DOI: 10.54254/2755-2721/132/2024.20704

Data processing method of silicon semiconductor vertex chamber based on impact point coordinate

Donghao Hong[1,#], Jingtong Zhen[2,#], Yuxuan Zhou[3,#], Zifeng Qiu[4,#,*]

[1] Department of Science & Engineering, University of Liverpool, Liverpool, L3 5TR, United Kindom, sgdhong2@liverpool.ac.uk
[2] School of United World College South East Asia, Singapore, 139654, Singapore, zhen148433@gapps.uwcsea.edu.sg
[3] School of Advanced Technology, University of Xi'an Jiaotong-Liverpool, Suzhou, 215000, China, Yuxuan.Zhou23@student.xjtlu.edu.cn
[4] SWJTU-Leeds joint school, Southwest Jiaotong University, Chengdu, 611756, China, el23qz2@leeds.ac.uk

#co-first authors
*corresponding author

Abstract. This paper presents a comprehensive study on the reconstruction of vertex positions and particle tracks using heavy ion collision data from the Large Hadron Collider (LHC). Nonlinear particle trajectories, background noise, and scattering effects complicate accurate vertex prediction and trajectory reconstruction in real particle detectors. The solution proposed in this paper is integrating geometric methods with the DBSCAN density clustering algorithm to overcome these issues. This method effectively improves the precision of vertex finding and particle track reconstruction. In the meantime, a geometrical technique and probabilistic quantization are employed to predict the motion radius of the particle more accurately. This framework has been independently confirmed in an analysis of >10,000 collision points, which proves the framework's robustness. Compared with the benchmark method, the framework only relies on the coordinates of the hit point recorded by the detector, which not only improves the universality of the algorithm application but also reduces the requirements on the equipment and the complexity of implementation and effectively improves the precision of vertex and trajectory reconstruction. The results are for the vertexing accuracy (0.005 cm) and average tracking search precision (95.40%) with a radius determination error of 0.5 cm. The proposed method also stands valid when these results are evaluated.

Keywords: Vertex Reconstruction, Track finding, Radius estimation, DBSCAN Algorithm, Probability quantization

1. Introduction

The investigation of particle trajectories and vertex reconstruction has garnered significant attention in recent years, particularly in the context of high-energy physics experiments. These experiments, slamming heavy ions together at near-light speeds, result in enormous data sets that must be analyzed with great detail to reveal what's happening behind the physics. Data processing for vertex reconstruction, track finding, and radius estimation is significant in particle identification, property measurement of particles, and understanding the force that governs their interaction.

The data utilized in this study is derived from the parameterization of heavy ion collision data at the Large Hadron Collider (LHC). This includes the parametric folding of responses from the LHC experiments' tracker, EMCAL, and HCAL components. As illustrated in Fig.1, a random number generator produces events with varying impact points per layer based on the known distribution from the LHC.

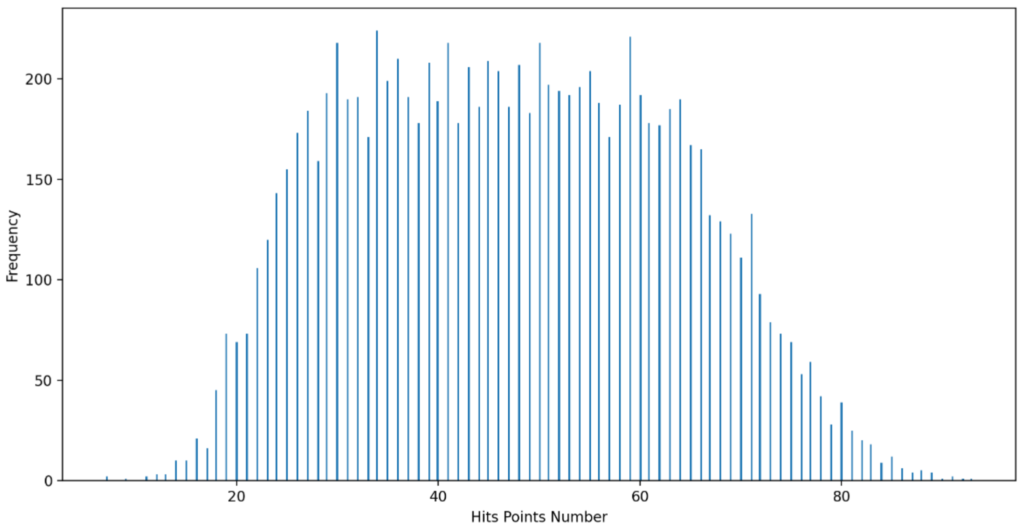

Figure 1. Number of hit points in each event

Previous studies have employed cellular automata, Legendre transform, and combined Kalman filter (CKF) methods for vertex reconstruction and track finding to predict vertices and reconstruct trajectories [1]. These algorithms typically start from a small trajectory segment or a fitted trajectory from adjacent detectors, gradually expanding candidate trajectories by sequentially adding measurement points [2]. In contrast, the vertex reconstruction technique based on data of collision points in the first two layers, and thus less sensitive to scattering errors, is used: it concerns a smaller amount of information and allows for faster calculation. An integrated geometric and DBSCAN (density-based spatial clustering for noise) algorithm is proposed in this study to enhance the accuracy. This hybrid approach leverages the strengths of both geometric methods and machine learning to improve vertex and trajectory reconstruction. In this paper, a method to reconstruct trajectory combining coordinate system transformation and DBSCAN is also put forward. Even in noisy environments, DBSCAN efficiently clusters collision points, while the geometric framework ensures accurate vertex prediction and particle trajectory reconstruction.

Under the influence of the axial magnetic field in the collider, charged particles exhibit circular motion on the central plane of the barrel detector. Accurately estimating the radius of this motion is crucial for measuring the momentum and charge of particles [3]. Previous studies often measure the radius by fitting the trajectory, a process prone to inaccuracies due to particle scattering effects across detector layers. This paper proposes a novel method for radius determination that minimizes the integrated scattering angle, allowing for a more direct and accurate measurement of the circular motion radius, thereby mitigating the impact of scattering on radius estimation.

The effectiveness of our proposed methods is validated through extensive simulations and analysis of actual experimental data from the LHC. Compared with the existing methods, this method only relies on the coordinates of the hit point recorded by the detector, which makes the algorithm more versatile. In addition to reducing the complexity of equipment requirements and implementation, the prediction accuracy of vertex reconstruction, track finding, and radius estimation is effectively improved. Our contributions include 1) an integrated approach combining geometric techniques with density clustering for accurate vertex and trajectory reconstruction; 2) a novel radius determination method that minimizes the impact of scattering effects; 3) extensive validation using a large dataset of experimental data; and 4) demonstration of the robustness of our methods under realistic experimental conditions.

2. Methodology

2.1. Vertex Reconstruction

As shown in Fig.2, during the impact on the first two layers, the deflection of particles under the influence of the magnetic field is minimal, resulting in negligible changes in particle speed within the detector. Consequently, the travel distance between the first two layers can be considered nearly equal. Additionally, the angle between the coordinates of the impact points on the first two layers and the vector formed by the vertex is tiny. In the magnetic field, through the study of particle trajectories, CEPC drift chamber simulation proves that particle deflection is negligible [4]. Thus, the particle trajectory can be approximated as a straight line. This simplification reduces computational complexity and provides a reference line for vertex reconstruction [5,6].

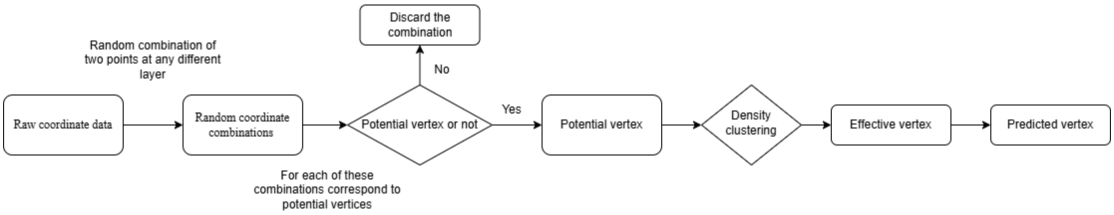

Figure 2. Flow chart for vertex finding

The direction of particle motion between the vertex and the vector formed by the impact points on the first two layers should remain approximately the same. The corresponding vertices should cluster within a smaller spatial range for actual impact points originating from the same vertex in an event. In such cases, the coordinate density near the vertex is significantly higher than in other locations, aiding in accurate prediction. Therefore, this study aims to predict vertex positions using geometric methods and density clustering.

Due to the limited availability of coordinate data, distinguishing noise points from actual impact points is not directly feasible, nor can the correct combination of real impact points be immediately confirmed. All impact point coordinates are projected onto the X-Y plane to streamline the analysis. This projection simplifies the spatial relationships among points and facilitates the identification of valid combinations between inner and outer layer points.

The first step involves identifying the combination of inner and outer layer points to obtain a set of impact point pairs. Among these, only the combinations where the resultant line is within 0.1 cm of the origin of the X-Y plane and the distance between the two points in the pair is between 2 and 4 cm are retained. For these combinations, the symmetry of the outer impact point relative to the inner impact point is used to predict the vertex, referred to as the potential vertex. These potential vertices are distributed around the Z-axis.

Next, density clustering is applied to the screened potential vertices to determine whether other potential vertices are within a specific range around each potential vertex [7]. It is considered valid if the number of potential vertices around a vertex reaches a set threshold. Then, predicted vertex coordinates are considered as coordinates corresponding to the smallest sum of Euclidean distances from all valid vertices in space [4,8].

2.2. Track finding

In this study, the primary objective of track finding is to reconstruct particle tracks from the coordinates of hits recorded by the detector. Each track has a vertex on the z-axis and five distinct hits across five detector layers. Track finding is a crucial step for subsequent track fitting and momentum reconstruction. This paper utilizes geometric fitting techniques and the machine learning algorithm DBSCAN [9] to cluster hits and identify potential tracks. The data employed in this method includes the coordinates of hits, the corresponding layer numbers, and the vertex coordinates for each event, as obtained from the vertex reconstruction process.

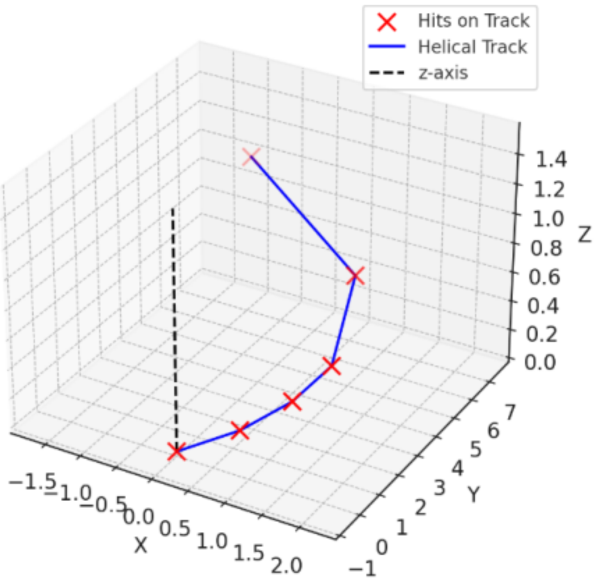

Figure 3. Sample helical track

For the hit coordinates, we calculate the ratio of their radial distance to their vertical displacement from the starting point, denoted as k. Analysis of the hit trajectories reveals that for all hits on the same track, the k value forms clusters with a gradually increasing or decreasing trend and an average difference of ± 0.02. As illustrated in Fig.3, this pattern suggests that the track may follow a spiral curve on a slightly expanding or narrowing cone [10] due to the influence of the magnetic field, external forces, particle scattering, and other potential impact factors.

Based on this observation, we use the k value of each hit on the first layer as the center of a potential track cluster, assigning a label to each cluster corresponding to its center. Beginning with the second layer, the DBSCAN algorithm is applied to identify the k values of the points closest to each cluster, layer by layer. Further analysis indicates that the average difference between k values with each event is influenced by the magnitude of these values, with the difference increasing as the k values rise. Consequently, the epsilon parameter is set to be the center of each cluster, and the minimum number of points (minPts) is set to four.

However, due to noise, some events yield redundant and incorrect trajectory combinations of hits. An additional parameter is introduced to enhance clustering accuracy, limiting the distance from hits on each layer to the cluster center. During this process, if a potential cluster center fails to identify any suitable hits on any layer that belongs to its cluster, it is classified as a noise point and excluded from the analysis. This approach effectively minimizes the impact of noise on clustering efficiency and accuracy.

2.3. Radius estimation

The scattering effect poses a significant challenge in calculating the radius of a particle's circular motion. As particles pass through each detector layer, their velocity direction changes, leading to variations in the center of their trajectory circle, while the magnitude of their velocity remains nearly constant. This consistency in velocity allows us to address errors introduced by the scattering effect.

DOI: 10.54254/2755-2721/132/2024.20704

To accurately estimate the radius, we iterate through a range of potential radii in fixed increments. For each guessed radius, we determine the center of the trajectory between two adjacent detector layers by using the coordinates of the corresponding collision points and the guessed radius. As illustrated in Fig.4 and combining property using the formula $L \approx r\theta,\ \alpha = \theta,$ it can be found that the distance between the centers of adjacent arcs divided by the guessed radius gives the scattering angle α of the particles in that layer.

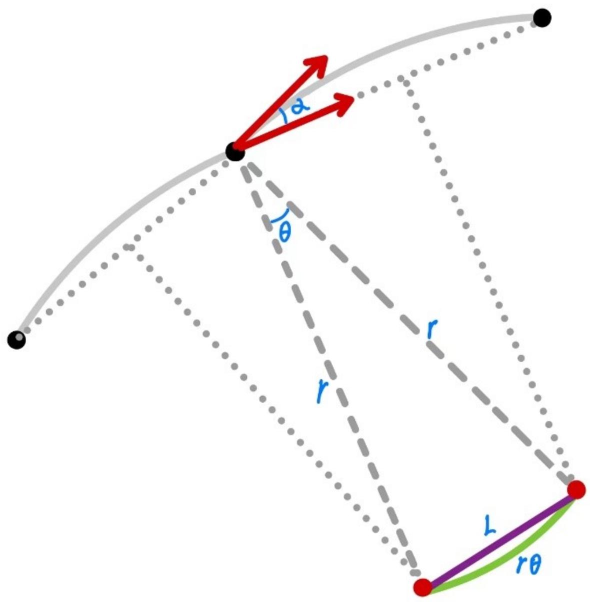

Figure 4. The scattering schematic diagram shows properties in scattering (L≈rθ, α=θ)

An analysis of the particle scattering angles derived from the Large Hadron Collider (LHC) experiment data (Fig. 5 and Fig. 6) reveals that the distribution of the scattering angle α is independent of the particle's circular motion radius (Fig. 5). This finding indicates that the scattering angle distribution is a universal characteristic, unaffected by individual particle properties. As shown in Fig.6, the scattering angle follows a Gaussian distribution with a mean (μ) of 0.00 and a standard deviation (δ) of 0.02.

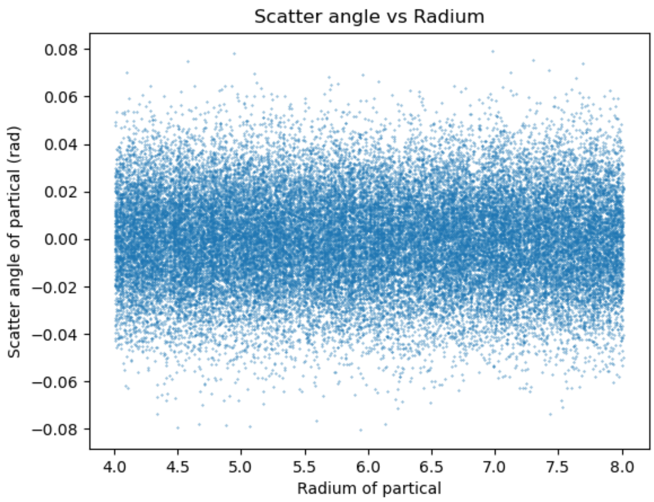

Figure 5. Relation between radium of particle

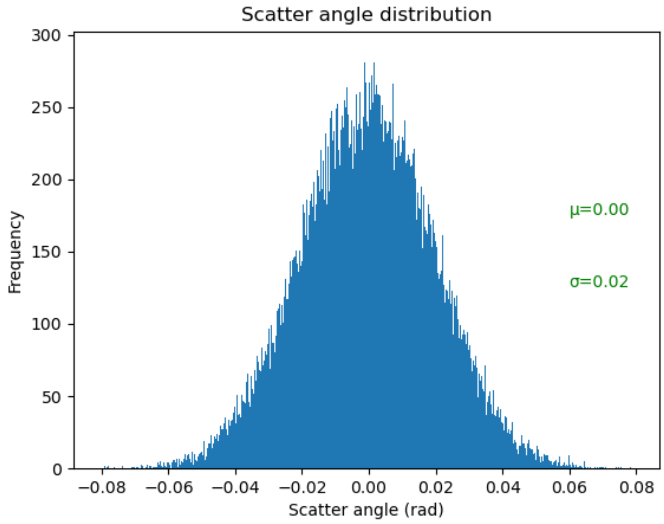

Figure 6. Scatter angle distribution

Given that the scattering angle of the particle α follows this Gaussian distribution, we quantify the probability of a particle experiencing a specific scattering deflection angle using the function $w = e^{-\frac{\alpha^2}{2\delta^2}}$, where the α is the scattering deflection angle and $\delta = 0.02$, according to Gaussian density function [11].

We calculate the average probability of the five scattering angles for each guessed radius and use this value to represent the trajectory's likelihood corresponding to that radius. The radius that maximizes the overall trajectory probability is considered the most accurate, as it minimizes the cumulative scattering deflection angle while ensuring the trajectory passes through the vertex and all impact points in a circular motion. This guessed radius is, therefore, regarded as the final result.

3. Experimental Model Evaluation

3.1. Vertex Reconstruction

The accuracy of vertex predictions is evaluated by measuring the Euclidean distance between the predicted and accurate vertex coordinates. The primary metrics used in this assessment are the mean prediction error and the error distribution across different events.

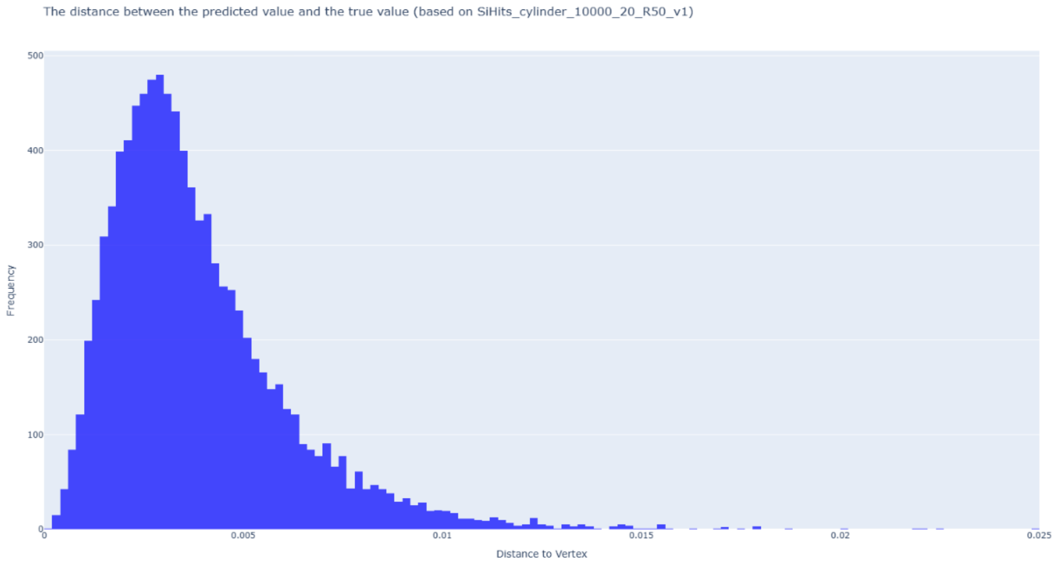

Figure 7. The error of the predicted result

As shown in Fig.7, the error distribution varies across events. The predicted vertex coordinates are within 0.005 cm of the actual values for most events. However, in a few cases, the errors can be around 0.01 cm or even 0.02 cm in sporadic cases. These errors occur because the actual particle motion is not an approximately straight line but an arc [12]. This discrepancy leads to a specific deviation of the vertex coordinates in the direction of the Z axis, thereby impacting the accuracy of the vertex prediction.

To address this issue, recurrent neural networks (RNN) and long short-term memory (LSTM) models are proposed as potential solutions [13]. When trained on large datasets, these models can process sequences of variable lengths and capture particle motion's nuanced dynamics. By training on a large amount of data, the RNN or LSTM model may eliminate the influence of the arc-shaped trajectory and measurement errors on prediction results.

3.2. Track finding

As depicted in Fig.8, the error rate of the clustering result is within 0.075 for most events. Across 2000 events, the clustering model achieves an average accuracy of 95.40% with an average error rate of 4.60%, but for rare cases, the error rate reaches nearly 0.4.

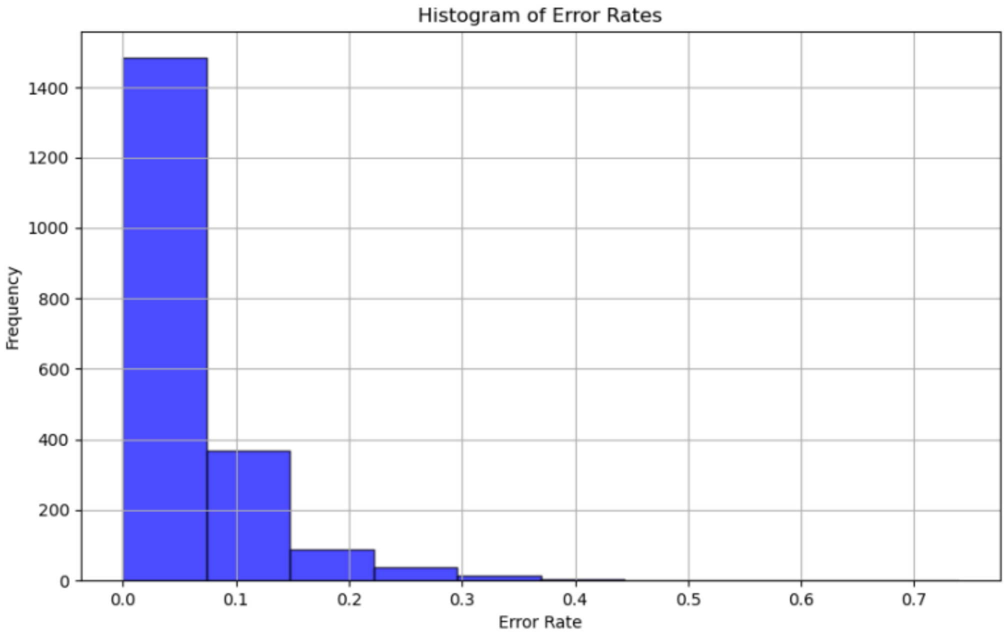

Figure 8. The error rate of the clustering result

Our analysis of events with relatively high error rates reveals that the primary sources of error are irregular particle scattering and noise points that are difficult for the model to identify. To mitigate these issues, preprocessing the data using a multilayer perceptron (MLP) [14] could eliminate some noise points. Optimizing parameter selection for each event may help reduce the error rate in these rare and extreme cases.

3.3. Radius estimation

To evaluate the model's performance, we annotated the initial momentum P and the scale factor $\frac{1}{qB}$ for each trajectory particle. We then compared the model's measured radius with the theoretical radius calculated by the formula $R = \frac{P}{qB}$ to determine the measurement error distribution (Fig. 9). The results show that a significant portion of the data (38.35%) had an error of less than 0.1, most of the data (72.40%) had an error of less than 0.25, and nearly all the data (93.47%) had an error of less than 0.5 (Table 1).

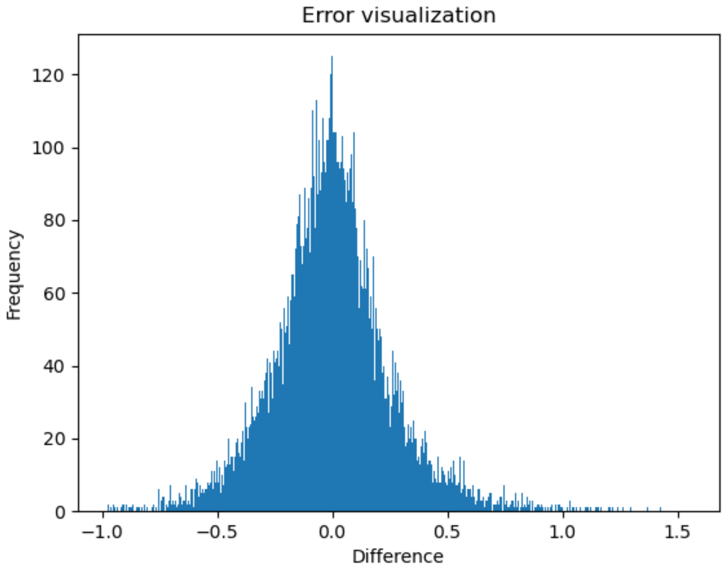

Figure 9. Differences between predicted radium and theoretical radium

The primary source of error stems from the fact that the actual radius does not always correspond to the value that maximizes the likelihood of the entire trajectory. Instead, it follows a Gaussian distribution centered around this value. Our model measures the mean of this Gaussian distribution, but the actual radius may fall elsewhere within the distribution, leading to a discrepancy between the measured and accurate radii.

Given the uncertainty associated with the actual radius, which lies within a Gaussian distribution when measured solely based on vertex coordinates, we propose adding additional parameters—such as the velocity direction at each layer—to enhance accuracy. This approach would provide more information to refine the estimation of the actual radius, leading to a more precise measurement.

Table 1. The percentage of events included within each error range

Error range	The proportion of events included
$difference\ in \pm 0.1$	*38.3%*
$difference\ in \pm 0.2$	*64.28%*
$difference\ in \pm 0.3$	*79.35%*
$difference\ in \pm 0.5$	*93.54%*
$difference\ in \pm 0.8$	*98.9%*
$difference\ in \pm 1.0$	*99.71%*

4. Conclusion

This study proposes a comprehensive approach for vertex reconstruction, track finding, and radius estimation in high-energy particle collisions, primarily using data from the Large Hadron Collider (LHC).

Combining geometric methods with density clustering techniques, notably the DBSCAN algorithm, several critical challenges are addressed in vertex reconstruction and track finding, including noise and particle scattering effects. The results show that the method significantly improves the accuracy of vertex reconstruction. Most of the errors are within 0.005cm. In addition, the track-finding accuracy is also improved, and the clustering accuracy reaches 95.40%.

For radius estimation, to more directly and accurately measure the radius of the particle's circular motion in the case of particle scattering interference, this paper introduces an acceptable radius estimation method considering the Gaussian distribution of scattering Angle. The radius estimation accuracy of 0.5 cm is realized through geometric method and probability quantization. The accuracy of particle trajectory reconstruction is further improved.

These findings contribute to the existing body of knowledge and provide practical solutions for improving collision data analysis in high-energy physics experiments. The study highlights the importance of combining traditional geometric methods with advanced machine learning techniques to overcome the limitations of existing methods.

The paper also discusses the limitations of the current approach, particularly the assumption of linear trajectories in vertex reconstruction, the presence of a relatively large percentage of noise in track finding, and the minimum measurement accuracy limit determined by uncertainty in scattering effect in radius estimation. It suggests future research directions that could involve refining this assumption or exploring more complex modeling techniques. Overall, our framework presented in this study has the potential to advance high-energy physics by offering a more accurate and reliable method for vertex reconstruction, track finding, and radius estimation.

Acknowledgment
Donghao Hong, Jingtong Zhen, Yuxuan Zhou, Zifeng Qiu contributed equally to this work and should be considered co-first authors.

References

[1] Bertacchi, V., Bilka, T., Braun, N., Casarosa, G., Corona, L., Cunliffe, S., Dattola, F., De Marino, G., De Nuccio, M., De Pietro, G., Dong, T.V., Dujany, G., Ecker, P., Eliachevitch, M., Fillinger, T., Frost, O., Frühwirth, R., Gebauer, U., Glazov, S., Gosling, N., Guo, A., Hauth, T., Heck, M., Kaleta, M., Kandra, J., Kleinwort, C., Kuhr, T., Kurz, S., Kvasnicka, P., Lettenbichler, J., Lueck, T., Martini, A., Metzner, F., Neverov, D., Niebuhr, C., Paoloni, E., Patra, S., Piilonen, L., Praz, C., Prim, M.T., Pulvermacher, C., Racs, S., Rad, N., Rados, P., Ritter, M., Rizzo, G., Rostomyan, A., Scavino, B., Schlüter, T., Schwenker, B., Spataro, S., Spruck, B., Svidras, H., Tenchini, F., Uematsu, Y., Webb, J., Wessel, C., Zani, L., 2020. Track finding at Belle II. ScienceDirect, Computer Physics Communications, pp.2.

[2] Billoir, P., Qian, S., 1990. Simultaneous PA recognition and C fitting filtering. Nuclear Instruments and Methods in Physics Research A, 294, pp.219-228. North-Holland. Received 6 March 1990.

[3] Steven Weinberg; Particle physics, from Rutherford to the LHC. Physics Today 1 August 2011; 64 (8): 29–32. https://doi.org/10.1063/PT.3.1216

[4] Liu, M. Y., Li, W. D., Huang, X. T., Zhang, Y., Lin, T., & Yuan, Y. (2024). Simulation and reconstruction of particle trajectories in the CEPC drift chamber. Nuclear Science and Techniques, 35(8), 128.

[5] CMS collaboration. (2014). Description and performance of track and primary-vertex reconstruction with the CMS tracker. Journal of Instrumentation, 9(10), P10009.

[6] Li, J., Qu, Z., & Xu, W. (2024). New vertex reconstruction algorithm for measuring gamma rays with the AMS tracker. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1058, 168835.

[7] Yin, L., Wang, Y., Chen, H., & Deng, W. (2022). An improved density peak clustering algorithm for multi-density data. Sensors, 22(22), 8814.

[8] Qasim, S. R., Kieseler, J., Iiyama, Y., & Pierini, M. (2019). Learning representations of irregular particle-detector geometry with distance-weighted graph networks. The European Physical Journal C, 79(7), 1-11.

[9] Kumar, K. M., & Reddy, A. R. M. (2016). A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method. *Pattern Recognition*, *58*, 39-48.

[10] Ye, Y., Zhu, J., & Wang, G. (2004). Geometric studies on variable radius spiral cone-beam scanning. Medical Physics, 31(6), 1473-1480.

DOI: 10.54254/2755-2721/132/2024.20704

[11] Ribeiro, M.I., 2004. Gaussian probability density functions: Properties and error characterization. Institute for Systems and Robotics, Lisboa, Portugal.
[12] DeZoort, G., Thais, S., Duarte, J., Razavimaleki, V., Atkinson, M., Ojalvo, I., ... & Elmer, P. (2021). Charged particle tracking via edge-classifying interaction networks. Computing and Software for Big Science, 5, 1-13.
[13] Mondal, S., & Mastrolorenzo, L. (2024). Machine Learning in High Energy Physics: A review of heavy-flavor jet tagging at the LHC. arXiv preprint arXiv:2404.01071.
[14] Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2013). The effect of data pre-processing on optimized training of artificial neural networks. Procedia Technology, 11, 32-39.