on skeletons in digital image processing.* In the international conference proceedings of IEEE Computer Society, 260-269, 2009

[25] R. Fabbri, L. Costa, J. Torelli, O. Bruno*. 2D Euclidean distance transform algorithms: A comparative survey.* ACM Computing Surveys 40, 1, 2008

[26] J. Ziegler, M. Ziegler, J. Biersack. *SRIM–The stopping and range of ions in matter*. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 268, no. 11, 1818-1823, 2010

[27] Bortfeld, Thomas. *An analytical approximation of the Bragg curve for therapeutic proton beams*. Medical physics 24: 2024, 1997

[28] L. Pinsky. *Converting Medipix2 (TimePix) Outputs to Dose and Dose Equivalents*. Technical Report 2011

[29] A. Hoecker, S. Peter, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes. *Tmva-toolkit for multivariate data analysis*. arXiv preprint physics/0703039, 2007

[30] R. Brun, F. Rademakers. *ROOT—an object oriented data analysis framework*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and

Associated Equipment 389, no. 1, 81-86, 1997

[31] S. Gerdung, P. Pihet, J. E. Grindborg, H. Roos, U. J. Schrewe, H. Schuhmacher. *Operation and application of tissue equivalent proportional counters*. Radiation protection dosimetry 61, no. 4, 381-404, 1995

[32] J. Zhong, R. Huang, S. Lee. *A program for the Bayesian Neural Network in the ROOT framework*. Computer Physics Communications 182, no. 12, 2655-2660, 2011

[33] Y. Freund, R. Schapire, N. Abe. *A short introduction to boosting*. Journal-Japanese Society For Artificial Intelligence 14, no. 771-780, 1999

Publications

[1] **Hoang S.**, Vilalta R., Pinsky L., Kroupa M., Stoffle N., Idarraga J., *Data Analysis of Tracks of Heavy Ion Particles in Timepix Detectors.* The 15th International Workshop on Advanced Computing and Analysis Techniques in Physics (ACAT-13), Beijing, China, 2013.

[2] **Hoang S.**, Pinsky L., Vilalta R. (2012), *LET Estimation for Heavy Ion Particles based on a Timepix-based Si Detector*. International Conference on Computing in High Energy and Nuclear Physics (CHEP-12), New York, USA.

[3] Stoffle N., Pinsky L., **Hoang S.,** Idarraga J., Kroupa M., Jakubek J., Turecek D., Pospisil S., *Initial Results on Charge and Velocity Discrimination for Heavy Ions using Silicon Timepix Detectors.* In the 14th International Workshop on Radiation Imaging Detectors, Figueira da Foz, Coimbra, Portugal, 2012.

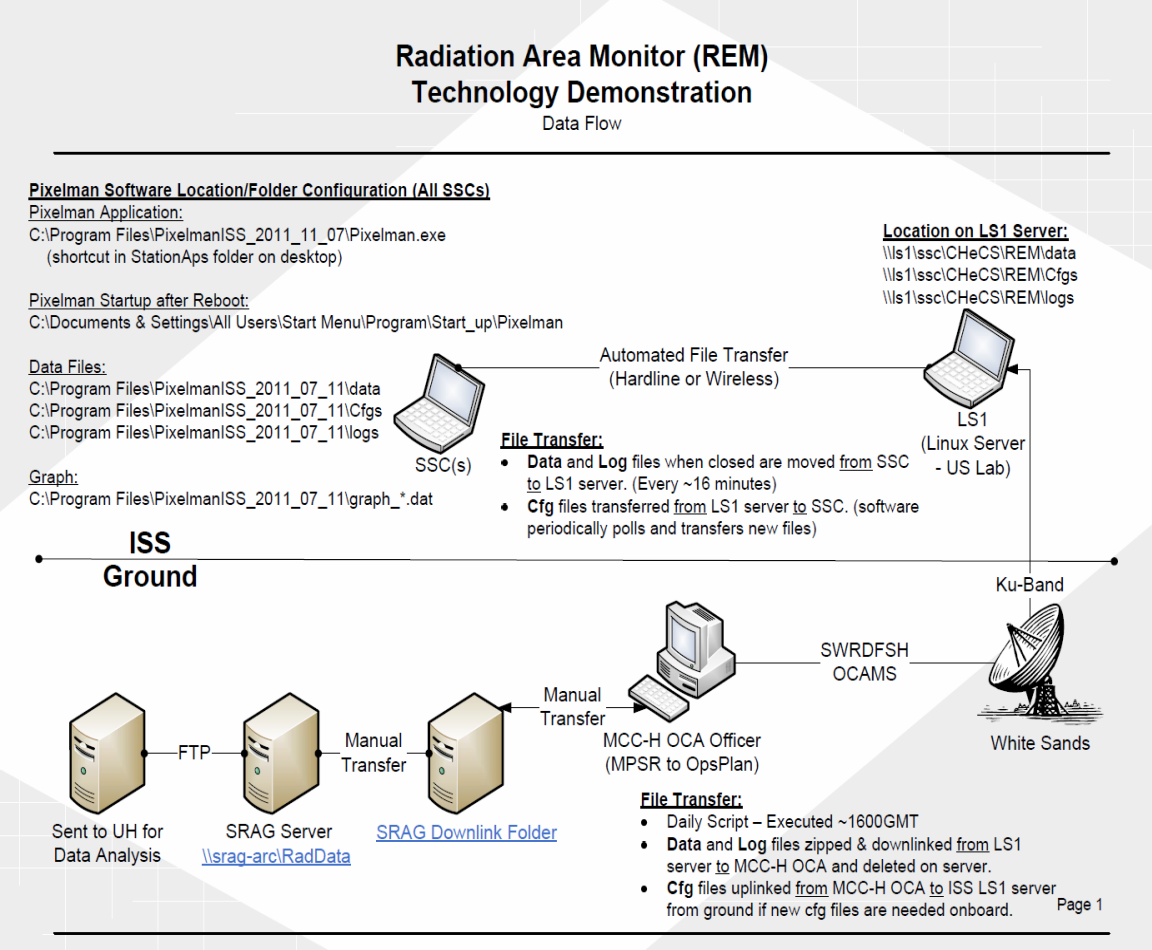
[4] Pinsky L., Empl A., **Hoang S.**, Stoffle N., Jakubek, J., Vykydal Z., Turecek D., Pospisil S.; Kitamura H., Ploc O., Uchihori Y., Yasuda N., Amberboy C., Hauss J., Lee K., Semones E., Zapp N., Parker R., and Cooke D., *Preparing for the First Medipix Detectors in Space*. Proceedings of the IEEE Aerospace Conference, Big Sky, Montana, March, 2012.

[5] Vilalta R., Kuchibhotla S., **Hoang S.**, Valerio R., Ocegueda F., and Pinsky L. (2012), *Classification of Sources of Ionizing Radiation in Space Missions: A Machine Learning Approach.* Acta Futura 5.

[6] Vilalta R., Kuchibhotla S., Ocegueda-Hernandez F., **Hoang S.**, Pinsky L. (2011) *Machine Learning for Identification of Sources of Ionizing Radiation During Space Missions.* Workshop on AI in Space: Intelligence Beyond Planet Earth. International Joint Conference on Artificial Intelligence (IJCAI-11), Barcelona, Spain.

[7] Stoffle N., Pinsky L., Empl A., **Hoang S.**, Pospisil S., Jakubek J., Turecek D., Vykydal Z. (2011), *Simulation of Van Allen Belt and Galactic Cosmic Ray Ionized Particle Tracks in a Si Timepix Detector*. International Comic Ray Conference, Beijng, China.

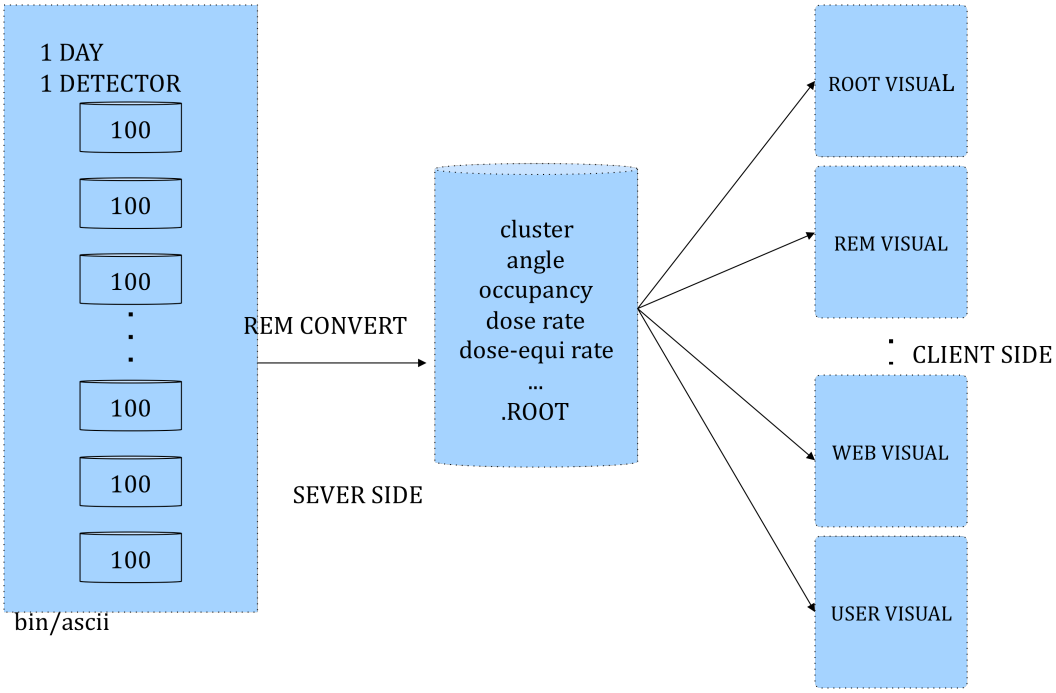
APPENDIX A
RADIATION AREA MONITOR
TECHNOLOGY DEMONSTRATION
DATA FLOW



APPENDIX B
RADIATION AREA MONITOR

DATA ANALYSIS

ARCHITECTURE



APPENDIX C
REMANALYSIS

SOFTWARE MANUAL

Son Hoang

University of Houston

Email: smhoang@uh.edu

1.1. Introduction

The REMAnalysis software is a data analysis framework for processing ISS, HIMAC, NSRL data generated from Timepix detectors. It was based upon the ROOT library [30], a popular framework for High Energy Physics (HEP) data processing and analysis. REMAnalysis was developed in collaboration with the University of Houston, Physics Department, and Pattern Analysis Laboratory of Computer Science Department. It is also sponsored by NASA Johnson Space Center.

The software provides data conversion, data analysis and data taking from the Timepix chip. It includes algorithms to find blobs, to calculate angles, LET, dose, dose-equivalent of heavy-ion particles in ISS and experimental data at HIMAC and NSRL. It also includes pattern recognition algorithms to identify δ-ray, interaction, stopping and to classify types of heavy ions. The graphical interfaces and statistical tools allow efficient use of doing statistics during a long period of time. Comparison between REM units and TEPC are also supported.

1.2. Overall features

- Display images of frame by frame of Timepix output data (charge, energy at each pixel).
- Display analysis of track by track of charged particles (angle, LET, dose, dose-equivalent).
- Display advanced features of tracks (delta-ray, skeleton, distance transform).
- Perform pattern recognition and classification (interaction, stopping, ion type prediction).
- Display statistics over multi-frames (angle distribution, LET distribution, dose, dose-equivalent distribution).
- Display comparison among REM units and TEPC.
- Display dose, dose-equivalent plots over longitude and altitude positions.

1.3. Installation

The software was developed in multiple platforms (linux, mac osx, window). ROOT is the only dependency for installing REMAnalysis. Please refer to the ROOT webpage http://root.cern.ch/drupal/ to get the documentation and learn how to install ROOT from source. If you do location independent installation, remember to add $ROOTSYS/bin/ to PATH and $ROOTSYS/lib/ to LD_LIBRARY_PATH by the following command:

source $ROOTSYS/bin/thisroot.sh

Here, $ROORSYS is the directory where you install ROOT

The REMAnalysis software is being developed under subversion control with host REM Server from TLC2 of University of Houston. For the time being, this software is not

allowed to be public. You need to request an ssh account from REM Server (rem.tlc2.uh.edu) by sending an email to smhoang@uh.edu. This account allows you to download ISS, HIMAC, NSRL raw data, and processed ROOT File after conversion. It also allows you to check out and commit code of REMAnalysis. You can use svn to check out the software as follow:

svn co svn+ssh://youraccount@rem.tlc2.uh.edu/mnt/rem_share/svn_repos/REM/REMAnalysis

Input your password (2 times).

The software will be downloaded and controlled under subversion. Change your directory into REMAnalysis, and compile the software using "make". An executable file with name REMAnalysis will be generated. Later on, you can update the software when a new version is available by using "svn update" (then, compile again). If the proxy at your working place blocks ssh protocol. You can request an https account in order to check out and commit REMAnalysis. Please use the following command:

svn co https://rem.tlc2.uh.edu/svn/REM/REMAnalysis

Input your username and password

1.4. Getting started

To start up the software

Change your current directory into REMAnalysis folder, then run the executable file by using the command below:

./REMAnalysis

Figure 1 shows the screen appearance that is displayed when software is started up.

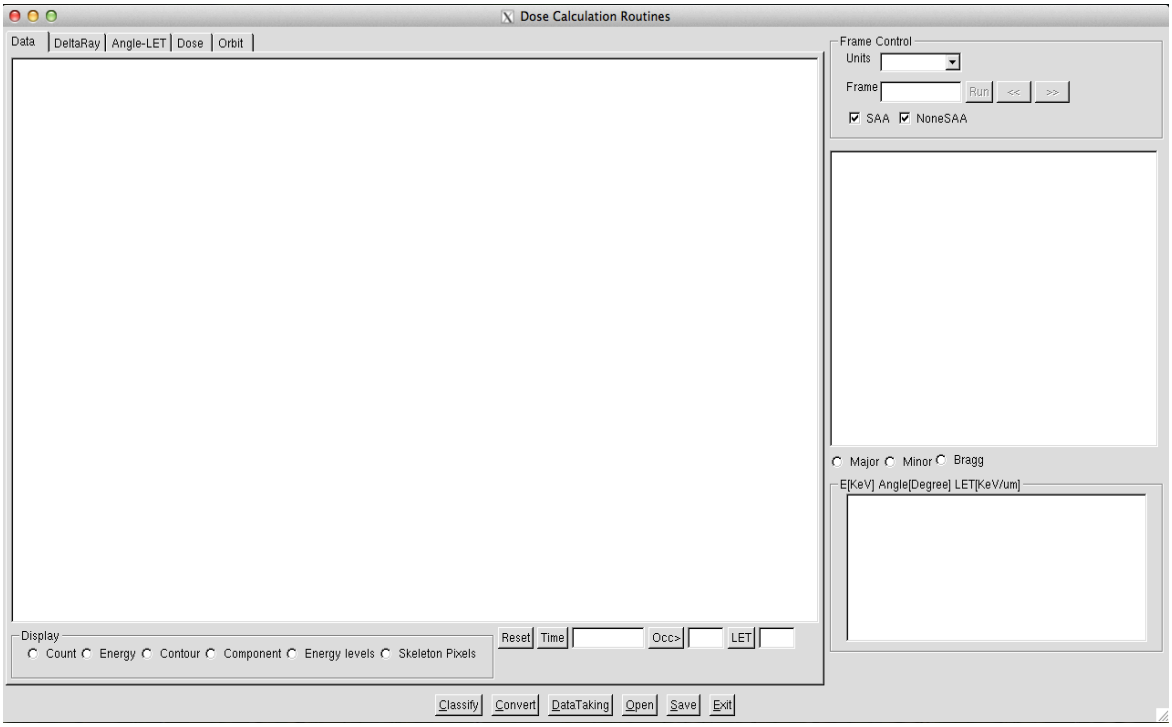


Figure 1. REMAnalysis main panel

1.5. Functions of tabs and buttons of main panel.

REMAnalysis mainly consists of three panels: left, right, bottom. The left panel consists of 5 tabs Data, δ-ray, Angle-LET, Dose, Orbit with corresponding controls in each tab. The right panel consists of Frame control, small canvas and cluster property-list box. Small canvas displays profiles of individual clusters appearing on the left panel. It also shows the frame corresponding to orbit position when clicking on world orbital map. Cluster property-list box displays individual analysis of clusters. The bottom panel includes buttons to open new windows for doing classification, conversion, data taking, opening converted ROOT files to process, saving images from main canvas of tabs and exiting the program.

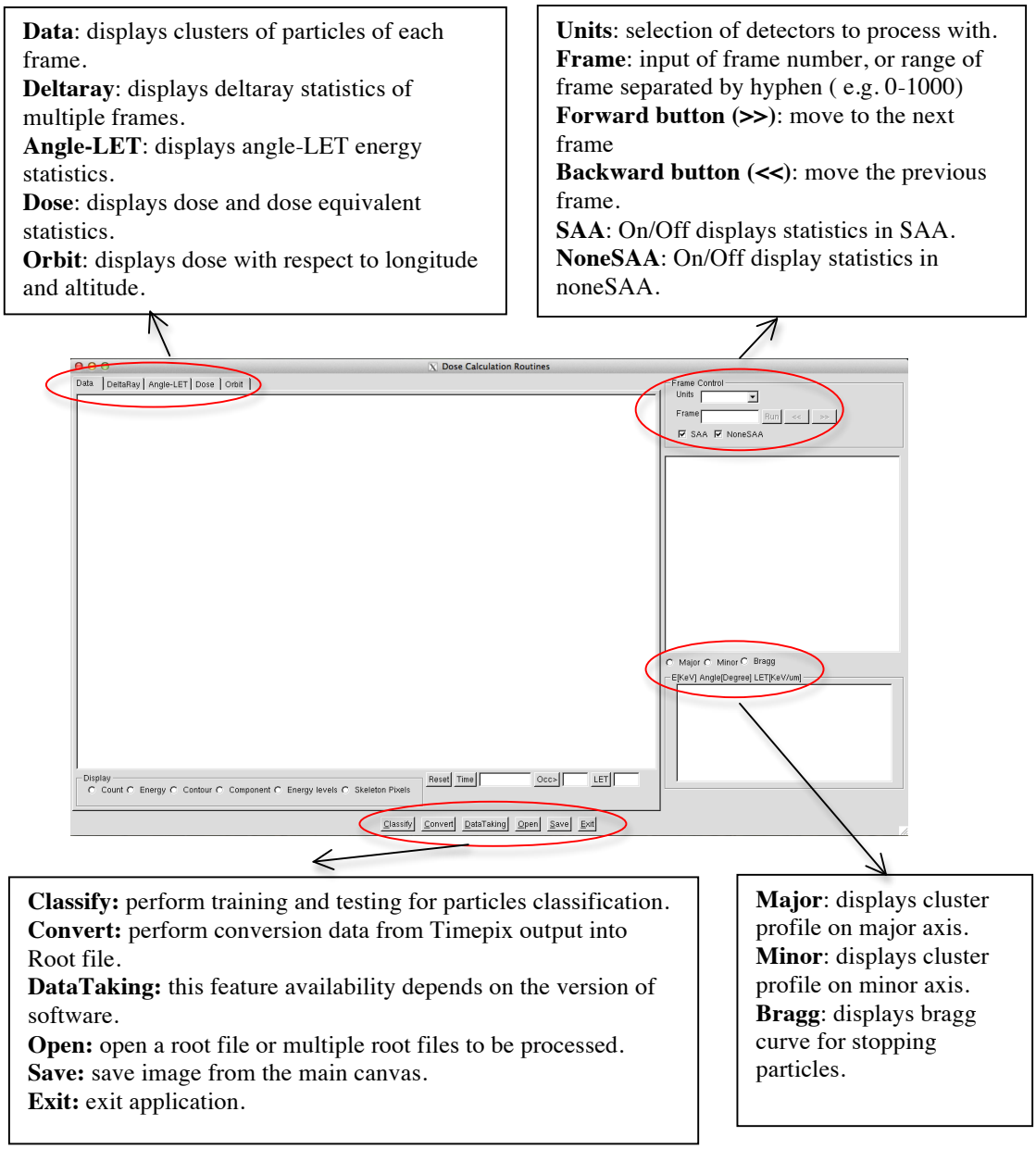


Figure 2. Functional operators on main panel

1.6. Data tab

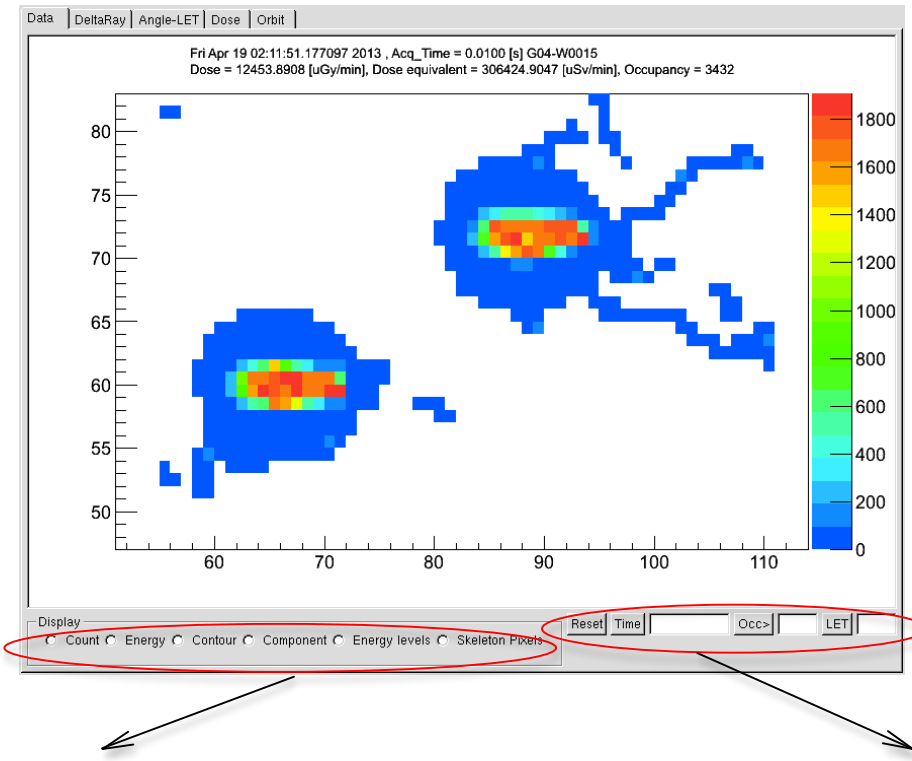
Frames containing clusters of particles will be displayed on the data canvas of this tab when opening a new ROOT file. Use forward or backward button to switch between frames. Frames can be zoomed by dragging the mouse from starting point to end point in

the data canvas; or be un-zoomed by double click on the data canvas or using Reset button. A result cluster with energy information at each pixel will be displayed when clicking on the item of the cluster property list box (lower right corner). A result cluster with energy information at each pixel will be displayed when clicking on the button "LET" with a provided ranges from start to end of LET of particles.

A result frame will be displayed when clicking on the button "Occ" or "Time" with a provided percentage minimum occupancy (e.g. 1, 1.5), or timestamp string (e.g. "Fri Apr 19:11:51"). Note that there are 256x256 pixels in total, the occupancy in this case is calculated by the number of positive-energy pixels divided by 65536. The timestamp string should follow the pattern of the time string in the title.

The title of data canvas provides metadata of the frame displayed on the canvas. The first line shows timestamp string at the moment when taking the data, the acquisition time in second, the chipID. The second line shows dose rate in uGy/min, dose-equivalent rate in uSv/min and the occupancy of this frame.

In "display" control of this tab, one can switch between radio buttons to see information of each cluster and its corresponding pixels. The "count" radio button shows the number of charges before calibration at each pixel. The "energy" radio button shows the calibrated energy at that pixel. The "component" radio button shows delta-ray, skirt, core of the cluster in different colors. The "energy level" shows the closet distance between a pixel and the cluster boundary after using distance transform. The "skeleton pixel" shows the skeleton of cluster.



Count: display the charge and pixel of each particle.
Energy: display energy and pixel of each particle after calibration.
Contour: display contour of energy of each particle.
Component: display components (deltaray, skirt, core) of each particle.
Energy levels: display levels of pixels in a cluster.
Skeleton Pixels: display skeleton of cluster.

Reset: unzoom the frame data.
Time: display the frame at the input time stamp.
Occ > : display the frame in which occupancy higher than the input value.
LET: display the frame in which LET of cluster between the input value (eg. 1.3 – 1.5)

Figure 3. Data tab

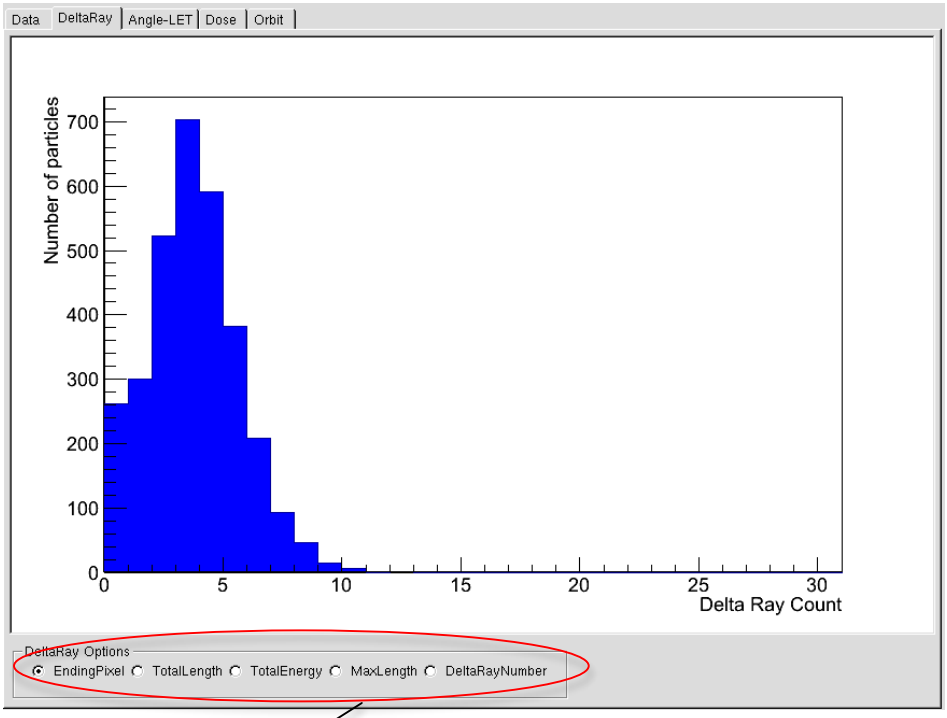
1.7. δ-ray tab

An energetic secondary electron (δ-ray) is produced through the net coulomb interaction of a matter with a charged particle. δ-ray electrons create a noticeable track of their own and can leave the immediate area of the primary particle's track when they have sufficient energy. This tab displays the statistics of delta-ray information of all particles appearing in frames specified in the textbox “Frame” of Frame Control. 4 features of δ-rays are

being investigated:

- Number of delta-ray (ending pixel)
- Total length of delta-ray (in term of number of pixels)
- Total energy of delta-ray (in KeV)
- Maximum length of delta-ray.

Input a range of the frame in the frame textbox in the Frame Control (eg. 0-1000), or 0 to process all of frames. Then click Run to generate statistics for plotting delta-rays information.



Endingpixel: display statistics of the ending pixel (number of deltarays).
TotalLenghth: display statistics of total length of deltarays in terms of pixel.
TotalEnergy: display statistics of total energy of deltarays.
MaxLength: display statistics of maximum length of deltarays.
DeltaRayNumber: display statistics of number of disconnected deltarays.

Figure 4. δ-ray tab

1.8. Angle-LET tab

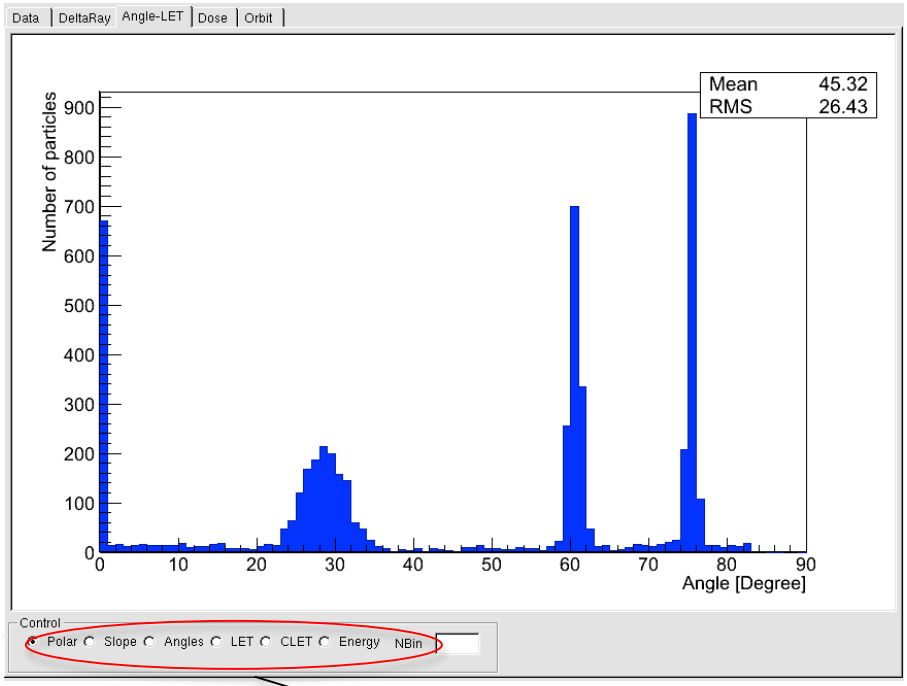
An energetic particle arbitrarily penetrates the detector layer in different angles ranging from 0 degree to 90 degree with respect to the normal direction and from -90 degree to 90 degree for azimuth direction. This is because there is an ambiguity in terms of the actual physical particle's direction of motion along the track. By default we assume that the particle direction is incident from above the sensor layer. Using the Timepix-based detector, LET can be measured from the total energy deposited, divided by the path length traversed by the particle as it passes through the Si layer. The traversed path length can be calculated once the incident angle is determined.

Ionization along the primary particle track causes the creation of electrons and holds, which then drift under the influence of the applied bias voltage. As the carriers drift, they interact with opposing charge carriers and create the recombination of high densities of electrons and holes (the so-called "plasma-effect"). Recombination will result in a reduced amount of energy in the measurement relative to what was deposited in the silicon by the incident ion. It will also affect the uncertainty in the LET calculations. We propose a smoothing technique to compensate energy loss due to this plasma-effect.

Some particles with slow energy, especially protons, do not penetrate the Si layer of Timepix chip. These stopping particles will screw up the angle and LET calculation. We will identify these particles and make a correction for their LET using SRIM simulation.

This tab displays statistics of polar angle, azimuth angle, LET, corrected LET and total energy for particles. Input a range of the frame in the frame textbox in the Frame Control

(eg. 0-1000), or 0 to process all of frames. Then click for generating plots.



Polar: display statistics of the polar angles with respect to the vertical direction.
Slope: display statistics of azimuthal angles.
Angles: display statistics of polar and azimuthal angles.
LET: display statistics of LET.
CLET: display statistics of corrected LET for stopping, saturation, volcano effect.
Energy: displays statistics of total energy.
NBin: number of bins.

Figure 5. Angle-LET tab

1.9. Dose tab

Dose is defined as the energy deposited by a source of radiation per unit mass of traversed matter. Dose-equivalent is the most common form to express the biological effects of the dose caused by a particular type of particle for radiation protection purposes. To calculate the Dose-equivalent, each type of potentially incident radiation is

given a Quality Factor, also referred to as Relative Biological Effectiveness (RBE). As proposed by the National Council on Radiation Protection [NCRP 153 (2008)] regarding heavy-ions in the space radiation environment, the Quality Factor is a function of the LET of the traversing particle.

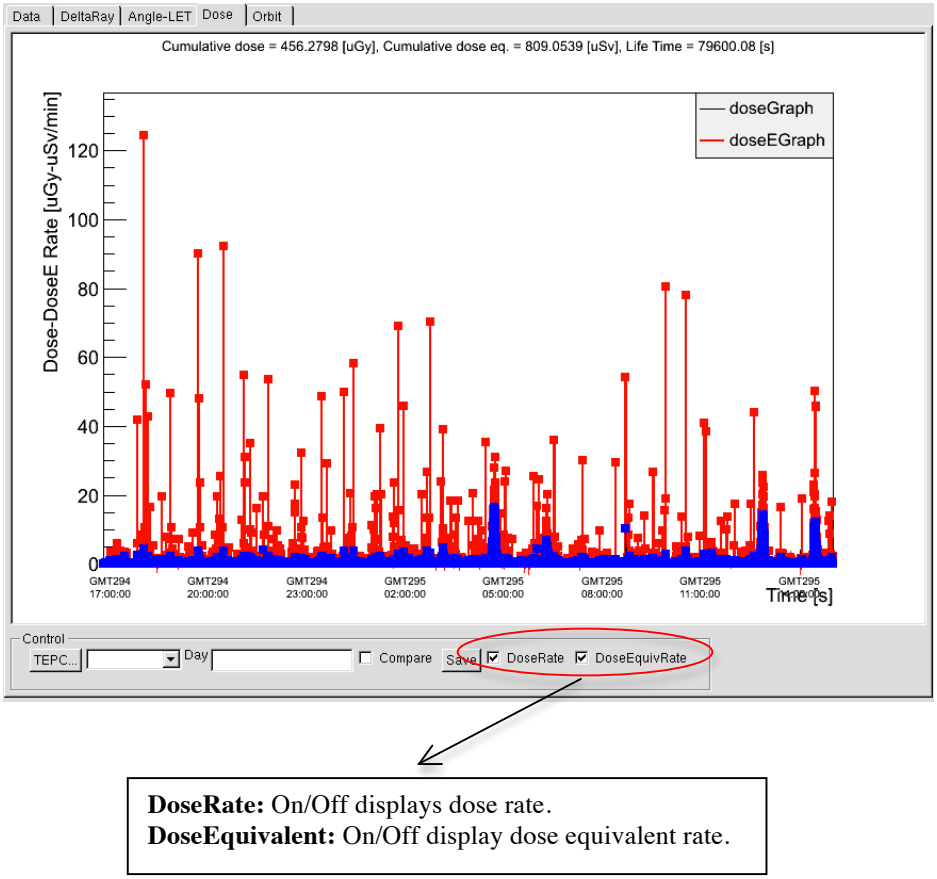
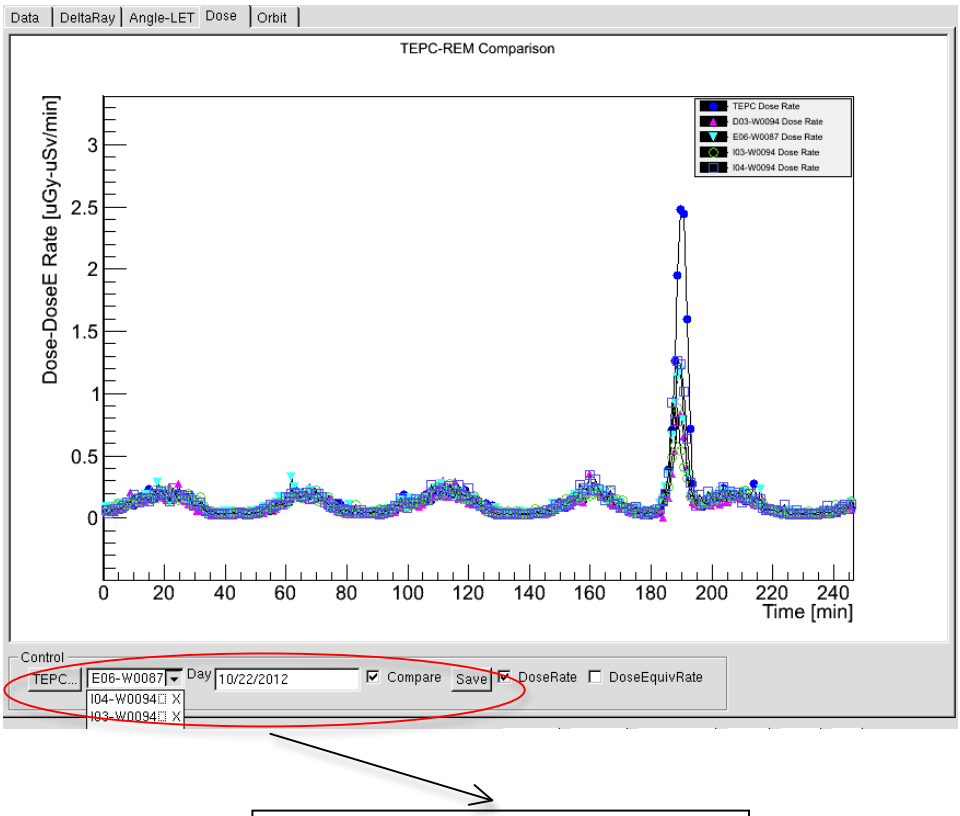


Figure 6. Dose, Dose-Equivalent

Figure 6 displays Dose and Dose-equivalent of each frame (corresponding to each timestamp) and plot them in uGy/min for Dose and uSv/min for Dose-Equivalent on the same canvas. Users can click on the points of Dose or Dose-Equivalent to get directly to corresponding frame displayed in data tab. Input a range of the frame in the frame

textbox in the Frame Control (eg. 0-1000), or 0 to process all of frames. Then click Run for generating plots.



TEPC: browse TEPC file.
ComboBox: enable / disable detectors.
Day: Input date to compare (mm/dd/yyyy)
Compare check: on/off comparison.
Save: Save comparison to file.

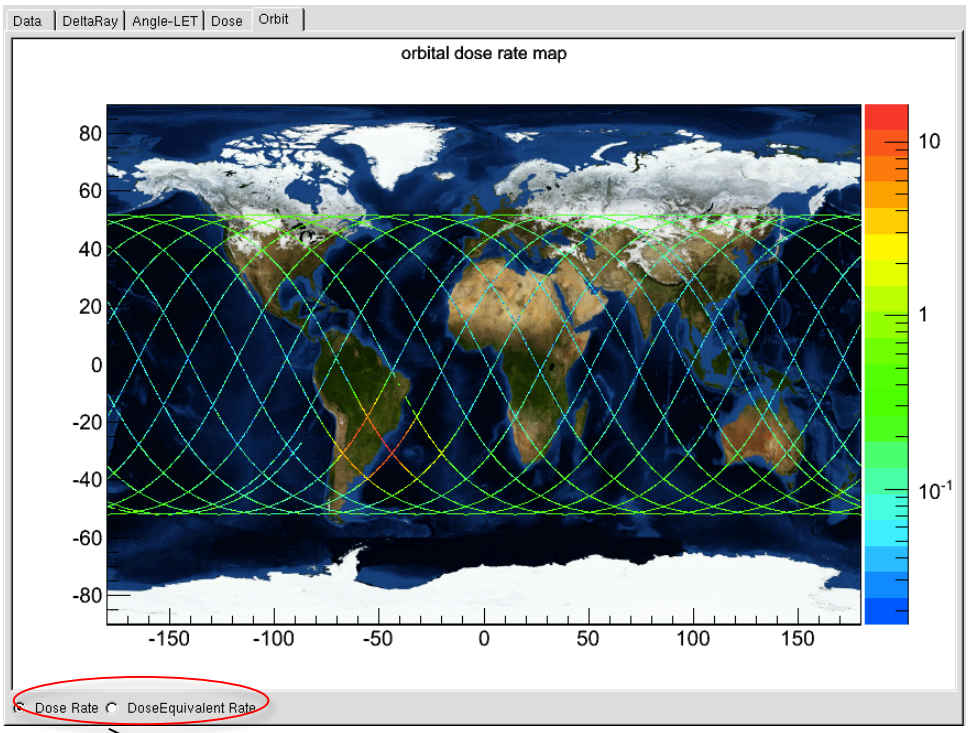
Figure 7. TEPC, REM units comparison

This tab also allows users to compare dose, dose-equivalent between TEPC and REM units or among REM units together (Figure 7). The Tissue Equivalent Proportional Counter (TEPC) is an automatic microdosimetry system used to characterize spatial radiation environment. The comparison between TEPC and REM units is necessary to

evaluate the performance of REM units and algorithms. Users can browse a TEPC file, select detectors (on/off), specify the day and click “Compare” button to generate plots.

1.10. Orbit tab

This tab displays dose, dose-equivalent maps of ISS REM data versus ISS latitude and longitude. We calculate ISS Orbit (longitude and latitude) based on the timestamp when taking the data frame and relied upon the SGP4 NORAD orbit propagator code. Input a range of the frame in the frame textbox in the Frame Control (eg. 0-1000), or 0 to process all of frames. Then click Run to generate plots.



DoseRate: displays dose rate on orbit.
DoseEquivalent Rate: display dose equivalent rate on orbit.

Figure 8. Dose tab

1.11. Conversion

The data are taken from Timepix chip and saved to file under Pixelman file format or BIRD file format. These output data need to be preprocessed, calibrated to convert from ToT value to energy value in KeV, and saved in ROOT format before starting analysis. Users can combine and convert multiple files at the same time and output one ROOT file only. The statistics plots (angle and LET) will be displayed during the conversion process. Click on the "Convert" button at the bottom of the main panel to open the following window.

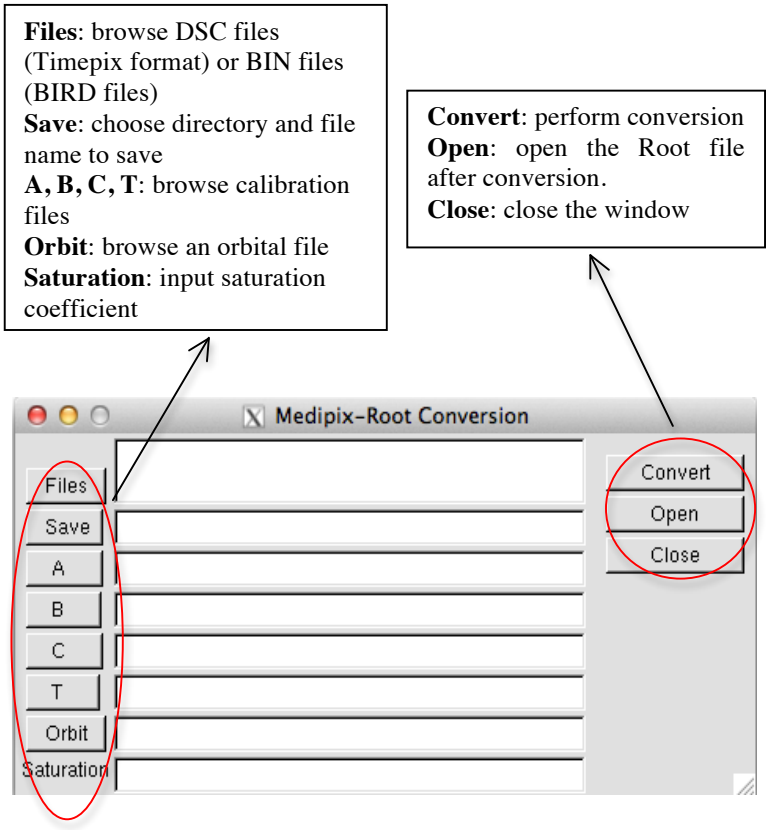


Figure 9. Conversion panel

1.12. Data taking

REMAnalysis is able to connect with Pixelman by using its libraries and APIs to do data taking. As the Pixelman libraries were built for different operating systems, a corresponding configuration needs to be done in order to perform this functionality. Note that license is required for using Pixelman with detectors without Fitpix interfaces.

In case REMAnalysis is not configured correctly, or the Pixelman libraries are not available, a message box will be displayed as follows.

For those who own a Fitpix interface, remember that it needs to be plugged into your computer prior to connecting any other detectors. The REMAnalysis then has to be run in root mode. Please use the following command:

sudo ./REMAnalysis

Then clicking on the button "DataTaking" at the bottom of the main panel. If the right prefix of the saved file name is provided ("mass_ionname_energy_angle_location", see instructions below), the final output files are automatically saved with the name convention:

"automicmass_ionname_energy_angle_biasvoltage_Ikrum_location_monthyear"

E.g. If the prefix is 004_He_0180_00_HIMAC, then the final filename is saved as:

004_He_0180_00_015_01_HIMAC_072013

The above example is for Helium 4, energy 180, angle 0, bias voltage 15, Ikrum 1, taken at HIMAC in July 2013.

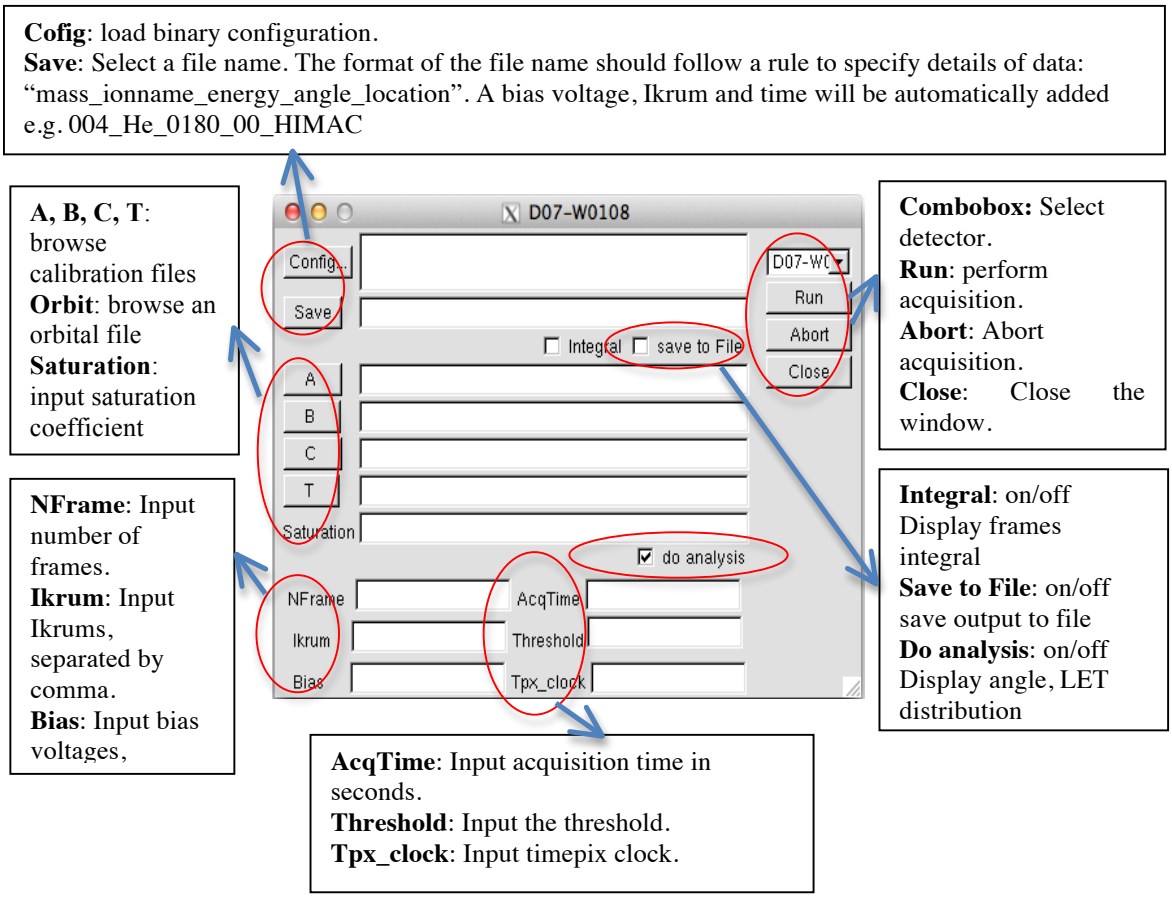


Figure 10. Data taking panel

1.13. Panel opening

After conversion, users can open a ROOT file or multiple ROOT files simultaneously using Open button at the bottom of the main panel. Note that only files with extension ".root" will be displayed. Depending on the number of different detectors generating these ROOT files, Combobox in Frame control will display corresponding number of items. A first frame will be displayed on the data tab.

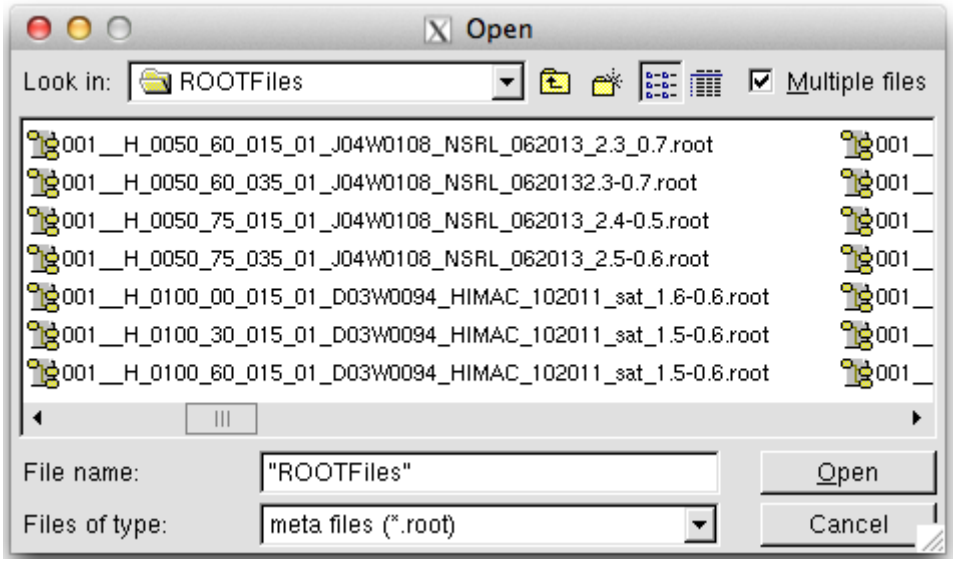


Figure 11. Open panel

1.14. Cluster properties

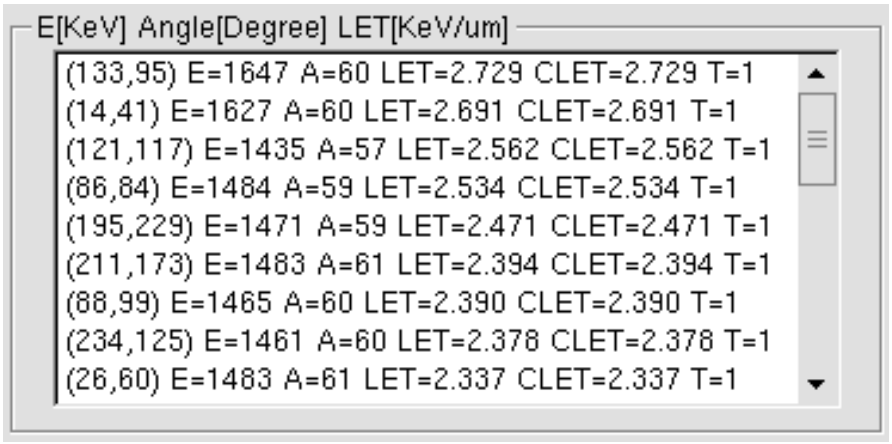


Figure 12. Cluster Properties

The list box at the right bottom corner shows properties of the particles on the canvas including coordinates, total energy (KeV), angles, LET, corrected LET, predicted type of particle. Note that the type value of a particle is equal to its atomic mass. If type = 0, it's predicted as light particles such as photon (x-ray, g-ray). If type=-1, it's predicted as interaction or overlapping. If type=-2, it's a boundary cluster. If type=-4, it's predicted as

stopping. A combination of these values will make the type have the value of -3 (interaction and boundary), -5 (interaction and stopping), -7 (interaction, boundary, stopping), -6 (boundary, stopping). Users can click on item of the list box to see structure of cluster.

1.15. REM Server at UH

Software, ISS data, HIMAC data, and NSRL data are being stored in REM Server at UH. The ISS data are transferred and updated on daily basis from ISS and NASA Server. The HIMAC, NSRL data are updated whenever an experiment is taken. Users with ssh account can access these data via a symbol link named "rem_share" which is automatically mounted at the home directory. Users can use the following command to get access to REM Server:

ssh youraccount@rem.tlc2.uh.edu

Input your password

The ISS data and processed ROOT files are stored at:

/mnt/rem_share/On-orbit_data or ~/rem_share/On-orbit_data

The HIMAC data are stored at:

/mnt/rem_share/HIMAC_data or ~/rem_share/HIMAC_data

The NSRL data are stored at:

/mnt/rem_share/NSRL_data or ~/rem_share/NSRL_data

Users can download files by using command line from your computer's terminal, for example:

scp –r youraccount@rem.tlc2.uh.edu/mnt/rem_share/On-orbit_data/yourfile yourlocalfolder