

# Reconstructing charged-particle trajectories in the PANDA Straw Tube Tracker using the LOfcal Track Finder (LOTF) algorithm.

Simon Gazagnes<sup>1,2,3a</sup>, Nasser Kalantar-Nayestanaki<sup>1</sup>, Johan G. Messchendorp<sup>1,4</sup>, Jenny Regina<sup>4</sup>, Tobias Stockmanns<sup>5</sup>, and Michael H.F. Wilkinson<sup>2</sup>, on behalf of the PANDA collaboration

<sup>1</sup> Energy and Sustainability Research Institute Groningen, University of Groningen, 9700 CC Groningen, The Netherlands

<sup>2</sup> Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, 9700 AK Groningen, The Netherlands

<sup>3</sup> Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA

<sup>4</sup> GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291 Darmstadt, Germany

<sup>5</sup> Forschungszentrum Jülich, Institut für Kernphysik, 52428 Jülich, Germany

Received: date / Revised version: date

**Abstract.** We present the LOfcal Track Finder (LOTF) algorithm, a method that performs charged-particle trajectory reconstruction using the Straw Tube Tracker, one of the central trackers of the antiProton ANnihilation at DArmstadt (PANDA) detector. The algorithm builds upon the neighboring relations of the tubes to connect individual hits and form track candidates. In addition, it uses a local fitting procedure to handle regions where several tracks overlap and utilizes a system of virtual nodes to reconstruct the z-information of the particle trajectories. We generated 30,000 events to assess the performance of our approach and compared its global track assignment efficiency with respect to two other track reconstruction methods. LOTF has (1) an average of 85% of *found* tracks, (2) the largest number of *Fully Pure* tracks, (3) the lowest amount of incorrect reconstructions, and (4) is significantly faster than the other two approaches. Further, we compared the z-reconstruction performance with one of the two alternative methods and show that LOTF improves the median z-error by a factor of 8.7. Finally, we tested our method using 3,750 data sets composed of 4 events each, demonstrating that our approach handles cases in which events are mixed. The *raw* (without parallelization) average reconstruction rate is about 68,000 hits/s, which makes the present algorithm promising for online data selection and processing.

## 1 Introduction

Observing exotic particle states is important in characterizing the properties of the fundamental building blocks of matter. Producing unexpected or not yet observed particle states with a small cross-section requires a vast number of particle collisions to create a detectable signal above the background noise. This complex task is now achievable with modern particle detectors that operate at very high luminosities and interaction rates. The upcoming antiProton ANnihilation at DArmstadt experiment (PANDA<sup>1</sup>, [1] [2]), that will be installed at the Facility of Antiproton and Ion Research (FAIR<sup>2</sup>, [3]), will provide such a setup. It will study interactions between protons or nuclei and antiprotons ( $\bar{p}p$ ,  $\bar{p}A$ ) using an antiproton beam with a momentum between 1.5 and 15 GeV/c, provided by the High Energy Storage Ring (HESR, [4]) that impinges on a target of hydrogen or noble elements. The

$\bar{p}p$  and  $\bar{p}A$  collisions typically produce a wide variety of particles which makes them suited to explore *e.g.* exotic particle states. PANDA will work with two main operational modes: the high-resolution mode with interaction rates of up to 2 MHz in [2], and the high-luminosity mode with interaction rates of up to 20 MHz and data rates up to 200 Gigabytes per second [5]. Traditionally, particle-collision experiments use hardware triggers to reduce the data stream. However, this is not foreseen with PANDA which will exploit software-based event filtering using on-line track reconstruction algorithms to select physically relevant events. Hence, efficient and fast track reconstruction algorithms are needed to identify the particle trajectories in a time frame matching the detector acquisition rate.

Most state-of-the-art track reconstruction algorithms are composed of two steps: a *track finding* step, where the individual hits are grouped into track candidates using Pattern Recognition techniques, and a *track fitting* step, where the particle properties are determined based on the tracks found. These two steps are often performed con-

<sup>a</sup> email: [gazagnes@utexas.edu](mailto:gazagnes@utexas.edu) (corresponding author)

<sup>1</sup> <https://panda.gsi.de/>

<sup>2</sup> <https://fair-center.eu/>

currently such that the particle properties (*e.g.* their momentum) are estimated during the reconstruction. Several techniques exist to perform the *track finding* step, usually categorized into *local* or *global* approaches. *Local* approaches, such as the Kalman filter [6], reconstruct particle paths by iteratively connecting hits one by one. *Global approaches*, such as the Hough transform [7], tackle the reconstruction by determining all track candidates simultaneously using the entire set of detections at once.

The performance of a given event reconstruction algorithm is usually measured in terms of (i) its *efficiency*, *i.e.* the fraction of real tracks that have been accurately identified; (ii) its *purity*, *i.e.* its ability to discard false hits or handle noisy data; and (iii) its computational speed. In practice, the design of the experiment itself must also be accounted for when implementing a track reconstruction algorithm. This is because specifics related to the efficiency of the data acquisition system and instrumental noise are also crucial aspects to consider. In general, all these aspects are tightly related to each other. For example, designing a very efficient track reconstruction procedure is often too slow to enable fast decision-making in real-time, and vice-versa. Hence, in-situ track reconstruction algorithms require an optimal balance between efficiency/purity and computational time, and must be optimized for the data acquisition it will work with.

In the context of the PANDA experiment, several strategies were developed over the past years. These methods build upon techniques such as the Hough transformation [8, 9], the Recursive Annealing Fit [10], the cellular automaton with Riemann mapping [11], the triplet finder [12], attribute-space-connected morphological filters [13, 14], or deep-learning [15, 16]. Additionally, a significant effort focused on parallelizing these algorithms using *e.g.*, FPGAs [17] or GPUs [8, 18]. Finally, some algorithms focused on supplementing track finders, *e.g.* the PZFINDER [10, 19] that aims at extracting the longitudinal position of the charged particles.

In this paper, we present the Local Track Finder (LOTF) algorithm, a novel method to perform track reconstruction using the Straw Tube Tracker (STT, [20]), an essential sub-detector for the track reconstruction and event identification of PANDA. The focus of LOTF is to enable an efficient in-situ track reconstruction. We developed a local approach using the neighboring relations of the tubes to connect individual hits and form track candidates (*e.g.*, similar to [11, 12, 14]). Further, LOTF builds upon a simple parametric fitting to disentangle individual trajectories in regions where several tracks overlap. Finally, it uses a system of virtual nodes, a technique originally proposed by Babai *et al.* [14], to reconstruct the z-information of the charged-particle trajectories. Its low computational complexity enables a fast and efficient track reconstruction, which makes it promising for online event selection.

This paper is organized as follows: Section 2 presents the PANDA central STT detector. Section 3 describes the implementation of our algorithm. Section 4 assesses the performance of our approach and Section 5 presents our

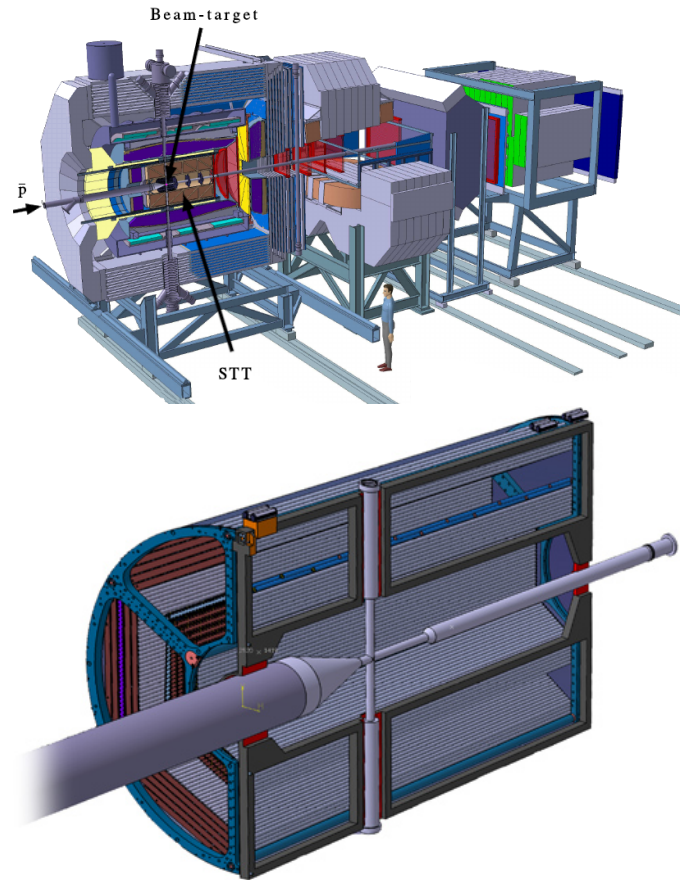


Fig. 1: Top: a schematic representation of the PANDA detector system with the beam-target and STT highlighted, adapted from [21]. In this work, we perform track reconstruction using the information from the STT, located around the beam-target interaction point. Bottom: a schematic representation of the STT, adapted from [20].

conclusions and discusses future prospects. Appendix 5 details the main algorithms implemented in LOTF.

## 2 The PANDA Straw Tube Tracker

The PANDA tracking system consists of central and forward trackers. The algorithm developed in this work performs track reconstruction using the STT system, which full description is provided in [20]. We detail here the most important characteristics.

In PANDA, the STT is a cylindrical volume located around the beam-target interaction point (see Figure 1), such that the target beam, which has a full-width diameter of  $\sim 3$  mm, will pass through its center. The angular acceptance of the STT is from  $20^\circ$  to  $140^\circ$ . Since PANDA will use a central magnetic field of 2T generated by a superconducting solenoid, particles with a transverse momentum larger than 50 MeV will transit, partly or fully, through the STT.

Regarding its configuration, the STT cylindrical shape is characterized by an inner radius of 15 cm, an outer

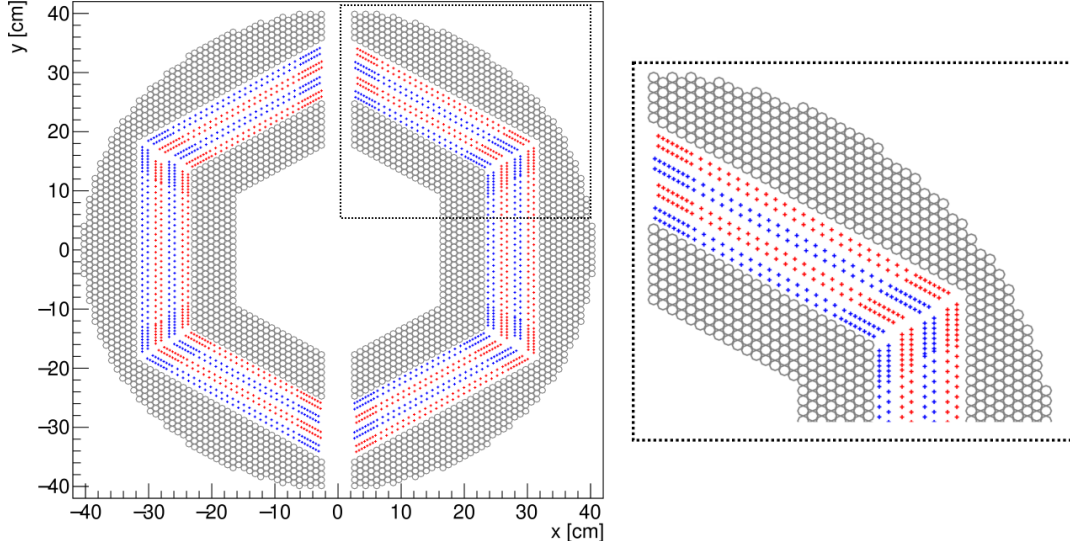


Fig. 2: The x-y projection of the tube positions in the STT (left), with a zoom-in on the top left section (right). The grey circles indicate the center-point coordinates of the axial tubes in the volume. The red and blue crosses show the center-point coordinates of the skewed tubes with an angle of  $+2.9$  and  $-2.9$  degrees, respectively.

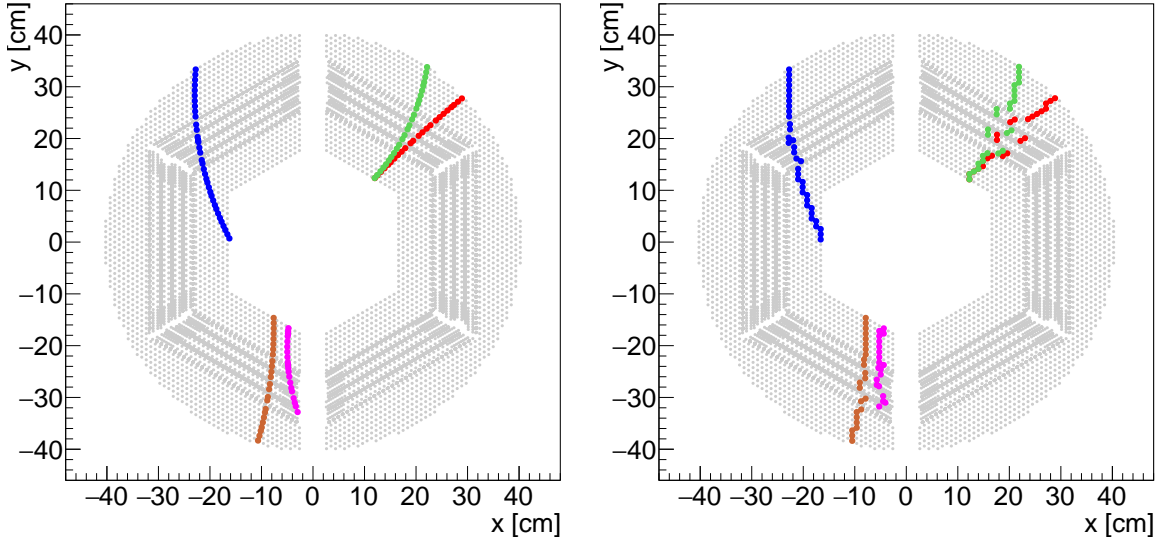


Fig. 3: Left: the x-y trajectories of 5 particles (in different colors) in a single event simulation. Right: the digitized version of the middle panel, showing the tubes that fired (*i.e.* were hit) for each track.

radius of 42 cm, and a length of 165 cm. Within this volume are arranged 4,224 straw tubes organized in 27 plane-parallel layers, with layer 0 being the closest to the beam-target interaction point. Each tube is 140 cm long with a 10 mm internal diameter and can be mounted in two different ways. In the innermost and outermost layers, the tubes are positioned parallel to the z-axis (beamline). In the eight middle layers, the straw tubes are skewed by  $-2.9$  and  $+2.9$  degrees with respect to the z-axis. The presence of these skewed tubes enables one to derive the z-information of the particle trajectory (see Section 2.2).

Each straw tube is filled with an Ar/CO<sub>2</sub> gas mixture and contains a thin wire along its axis. An electric

field is generated in the gas-filled area such that, when a charged particle passes through the tube, the gas becomes ionized. The gas ionization generates a current in the wire which is used as a readout signal (hereafter, we refer to these tubes as *activated* or *hit*). The time it takes the electrons to reach the wire is called the drift time. In general, the drift time is used to infer the *isochrone radius*, *i.e.*, the radius of the cylindrical surface that represents the possible positions that the particle might have traversed along the tube. Then, the exact particle path can be reconstructed by fitting a trajectory that is tangential to all the isochrone radii (see, *e.g.*, Figure 3 in [10]). The track reconstruction algorithm presented in this work does not

use drift time information. This choice is motivated by two aspects. First, recovering the particle hit positions along each tube is not vital for identifying track candidates because one can simply group hits based on their neighborhood relations. The reconstruction of the exact particle trajectory can be performed later when one needs to extract tighter constraints on the particle momentum. Second, accounting for the isochrone radii involves an additional computational cost that can be prohibitive for online track reconstruction.

We must note that the drift time has important global implications for the track reconstruction in PANDA. If the event rate is smaller than the maximal drift time of the straw tubes ( $\approx 250$  ns), hits belonging to successive events can overlap during the readout of the STT data, which is referred to as event-mixing. Event-mixing has a relatively small influence at 2 MHz (1 event every 500 ns) but has strong repercussions for track reconstructions based on the high-luminosity mode (20 MHz, 1 event every 50 ns). We discuss this aspect further in Section 4.2.

In this work, we represent the STT volume using a graph where each node represents a single tube. Each node is parameterized by the half-length, the three-dimensional direction vector, and the coordinates of the center point of the tube it represents. Additionally, these nodes contain the list of the tubes' direct neighbors whose number can vary between 2 and 22 (skewed tubes or axial tubes on the edges of the STT volume have a varying number of neighbors). In the following sections, we present successively the x-y and z-y projections of the STT volume.

## 2.1 x-y projection

Figure 2 shows the x-y projection of the STT volume based on the center-point coordinates of all the tubes. The tubes in the intermediate layers that have a skew angle with respect to the beam of  $+2.9$  and  $-2.9$  degrees are indicated with red and blue crosses, respectively. Figure 2 presents a simulated event with five particle trajectories projected on the x-y plane of the STT. To generate this event, we used the PANDAROOT software developed by the PANDA collaboration [22] which contains a simulation and digitization package. We used the FRITIOF (FTF) generator with an antiproton beam momentum of  $1.5$  GeV/c. The left panel shows the simulated trajectories of the five charged particles resulting from a  $\bar{p}p$  interaction, according to the transport model GEANT4 [23–25]. The right panel shows the digitization of the particles traveling along these paths based on the center-point coordinates of the tubes that are activated for each particle path. The exact hit positions along the skewed tubes can vary by  $\pm 3$  cm compared to the center-point coordinates. This causes the x-y projection of these digitized tracks to look discontinuous around the intermediate layers (*i.e.*, the green and red tracks in the right panel).

An accurate track reconstruction in the x-y plane is crucial to determine the trajectory curvature and reconstructing the transverse momentum component of a particle. In the next section, we show the STT coordinates

projected on the z-y plane, which is used to reconstruct the z-information of the particle tracks and derive the longitudinal momentum component.

## 2.2 z projection

The reconstruction of the particle tracks in the z direction is much more complex than in the x-y plane. This is because, in the global coordinate system of the detector, the z center-point position of almost all the straw and skewed tubes is at  $z = 35$  cm (the interaction point is at  $x, y, z = (0, 0, 0)$ ). A small number of skewed tubes, located at the sector boundaries, have a shorter length, hence they have a different z-coordinate compared to the full-length tubes.

For any activated tube, we do not have information about the exact z-coordinate of the particle hit. To refine the amount of information available in the z direction and improve the reconstruction of the z-information, we use a system of *virtual nodes*, a technique introduced in Babai *et al.* [13]. The concept of virtual nodes builds upon the design of the STT which includes layers of tubes having different slopes in the z direction. For two neighboring tubes having different z-slopes, there exists a *virtual* intersection point in the three-dimensional space that corresponds to the point where the distance between the two tubes is minimal. This virtual point is represented schematically in the top panel of Figure 4. The x, y, and z coordinates of this virtual point are strictly defined such that, if these two tubes are activated, one can assume that the particle passed through this virtual intersection point. This assumption is not always correct as it depends on the particle incident trajectory, yet it provides a first estimate to recover the particle z-information.

The procedure to determine the x, y, and z coordinates of the *virtual* nodes is the following: for all pairs of neighboring tubes, we extract the direction vectors of the two tubes and find the unit vector perpendicular to both lines. We then determine the point on this perpendicular line that is equidistant to both tubes. We take the coordinates of this point as the coordinates of the virtual node for this particular pair of tubes. All virtual nodes are set with only two neighbors which are the pair of tubes used to derive their coordinates. A virtual node is set as activated if both these tubes are activated.

The bottom panels of Figure 4 present the extended grid, in the x-y (left panel) and z-y (right panel) planes, including the virtual nodes in green. For the x-y plane, virtual nodes span in the inter-layers between the axial,  $+2.9$ , and  $-2.9$  degrees skewed tubes. For the z-y plane, these nodes substantially help to refine the discretization<sup>3</sup>. Hence, they provide key information for the reconstruction of the particle trajectories in the z-direction. We must note that, while these virtual nodes were already introduced by Babai *et al.* [13, 14], the authors used a different approach to compute their coordinates. For each pair

<sup>3</sup> In all this work, we only use the x-y and z-y plane to visualize the track reconstruction.



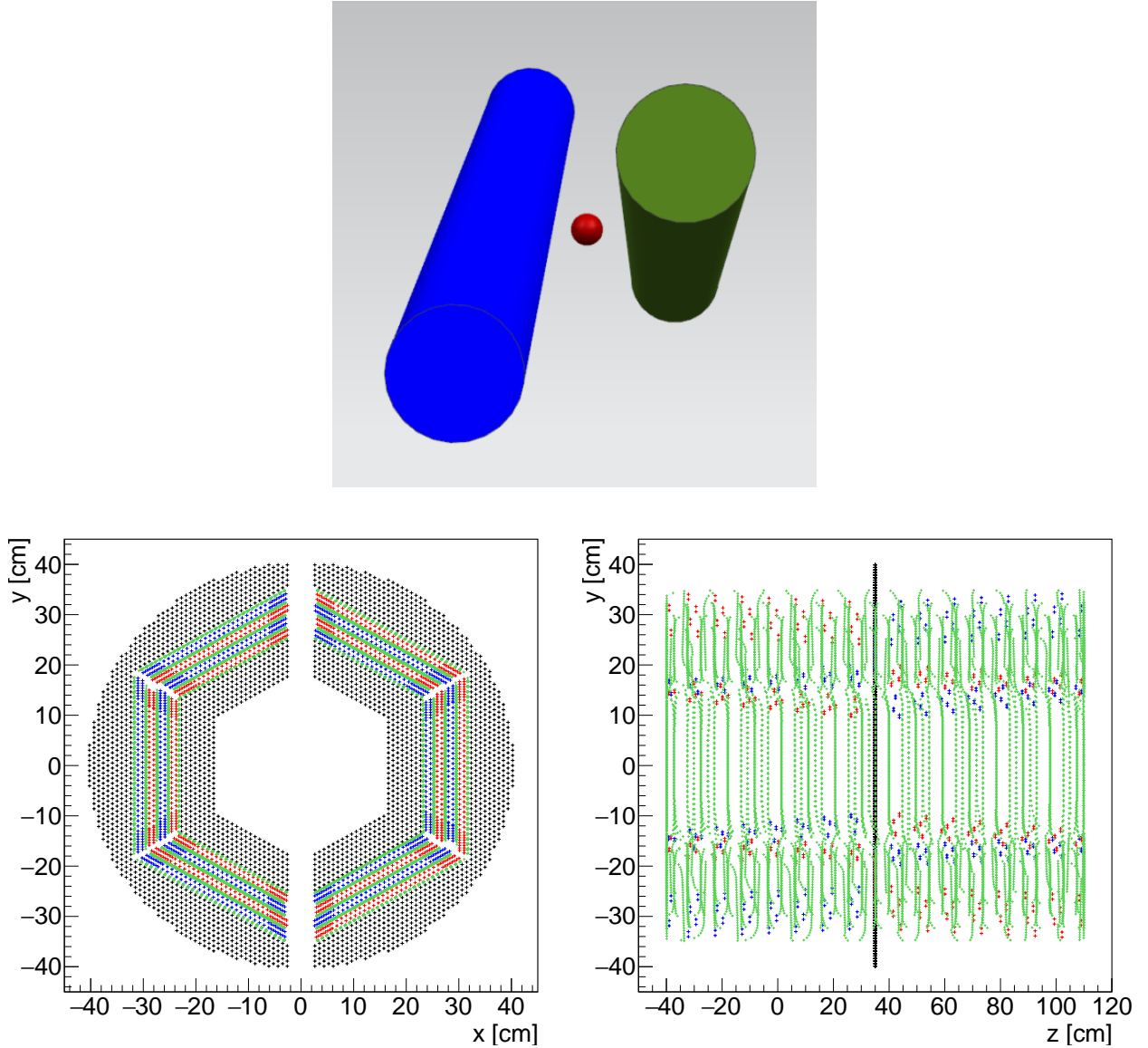


Fig. 4: Top: a schematic view of the position of a virtual node (in red) for a pair of tubes having different  $z$ -slopes (Figure from [13]). Bottom left: the  $x$ - $y$  projection of the extended grid. Black crosses indicate the center-point coordinates of the axial tubes, and red and blue crosses show the center-point coordinates of the skewed tubes with an angle of  $+2.9$  and  $-2.9$  degrees, respectively. The virtual nodes appear in green. Bottom right: same but for the  $z$ - $y$  projection. Because almost all the straw and skewed tubes span over the entire longitudinal length of the STT, their  $z$  center-point position is at  $z = 35$  cm (the beam-interaction point is at  $z = 0$ ). A small number of skewed tubes, located at the sector boundaries, have a shorter length, hence, they have a different  $z$ -coordinate.

of tubes with different slopes, the authors defined a two-dimensional plane using the central-point  $x$ - $y$  coordinates of each tube. Then, they computed the virtual point coordinates as the coordinates of the point at the center of this two-dimensional plane. The positions of the virtual nodes using their approach significantly differ from our method which assumes that the particle passed through the point where the distance between both tubes is minimal. As a consequence, the  $z$ -coordinates of the virtual nodes in [14] lie in a narrower range, and this has important shortcomings for the  $z$ -reconstruction. In Section 3.1, we show that

our approach provides a better reconstruction of the  $z$ -information.

The  $z$ - $y$  projection of the particle tracks shown in Figure 3 is presented in Figure 5. The left panel presents the simulated trajectories of the five particles in the  $z$ - $y$  plane and the right panel displays the  $z$ - $y$  coordinates of the virtual nodes (stars) that are activated. The comparison of the two panels illustrates that virtual nodes only provide a first estimate for the particle  $z$ -transit. Section 3.2 further details how we build upon these nodes to reconstruct tracks that resemble the simulated ones.

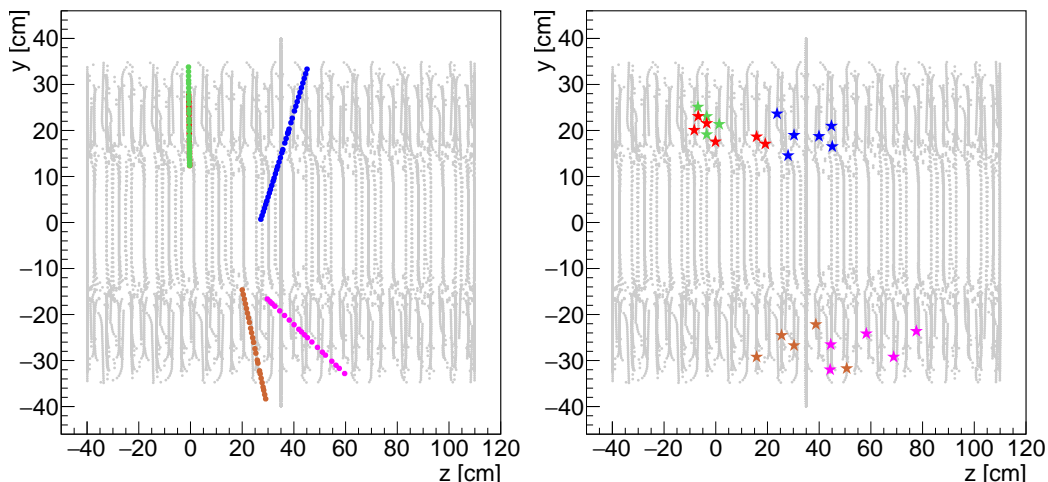


Fig. 5: Left: the simulated paths in the  $z$ - $y$  plane for the five charged particles in the event shown in Section 2.1. The red and green tracks are overlapping in the upper left part. Right: the coordinates of the virtual nodes (stars) that are *activated* for each track.

### 3 The LOTF algorithm

In this work, we aim at providing an algorithm suitable to perform an in-situ track reconstruction using the STT information, both in the high-resolution (2 MHz interaction rate) and high-luminosity (20 MHz interaction rate) operational modes. Performing a real-time reconstruction is challenging given the diversity of physical processes, involving complex trajectories' morphology with long-lived particles and tracks from secondary decay particles. Hence, the tracking system must have both high track reconstruction efficiency and a high reconstruction rate matching the data acquisition rate. To achieve these requirements, LOTF builds upon low-computational complexity techniques, selected for their potential in extracting complex, irregular, and potentially overlapping patterns in the STT. Additionally, while some track reconstruction methods perform the track reconstruction in a different parameter space to (*e.g.*, the Hough space [8] or orientation space [14]), LOTF only uses the Cartesian space (*i.e.*, the  $x$ ,  $y$ , and  $z$  coordinates of the different hits), hence avoiding potential overheads due to space transformations.

Section 3.1 details the  $x$ - $y$  reconstruction strategy and Section 3.2 describes the procedure to recover the  $z$ -information of the charged-particle trajectories.

#### 3.1 $x$ - $y$ reconstruction

Our approach to reconstructing the particle trajectories in the  $x$ - $y$  plane consists of three steps:

- The *connect* phase identifies track candidates based on hits located in the inner- and outer-most tube layers.
- The *fitting* phase extends the reconstruction in regions where several track candidates overlap using a parametric fitting procedure.

- The *merging* phase combines track candidates to produce the final reconstructed tracks.

We describe these phases in the following sub-sections.

##### 3.1.1 The *connect* phase

Algorithm 1 in Appendix A.1 details the procedure used during the *connect* phase, and we summarize the main steps here. Algorithm 1 aims at identifying portions of tracks that are isolated. To achieve this, we extract the activated nodes that have less than five neighbors (we consider as neighbors the activated tubes directly surrounding the current node) and that belong to the inner- and outer-most layers of the STT (we refer to these layers as the “limit layers”). We note that the choice of using five neighbors as a threshold is empirical, but we find that it generally characterizes configurations where tracks are overlapping. Additionally, we extract nodes located in limit layers because they usually correspond to a track extremity. We start from these nodes and iteratively look among the surrounding neighbors to identify the track continuation. These neighbors are added to the current track if there is a low complexity in assessing whether they belong to the same track or not. We consider a case to be complex if the current node

1. has too many neighbors (more than 5),
2. has neighbors in multiple directions (*i.e.*, located on all surrounding layers),
3. has no other activated neighbors surrounding it.

We note that #1 and #2 usually suggest that there might be several adjacent or overlapping tracks around the region where the current node is located. These cases are complex and are resolved during the *fitting* phase. On the other hand, one might encounter case #3 if a track is complete or if there exists a gap such that the next

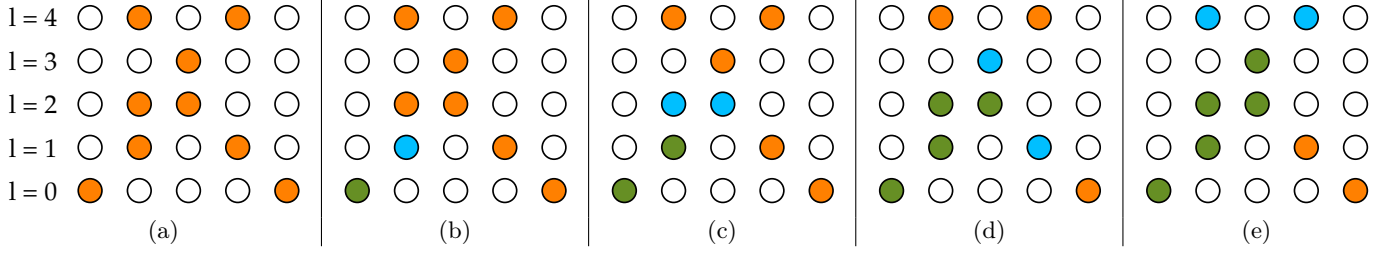


Fig. 6: A schematic example showing the functioning of the *connect* procedure. In practice, our algorithm is applied to the STT hexagonal geometry where each tube can have up to 22 neighbors. In this schematic, for simplicity, we define a squared geometric configuration where each tube (empty black circle) can have at most 8 neighbors. Orange circles are activated tubes, green circles are tubes that belong to the track being reconstructed, and blue circles show the current list of activated neighbors at each step. The left panel shows the index of each layer  $l$ . In (a), we show the set of activated tubes. In (b), we start from a node in the inner layer ( $l = 0$ ). This node has only one active neighbor, thus the case is trivial and we add this neighbor to the current track. In (c), we have two neighbors on layer 2. Both these nodes are adjacent and located consistently with respect to the current direction of the track, hence we connect both to the current track. In (d), we have neighbors both in the upper and lower layer. Given that we were adding nodes on increasing layer number (from  $l = 0$  to  $l = 2$ ), we connect only the node that belongs to layer 3. In (e), we encounter a complex case where we have two neighbors on the next layer, but these nodes are not adjacent. Finding which of these nodes should be added is not trivial. The *connect* procedure stops and the reconstruction will continue during the *fitting* phase.

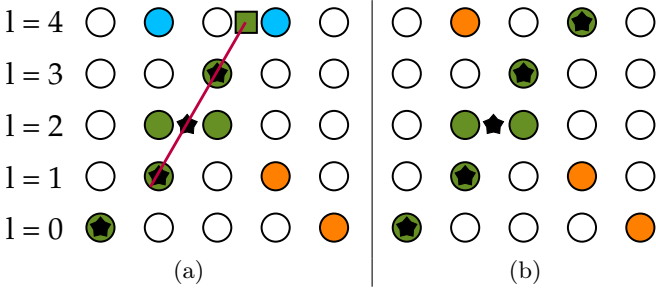


Fig. 7: The continuation of the schematic example presented in Figure 6. Orange circles show activated tubes, green circles represent tubes that belong to the track being reconstructed, and blue circles show the current list of neighbors. The black stars show the position of the *anchor* nodes for the green track. On layer  $l = 2$ , the anchor node position corresponds to the average of the coordinates of the two green nodes. In (a), we fit a parametric line using the position of the three most recent anchors in the track. The green square shows the predicted position of the next hit given the line equation and the layer distance. The right blue node has the smallest euclidean distance to this prediction, hence it is connected to the track in (b).

neighbors are not direct neighbors of the last-added node. To disentangle both cases, we label the track as complete if it starts and ends in the limit layers of the STT. If this is not the case, we flag the track to be investigated during the *merging phase*, as there might exist another track candidate it can be connected to.

The *connect* procedure is sufficient to find all “simple” isolated tracks with no or few complex neighboring cases entirely. However, it does not perform a complete reconstruction for adjacent or overlapping tracks which are only partially found at the end of this phase. Figure 6 shows

a schematic representation of a track reconstruction using the *connect* procedure.

### 3.1.2 The *fitting* phase

During this phase, we resume the reconstruction of partially reconstructed tracks. Algorithm 2 in Appendix A.2 details the fitting procedure and we present the main steps here. As mentioned in the previous section, the *connect* procedure provides incomplete track portions in regions where multiple particle trajectories might cross or be too close to each other. In such configurations, finding the track continuation based on the list of neighbors around the last node added to the track is not trivial. We use a linear fitting approach based on a system of *anchors* nodes to resolve these complex cases. *Anchor* nodes are supplemental nodes representing at maximum three individual tubes on the same layer. Since several adjacent tubes on the same layer might be activated by the transit of a single particle, anchor nodes provide a simple way to join these multiple detections in a single *anchor* (we define the size of an anchor node as the number of tubes it represents).

Anchor nodes are determined on the fly during the track reconstruction. We construct a new anchor node each time a node from a different layer than the previous node is added to the track. The coordinates of a given anchor node are obtained by taking the average of the coordinates of the nodes it contains. In general, the *anchor* system virtually smooths the current track candidate by only taking a smaller subset of nodes. Since the spatial distribution of these anchor nodes is smaller than for the individual tubes, they are well suited for the fitting procedure used to identify the best track continuation.

The fitting strategy is implemented as follows: we use the x-y coordinates of the three closest anchor nodes (at most) to the track extremity and fit a parametric line in

the x-y plane, which equation is given by  $x = x_0 + x_1 * t$  and  $y = y_0 + y_1 * t$  with  $t$  the independent variable. The choice of a linear fit might appear counter-intuitive since particles have circular motions, whose curvature depends on their momenta. Yet, as emphasized by, *e.g.*, Figure 3, straight lines provide a fair approximation for the particle trajectories on small scales (three different layers at most). Importantly, this approach has a small computational complexity which makes it promising to minimize the overall computational time. We note that we also tested a quadratic fitting approach, and found that the computational overhead was negligible when using a limited number of points. Nevertheless, we did not find significant improvements in the fit quality, and this approach did not perform well for short track candidates (less than three anchor nodes, hence not enough information to constrain the fit), or when applied to low-momentum particles circling within the STT volume. Therefore, we use the linear fitting approach as the default option in LOTF, but the current code implementation enables the user to switch to a quadratic fit if preferred.

The resulting line equations enable us to predict the position of the next hit based on the current portion of the track identified. Using this prediction, we look among all potential neighbors to find the one closest to it. We add a fixed threshold of 5 cm to the maximal distance a node can have compared to the predicted position. 5 cm is roughly two times the maximum distance that can be found between the center point of two neighboring tubes. Hence, using this threshold avoids adding nodes unrealistically far from the current track extremity. If a neighboring node successfully meets this criterion, and if it was not already assigned to another track candidate, it is added to the reconstructed track. On the other hand, if the node was already connected to another track candidate, we test whether both tracks should be merged according to their spatial orientation. To achieve this, we derive a local direction vector around the track extremities such that both tracks are flagged to be merged if the angle between their direction vectors is larger than 110 degrees (180 degrees means that both vectors are parallel). We note that this criterion has been chosen empirically as the value that works best after a visual inspection of a few events. Using a higher (lower) value decreases (increases) the likelihood that two tracks are merged. In general, we find that varying this criterion by  $\pm 20$  degrees has a limited impact on the reconstruction performance of event-based data with low track multiplicity because the number of complex merging cases (*i.e.*, cases where a single track candidate could be merged to different tracklets) remains small. Nevertheless, we have not extensively tested the limit of this value in the context of event-mixing data sets where the number of overlapping particle trajectories can be much larger (see Section 4.2). We plan to work on optimizing this criterion in future developments focused on improving the high-luminosity track reconstruction performance.

For any track, the fitting procedure ends if (i) there are no further neighbors; (ii) no good neighbor is matching

the predicted position of the next hit; or (iii) we found a neighboring track to connect with the current one.

Figure 7 illustrates the fitting procedure using the schematic configuration shown in Figure 6.

### 3.1.3 The *merging* phase

Before merging the tracks that were flagged during the *fitting* phase, we perform a global check of all tracks to identify particle trajectories that might still be incomplete. As specified in Section 3.1.1, we assume that a track is complete if its two extremities are on the boundaries of the STT. Incomplete tracks, as defined using this criterion, are either tracks whose extremities lie in the intermediate layers of the STT or tracks having a hit gap such that the remaining portion of the track to be connected is not directly neighboring the current track extremity. In the latter scenario, we typically reconstruct several track candidates that correspond to a single simulated trajectory. To determine whether a track candidate should be extended despite not having direct neighbors, we look for other track candidates within a radius of 5 cm around its extremities. As noted in Section 3.1.2, this distance represents a gap of at least two tubes between the two track extremities. If we find one neighboring track within this radius, and if its direction vector is consistent with the current track (using the same criterion as detailed previously, *i.e.*, the angle between their direction vectors is larger than 110 degrees), both are flagged to be merged.

Once we tested all tracks, for each pair of tracks to be merged, their respective nodes are combined consistently into a new path candidate.

### 3.1.4 Example of x-y reconstructions

In this section, we visualize two cases of track reconstruction in the x-y plane using the LOTF algorithm. These examples are simply for illustration, a complete performance assessment is presented in Section 4.

Figure 8 details the different steps of the track reconstruction algorithm for an event with five tracks (hereafter referred to as event A). Out of these five trajectories, one is isolated (upper left sector), two others are overlapping (upper right sector), and two are adjacent such that they have common neighbors (lower left sector). The bottom left panel presents the reconstruction status after the *connect* phase. The bottom left panel shows that the blue trajectory is already purely reconstructed (we refer to a track as purely reconstructed when it contains all and only the hits from a single simulated track) after the *connect* phase. The four tracks that are adjacent or overlapping are only partially reconstructed.

After the *fitting* phase, the four incomplete tracks have been expanded such that they are either complete or flagged to be merged during the next phase (the black stars show the pair of tracks flagged for merging).

Finally, at the end of the *merging* phase, the brown and pink tracks are purely reconstructed, and the red and



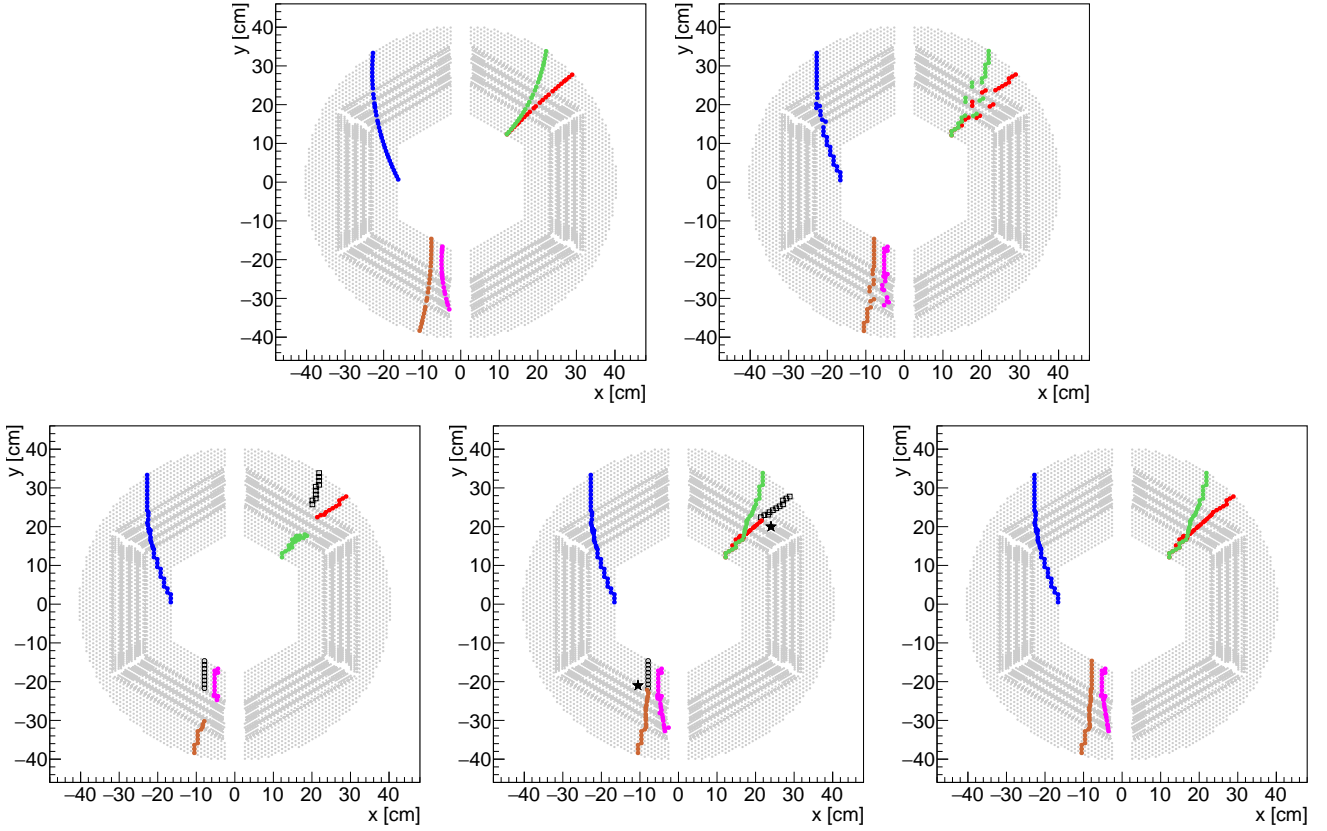


Fig. 8: An example of event reconstruction. Top row: the simulated paths of five particles shown in different colors (left) and the digitized tracks in the STT (right). Bottom row, left to right: the status of the reconstruction after the *connect* phase (each different marker/color represents a different track candidate), the reconstruction after the *fitting* phase (the black stars are set near tracks that will be merged in the third phase), and the final x-y reconstruction after the *merging* phase.

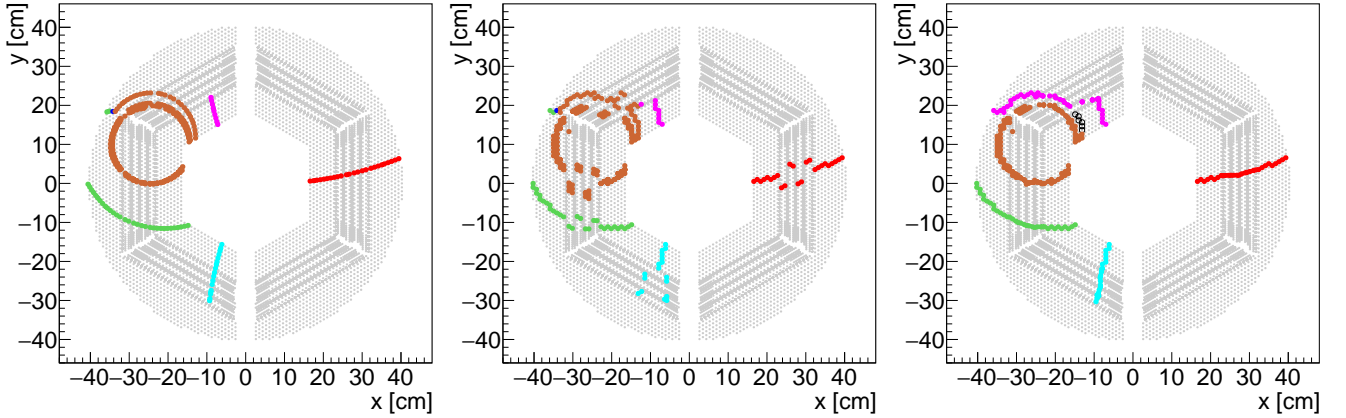


Fig. 9: Another example of track reconstruction for an event with five tracks. Left to right: the simulated particle trajectories, the digitized tracks, and the x-y reconstruction using the LOTF algorithm.

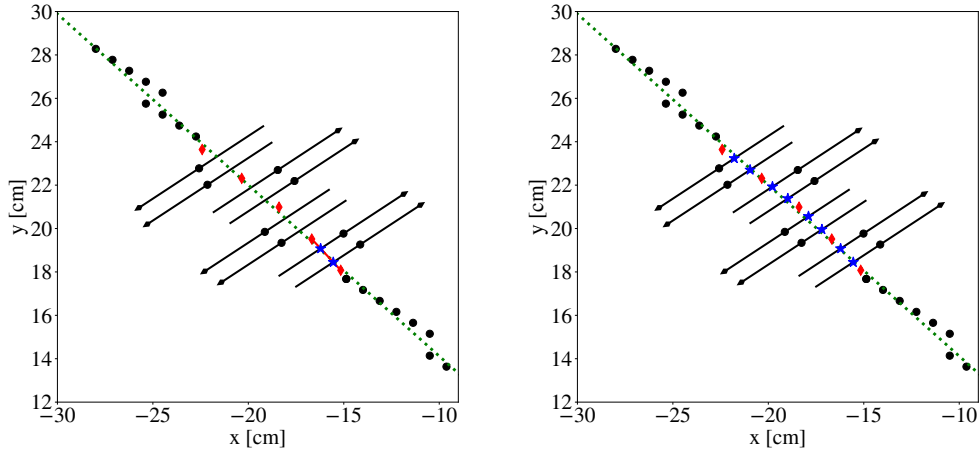


Fig. 10: An example illustrating how we use virtual nodes to refine the x-y hit coordinates along the skewed tubes. The simulated particle trajectory is shown with a green dotted line. The center-point coordinates of the activated tubes are indicated with black circles. The black arrows represent the skewed tubes, with the head pointing in different directions to denote the  $+2.9$  and  $-2.9$  degrees skew angles. Virtual nodes are shown with red diamonds. Left: we compute the intersection between the segment determined by two virtual nodes (red line) and the direction vector of the skewed tubes to refine the particle hit positions (blue stars) along the skewed tubes. Right: the final refined coordinates of the particle hits along all skewed tubes. For each skewed tube, determining the x-y hit coordinates precisely enables us to derive the z-information of the particle trajectory around these layers.

green reconstructions hold 94% of all the hits from their respective simulated tracks. Additionally, the shape of the reconstructed tracks is consistent with the simulated trajectories shown in the top left panel. This is because, as further detailed in the next section, we re-determine each node's coordinates in order to correct for the discontinuities around the skewed layers.

Figure 9 shows another event, hereafter referred to as event B, where a low-momentum secondary electron describes a circle in the STT volume (brown trajectory). The LOTF algorithm provides an almost complete reconstruction of the trajectory of this particle: more than 76% of the nodes in the simulated track have been recovered in a single reconstructed track, and only the outer part of the loop has been wrongly connected to another track (in purple). The connection between the purple track and part of the brown trajectory is a consequence of our merging approach that considers tracks incomplete if they end in the middle layers of the STT. We plan to improve this issue in the future by refining the criterion that decides whether two tracks should be merged.

### 3.2 z coordinates reconstruction

By construction, the ability of the algorithm to accurately recover the z-information for each track depends on its x-y reconstruction performance. This is because, as detailed in Section 2.2, the z-reconstruction is based on the virtual nodes system, where each activated virtual node must be correctly assigned to the reconstructed track it belongs to. To achieve the z-reconstruction on the fly without using the tubes' drift time information, we use the well-defined

coordinates of the virtual nodes to approximate the particle incident trajectory around the skewed tubes.

We implemented the procedure CORRECTSKEWEDXY, the pseudo-code of which is given in Appendix A.3, to approximate the particle incident trajectory around the skewed layers. We describe its functioning here. Between two virtual nodes located on different layers, there are always two layers of skewed tubes. Hence, the local particle incident trajectory can be approximated using the line determined by two virtual nodes. The intersection between this line and the slope vector of the skewed tubes provides an estimate of the exact x-y position of the particle hit along the skewed tubes. An example detailing this procedure is presented in Figure 10.

Given the relatively small skew angle of these tubes, the typical uncertainty on the z coordinate determined in this way is significantly larger than the typical error on the x and y coordinates. For example, a  $\pm 7.5$  mm difference in the x-y estimates will lead to an uncertainty of  $\pm 15$  cm in the estimated z coordinate. However, as the number of z-hits recovered for each track increases, the error on the reconstructed z-momentum will decrease. We note that the determination of the z-information for a particle trajectory is only possible if the particle crossed at least two layers of tubes with different z-inclinations.

To obtain the z-information consistently for all the nodes in a track, we fit a parametric line in the z-y plane ( $y = y_0 + y_1 * t$  and  $z = z_0 + z_1 * t$  with  $t$  the independent variable)<sup>4</sup> using all available z-coordinates from the virtual nodes and skewed tubes. e note that charged

<sup>4</sup> The choice of using the z-y plane instead of the z-x plane is arbitrary, and has no impact on the reconstruction results.

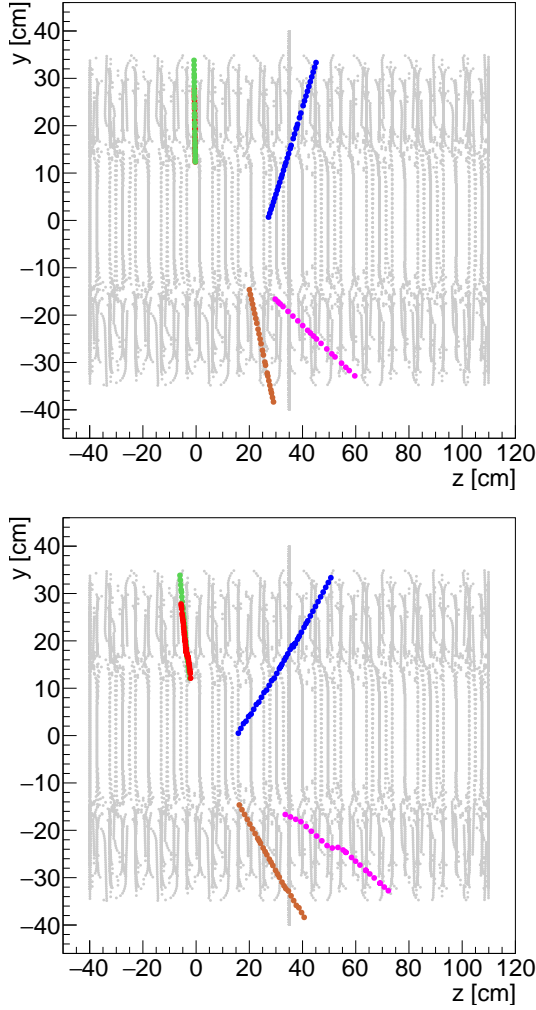


Fig. 11: The result of the reconstruction of the  $z$ -information for the event A. Top: the simulated trajectories of the particles in the  $z$ - $y$  plane. Bottom: the  $z$ -reconstruction obtained with the LOTF algorithm using an interpolation system that fits a  $z$ -parametric line.

particles moving in a solenoidal magnetic field follow helical trajectories, hence, their projection onto the  $y$ - $z$  plane repeats periodically rather than following a straight line. Nevertheless, for most particle tracks that are relevant in the context of the PANDA experiment, the radius of the helix is much larger than the traversed STT dimensions. As a result, the curvature of the particle paths is not very noticeable when projected onto the  $y$ - $z$  plane so the trajectories can be approximated by straight lines.

We note that there exists some scatter in the reconstructed hit-coordinates, and these values are typically contained within a relatively small coordinate range ( $\sim 10$  cm for the  $y$  coordinates, see *e.g.*, Figure 5). Additionally, we also lack the  $z$ -information in the inner part of the STT (between the beam-target interaction point and the STT inner radius of 16 cm). This means that additional constraints must be used to avoid getting unre-

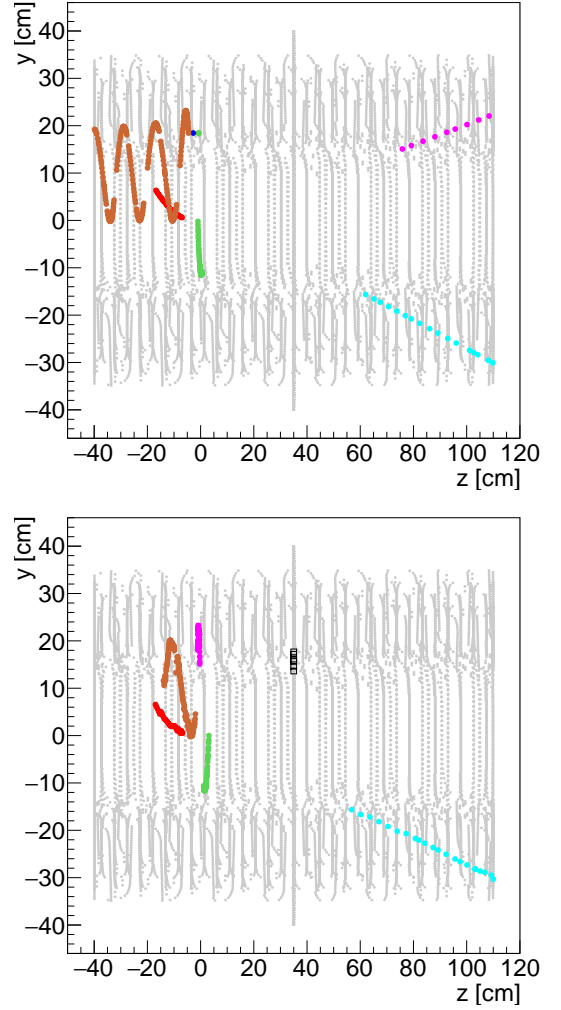


Fig. 12: Same as Figure 11 but for the event B. The  $z$ -information of the purple track is not correctly reconstructed because its  $x$ - $y$  reconstruction includes nodes from two independent trajectories (see the discussion in the text). Additionally, we only recover a single loop from the brown trajectory because this particle passed through the same tubes several times. The LOTF algorithm is not able to separate the distinct  $z$ -information from the individual loops that compose the full helical trajectory.

alistic  $z$ -reconstruction. In this work, we choose to add the point  $(x,y,z) = (0,0,0)$  to the list of values used for the  $z$ -reconstruction, but only for tracks that do not curl within the STT volume (*i.e.*, tracks that have a clear inner-to-outer or outer-to-inner direction). This assumes that the origin of these tracks is close to the beam-target interaction point, which is a fair assumption for stable particles such as charged pions, protons, or antiprotons, that transverse the STT, and short-lived particles, which decay almost instantly at the primary interaction point. However, particles with a medium lifetime like ground state hyperons can travel several tenths of centimeters before they decay into stable particles. Here the assumption that the

decay products come from the primary interaction point is not true. In these cases, our approach is not optimal and we plan to remove this assumption using the inner Micro Vertex Detector [MVD, 26] that will allow us to improve the accuracy of the z-reconstruction for both the long-lived particles and their decay products.

Figure 11 shows the z-reconstruction for event A. The mean z-error (reconstructed minus simulated values) for the five tracks is  $5.4 \pm 2.5$  cm which is relatively small compared to the typical z-extent of the simulated tracks.

Figure 12 shows the z-reconstruction for event B. The green, red, and cyan tracks have a mean z-error of 2.4, 0.1, and 2.1 cm, respectively. The trajectory of the low-momentum secondary electron (brown) is only partially recovered in the z plane, yielding an average z-error of 16.4 cm. This particle passed through the same tubes several times, and LOTF is not able to disentangle the individual loops that compose the full helical trajectory. Finally, the purple track has a mean z-error of 92 cm, emphasizing that the z information is poorly reconstructed. This is because the x-y reconstruction for this trajectory has merged portions of two independent simulated tracks (see Figure 9) such that the z-interpolation used coordinates from two separate trajectories. Consequently, the final z-reconstruction for this track is significantly different than the ground truth. This example emphasizes that the performance of the z-reconstruction crucially depends on the accuracy of the x-y reconstruction performed in the first place.

## 4 Performance

The purpose of LOTF is to enable in-situ track reconstruction with the STT. To meet the requirements of the PANDA physics experiment and have the highest chance of selecting events with rare or specific particle states, an optimal track reconstruction algorithm must have (1) large reconstruction efficiency (2) good reconstructed momentum resolution ( $1-2\%$  [27]), and (3), high reconstruction rates to match the expected 160 Mhits/s and 1,600 Mhits/s generated in the STT for the high-resolution and high-luminosity modes. In this work, we only assess the efficiency and reconstruction rate of the algorithm because both are most important for online track reconstructions. The reconstruction of the particles' momenta, which is performed after the track reconstruction and can be refined using the information from other trackers, will be investigated in a separate work.

We assess the performance of the LOTF algorithm based on 30,000 events generated using the FTF background generator of the PANDAROOT software<sup>5</sup>. We simulated 15,000 of these events using an antiproton beam momentum ( $\bar{p}_{\text{beam}}$ ) of 3 GeV/c, and 15,000 events using a beam momentum of 15 GeV/c. We used two different beam-momentum values to generate events with different track multiplicities and particle momentum,  $p$ . The left panel of Figure 13 shows the frequency spectrum of the number of

tracks per event. Most of the events have a track multiplicity between 1 and 4. On average, the simulation using a 15 GeV/c antiproton beam momentum has a larger track multiplicity per event than that of the simulation with  $\bar{p}_{\text{beam}} = 3$  GeV/c. The total number of STT tracks generated is 56,190 and 68,052 for the low and high beam momentum simulations, respectively. For the track analysis, we do not consider tracks that have fewer than 6 STT hits. The number of simulated tracks with at least 6 hits is 50,124 and 57,481 for the low and high beam momentum simulations, respectively.

The right panel of Figure 13 shows the distribution of particle momentum simulated for each case. As expected, the distribution of particle momenta is more extended towards large  $p$  when the antiproton beam momentum is 15 GeV/c. We note that the maximum  $p$  obtained with the latter simulation is around 11 GeV/c, which is smaller than  $\bar{p}_{\text{beam}}$ . This is because particles with larger momenta have very small scattering angles such that they do not cross the STT volume. Additionally, the momentum of each particle also depends on the number of particles that were formed during the interaction such that the distribution of individual  $p$  is not immediately proportional to  $\bar{p}_{\text{beam}}$ . Overall, the majority of primary tracks originate from charged pions  $\pi^+$  and  $\pi^-$  but the type of simulated particles varies in the simulations such that we cover a large spectrum of event types. Of the 30,000 events, 21,376 ( $\approx 71\%$ ) have overlapping tracks that are interesting for the performance assessment.

In Section 4.1, we compare the performance of the LOTF algorithm to the recent track reconstruction approach presented in Babai *et al.* [14], and to the current track reconstruction method, BARRELTRACKFINDER, implemented in the PANDAROOT software. In Section 4.2, we assess the algorithm's performance using data sets composed of 4 successive events to assess the impact of event-mixing. Finally, in Section 4.3, we scrutinize the timing performance of LOTF.

### 4.1 Track reconstruction performance

Here we focus on assessing the performance of LOTF in assigning the correct nodes to the correct track (Section 4.1.1), while simultaneously quantifying the corresponding average z-reconstruction error (Section 4.1.2).

#### 4.1.1 Comparison with other methods

In the context of cylindrical straw tube detectors, several track reconstruction approaches have been developed over the past years (e.g. [9, 10, 14, 17, 18, 28–30]). In this work, we compare LOTF to two of these approaches: the track reconstruction technique from Babai *et al.* [14] (referred to as BABAI+20) and the BARRELTRACKFINDER algorithm [30] implemented in PANDAROOT [22].

The BARRELTRACKFINDER algorithm builds upon conformal mapping [31]; circles in the x-y plane are transformed into straight lines and hits are merged based on

<sup>5</sup> We used PANDAROOT v12.0.1



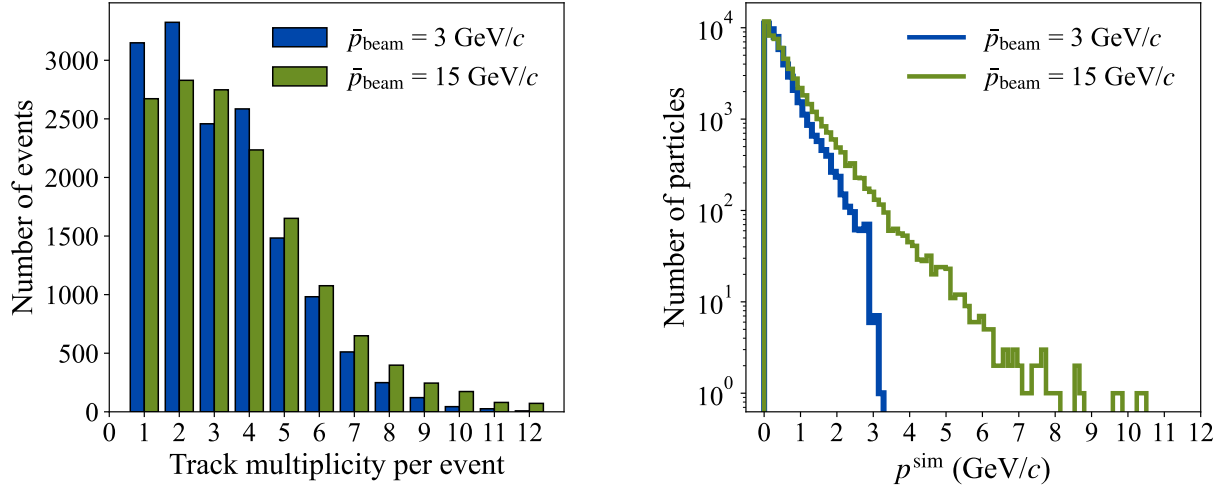


Fig. 13: Left: histogram representing the distribution of particle tracks per event in our experimental setup. A total of 30,000 events were simulated in the form of two times 15,000 events generated with an antiproton beam momentum ( $\bar{p}_{\text{beam}}$ ) of 3 and 15 GeV/c. Right: distribution of the particles' momentum for the 50,124 and 57,481 particles that reached the STT and have more than 5 STT hits in the simulations with a beam momentum of 3 GeV/c and 15 GeV/c, respectively.

$\bar{p}_{\text{beam}} = 3 \text{ GeV}/c$									
Method	Simulated	Reconstructed	Found	F-Pure	F-Impure	P-Pure	P-Impure	Ghosts	Clones
LOTF		58,701	43,130 <i>86.0%</i>	26,276 60.9%	3,266 7.6%	3,395 7.9%	10,193 23.6%	5,577 12.9%	9,994 23.1%
BABAI+20	50,124	58,868	40,338 <i>80.5%</i>	16,849 41.7%	1,973 4.9%	9,071 22.5%	12,445 30.9%	6,670 16.5%	11,860 29.4%
BARRELTRACKFINDER		57,869	39,727 <i>79.3%</i>	8,802 22.1%	0 0.0%	16,372 41.2%	14,553 36.7%	7,672 19.3%	10,470 26.4%
$\bar{p}_{\text{beam}} = 15 \text{ GeV}/c$									
Method	Simulated	Reconstructed	Found	F-Pure	F-Impure	P-Pure	P-Impure	Ghosts	Clones
LOTF		63,481	48,344 <i>84.1%</i>	29,821 61.7%	4,489 9.3%	3,469 7.1%	10,565 21.9%	6,731 13.9%	8,406 17.4%
BABAI+20	57,481	62,405	44,439 <i>77.3%</i>	18,357 41.3%	2,529 5.7%	10,518 23.7%	13,035 29.3%	8,075 18.2%	9,891 22.2%
BARRELTRACKFINDER		81,579	51,269 <i>89.2%</i>	8,599 16.8%	0 0.0%	20,646 40.3%	22,024 42.9%	16,397 32.0%	13,913 27.1%

Table 1: A summary of the performance of the three track reconstruction methods: LOTF (this work), BABAI+20, and BARRELTRACKFINDER. The top and bottom tables present the results of the tracks reconstruction for the low and high beam momentum simulations, respectively. The column “Simulated” gives the number of simulated tracks that have at least 6 hits in the STT, and the column “Reconstructed” shows the total number of reconstructed tracks for each approach. The column “Found” displays the number of *found* tracks, defined as the sum of all reconstructed tracks that are either F-Pure, F-Impure, P-Pure, or P-Impure (ranks 1 to 4, the description of these ranks is given in the text). The percentage shown in italic is the ratio of the number of *found* tracks over the number of simulated tracks. The columns “F-Pure” to “Clones” detail the number of reconstructed tracks in each rank. All the percentages shown in these columns correspond to the ratio of the number of tracks in a given rank over the number of *found* tracks (*i.e.*, percentages from the columns “F-Pure” to “P-Impure” add up to 100).

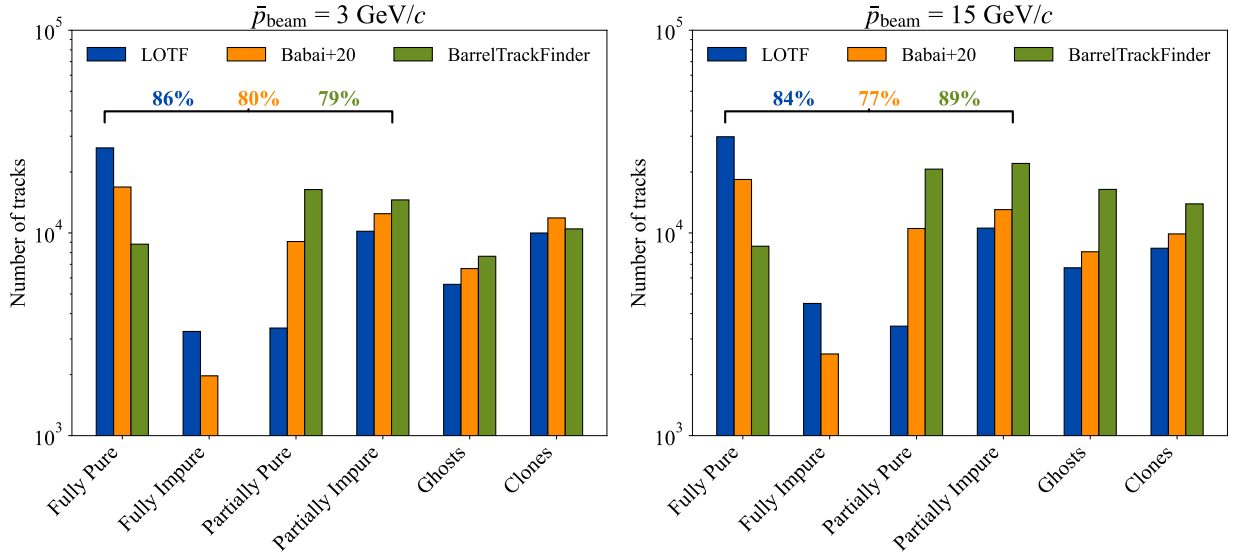


Fig. 14: Left: histogram presenting the distribution of reconstructed tracks in each rank (defined in the text) for the 15,000 events with  $\bar{p}_{\text{beam}} = 3 \text{ GeV}/c$  (50,124 simulated tracks with more than 5 STT hits). Right: same but for the 15,000 events with  $\bar{p}_{\text{beam}} = 15 \text{ GeV}/c$  (57,481 simulated tracks with more than 5 STT hits). The percentages shown in each panel represent the ratio of the number of *found* tracks (*i.e.*, being either *Fully Pure*, *Fully Impure*, *Partially Pure*, or *Partially Impure*) over the number of simulated tracks for the three different methods. LOTF has 86% and 84% of *found* tracks in the simulations with  $\bar{p}_{\text{beam}} = 3 \text{ GeV}/c$  and  $\bar{p}_{\text{beam}} = 15 \text{ GeV}/c$ , respectively. Additionally, it has the largest number of *Fully Pure* tracks (perfectly reconstructed tracks) and the lowest number of *Ghosts* and *Clones* (incorrect reconstructions) in both cases.

their connections in this new plane. The BARRELTRACK-FINDER can perform a global track reconstruction including information from several PANDA detectors. For a fair comparison, we only use it locally by including the STT information.

BABAI+20 performs charged-particle trajectory reconstruction using attribute-space-connected morphological filters [32]; the original data set is transformed into an orientation-based space where hits are grouped into tracks depending on which orientation subspace they belong to. The BABAI+20 approach is different from the local linear fitting procedure used in LOTF, yet, LOTF builds upon several techniques introduced by BABAI+20 such as the construction of a connected graph representing the STT geometrical information and the use of virtual nodes. Hence, the comparison to BABAI+20 is relevant to explore the performance improvements enabled by LOTF.

To assess the track reconstruction performance, we use the metric extended and described in [29] and implemented in the PANDAROOT software: the reconstructed tracks are assigned a *rank* which characterizes the quality of the reconstruction. These ranks are defined as follows:

Rank 1. *Fully Pure* track: the reconstructed track includes only the hits belonging to a single simulated track, and all hits from the latter have been found. In other words, the reconstructed track perfectly matches a given simulated track.

Rank 2. *Fully Impure* track: the reconstructed track holds all the hits from a simulated track, but also includes some impurities, *i.e.*, hits belonging to other

simulated tracks. The fraction of *impure* hits must not be larger than 30%.

Rank 3. *Partially Pure* track: the reconstructed track holds at least 70% of all the hits from a simulated track and only includes hits from that track (*i.e.*, no impurities allowed).

Rank 4. *Partially Impure* track: at least 70% of all the hits in the reconstructed track originate from a single simulated track. A Rank 4 track might include impurities.

Rank 5. *Ghost* track: less than 70% of the hits in the reconstructed track originate from the same simulated track.

Rank 6. *Clone* track: if several reconstructed tracks match with the same simulated track, only the reconstructed track with the largest intersection set is defined as the *true* reconstructed track, the other reconstructions are clones.

All the reconstructed tracks can only belong to one of these categories. Rank 1 is the best achievable rank, the total number of tracks having a rank comprised between 1 and 4 are regarded as *found* tracks while the number of *Clones* and *Ghosts* tracks (ranks 5 and 6) are regarded as incorrect reconstructions. An optimal track reconstruction algorithm should have the largest number of *found* tracks, the lowest number of *Clones* and *Ghosts*, and the largest number of *Fully Pure* tracks.

Using this metric, we compare the three algorithms in reconstructing the 50,124 and 57,481 tracks simulated with the FTF generator having at least 6 hits in the STT

in the low and high beam momentum simulations, respectively. Table 1 summarizes the number and percentages of reconstructed tracks in each category for the three methods. The percentage of *found* track is derived with respect to the total number of simulated tracks, and the percentages of tracks in each rank are derived with respect to the number of *found* tracks. Hence, for tracks that are either *Fully Pure*, *Fully Impure*, *Partially Pure*, or *Partially Impure*, these percentages represent the fraction of *found* tracks in each category. For the *Clones* and *Ghosts*, these percentages represent the fraction of incorrect tracks reconstructed in addition to the *found* tracks.

Figure 14 shows these results in the form of histograms for the two simulation data sets. LOTF has the largest percentage of *found* tracks in the simulation with  $\bar{p}_{\text{beam}} = 3$  GeV/c (86%), and the second largest in the simulation with  $\bar{p}_{\text{beam}} = 15$  GeV/c (84%, while the BARRELTRACKFINDER performs better with 89%). Interestingly, for the BARRELTRACKFINDER approach, the fraction of *found* tracks increases in the simulation with  $\bar{p}_{\text{beam}} = 15$  GeV/c compared to the simulation with  $\bar{p}_{\text{beam}} = 3$  GeV/c (89% against 79%), while LOTF and BABAI+20 have the opposite behavior (a drop of 2 to 3% in the number of *found* tracks for both). We note that BARRELTRACKFINDER has a factor of 1.5 to 2.5 times more *Ghosts* and *Clones* in the simulation with  $\bar{p}_{\text{beam}} = 15$  GeV/c than LOTF and BABAI+20, suggesting that its better performance comes at a cost of having a larger number of incorrect reconstructions.

Figure 14 also highlights that LOTF holds the largest number of *Fully Pure* tracks, a number 1.6 to 4 times larger for LOTF than for BABAI+20 and BARRELTRACKFINDER, respectively. The fraction of these tracks in the low (high) beam momentum simulation case is 60.9% (61.7%) for LOTF, 41.7% (41.3%) for BABAI+20, and 22.1% (16.8%) for BARRELTRACKFINDER. Additionally, LOTF has the lowest amount of *Ghosts* and *Clones* of all three methods for both data sets.

Finally, to process the 30,000 events using an Intel i9-11900H CPU at 2.50 GHz, LOTF took 44.3 s, BARRELTRACKFINDER 590.1 s, and BABAI+20 41,301.4 s. An in-depth analysis of the LOTF processing time as a function of the different reconstruction phases is presented in Section 4.3.

#### 4.1.2 z-reconstruction performance

In this section, we investigate the z-reconstruction performance of LOTF. We define the “z-error” as  $z_{\text{rec}} - z_{\text{sim}}$ , where  $z_{\text{rec}}$  and  $z_{\text{sim}}$  are the reconstructed and simulated z-coordinates of the particle hits, respectively. To compute these values, we match, for each event, each simulated track with the reconstructed track that has the largest intersection set (larger number of matching hits). We then compute the z-errors for each node in the intersection set. LOTF reconstructs the z-information using virtual nodes, a technique introduced in Babai *et al.* [14]. Hence, we choose to compare the z-reconstruction performance of both approaches. We do not include BARRELTRACKFINDER in

this comparison as there is no simple way to extract the reconstructed z coordinates for the latter method. We leave an in-depth analysis of the particles’ longitudinal momentum reconstruction, including a comparison with recent promising approaches (*e.g.*, the PzFINDER [10]), for future work.

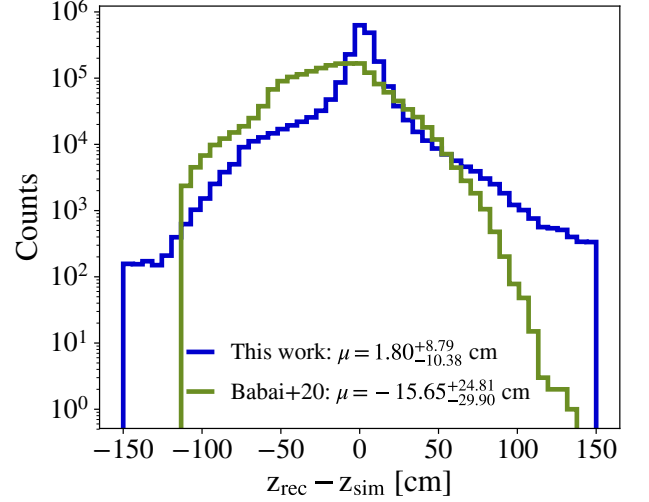


Fig. 15: The distribution of the z-errors defined as the difference between the z-values of the reconstructed tracks ( $z_{\text{rec}}$ ) and the z-values in the simulation ( $z_{\text{sim}}$ ). The blue curve shows the distribution for the LOTF algorithm and the green curve shows that of BABAI+20. The median and associated 16<sup>th</sup> and 84<sup>th</sup> percentile errors is  $1.80^{+9.79}_{-10.38}$  cm for LOTF and  $-15.65^{+24.81}_{-29.90}$  cm for BABAI+20. LOTF improves the median z-error by a factor of 8.7 compared to BABAI+20.

Figure 15 compares the distribution of the z-errors for BABAI+20 and LOTF for all the reconstructed tracks in the 30,000 events. As noted in Section 3.2, the z-information can only be reconstructed if the particle crossed at least two skewed layers in the STT. Among the 50,124 and 57,481 tracks from the low and high beam momentum simulations considered in this experiment, 43,583 and 48,712 fulfill this requirement, respectively, and are used for assessing the z-reconstruction performance. The BABAI+20 yields a median z-error of  $-15.65^{+24.81}_{-29.90}$  cm, a value 8.7 times larger than the  $1.80^{+9.79}_{-10.38}$  cm median z-error obtained with LOTF. Additionally, 68.3% of the reconstructed z-values are within  $\pm 10$  cm of the simulated values for LOTF, while this percentage is only 26.4% for BABAI+20.

The wings of the distribution of the z-errors with LOTF are wider than for BABAI+20. The number of tracks that have an average z-reconstruction error of more than 110 cm is 77 for LOTF, but only one for BABAI+20. The absence of very large under- or over-estimation of the z-coordinates for the BABAI+20 can be explained by the different approaches used by the authors to derive the virtual node coordinates. As mentioned in Section 3.2, the

virtual node z-coordinates lie in a limited range, typically between 0 to  $\sim 60$  cm, such that the z-error distribution is typically contained between -120 and +120 cm. While the number of cases where LOTF gives very large under- and over-estimation of the z-coordinates is low ( $< 0.1\%$  of all tracks), these cases could be prohibitive during actual experiments because they would lead to wrong estimates of the particles' longitudinal momentum component. Nevertheless, in practice, including information from the inner detectors such as the MVD would enable significant improvements in the z-reconstruction performance, and this aspect will be explored in future work.

Overall, these results highlight that LOTF recovers the z-coordinates with a relatively small error compared to the overall STT longitudinal dimension ( $\sim 150$  cm). This aspect is promising for extracting an accurate estimate of the particles' longitudinal momentum component.

## 4.2 Towards high-luminosity experiments

With the foreseen interaction rates and the relatively slow readout of the STT, one expects a large overlap in events for the PANDA high-luminosity mode. In theory, given the maximum drift time of the tubes (250 ns) and the expected 20 MHz interaction rate (two successive events separated on average by 50 ns), hits belonging to 4 to 5 independent events might overlap and make the reconstruction process harder due to the larger number of overlapping tracks. In practice, taking into account that events will be processed in bursts of no less than 2,000 ns and the 400 ns beam gap, we expect that on average 32 events will mix at 20 MHz (3.2 at 2 MHz).

The burst building scheme of PANDA implies that the data flow from the STT detector will be grouped in a time-based manner where track reconstruction algorithms not only have to find the tracks in real-time but also group these tracks into separate events. Therefore, any in-situ track reconstruction algorithms must account for the time-stamps of the hits as an additional dimension (see [33] for a recent review on time-based reconstruction algorithms). Performing track-reconstruction in a time-based manner has a supplemental computational cost which can be, when all detectors' information occurring within 2,000 ns is processed, up to  $\approx 1.5$  times higher than the typical event-based time reconstruction [11].

In this Section, we explore the impact of event mixing on the LOTF performance discussed in the previous section. To mimic these effects, we use the 15,000 events generated with an antiproton beam momentum of 3 GeV/c and re-arrange them into 3,750 data sets of 4 events each. Note that this is an idealization assuming that the data collected is in an event-based manner, *i.e.*, we do not consider the hits time-stamps as LOTF is not yet able to deal with this 4<sup>th</sup> data-dimension. Yet, this setup enables a fair initial investigation of the ability of our algorithm to disentangle overlapping tracks in more complex data sets.

Figure 16 shows an example of track reconstruction based on one data set encompassing hits from 4 successive events. As can be seen in this figure, the number of

overlapping tracks greatly increases in this setup, which significantly impacts the overall quality of the track reconstruction.

To assess the performance of the LOTF algorithm for these more complex cases, we use the metric described in Section 4.1.1. Figure 17 compares the distribution of track ranks for the set of reconstructed tracks based on the original simulation with a beam momentum of 3 GeV/c with 15,000 events and based on the new setup composed of 3,750 data sets of 4 events each. The percentage of *found* tracks decreases from 86% to 73% for the setup mimicking the impact of event-mixing. The number of *Fully* and *Partially Pure* tracks decreases while the number of *Fully* and *Partially Impure* tracks increase. This is expected since, as more and more simulated tracks overlap, the reconstructed tracks are more likely to include hits from several tracks.

The number of *Clones* remains fairly constant and we note a significant increase in the number of *Ghost* tracks. *Ghost* tracks are the consequence of the algorithm combining too many hits belonging to different simulated tracks. These tracks have not been matched to any simulated tracks, and emphasize the larger complexity of accurately reconstructing individual tracks when the likelihood that tracks overlap is higher. We note that accounting for the hits time-stamps can greatly improve the efficiency of the track reconstruction and significantly reduce the number of *Ghost* tracks because it adds an additional constraint when selecting a group of hits that might belong to the same track (see *e.g.*, [34]). Extending LOTF implementation to handle the time dimension is one of our priorities.

Overall, a loss of performance is expected given the more complex nature of the event-mixing case, yet the algorithm still performs well in identifying a large fraction of the tracks with decent accuracy. These results are promising for future PANDA experiments operating at interaction rates of up to 20 MHz.

## 4.3 Speed of the algorithm

Here, we assess the speed performance of the LOTF algorithm. We test our method using a laptop with a single core of an Intel i9-11900H CPU at 2.50 GHz. We measure the time and reconstruction rate (hits/s) for all events and explore its evolution as a function of the number of hits per data set processed. We consider the timings of all the 33,750 events (30,000 single events plus 3,750 data sets with 4 events concatenated) processed in this work. The total number of tracks considered is 179,523, and each track holds on average 21 hits. Over the 33,750 data sets considered, only 424 ( $\approx 1\%$ ) have more than 500 STT hits. These cases typically correspond to the data sets with event-mixing used in Section 4.2. To have a complete assessment of the computational footprint of our algorithm, we decompose the total time of the track reconstruction in several processing steps:

- “filling graph”: the time needed to fill the STT graph with the collected hits;
- “connect”: the time needed to complete the *connect* phase;



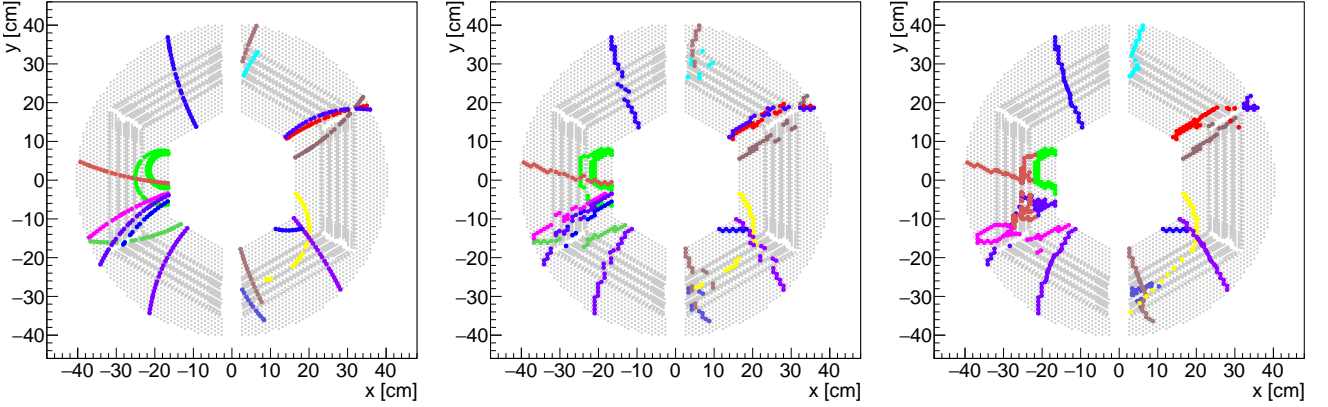


Fig. 16: An example of data set used to test our algorithm in Section 4.2. We concatenated 4 successive events to simulate the effects of event-mixing. Left: The simulated trajectories. Middle: The corresponding hits in the STT. Right: The reconstruction using our algorithm.

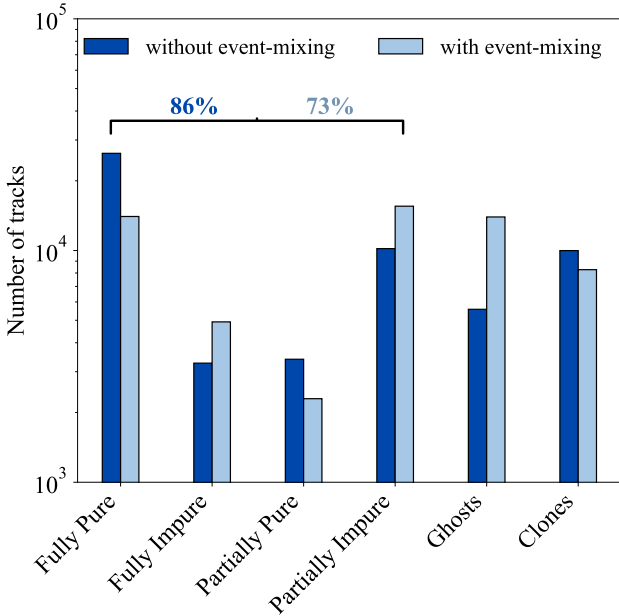


Fig. 17: Histogram presenting the distribution of reconstructed tracks in each rank for the 50,124 simulated tracks in the 3 GeV/c antiproton beam momentum simulation. The dark blue distribution shows the results for the 15,000 events processed one by one (no event-mixing), and the light blue distribution shows the results for the setup where hits from 4 successive events are processed at once (mimicking the effect of event-mixing). The percentage shown represents the fraction of reconstructed tracks that are *found* (i.e., having ranks between 1 and 4, see Section 4.1.1 for details).

- “fitting”: the time needed to complete the *fitting* phase;
- “merging”: the time needed to complete the *merging* phase;
- “z-reconstruction”: the time needed to perform the z-reconstruction.

In all time measurements presented hereafter, we do not consider input/output overheads. The top panel of Figure 18 shows the evolution of the average wall-clock time for each processing step as a function of the number of hits in the STT per data set.

The most time-consuming phase is the connect phase where hits are connected one by one as a function of their spatial and neighborhood relations. In particular, this phase takes more than 50% of the total reconstruction time for data sets with less than 300 hits. The contribution of the fitting phase is not significant for events with less than 100 hits, but rapidly increases as the number of hits per data becomes larger ( $\sim 30\%$  of the total time for  $n_{\text{hits}} > 300$ ). This is somewhat expected as data sets with a larger number of STT hits likely have a larger number of overlapping tracks as well, such that an increasing amount of time must be spent during this phase to reconstruct these complex tracks. Finally, we note that the merging, filling graph, and z-reconstruction time are the least time-consuming processing steps, mostly because they involve very simple tasks.

In [14], the authors estimate the complexity of their method to be  $O(n \log(n))$ . We find here that the total reconstruction time is fairly consistent with this function. Additionally, we tried to empirically estimate the time dependence of each processing step on the number of STT hits by fitting a power law to the different curves observed in Figure 18. We find all processing steps have a computational complexity close to  $O(n^{1.5})$ , with the exception of the filling graph whose complexity is  $O(n)$ . We must note that these values should only be considered as rough estimates. Indeed, empirically estimating the asymptotic behavior of our method for large  $n$  is complex because we are dealing with modest values of  $n$  ( $< 1000$ ).

The lower panel of Figure 18 which presents the evolution of the reconstruction rate as a function of the number of STT hits shows similar trends. Interestingly, the reconstruction rate of all processing steps remains fairly constant when the number of hits considered is larger than 200. This scaling behavior is promising for future develop-

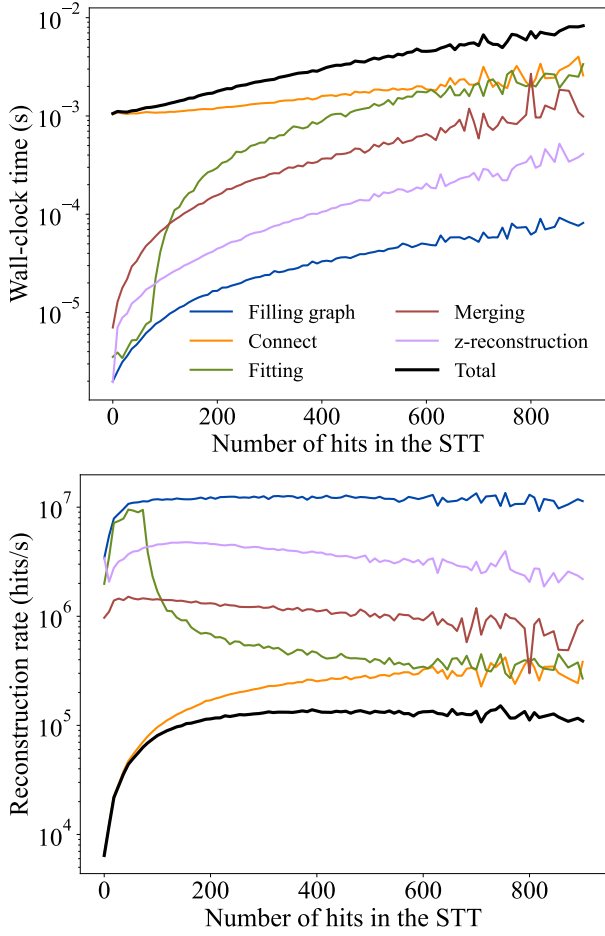


Fig. 18: The evolution of the wall-clock time (top panel) and reconstruction rate (in hits/s, bottom panel) as a function of the number of STT hits in the 33,750 events processed in this work. We decomposed the total time based on the five major processing steps performed by LOTF (see details in Section 4.3). The connect phase takes more than 50% of the total reconstruction time for data sets with less than 300 hits. The fitting phase contribution is negligible for data sets with less than 100 STT hits, but becomes more and more significant as the number of hits increases, reaching 30% of the total time for data sets with more than 300 hits. Overall, the average reconstruction rate is 67,849 hits/s ( $\sim 107$  tracks/s assuming 21 hits per track) and remains constant as the number of hits in the STT increases.

ments that will aim to limit our method's computational footprint using an appropriate parallelization scheme (see discussion hereafter).

The average processing time per data set is about 1.5 ms for the case where events are processed one by one, and 2.3 ms for the 3,750 data sets where 4 events are processed simultaneously. The average reconstruction rate is 67,849 hits/s or  $\sim 107$  tracks/s assuming an average of 21 hits per track based on our experiment. This rate is lower for data sets with a small number of hits (between

10,000 and 50,000 hits/s for  $10 < n_{\text{hits}} < 100$ ), but reaches a plateau at 124,345 hits/s for data sets with a number of STT hits larger than 200. As a comparison, running with the lowest and highest luminosities, PANDA is expected to generate around 160 Mhits/s and 1,600 Mhits/s in the STT for the high-resolution and high-luminosity modes, respectively. Hence, an improvement of more than a factor of 3 and 4 orders of magnitude is required to perform efficient online track reconstructions for both operation modes.

We are confident that future developments will enable us to get closer to the requirements set by the expected interaction rates of PANDA. Currently, LOTF does not include any parallelization scheme. The most time-consuming phase, the connect phase, involves simple and separable series of tasks that can be easily distributed over several processes. Parallelizing the fitting phase would be slightly more complex as all operations are not entirely separable, yet, decent gains should still be reachable by distributing smaller operations (*e.g.*, the selection of the best track continuation based on the fit prediction). Additionally, a GPU-based (*e.g.* as in [18, 28]) or FPGA-based implementation [17] could help to further increase the average reconstruction rate to meet the requirements for the PANDA in-situ track reconstruction. The latter approach is particularly promising as the authors report an improvement of up to 3 orders of magnitude compared to classical CPU-based approaches. Alternatively, since each event can be reconstructed independently, one could employ this method on a machine with 100 (1,000) cores and meet the requirements set by the foreseen interaction rates of the high-resolution (high-luminosity) operational mode.

## 5 Summary

Designing fast and efficient track reconstruction algorithms is crucial to meet the requirements of modern particle detectors operating at very high interaction rates. In this work, we presented the Local Track Finder (LOTF) algorithm that performs fast track reconstructions using the data collected by the Straw Tube Tracker embedded in the upcoming PANDA experiment. The functioning of our algorithm uses a local approach that connects single isolated hits to form tracks. Further, it builds upon a parametric linear fitting method to refine the tracks in regions where several particle trajectories overlap. Additionally, we utilize the virtual nodes system introduced in Babai *et al.* [13, 14] to perform the z-reconstruction. Our approach does not depend on the drift time of the tubes and only requires the STT geometry, including the neighborhood relation of all tubes (similar to [11, 12, 14]).

In Section 4.1, we generated 30,000 events (107,605 particle trajectories with at least 6 STT hits) using an antiproton beam momentum of 3 and 15 GeV/c to assess the performance of our algorithm. We compare our method to the method of Babai *et al.* [14] and to the method implemented in the PANDAROOT software (BARRELTRACKFINDER) using the standard PANDA track quality assessment metric. We showed that, in the

low and high beam momentum simulations, LOTF has (1) respectively 86% and 84% of *found* tracks (*i.e.*, tracks having a rank between 1 to 4), (2) the largest number of *Fully Pure* tracks (rank 1, perfect reconstructions), (3) the lowest amount of *Ghosts* and *Clones* (ranks 5 and 6, incorrect reconstructions), and (4), is faster by a factor of  $\sim 13$  and 940 compared to BARRELTRACKFINDER and BABAI+20, respectively. Additionally, we emphasized that our z-reconstruction approach leads to an average z-error of  $1.80^{+9.79}_{-10.38}$  cm for all tracks which is promising for extracting the particles' longitudinal momentum component.

In Section 4.2, we mimicked the effects of event-mixing to further explore the foreseen performance of our algorithm for experiments using the PANDA high-luminosity mode (20 MHz interaction rate, involving a significant overlap between events). Using the simulation with 15,000 events simulation generated with an antiproton beam momentum of 3 GeV/c, we created 3,750 data sets combining each of the STT hits from 4 successive events. We showed that the percentage of *found* tracks decreases (86% to 73%) in the setup mimicking event-mixing. Yet, LOTF still performs well in identifying a large fraction of the tracks with a decent accuracy according to the quality assessment metric used in this work, which is promising for dealing with such complex event-mixing data sets.

In Section 4.3, we scrutinized the time and processing rate of our method. We obtained an average of 1.5 ms per data set for data sets composed of a single event and 2.3 ms per data set for data sets composed of 4 successive events. The average processing rate is about 68,000 hits/s for both cases. The current efficiency and raw computational speed (no parallelization included) are promising in enabling a fast in-situ event selection with PANDA. To make this method a robust state-of-the-art algorithm for PANDA, we are planning several improvements. We aim at decreasing the current processing time per event by 2 orders of magnitudes by optimizing and parallelizing the phases having a large computational footprint (*e.g.*, the *connect* and *fitting* phases) using GPUs. Then, in practice, since each event can be reconstructed independently, another improvement of two and three orders of magnitude should be achievable by employing this method on a machine with 100 (1,000) cores, and therefore, enable an efficient in-situ event selection for experiments working with the PANDA high-resolution (high-luminosity) mode.

Finally, in the future, we will aim to include the data collected by the MVD [26] and EMC [35] detectors in the reconstruction to provide further constraints on the particle trajectories inside and outside of the STT volume. This effort will enable us to improve the efficiency of our algorithm in identifying and reconstructing overlapping particle trajectories. However, we must ensure that these improvements do not significantly impact the current speed of the reconstruction process.

## Acknowledgements

This paper is based on research developed in the DSSC Doctoral Training Programme. The authors would like to thank the referees whose comments greatly helped improve the quality of this manuscript. SG thanks Mohamad Babai for sharing his code, and for the very useful discussions throughout this work. SG also thanks the PANDA collaboration for the development of the PANDA-ROOT software. SG acknowledges the use of the python packages NUMPY [36], MATPLOTLIB [37], and PANDA [38].

## Declaration

### Funding

This work was funded by the Centre for Data Science and Systems' Complexity, University of Groningen.

### Conflicts of interest/Competing interests

Not applicable

### Availability of data and material

The simulated data sets underlying this paper are available upon request to the corresponding author.

### Code availability

The code of the LOTF algorithm is available at <https://github.com/sgazagnes/lotf>.

## References

1. B. Ketzer & the PANDA collaboration. *Int. J. Mod. Phys. A* **21** (2006).
2. Barucca, G. & the PANDA collaboration. *Eur. Phys. J. A* **57** (2021).
3. Durante, M. *et al. Phys. Scripta* **94** (2019).
4. Frankfurt, L. *et al. Eur. Phys. J. A* **56** (2020).
5. Erni, W. & the PANDA collaboration. arXiv: 0903.3905 (2009).
6. Kalman, R. E. *Journal of Basic Engineering* **82** (1960).
7. Hough, P. V. in *Proc. of the International Conference on High Energy Accelerators and Instrumentation* (1959).
8. Bianchi, L. *et al. in Journal of Physics: Conference Series* **898** (2017).
9. Aliche, A., Stockmanns, T. & Ritman, J. *EPJ Web Conf.* **251** (2021).
10. Ikegami Andersson, W. *et al. Computing and Software for Big Science* **5** (2021).

11. Regina, J. & Ikegami Andersson, W. in *Connecting the Dots and Workshop on Intelligent Trackers* (Oct. 2019). arXiv: 1910.06086 [nucl-ex].
12. Adinetz, A. et al. *Procedia Computer Science* **29** (2014).
13. Babai, M. et al. in *Mathematical Morphology and Its Applications to Signal and Image Processing* **9082** (Springer, 2015).
14. Babai, M. et al. *Pattern recognition* **106** (2020).
15. Esmail, W. et al. arXiv: 1910.07191 (2019).
16. Akram, A. & Ju, X. *arXiv e-prints*. arXiv: 2208.12178 [hep-ex] (Aug. 2022).
17. Liang, Y. et al. *IEEE Transactions on Nuclear Science* **64** (2017).
18. Herten, A. & The PANDA collaboration. in *GPU Computing in High-Energy Physics* (2015).
19. Ikegami Andersson, W. PhD thesis (Uppsala University, Department of Physics and Astronomy, 2020). ISBN: 978-91-513-0850-0.
20. Erni, W. & the PANDA collaboration. *European Physical Journal A. Hadrons and Nuclei* **49** (2013).
21. Schwarz, C. et al. *Journal of Instrumentation* **7** (2012).
22. Spataro, S. & the PANDA collaboration. *Journal of Physics: Conference Series* **331** (2011).
23. Agostinelli, S. et al. *Nuclear Instruments and Methods in Physics Research Section A* **506** (2003).
24. Allison, J. et al. *IEEE Transactions on Nuclear Science* **53** (2006).
25. Allison, J. et al. *Nuclear Instruments and Methods in Physics Research Section A* **835** (2016).
26. Erni, W. & the PANDA collaboration. arXiv: 1207.6581 (2012).
27. Smyrski, J. *Physics Procedia* **37**. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011) (2012).
28. Schumann, J. FH Aachen, Campus Jülich, Bachelorarbeit, 2013. BA (FH Aachen, Campus Jülich, 2013).
29. Regina, J. PhD thesis (Uppsala University, Nuclear Physics, 2021). ISBN: 978-91-513-1373-3.
30. Karabowicz, R. *GSI SCIENTIFIC REPORT* **8** (2010).
31. Hoppner, C. et al. *Nucl. Instrum. Meth. A* **620**. arXiv: 0911.1008 (2010).
32. Wilkinson, M. H. F. *Image Vis. Comput.* **25** (2007).
33. Papenbrock, M., Ikegami Andersson, W. & Regina, J. *EPJ Web Conf.* **245** (eds Doglioni, C. et al.) (2020).
34. Regina, J. in *Connecting the Dots Workshop Proceedings* (Aug. 2020).
35. Erni, W. & the PANDA collaboration. arXiv: 0810.1216 (2008).
36. Harris, C. R. et al. *Nature* **585** (2020).
37. Hunter, J. D. *Computing in Science Engineering* **9** (2007).
38. McKinney, W. in *Proceedings of the 9th Python in Science Conference* (eds Stéfan van der Walt & Jarrod Millman) (2010).

## A Algorithms

In this section, we present the code of the algorithms used in the present work.

### A.1 The *connect* phase

---

#### Algorithm 1 Pseudo-code of the *connect* phase

---

```

1: procedure FINDTRACKLETS(STTgraph)
2:   hit_queue ← FINDSTARTINGHITS(STTgraph)
3:   for all hit in hit_queue do
4:     CurHit ← hit
5:     PathCandidate Cand ← INIT(CurHit)
6:     CurDir ← 0
7:     Cond ← true
8:     neighbors ← FINDNEIGHBORS(CurHit)
9:     while Cond do
10:      if neighbors.SIZE() = 1 then
11:        Cand.INSERT(neighbors[0])
12:        CurDir ← FINDDIR(CurHit, neighbors[0])
13:        CurHit ← neighbors[0]
14:        neighbors ← FINDNEIGHBORS(CurHit)
15:      else if size(neighbors) > 1 then
16:        SORTPERLAYER(neighbors, upL, downL, sameL)
17:
18:        if (upL.SIZE() > 0 and downL.SIZE() > 0
19:          and (sameL.SIZE() > 0) or CurDir = SAME)
20:          then
21:            Cond ← false
22:          else if (size(upL) > 0 and downL.SIZE() > 0)
23:            then
24:              if CurDir = UP then
25:                neighbors.REMOVE(downL)
26:              else if CurDir = DOWN then
27:                neighbors.REMOVE(upL)
28:              end if
29:              if Cond = true and ARECONNECTED(neighbors) then
30:                Cand.INSERT(neighbors)
31:                CurDir ← FINDDIR(CurHit, neighbors)
32:                neighbors ← FINDNEIGHBORS(neighbors)
33:                CurHit ← neighbors[-1]
34:              end if
35:            end if
36:          else
37:            Cond ← false
38:          end if
39:        end while
40:        Cand.neighbors.INSERT(neighbors)
41:        if ONLAYERLIMIT(Cand.headNode) and ONLAYERLIMIT(Cand.tailNode) and neighbors.SIZE() = 0
42:          then
43:            Cand.status ← finished
44:          else
45:            Cand.status ← ongoing
46:          end if
47:      end for
48: end procedure

```

---

In Section 3.1.1, we briefly presented the *connect phase*. In this appendix, we present and detail the corresponding pseudocode in Algorithm 1. The function `FINDSTARTINGHITS` looks for the edges of the track by extracting all nodes that belong to the inner- or outermost layer of the STT (the “limit layers”). Every time we find such a node, we define a new track candidate using the structure *PathCandidate* and set the current node as the tail of the track. Neighbors of this node are extracted and stored in the *neighbors* vector. We use a variable *CurDir* to determine the current orientation of the track, based on the layer of the nodes that have been added so far (layer 0 being the closest to the beam-target interaction point). The *CurDir* variable is either *UP*, *DOWN*, or *SAME* depending on whether the layer index difference of the two most recently added nodes is 1, -1, or 0, respectively.

We connect neighboring activated tubes iteratively to a track using a loop that successively looks for the next available neighbors around the most recently added node. If we have a single activated tube, it is automatically connected to the track. Otherwise, we use the procedure `SORTPERLAYER` to sort all the available neighboring nodes based on their layer index. We then add only neighbors that are consistent with the track direction and are adjacent to each other (function `ARECONNECTED`). If the layer configuration of all neighbors is too complex (*i.e.* we have neighbors on all different layers and the track direction is unclear), we pause the track reconstruction and store the current list of neighbors. Then, the track reconstruction is resumed during the *fitting phase* to resolve these complex cases.

The track reconstruction also stops when we have no more neighboring nodes in sight. For the latter case, we test whether the current track looks complete by testing if its tail node (the first node added) and head node (the last node added) are on layer limits. This criterion assumes that a track transiting through the entire STT volume or circling in it is finished. If a track has no more neighboring nodes but one of its extremities ends in the middle of the STT, we flag it as *ongoing* for later investigation. Indeed, there might exist another track candidate, not directly neighboring, that it can be connected to.

## A.2 The *fitting phase*

Algorithm 2 details the steps described in Section 3.1.2. For all the tracks flagged as *ongoing*, we use a local fitting approach based on a system of *anchor* nodes to look for the next best node to add to the track. The function `EXTRACTLAST3ANC` recovers the last three anchors in the track that are used to fit the parametric equation system for the x and y coordinates (function `FITXYLINE`). The variables *xPred* and *yPred* are the predicted coordinates of the next nodes according to the parametric equations. For all the neighboring nodes, we compute the distance (labeled *hitDist*) between the node position and the predicted coordinates of the next hit using the function `COMPUTEDIST`. Additionally, we use the anchors to determine the track direction (*CurDir*) which is derived based on

---

### Algorithm 2 Pseudo-code of the *fitting phase*

---

```

1: procedure FITNEXTHIT(STTgraph, PathCandidate
   tracklets)
2:   for all track in tracklets with track.status = ongoing
   do
3:     neighbors ← track.neighbors
4:     DistDir ← New 2-D tuple array
5:     Cond ← true
6:     while Cond do
7:       for all hit in neighbors do
8:         Anchors ← CurTrack.EXTRACTLAST3ANC()
9:         xInterp, yInterp ← FITXYLINE(Anchors)
10:        CurDir ← FINDDIR(Anchors)
11:        xPred, yPred ← HITPOS(xInterp, yInterp)
12:        hitDist ← COMPUTEDIST(hit, xPred, yPred)
13:        hitDir ← COMPAREDIR(CurDir, hit)
14:        DistDir.INSERT(hitDist, hitDir)
15:       end for
16:       bestHit ← FINDBESTHIT(DistDir)
17:       if bestHit = -1 then
18:         Cond ← false
19:       else
20:         hitTrack ← FINDTRACKOF(bestHit)
21:         if hitTrack ≠ -1 and
           CHECKMERGING(track, hitTrack) = 1 then
22:           ADDTRACKFORMERGING(track, hitTrack)
23:           Cond ← False
24:         else
25:           track.INSERT(bestHit)
26:           neighbors ← FINDNEXTNEIGHBORS(bestHit)
27:         end if
28:       end if
29:       if neighbors.SIZE() = 0 then
30:         Cond ← false
31:       end if
32:     end while
33:     if LAYERLIMIT(track.headNode) and LAYER-
       LIMIT(track.tailNode) then
34:       track.status ← finished
35:     end if
36:   end for
37: end procedure

```

---

the layers the anchors belong to (similar to the method used during the *connect phase*). Once we tested all the hits, we look for the best possible fit in the list by taking the node with the minimal distance while having a consistent direction with respect to the *CurDir* variable. This is done using the function `FINDBESTHIT`. As detailed in Section 3.1.2, we set a criterion on the maximum distance acceptable such that the function returns -1 if no neighboring nodes are sufficiently close to the predicted node position.

When a good node is found, if it already belongs to another track, we check whether these tracks should be merged using the function `CHECKMERGING` which computes the intersection angle between two tracks. If the `CHECKMERGING` test is successful (see Section 3.1.2), the fitting phase for this particular track is stopped, and both



tracks are flagged as *ToMerge* such that no more hits are added.

On the other hand, if CHECKMERGING returns false, or if the best node found does not belong to any track, we connect this node to the current track. We repeat the steps above updating the list of neighbors to include the neighbors of the most recently added tube. The algorithm continues until the current track is flagged for merging, or until the list of next neighbors is empty.

### A.3 z-reconstruction

In this section, we detail the pseudo-code for the procedures used for the z-reconstruction in Algorithm 3. The CORRECTSKEWEDXY procedure is the function used to estimate the exact hit position along the tube (see Figure 10). The procedure INTERPOLATEZCOORD is the function used at the end of the track reconstruction to fit a parametric line in the z-direction and re-estimate consistently the z coordinates of all the nodes in the track. The functioning of the two procedures is detailed in Section 3.2.

---

#### Algorithm 3 Pseudo-code of the z-reconstruction step

---

```

1: procedure CORRECTSKEWEDXY(STTgraph, PathCandidate track, GridNode hitToInsert)
2:   if hitToInsert is a virtual node then
3:     GridNode lastVirt  $\leftarrow$  EXTRACTVIRT(track)
4:     xDir  $\leftarrow$  hitToInsert.x - lastVirt.x
5:     yDir  $\leftarrow$  hitToInsert.y - lastVirt.y
6:     for all anchors in track added after lastVirt do
7:       anchor.x, anchor.y  $\leftarrow$  INTERSECT(xDir, yDir)
8:       anchor.z  $\leftarrow$  ESTIMATEZCOORD(anchor)
9:     end for
10:  end if
11: end procedure

12: procedure INTERPOLATEZCOORD(PathCandidate track)
13:  zArray  $\leftarrow$  EXTRACTALLANCHORSWITHZCOORD(track)

14:  if track.Dir = InnerToOuter then
15:    zArray.insertAtFirstPos(0)
16:  else if track.Dir = OuterToInner then
17:    zArray.insertAtLastPos(0)
18:  end if
19:  zInterp  $\leftarrow$  FITZLINE(zArray)
20:  for all nodes in track do
21:    nodes.z  $\leftarrow$  COMPUTEZFROMINTERP(zInterp)
22:  end for
23: end procedure

```

---