

# Refined drift chamber simulation in the CEPC experiment

Wenxing Fang<sup>1,\*</sup>, Xingtao Huang<sup>2</sup>, Weidong Li<sup>1</sup>, Tao Lin<sup>1</sup>, Mengyao Liu<sup>2</sup>, Ye Yuan<sup>1</sup>, Xueyao Zhang<sup>2</sup>, and Yao Zhang<sup>1</sup>

<sup>1</sup>Institute of High Energy Physics, Beijing 100049, People's Republic of China

<sup>2</sup>Institute of Frontier and Interdisciplinary Science, Shandong University, Qingdao, Shandong, China

## Abstract.

The Circular Electron Positron Collider (CEPC) is a future experiment aimed at studying the properties of the Higgs boson with high precision. This requires excellent track reconstruction and particle identification (PID) performance, which is achieved in the 4<sup>th</sup> conceptual detector design of the CEPC experiments by combining a silicon tracker and a drift chamber. The drift chamber not only improves track reconstruction but also provides excellent PID with the cluster counting method. To evaluate the performance of this design accurately, a detailed simulation is necessary. In this paper, we present a refined drift chamber simulation by combining Geant4 and Garfield++. However, traditional waveform simulation using Garfield++ is extremely time-consuming, which motivates us to develop a fast waveform simulation method using a neural network. We validate the method using real data from the BESIII experiment. The results demonstrate the effectiveness of our approach and provide valuable insights for future experiments.

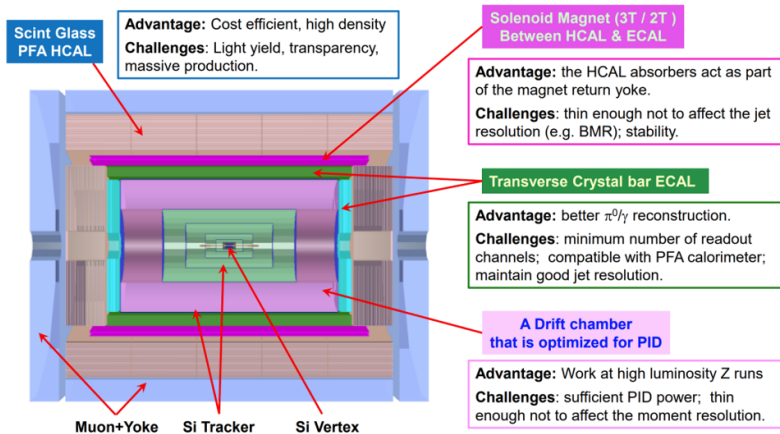
## 1 Introduction

The discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments at the LHC confirmed the predictions of the Standard Model (SM) and demonstrated its tremendous success. However, some unresolved issues in the SM, such as the existence of dark matter and energy and the asymmetry between matter and antimatter, suggest that there may be physics beyond the SM (BSM). The Higgs boson is a critical piece of the puzzle in discovering new physics, and several future collider experiments, such as FCC, CEPC, ILC, and CLIC, aim to measure its properties precisely. The CEPC is a proposed 100 km circular electron-positron collider developed by Chinese physicists [1]. The central mass of the energy is expected to be around 240 GeV for the  $e^+e^- \rightarrow ZH$  process, and the CEPC also aims to produce large amounts of W and Z bosons for precise electroweak physics studies.

The precise measurement of the Higgs boson imposes stringent requirements on detector performance, including a tracking efficiency close to 100%, momentum resolution less than 0.1%, and 2 sigma separation power for kaons and pions with momentum below 20 GeV. To meet these requirements, the 4<sup>th</sup> conceptual detector design of the CEPC combines a silicon tracker and a drift chamber. The Figure 1 shows the 4<sup>th</sup> conceptual detector along with descriptions of each sub-detector. The drift chamber is a vital component of this design,

---

\*e-mail: fangwx@ihep.ac.cn



**Figure 1.** The design of the CEPC 4<sup>th</sup> conceptual detector.

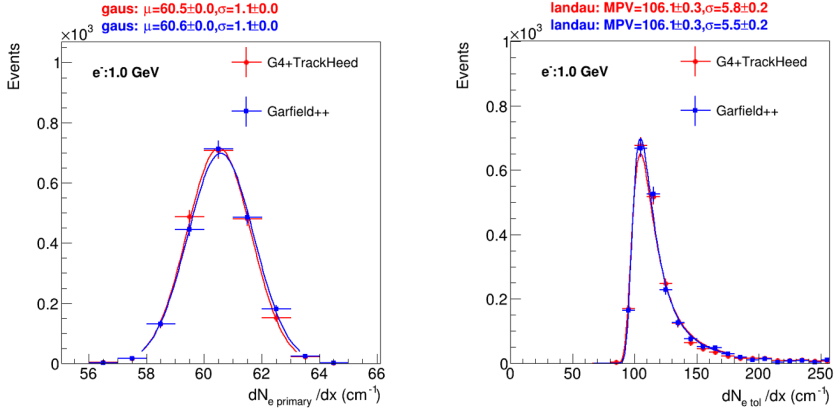
as it improves track reconstruction performance and enables excellent particle identification (PID) using the cluster counting method[2]. To evaluate the performance of tracking and PID, a detailed simulation of the drift chamber is essential. In the following section, we will describe a refined drift chamber simulation.

This paper is structured as follows: Section 2 will describe the integration of Garfield++ with Geant4. Section 3 shows the fast waveform simulation using neural networks. A summary is given in Section 4.

## 2 Integration of Garfield++ with Geant4

While Geant4 [3] is widely used for precise detector simulation, it has limitations in simulating the ionization process that occurs when a charged particle passes through a thin gas. Additionally, Geant4 cannot accurately simulate the response inside a drift chamber cell, including the drift of ionized electrons and ions, the avalanche process when ionized electrons are close to the signal wire, and the production of the induced waveform of the signal wire. Fortunately, Garfield++ [4] is commonly used for studying detector responses with simple geometries, and it can simulate the primary and secondary ionizations, the drift of ionized electrons and ions, and the induced waveform. However, Garfield++ does not simulate the multi-scattering process, making it unsuitable for larger detector simulations. To achieve a more accurate simulation of a drift chamber's response to charged particles, integrating Geant4 with Garfield++ is a recommended approach. By combining the capabilities of both tools, a more comprehensive simulation can be achieved.

Several methods have been proposed for integrating Geant4 and Garfield++ [5]. In this work, we propose a method that integrates Geant4 and Garfield++ at the G4Step level. The fundamental concept behind this method is to obtain information on the particle type, momentum, start position, and step length for each step during a Geant4 simulation. This data is then used in Garfield++ to simulate the response of the drift chamber cell. The energy deposition simulated by Garfield++ is then used to update the energy of the charged particle and initiate the next step of both Geant4 and Garfield++ simulations. To implement this method, we have created a Gaudi [6] tool within CEPCSW [7], which is the official offline software for the CEPC experiment. To support one-step simulations with a desired step length, we have



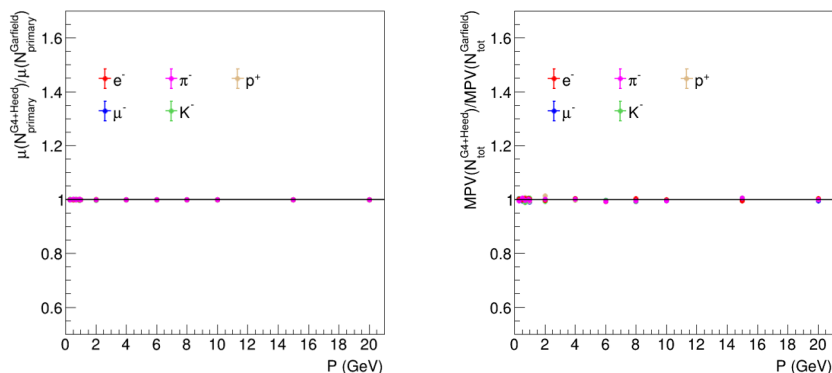
**Figure 2.** The distribution of the number of primary ionization (left) and total ionization (right) per centimeter for 1 GeV electron passing through the gas (50%He + 50% $C_4H_{10}$ ). Red histogram is for Geant4+TrackHeed simulation, blue histogram is for standalone Garfield++ simulation.

made an update to Garfield++. In this tool, the primary and secondary ionization simulations are performed using the TrackHeed class from Garfield++. Simulating waveforms involves simulating the avalanche process, which can be highly time-consuming for Garfield++. To address this, we have developed a fast waveform simulation using a neural network, which is described in Section 3. After the simulation of each step, the information about the primary and total ionization is saved into an event data model, which in our case is EDM4hep [8].

The performance of the simulation is checked. For example, Figure 2 shows the distribution of the number of primary ionization and total ionization per centimeter for 1 GeV electron passing through the gas which is 50%He + 50% $C_4H_{10}$ . From the result one can see the integration simulation is consistent with Garfield++ standalone simulation. Additionally, we evaluated the mean value of the primary ionization distribution and the most probable value of the total ionization distribution for particles with varying momenta. The results shown in Figure 3 demonstrate consistency between the two simulation approaches.

### 3 Fast waveform simulation

The waveform simulation is important as it will be used to reconstruct the number of primary ionizations (or clusters) which is the input of the cluster counting PID method. Unfortunately, the Garfield++ waveform simulation method is not suitable for the CEPC experiment due to its being highly time-consuming. To circumvent this issue, a fast waveform simulation is developed. Through the study of the simulated pulses from Garfield++, we found the pulse shape is similar for different pulses, the differences are in the start time and amplitude of the pulse which is dependent on the local position of ionized electron in the drift chamber cell. Therefore, the basic idea of the fast waveform simulation consists of two steps. The first step is to simulate the start time and amplitude of each pulse according to the local position of the ionized electron. The second step is to use a template of pulse and the simulated start time and amplitude of pulses to form the waveform (by piling up the pulses, the same approach used by Garfield++) for each drift chamber cell. The main effort of this work is focused on the first step which is non-trivial and will be described in detail in the following subsections.

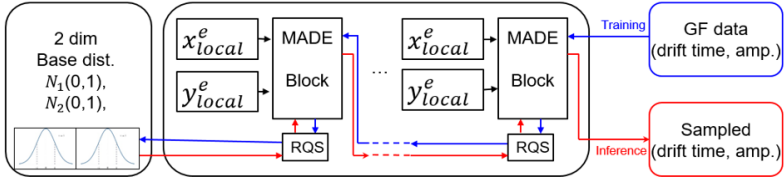


**Figure 3.** The ratio between Geant4+TrackHeed simulation and standalone Garfield++ simulation for mean value of the primary ionization distribution (left) and for the most probable value of the total ionization distribution (right).

### 3.1 The neural network model

The goal of this task is to simulate the amplitude and start time of the pulse (or the drift time of the ionized electron) according to the local  $x$  and  $y$  position of the ionized electron in a drift chamber cell. Traditionally, the histogram sampling method is used for this task, which involves dividing the local  $x$  and  $y$  positions into many fine bins and producing numerous 2D histograms of the amplitude and drift time distributions. However, this approach can be computationally expensive. In recent years, deep learning has developed rapidly, and it has become an area of advanced research for simulation or generation tasks. In high-energy physics, there have been many studies that use deep learning to simulate detector responses, such as CaloGAN [9] and CaloFlow [10], which have shown promising performance.

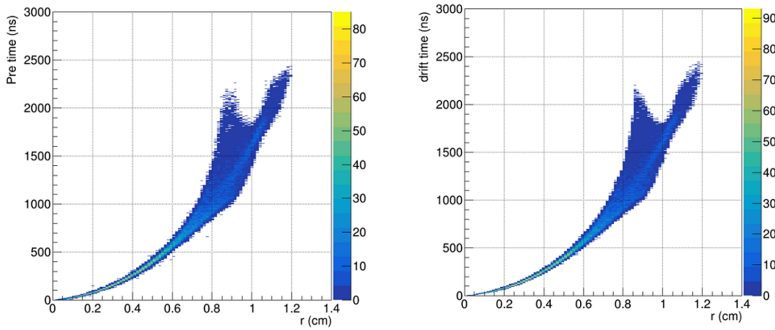
In this study, we developed a neural network model based on Normalizing Flow (NF) [11] to do the simulation. Compared to Generative Adversarial Networks (GAN) [12], NF has been shown to be more stable and convergent during training. The structure of our NF model is similar to that of CaloFlow [10], which uses rational quadratic spline (RQS) [13] for transformation and uses a Masked Autoencoder for Density Estimation (MADE) [14] block for learning the parameters of the RQS transformation. The training and inference flow of our model can be seen in Figure 4. During training, the input data from Garfield++ simulation, including the amplitude and drift time, are fed into the NF model, along with the local  $x$  and  $y$  information as conditional data for the MADE block. The MADE block learns the parameters of the RQS transformation, which is then used to transform the distributions of the amplitude and drift time. By stacking this workflow several times, the final transformed distributions of the amplitude and drift time converge to the base distributions, which are two Normal distributions in our case. During inference, one simply needs to sample from two independent Normal distributions and perform the inverse RQS transformations to obtain the simulated amplitude and drift time. The detailed structure and parameters of our model can be found in Figure 5. In our study, we have used the Garfield++ simulation data for training the NF model, and the Adam optimizer [15] was used for the optimization.



**Figure 4.** The schematic plot of the NF model, the blue lines are for training, the red lines are for inference.

Base distribution	Number of MADE blocks	Layer sizes			Number of RQS bins
		input	hidden	output	
2-dim Standard Normal	6	64	3×64	23	8

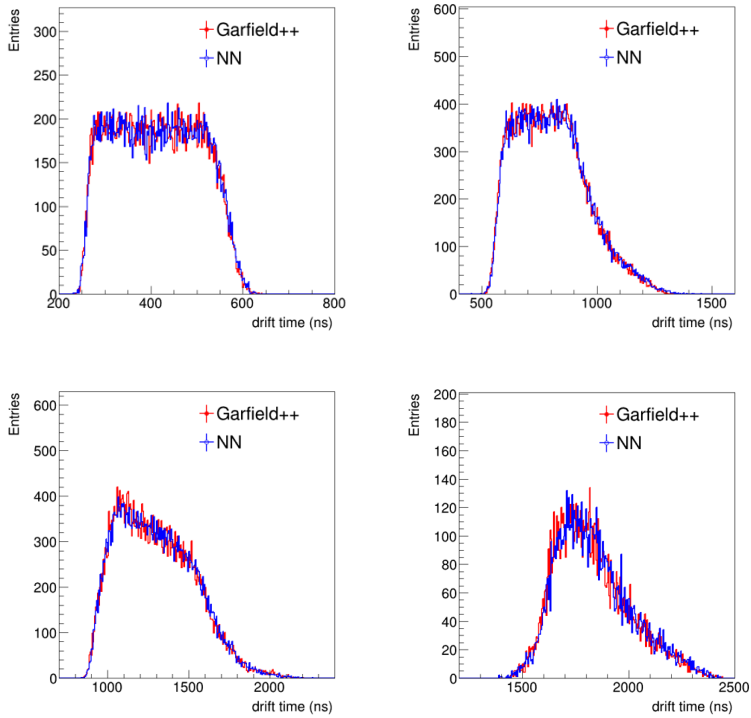
**Figure 5.** The parameters of the NF model.



**Figure 6.** The two-dimension distribution of drift time versus radius. The left plot is from the neural network simulation and the right plot is from Garfield++ simulation.

### 3.2 Results

We have validated the results of our waveform simulation using a comparison with the Garfield++ simulation. For example, we have compared the two-dimensional distributions of drift time versus radius (equal to  $\sqrt{x^2 + y^2}$ ), as shown in Figure 6. The results from the neural network simulation and Garfield++ simulation are very similar, indicating that our neural network simulation has high fidelity. The forking behavior of the drift time for large radii is caused by the inhomogeneous electromagnetic field in the cell edge region which is a known effect. In addition, Figure 7 shows the drift time distribution in different radius regions. The results from the neural network simulation are consistent with those from the Garfield++ simulation, further validating the accuracy of our waveform simulation.

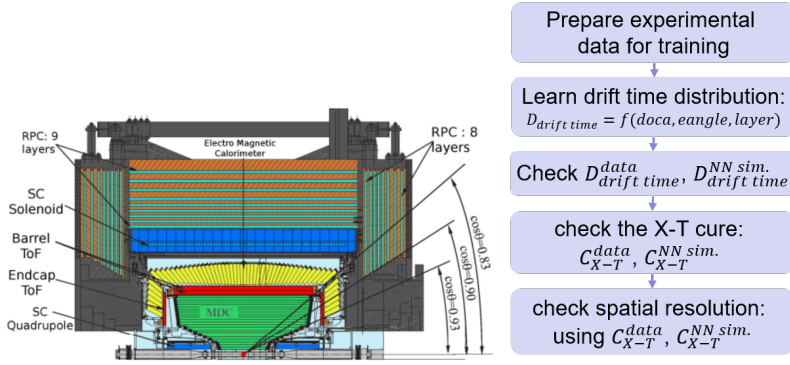


**Figure 7.** The comparison of drift time distribution between the Garfield++ and the neural network simulation for different radius regions. The top left is for a radius of 0.4-0.6 cm, the top right is for a radius of 0.6-0.8 cm, the bottom left is for a radius of 0.8-1.0 cm, and the bottom right is for a radius of 1.0-1.2 cm.

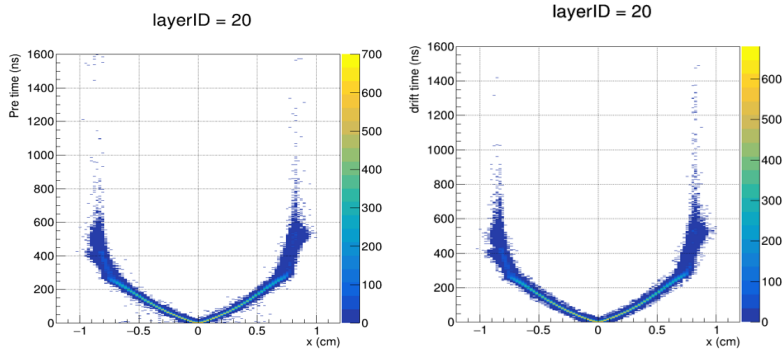
### 3.3 Method validation

To further validate our simulation method, we have performed a study using real data from the BESIII experiment [16]. The BESIII experiment also uses a drift chamber for tracking and particle identification (PID) using the  $dE/dx$  method. The detector display for BESIII can be seen in the left plot of Figure 8. In our study, we aim to simulate the drift time versus the entrance angle and distance of the closest approach (DOCA) of a track using real data. Compared to simulated data, real data is more complex and presents greater challenges in learning its distribution.

The workflow of our study can be seen in the right plot of Figure 8. First, we prepare the experimental data for training, which in our case is the radiative Bhabha event. We then construct a Normalizing Flow model with a similar structure to the one described in Section 3.1 to learn the drift time distribution according to the entrance angle, DOCA, and the layer number of the drift chamber cell. The cell geometry and high voltage of the signal wire vary across different layers, and these effects are taken into account in our simulation. An example of the drift time simulation result can be seen in Figure 9. The results from the real data and the neural network simulation are quite similar, indicating that our model can accurately learn the drift time distribution from the real data. After training, we obtain the X-T curves, which



**Figure 8.** Left plot: the display of the BES3 detector. Right plot: the workflow of this validation study.



**Figure 9.** The distributions of drift time versus radius from neural network simulation (left) and real data (right) for 20<sup>th</sup> layer of the drift chamber.

layerID	Near-wire-region ( $C_{X-T}^{data}$ v.s. $C_{X-T}^{NNSim}$ )		Far-wire-region ( $C_{X-T}^{data}$ v.s. $C_{X-T}^{NNSim}$ )	
	mean	sigma	mean	sigma
8	0.01096/0.01085	0.02137/0.02138	0.01162/0.01181	0.02674/0.02631
30	0.001055/0.001564	0.01404/0.01401	-3.1e-4/-5.1e-5	0.02060/0.02086

**Figure 10.** The compared the spatial resolution of using  $C_{X-T}^{data}$  and  $C_{X-T}^{NNSim}$  in the near wire region and far wire region.

are the mean of the drift time versus radius, using the experimental data ( $C_{X-T}^{data}$ ) and the neural network simulation data ( $C_{X-T}^{NNSim}$ ). We then check the spatial resolution of the reconstructed track separately using  $C_{X-T}^{data}$  and  $C_{X-T}^{NNSim}$ , as shown in Figure 10. The spatial resolution results are consistent when using these two X-T curves, indicating that our neural network model can accurately learn the drift time distribution from the real data and that our simulation method is feasible.

## 4 Summary

Accurately simulating the drift chamber detector is crucial for the success of the CEPC experiment. This paper presents a study on the integration of Garfield++ simulation with Geant4 simulation to achieve precise simulation of detector response in the drift chamber cells. In order to address the issue of waveform simulation being too time-consuming, a fast waveform simulation method using the neural network has been introduced which gives promising results. Furthermore, the method has been validated using real data from the BESIII experiment, demonstrating its feasibility.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No.12205322, the Jialin Xie Foundation of the Institute of High Energy Physics under Grant No.E35466U210, the Innovation Project of the Institute of High Energy Physics under Grant No.E15455U210. In addition, this work has been performed in collaboration with AIDAinnova (funded by the European Union's Horizon 2020 Research and Innovation programs under Grant Agreement No 101004761)

## References

- [1] T.C.S. Group, *CEPC Conceptual Design Report: Volume 2 - Physics & Detector* (2018), 1811.10545
- [2] F. Cuna et al., *Simulation of particle identification with the cluster counting technique* (2021), 2105.07064
- [3] *Geant4, a simulation toolkit*, <https://geant4.web.cern.ch/node/1>
- [4] *Garfield++, a toolkit for the detailed detector simulation*, <https://garfieldpp.web.cern.ch/garfieldpp/>
- [5] D. Pfeiffer et al., *Interfacing geant4, garfield and degrad for the simulation of gaseous detectors* (2019), 1806.05880
- [6] *Gaudi, a framework software*, <https://gaudi.web.cern.ch/gaudi/>
- [7] *CEPCSW, a CEPC offline software*, <https://github.com/cepc/CEPCSW>
- [8] *EDM4HEP, A generic event data model for future HEP collider experiments*, <https://github.com/key4hep/EDM4hep>
- [9] M. Paganini, L. de Oliveira, B. Nachman, *Calogan: Simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks* (2018), 1712.10321
- [10] C. Krause, D. Shih, *Caloflow: Fast and accurate generation of calorimeter showers with normalizing flows* (2023), 2106.05285
- [11] I. Kobzyev, S.J. Prince, M.A. Brubaker, *Normalizing flows: An introduction and review of current methods* (2021), 1908.09257
- [12] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Generative adversarial networks* (2014), 1406.2661
- [13] C. Durkan, A. Bekasov, I. Murray, G. Papamakarios, *Neural spline flows* (2019), 1906.04032
- [14] M. Germain, K. Gregor, I. Murray, H. Larochelle, *Made: Masked autoencoder for distribution estimation* (2015), 1502.03509
- [15] D.P. Kingma, J. Ba, *Adam: A method for stochastic optimization* (2017), 1412.6980
- [16] M. Ablikim et al., *Design and construction of the besiii detector* (2010), 0911.4960