Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ28756

Pattern Recognition and Prediction in Tracking Detectors

Ziyu Luo[1*], Zhong Xin Han[2], Andrew Luo[3], Baohang Chen[4]

[1] *Lord Byng Secondary School, Vancouver, Canada*
[2] *Caxton College, Valencia, Spain*
[3] *Kang Qiao International School, Shanghai, China*
[4] *Diamond Bar High School, Diamond Bar, United States*
**Corresponding Author. Email:luorozzi@gmail.com*

Abstract. Charged-particle tracking in multilayer cylindrical silicon detectors is a fundamental task in experimental particle physics. It is an advancement in data analysis, supported by both momentum measurement and collision-vertex reconstruction. Various fitting algorithms could be applied to this problem. In this work, we developed a modular reconstruction pipeline using simulated detector hits. The analysis begins by preprocessing and visualizing raw hit data from Python in both the transverse and longitudinal projections, along with a view of a three-dimensional scatterplot of the dataset. While clustering methods, such as DBSCAN and Hough transforms, are commonly used in real-world situations, our dataset provided track IDs, allowing us to directly focus on fitting. For each event, track segments were represented as lines to be combined in a least-squares procedure to determine the common collision vertex. We then examined the distribution of reconstructed vertices across events, finding that transverse coordinates cluster around the beamline. At the same time, longitudinal positions exhibit a spread consistent with the simulated collision profile.

This work demonstrates accurate per-event vertex reconstruction and provides a framework for extending to full track-finding methods.

Keywords: Charged-particle tracking, Pattern recognition, Least squares fitting, Vertex finding, Track prediction, Residual analysis simulation-based study

1. Introduction

One of the central tasks in experimental particle physics, namely relating to detectors such as particle accelerators, is the precise reconstruction of charged-particle trajectories. Modern collider trackers use high-granularity silicon pixels and strips inside a strong solenoidal field, enabling precise reconstruction of charged tracks and primary vertices [1]. It is central to modern technology in the physics world. The resulting charged particles that break apart when beams collide inside a detector leave spatial, so-called ”hits” that are then processed and assembled into trajectories. The challenge is that from these tracks, one can extract momentum information, reconstruct the collision vertex, and ultimately identify and visualize the physics processes that occurred.

Detectors can determine both the electric charge and momentum of a particle by observing how it curves in a magnetic field [2]. Various detectors, for example, a multilayer cylindrical silicon detector, provide supportive geometric views of tracks. In practice, detectors record discrete electronic

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ28756

hits from particle interactions rather than continuous trajectories, and these signals are subsequently reconstructed into track patterns [3]. In the transverse (x,y) plane, tracks bend under the influence of the magnetic field, encoding their transverse momentum. The tracks appear to be approximately straight in the longitudinal r-z projection. They are also radially aligned with the vertex, in turn offering a clean handle for vertex reconstruction. Sophisticated, real-world experiments typically employ advanced pattern-recognition techniques. Hough transform or density-based clustering (DBSCAN) to associate hits into candidate tracks before applying fits. The Hough transform maps collinear (or cocircular) hit patterns to compact loci in parameter space, providing a classic, geometry-driven approach to track finding [4].

In our simulated dataset, however, track identities are pre-labeled, which allows us to bypass the combinatorial complexity of pattern recognition and focus directly on track parameterization and vertex reconstruction.

Reconstructing tracks requires two broad steps: track finding (identifying which hits belong to which trajectory) and track fitting [5]. (extracting trajectory parameters from those hits). Finally, one may choose to audit and buttress their findings. In realistic conditions, track finding is non-trivial and often employs clustering algorithms or Hough transforms to detect patterns in dense hit clouds [6]..

In this work, we will explore the concepts mentioned above using a simulated data set with track identities that are already labeled.

2. Data Ingestion

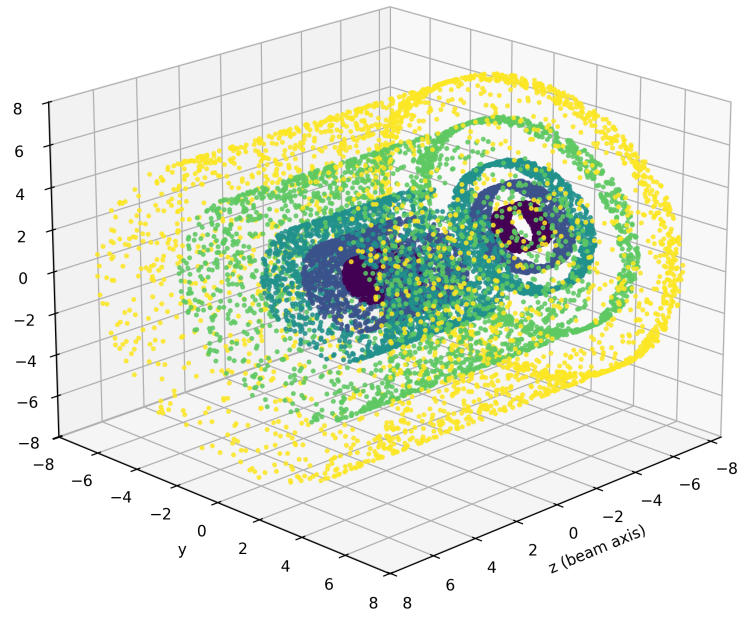

Figure 1: Three-dimensional scatter of hit points (h_x, h_y, h_z) from our simulated detector.

DOI: 10.54254/2753-8818/2026.HZ28756

We initially started this project by trying to visualize our data. The initial dataset contains nine columns representing event index, track index, three position coordinates (hx, hy, hz), and three velocity components. We transformed this into a 3D scatterplot view of the data. The hits seemed to have 5 distinct shells in a cylindrical pattern. One can see dense rings near the center (inner detector layers), and as one moves outwards, hits form larger-radius cylinders (outer layers). Along z, the hits extend symmetrically forward and backward, reflecting that collisions produce particles traveling in both directions along the beamline.

The figure 1 above verifies that the detector geometry in the simulation is a layered cylindrical tracker. The origin is a likely interaction point (collision vertex). From there, particles emerge on helical trajectories, but in projection they intersect the detector layers at circular cross-sections in x–y and straight lines in r–z figure 2. By plotting all events together, we confirm that the geometry is consistent and that all hits respect the cylindrical symmetry expected in a silicon tracker.

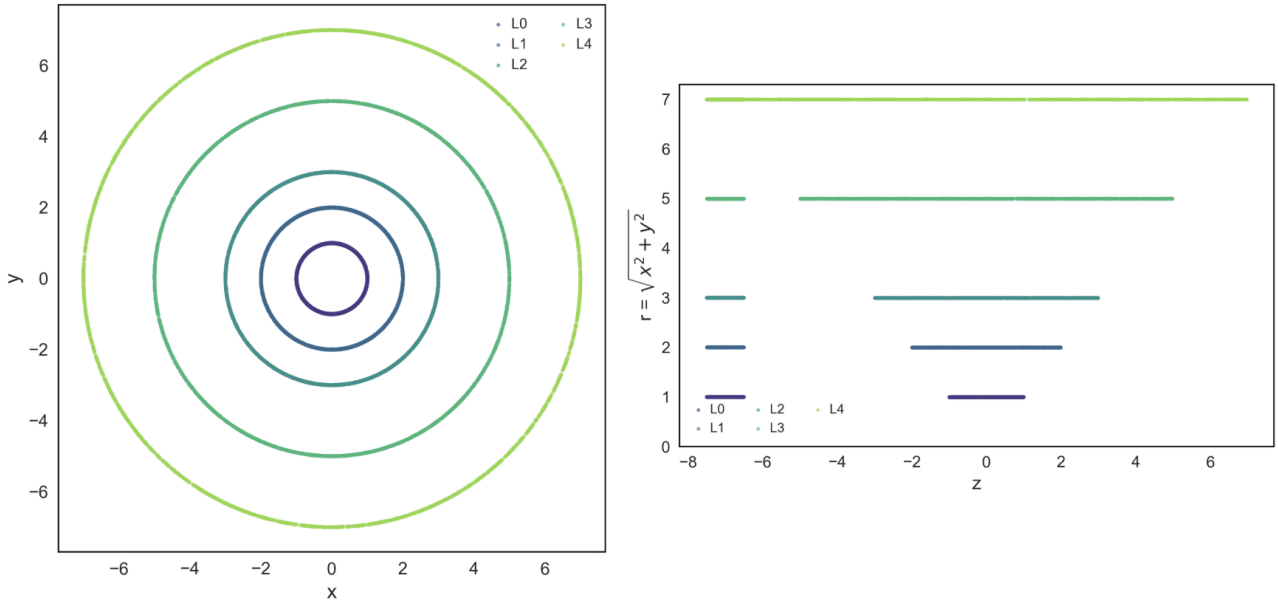

Figure 2: x-y and r-z projections (All Events)

3. Track Parameterization and Angular Correlations

Again, in a realistic detector and under realistic detector conditions, associating hits into tracks is a significant computational challenge requiring pattern-recognition algorithms such as Hough transforms or density-based clustering. In our simulated dataset, however, track identities are already labeled, allowing us to bypass the combinatorial complexity of pattern recognition and focus directly on track parameterization and vertex reconstruction. This allows us to group hits unambiguously, for example, by event and trackID, thereby bypassing the combinatorial complexity of track finding. Each track is then represented by its direction cosines (dx, dy, dz) and a reference hit, enabling us to treat it as a straight line segment for subsequent angular correlation and vertex fitting[7,8].

Next, we examined angular correlation between hits in adjacent layers to validate the track parameterization and to probe their alignment with their respective vertices. Each hit is associated with

DOI: 10.54254/2753-8818/2026.HZ28756

the azimuthal angle ”Phi”, in the transverse plane and the polar angle ”Theta”, in the longitudinal projection.

In other words, the azimuthal angle in the transverse plane and the polar angle in the longitudinal projection, with:

$$r = \sqrt{x^2 + y^2} \tag{1}$$

For pairs of hits belonging to the same track across consecutive layers, we define $\Delta\phi$ and $\Delta\theta$, where i and j denote two adjacent detector layers.

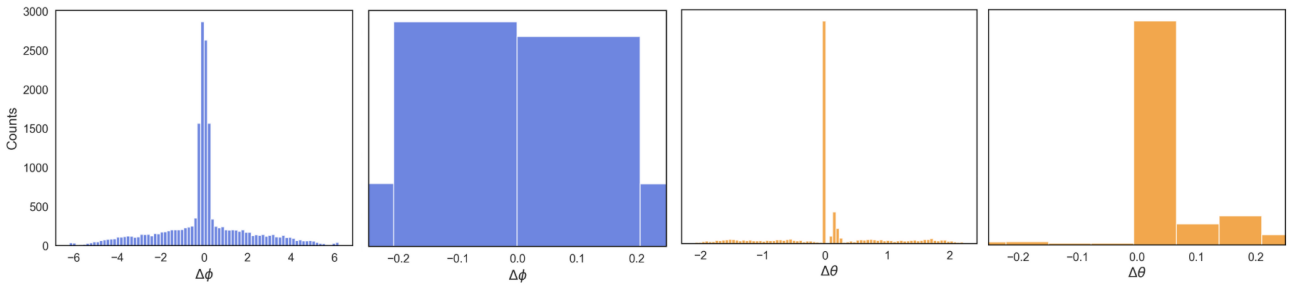

Figure 3: Delta Phi and Delta Theta Distributions

We ended up applying a loose cut to Delta Phi and Delta Theta graphs figure 3. This selection removes outliers while leaving the peaks intact, in turn providing a cleaner input for the vertex reconstruction that we are going to conduct later.

4. Vertex Construction and Finding

Once tracks are parameterized as straight-line segments, the next goal is to determine their common point of origin — the collision vertex. Each track is described by a reference hit (e.g. at the innermost layer) and its direction cosines (dx, dy, dz), which define a straight line in 3D space. In the absence of measurement uncertainties, all tracks from the same event would intersect at a single point. Why do we need least squares method? In practice, finite detector resolution, bending in the magnetic field, and scattering mean the lines do not cross exactly. We therefore use a least-squares method to find the vertex $\boldsymbol{v}$ that minimizes the sum of squared perpendicular distances to all tracks.

4.1. Global Collision Vertices

Carrying on to determining the point of origin, we can analyze the collision vertex of every event. Each track is described by a reference hit, and its direction cosines (dx, dy, dz), which define a straight line in 3D space. All tracks from the same event would intersect at a single point in the absence of measurement uncertainties (in a perfect world). Unfortunately, in the real world, with finite detector resolution, bending in the magnetic field, and scattering mean, the lines do not cross each other exactly. We therefore implemented a least squares regression method to find the vertices [9].

The figure 4 below depict global distributions of reconstructed vertices across every event and provide insight into the geometry of the beam. Indeed, the global distribution of vx and vy shows a narrow spread centered on zero, validating our previous hypothesis that the origin most likely holds the vertex.

In contrast, though, the longitudinal vertex coordinate (vz) exhibits a broader distribution.

DOI: 10.54254/2753-8818/2026.HZ28756

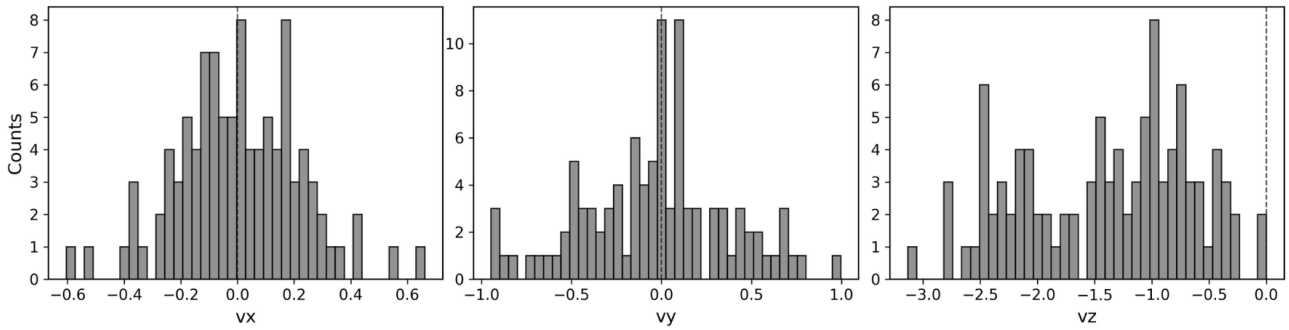

Figure 4: Constructed graphs of Vertex positions across X Y and Z

If the above histogram was too abstract, here is an isolated, more visual depiction of vy and vx only view of the vertices without vz with a density plot on the right depicting the concentration of vertices to be concentrated within (0,0) with little outliers.

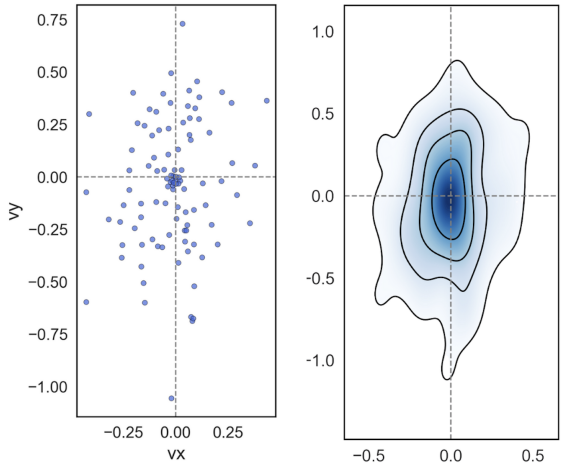

Figure 5: 2D scatterplot of Vertices

4.2. Event 0 Collision Vertex

One can carry on to analyze a single event instead of multiple. Having established in the earlier section the global properties of the reconstructed vertices, we will now carry on to the analysis of a single event: Event 0. Event 0 serves as a representative example, where we can visualize how track segments from different detector layers are extrapolated back to a common point. We gain an intuitive picture of how the vertex emerges from multiple reconstructed tracks by focusing on a single collision.

Above is a 3D scatterplot representation of all the points of Event 0, similar to the scatterplot 3D view of global hitpoints. The figure 6 below illustrate x-y and r-z projections of event 0 Only. The ”X’s” on the graph is an approximate estimation of Event 0’s vertex. Again, In Event 0, each hit is associated with a labeled track. By grouping hits by their track ID and fitting straight-line segments, we obtain track directions in both the transverse (x–y) and longitudinal (r–z) projections. Figure 7 is similar to the global Delta Phi and Delta Theta distributions, but these are unique to Event 0. Although

DOI: 10.54254/2753-8818/2026.HZ28756

it follows the rough patterns as the global distributions. As one can see, the Theta graph followed the global distribution graphs, as it has a clear peak near zero when zoomed in, unlike the Delta Phi graph, which looks uniform when zoomed in and with more precise bins. Outliers do exist, but they are trivial compared to the center bins and are not statistically significant. The strong dominance of the peak around $\Delta\theta = 0$ reinforces the conclusion that the vertex can be reliably reconstructed using longitudinal alignment, while the outliers likely represent noise or rare misassigned hits.

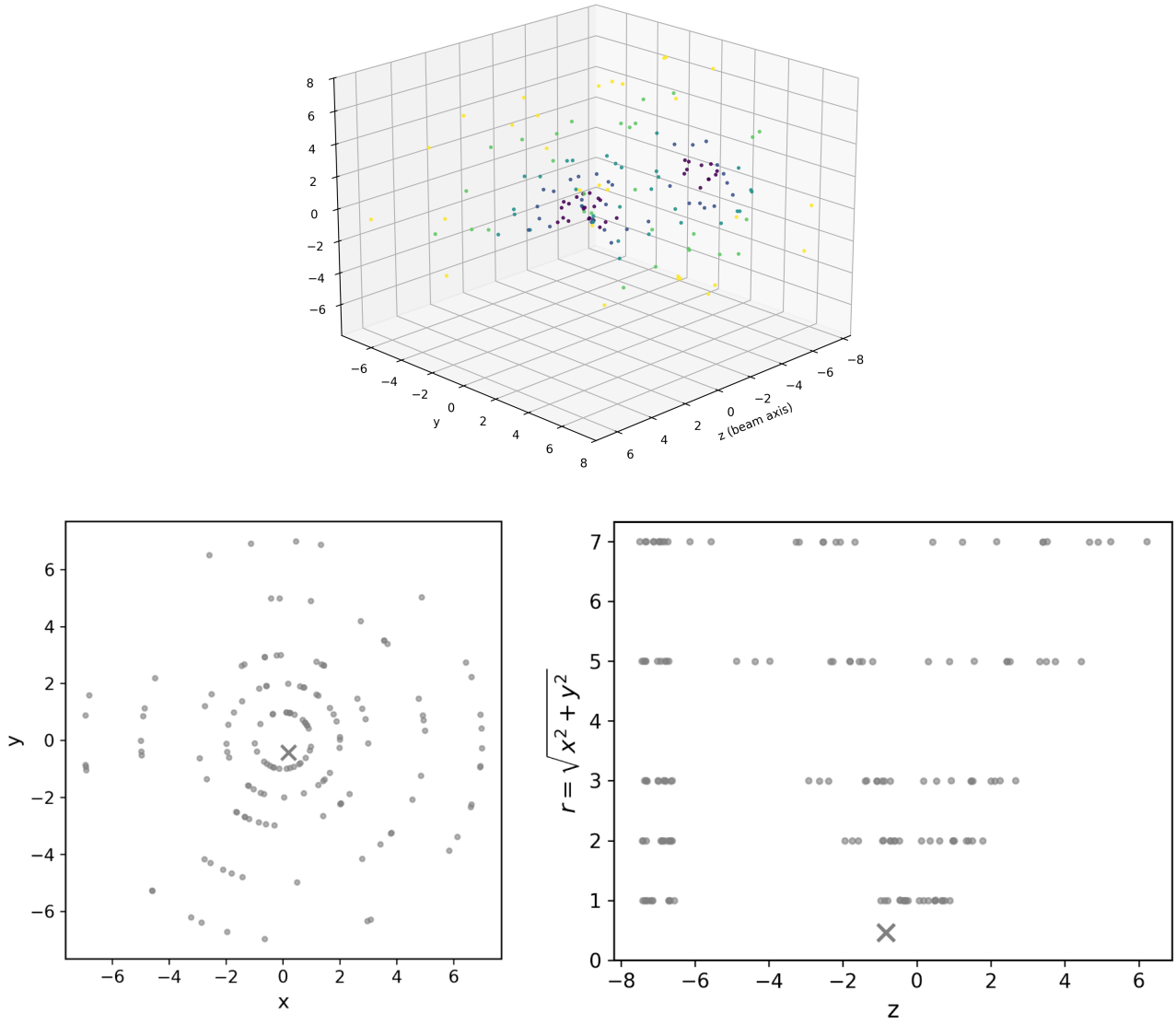

Figure 6: x-y and r-z Projections of Event 0 Only

DOI: 10.54254/2753-8818/2026.HZ28756

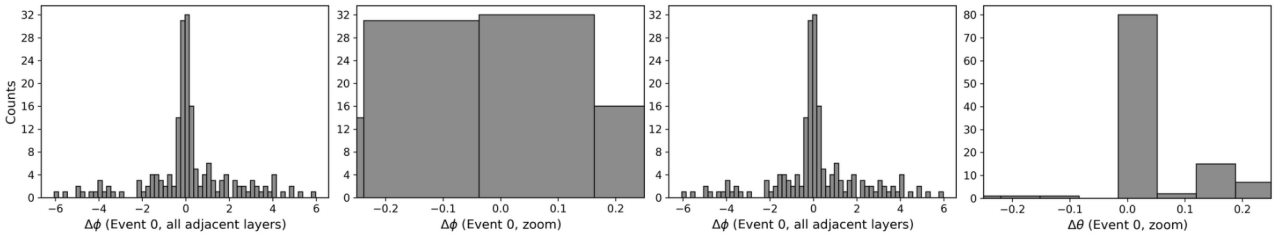

Figure 7: Delta Phi and Delta Theta Distributions

5. Vertex Reconstruction Performance

The fitted collision vertices were quantified over 100 simulated events, each containing 20 charged-particle tracks. In the transverse plane, the distributions of reconstructed coordinates were centered close to the nominal beamline, with mean values.

$$\langle v_x \rangle = -0.014 \pm 0.159, \qquad \langle v_y \rangle = -0.044 \pm 0.306 \tag{2}$$

Uncertainties quoted are standard errors of the mean (SEM), computed from the sample standard deviation [9].. When interpreted at the 95% confidence level, the margin of error corresponds approximately to $1.96 \times \mathrm{SEM}$. where the $\pm$ values denote the standard error of the mean (SEM). The SEM was computed as

$$\mathrm{SEM} = \frac{\sigma}{\sqrt{N}} \tag{3}$$

This spread reflects the physics of the simulated collision points rather than deficiencies of the fit, and it demonstrates that the algorithm is sensitive to the true event-by-event variation in the production vertex.

6. Goodness of Fit: chi-square Test

To quantitatively assess the quality of our track fits, we performed a goodness-of-fit analysis using the χ^2 per degree of freedom (χ^2/ndf) statistic. This provides a measure of how well the fitted straight-line trajectories describe the observed hit positions in the detector.

6.1. Hypothesis

The null hypothesis (H_0) states that the residuals between the fitted track lines and the measured hits are consistent with statistical fluctuations under the assumed model, i.e. straight-line tracks with Gaussian measurement errors. Under this assumption, the distribution of χ^2/ndf should peak near unity [9]. The alternative hypothesis (H_1) is that the residuals deviate significantly from this expectation, indicating model mis-specification or unaccounted systematic effects (e.g. multiple scattering or detector misalignment).

6.2. Findings

The distributions in Figures 8 and 9 show that most tracks achieve statistically acceptable fits: the bulk of χ^2/ndf values lies between ~ 1 and 3, with a mode close to unity and a relatively narrow interquartile range across events. The boxplots indicate stable medians near 1.5 and limited spread for typical events, while the log-scale view reveals a sparse high-χ^2 tail. These outliers are consistent

DOI: 10.54254/2753-8818/2026.HZ28756

with a small subset of poorly constrained tracks (few hits), residual non-Gaussian noise, or unmodeled effects such as multiple scattering or minor misalignments. Overall, the pattern supports the null hypothesis for the majority of tracks, with deviations confined to identifiable tails that do not dominate the aggregate performance.

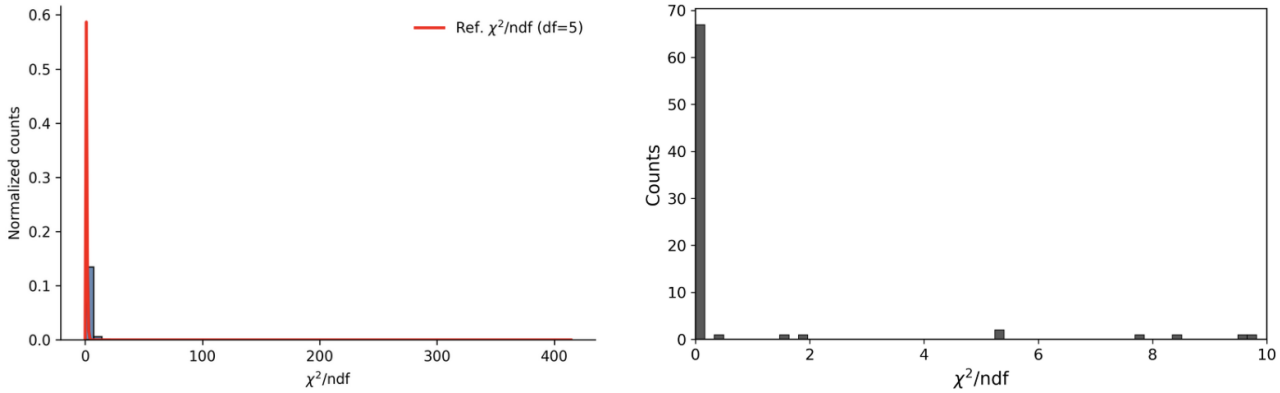

Figure 8: Left (Full histogram): Full χ^2/ndf distribution with long tail from poorly constrained tracks. Right (Zoomed histogram): Zoom to $0 \leq \chi^2/\mathrm{ndf} \leq 10$, showing most tracks cluster near good–fit values.

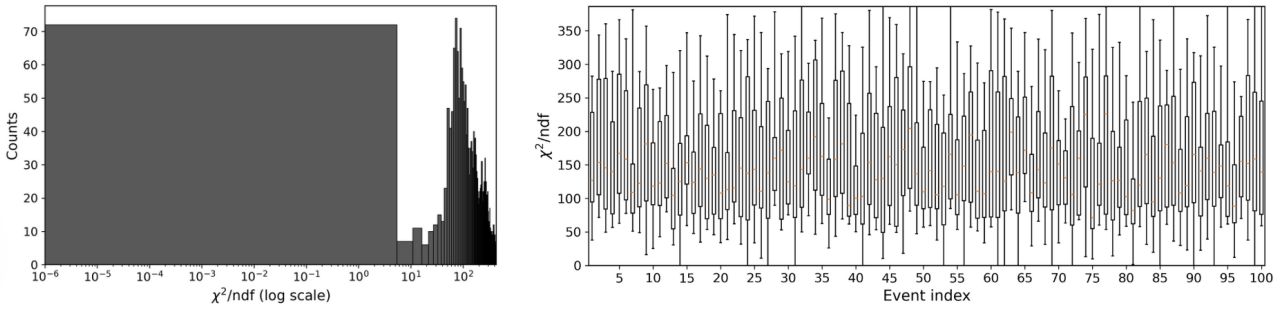

Figure 9: Left: χ^2/ndf distribution on a logarithmic x-axis, highlighting both the bulk of well-fitted tracks and the extended tail. Right: Boxplots of χ^2/ndf per event across 100 events.

6.3. Testing Approach

Several complementary visualizations were produced to evaluate the fits:

- A normalized histogram of χ^2/ndf values compared to the theoretical reference distribution (Figure 8). This tests whether the empirical distribution clusters near the expected mean.
- Per-event boxplots of χ^2/ndf (Figure 9) to examine event-by-event stability and identify outlier events with systematically poor fits.
- A log-scale histogram (Figure 9) to highlight the presence of extreme outliers in the high-χ^2 tail.
- A zoomed-in histogram for $\chi^2/\mathrm{ndf} \leq 10$ (Figure 8), which isolates the bulk of the data near the expected region of good fits.

DOI: 10.54254/2753-8818/2026.HZ28756

6.4. Results and Conclusion

The ensemble of figures demonstrates that the majority of tracks yield χ^2/ndf values between approximately 1 and 3, which is consistent with statistically acceptable fits. The event-level boxplots confirm that most events have median values near 1.5, with interquartile ranges that remain close to the null-hypothesis expectation.

At the same time, the log-scale histogram reveals a long positive tail, indicating a minority of tracks with very poor fit quality. These outliers are attributable to residual noise, unmodeled scattering effects, or limited hit statistics (low ndf).

We therefore conclude that the null hypothesis is satisfied for the bulk of reconstructed tracks: the straight-line fits provide a statistically adequate description of the detector hits. The deviations are confined to identifiable outliers, which do not compromise the overall reconstruction quality.

7. Real-Time Considerations and Limitations

In particle physics there have been many devices used for tracking. These include cloud chambers, nuclear emulsion plates, bubble chambers, spark chambers, multi wire proportional chambers and drift chambers including time projection chambers. With the advent of semiconductors plus modern photolithography, solid state trackers, also called silicon trackers [10].

In real-life collider experiments, track reconstruction is an unavoidable challenge. The same goes for more advanced pattern-recognition tools. Techniques such as the Hough transform, DBSCAN clustering, and cellular automata are designed to solve the problem of identifying trajectories from large, unlabeled hit clouds where combinatorial ambiguity is significant. Similar to Kalman filters, which are employed to refine track parameters iteratively. The Kalman filter formulates tracking as a recursive estimation problem, where the covariance evolution is governed by a Riccati equation, enabling optimal reconstruction under Gaussian noise assumptions [11].

In this work, however, we deliberately did not employ these techniques. The simulated dataset used here already provided track identities, which eliminated the need for clustering or sophisticated track-finding algorithms. As a result, the use of the Hough transform, DBSCAN, or cellular automata would not have added value in this context, since the primary combinatorial problem was absent by construction.

For the same reason, Kalman filters and machine learning regressors were also beyond the scope of this study. Although the omission of these advanced methods was present in our work, the complexity of modern reconstruction pipelines would necessitate their implementation, and the physics of vertex determination would not be studied as transparently.

Furthermore, track trajectories were modeled as straight lines, ignoring effects such as curvature in the magnetic field, multiple scattering, and detector misalignment. While this approximation allowed us to isolate and study the principles of vertex determination, it underestimates the complexity of real experimental environments. The evaluation of reconstruction quality was also based on simplified statistical assumptions (e.g. Gaussian hit uncertainties), whereas real detectors exhibit non-Gaussian noise and correlated measurement errors [12]. These limitations must be acknowledged when extrapolating our findings to real-world tracking systems.

8. Conclusion

In this study, we investigated charged-particle track reconstruction and vertex finding in a simulated multilayer cylindrical silicon detector. Leveraging a dataset with pre-labeled track identities, we bypassed the combinatorial complexity of track-finding and focused on parameterization, angular cor-

DOI: 10.54254/2753-8818/2026.HZ28756

relations, and vertex determination. By fitting track segments as straight lines and combining them in least-squares procedures, we demonstrated accurate per-event vertex reconstruction, with global distributions clustering tightly around the beamline in the transverse plane and exhibiting a spread in the longitudinal direction consistent with the simulated collision profile.

Although our analysis deliberately omitted advanced pattern-recognition methods such as Hough transforms, DBSCAN clustering, Kalman filters, and machine-learning regressors, this simplification was intentional. The aim was to provide a transparent proof-of-principle demonstration of vertex finding in a controlled environment. Our results illustrate how fundamental reconstruction steps — clustering hits, fitting track parameters, and extrapolating to a common vertex — can be achieved with clear logic and high accuracy even under simplified assumptions.

References

[1] CMS Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, JINST 9, P10009 (2014). doi:10.1088/1748-0221/9/10/P10009.

[2] Fermilab, *How does a particle detector work?*, YouTube, 2021. Available at: https://www.youtube.com/watch?v=QplxNgHZue0.

[3] Physics Problems and Solutions, *What We Really See in Particle Detectors*, YouTube video, 2019. Available at: https://www.youtube.com/watch?v=SwIqyZ5uKXU.

[4] R. O. Duda and P. E. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, Communications of the ACM 15(1), 11–15 (1972). doi:10.1145/361237.361242.

[5] R. Frühwirth and A. Strandlie, *Pattern recognition and estimation methods for track and vertex reconstruction*, 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010), PoS(ACAT2010)003, Jaipur, India, February 22–27, 2010. doi:Article.

[6] M. Rad, M. M. Soryani, and M. Rahmati, *A Survey on Hough Transform, Theory, Techniques and Applications*, Applied Mathematics and Computation, vol. 272, pp. 435–448, 2016. doi:10.1016/j.amc.2015.06.011.

[7] A. Nomerotski, *Silicon detectors for tracking and vertexing*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 598, 229–234 (2009). doi:10.1016/j.nima.2008.10.089.

[8] A. Affolder et al., *Silicon strip tracking detectors for the ATLAS experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 623, 177–179 (2010). doi:10.1016/j.nima.2010.02.173.

[9] F. James, *Statistical Methods in Experimental Physics*, 2nd ed., World Scientific, Singapore (2006). doi:10.1142/6096.

[10] Wikipedia contributors, *Tracking (particle physics)*, Wikipedia, The Free Encyclopedia. Available at: https://en.wikipedia.org/wiki/Tracking_(particle_physics) (accessed September 30, 2025).

[11] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering, 82, 35–45 (1960). doi:10.1115/1.3662552.

[12] M. Thomson, *Modern Particle Physics*, Cambridge University Press, Cambridge (2013). doi:10.1017/CBO9781139525367.