



Data-driven simultaneous vertex and energy reconstruction for large liquid scintillator detectors

Gui-Hong Huang¹ · Wei Jiang^{2,3} · Liang-Jian Wen² · Yi-Fang Wang² · Wu-Ming Luo²

Received: 15 December 2022 / Revised: 2 March 2023 / Accepted: 2 April 2023 / Published online: 17 June 2023

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2023

Abstract

High-precision vertex and energy reconstruction are crucial for large liquid scintillator detectors such as that at the Jiangmen Underground Neutrino Observatory (JUNO), especially for the determination of neutrino mass ordering by analyzing the energy spectrum of reactor neutrinos. This paper presents a data-driven method to obtain a more realistic and accurate expected PMT response of positron events in JUNO and develops a simultaneous vertex and energy reconstruction method that combines the charge and time information of PMTs. For the JUNO detector, the impact of the vertex inaccuracy on the energy resolution is approximately 0.6%.

Keywords JUNO · Liquid scintillator detector · Neutrino experiment · Vertex reconstruction · Energy reconstruction

1 Introduction

Neutrino studies have led to significant breakthroughs in particle physics and astrophysics. Neutrino experiments Super-Kamiokande and SNO proved that neutrinos have mass [1, 2], KamLAND, SNO and Daya Bay measured three of the oscillation parameters (Δm_{21}^2 , $\sin^2\theta_{12}$, $\sin^2\theta_{13}$) to a precision in the order of a few percent [3–5]. The detection of high-energy cosmic neutrinos by the IceCube experiment has opened a new window in neutrino astronomy [6].

Next-generation detectors aimed at determining neutrino mass ordering (NMO) and CP violation phase are under construction. The Jiangmen Underground Neutrino Observatory (JUNO) will be the largest liquid scintillator (LS) detector in the world with the primary goal of determining NMO [7]. This requires the precise reconstruction of reactor neutrinos to extract NMO information from the energy spectrum.

Reactor neutrinos are detected via inverse beta decay (IBD) in LS, wherein the outgoing positron and neutron produce correlated prompt and delayed signals, respectively. The energy of the incident neutrino can be deduced from the positron energy. In LS detectors, the reconstruction of the positron vertex and energy are strongly correlated. On the one hand, due to the nonuniform detector response, the precision of the vertex affects the energy nonuniformity, which is one of the main contributing factors to the energy resolution. On the other hand, the vertex resolution is highly energy-dependent, and positrons with larger energy emit more photons, resulting in a more accurate reconstruction of the vertex. Several studies have been conducted on the vertex or energy reconstruction of positrons in JUNO, including likelihood methods [8–11] and machine learning methods [12–14]. The basic strategy of the likelihood method is to obtain the expected charge or time response of PMTs first, which strongly depends on the vertex or energy. Given the observed charge or time information of PMTs, a maximum likelihood method is utilized to reconstruct the

This work was supported by the National Key R & D Program of China (No.2018YFA0404100), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. 12175257), the National Natural Science Foundation of China (No. 12175257) and the Science Foundation of High-Level Talents of Wuyi University (No. 2021AL027).

✉ Gui-Hong Huang
huanggh@wyu.edu.cn

Wu-Ming Luo
luowm@ihep.ac.cn

- ¹ School of Applied Physics and Materials, Wuyi University, Jiangmen 529020, China
- ² Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
- ³ School of Physical Sciences, University of Chinese Academy of Science, Beijing 100049, China

positron vertex or energy. However, in the previous energy reconstruction studies applied to JUNO [8, 9], the vertex was assumed to be known, and the electronic effects of the PMTs were not considered. The PMT time probability density function (PDF) of vertex reconstruction studies in [10, 11] was vertex-independent and relied on Monte Carlo simulations. In this study, we developed a simultaneous reconstruction of the positron vertex and energy using both the charge and time information of PMTs, as well as the required PMT response extracted from the calibration data, to improve the precision of reconstruction. Major updates with respect to previous studies are listed below.

- More realistic expected charge response of PMTs with all electronic effects included
- More realistic time PDF of PMT photon hits has been constructed based on ^{68}Ge calibration data rather than positron data from Monte Carlo simulations
- The dependence on the propagation distance of time of flight or effective refractive index of photons in LS is calibrated, leading to a more accurate time of flight
- More accurate time PDF of PMT photon hits which considers the dependence on the vertex radius and photon propagation distance
- Simultaneous reconstruction of vertex and energy with both charge and time information of PMTs

The remainder of this paper is organized as follows: Sect. 2 briefly describes the JUNO detector and the Monte Carlo samples. Section 3 presents the updates of crucial inputs to the reconstruction. Section 4 describes the proposed reconstruction method and its performance. Finally, Sect. 5 presents the conclusions.

2 JUNO detector and data samples

The JUNO detector consists of a central detector (CD), top tracker detector, and water Cherenkov detector. The target matter of the CD is 20k ton liquid scintillator filled in a 35.4-m acrylic ball, monitored by approximately 12612 20-inch MCP-PMTs, 5000 20-inch Dynode-PMTs and 25600 3-inch PMTs [15]. The LS is composed of PPO, LAB and bis-MSB [16]. In addition, JUNO has a comprehensive calibration system, which consists of the Cable Loop System (CLS), the Auto Calibrate Unit (ACU), the Guide Tube Calibration System (GTCS) and the Remotely Operated under-LS Vehicles (ROV). A schematic view of the CD and calibration system is shown in Fig. 1 [9]. Further details regarding the calibration system can be found elsewhere [17]. Only the CLS and ACU were used in this study.

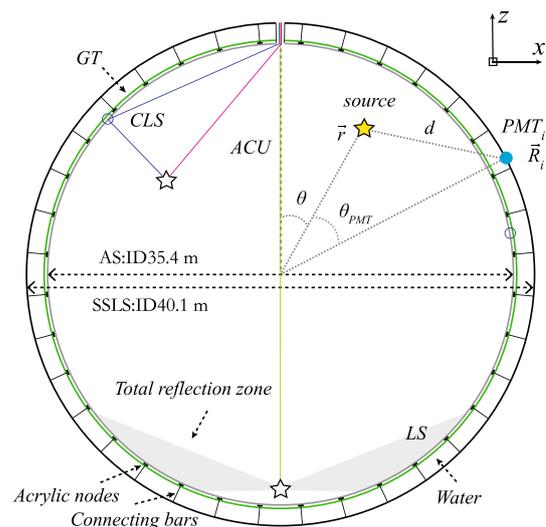


Fig. 1 Schematic view of the CD and the calibration system. The Z-axis is the vertical central axis of the CD. An example of the total reflection is also shown at the bottom. Upper-Right: definition of the three parameters of $\hat{\mu}$ in Sect. 3. \vec{r} is the calibration source position, \vec{R}_i is the i^{th} PMT position, θ_{PMT} is the angle between \vec{r} and \vec{R}_i , and d is the distance between \vec{r} and \vec{R}_i

Because the JUNO detector is still under construction,

Table 1 MC data samples. "op" stands for optical photons. "Pos." stands for position. "Energy" is deposited energy

Source	Type	Energy (MeV)	Pos	Stats
Laser	op	~ 1	296	10k/pos
^{68}Ge	γ	1.022	296	10k/pos
Positron	e^+	$(0,1,2,\dots,8)+1.022$	Uniform	450k/energy
Electron	e^-	1	296	10k/pos

Monte Carlo (MC) samples were simulated using custom Geant4-based (version 4.10.p02) offline software SNiPER [18]. The calibration and physics data samples are summarized in Table 1. Laser and ^{68}Ge calibration samples were used to construct crucial inputs for the vertex and energy reconstruction. The calibration positions (Fig. 2) were set as those in Case 5, as described in Ref. [9], and 10k events were simulated at each position. Nine sets of positron samples with discrete kinetic energies of $E_k = (0, 1, 2, \dots, 8)$ MeV were produced to evaluate the reconstruction performance. The statistics for each set were 450k, and events were uniformly distributed in the CD. Electron data were generated to elaborate the construction principle and performance of the time PDF of positrons.

For all samples, a realistic detector geometry was deployed. The optical parameters of the LS based on the measurements [16] were also implemented. Various

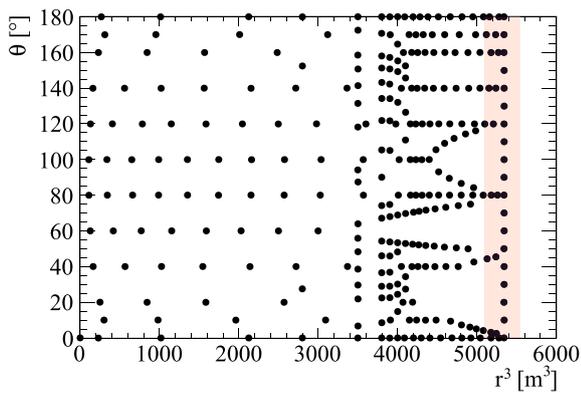


Fig. 2 Calibration positions on the X-Z plane of CD [9]. The red band corresponds to the region outside the FV with $17.2 \text{ m} < r < 17.7 \text{ m}$

Table 2 Electronic parameters of 20-inch PMTs

	Dynode-PMT	MCP-PMT
Charge resolution	$0.28 \pm 0.02 \text{ p.e}$	$0.33 \pm 0.03 \text{ p.e}$
Time transit spread	$1.1 \pm 0.1 \text{ ns}$	$7.6 \pm 0.1 \text{ ns}$
Dark noise	$15 \pm 6 \text{ kHz}$	$32 \pm 16 \text{ kHz}$

optical processes, including scintillation, Cherenkov process, absorption and re-emission, Rayleigh scattering, and reflection or refraction at detector boundaries, were simulated using Geant4 in the detector simulation. In addition, the electronic effects of PMTs, such as charge smearing, transit time spread, and dark noise (DN), were implemented using a toy electronic simulation. The PMT parameters were obtained from PMT testing [11, 19] and are summarized in Table 2. These values are still being refined, and electronics testing is ongoing [20]. Although the parameters are different PMT by PMT, they approximately follow Gaussian distributions for each type of PMT. One exception is the dark noise rate, which has a much wider and nonsymmetric spread.

3 Construction of the nPE map and time PDF of PMTs using calibration data

The basic strategy for the energy or vertex reconstruction is similar to that described in Ref. [9, 11]. For any positron event, the charge and time responses of all the PMTs strongly depend on the positron vertex and energy. First, the expected charge (referred to as the nPE map) and time PDF of PMTs were constructed using the calibration data. Given the observed charge

and time information of PMTs, a likelihood function was built and utilized to reconstruct the energy or vertex. As mentioned in introduction, with respect to Ref. [9, 11], a few important updates regarding the expected charge and time responses of PMTs were implemented in this study, and the details are described in this section.

3.1 Realistic nPE map with full electronic effects

One of the crucial components of the energy reconstruction in Ref. [9] is the nPE map denoted by $\hat{\mu}^L(r, \theta, \theta_{\text{PMT}})$. This describes the expected number of LS photoelectrons per unit of visible energy. The visible energy is defined as PE_{total}/Y_0 , where PE_{total} is the total number of PEs and Y_0 is the constant light yield defined in Ref. [17]. The definitions of $r, \theta, \theta_{\text{PMT}}$ are shown in Fig. 1. A data-driven method for constructing the nPE map was introduced in [9], in which electronic effects were omitted. In a realistic case with full electronic effects, the observable of PMTs changes from nPE to charge due to the charge smearing of every photoelectron. Dark noise also contributes to the total charge of each PMT. Consequently, the formula for $\hat{\mu}^L$ must be modified accordingly after considering these two effects. The updated formulas are given by Eq. 1,

$$\hat{\mu}^L(r, \theta, \theta_{\text{PMT}}) = \frac{1}{E_{\text{vis}}} \frac{1}{M} \sum_{i=1}^M \frac{\bar{q}_i}{\hat{Q}_i} - \mu_i^{\text{D}}, \mu_i^{\text{D}} = \text{DNR}_i \times L, \quad (1)$$

where i runs over PMTs with the same θ_{PMT} and DE_i is the relative detection efficiency, with the mean relative detection efficiency normalized to 1. Photoelectrons from the LS have a strong temporal correlation, whereas dark noise photoelectrons occur randomly in time. The length of the full electronic readout window is $L_{\text{FADC}} = 1250 \text{ ns}$. To reduce the impact of dark noise, the signal window $[t_r^{\text{A}}, t_r^{\text{B}}]$ is set according to the residual time distribution (see Eq. 4) of the positrons to exclude most dark noise. These dark-noise photoelectrons within the signal window contaminate the photoelectrons from the physical signal; thus, the expected number of dark-noise photoelectrons μ_i^{D} must be subtracted. Here, μ_i^{D} is proportional to the dark noise rate DNR_i and the signal window length $L = t_r^{\text{B}} - t_r^{\text{A}}$, which was optimized to 280 ns by scanning its value from 160 to 540 ns and selecting the one with the best effective energy resolution [7]. In contrast, the average detected nPE \bar{n}_i was estimated using $\frac{\bar{q}_i}{\hat{Q}_i}$, where \bar{q}_i is the average recorded charge inside the signal window and \hat{Q}_i is the expected average charge of one photoelectron. Except for the two changes in Eq. 1, the construction procedure for the nPE map is the same as that in Ref. [9].

3.2 Calibration of the effective refractive index of photons in LS

In addition to charge, another important observable of PMTs is the hit time of the photons. This can be approximately expressed as Eq. 2:

$$t'_h = t_0 + t_{LS} + t_{tof} + t'_{TT} + t_d, \quad (2)$$

The different components are shown in Fig. 3 with a simple illustration, where t_0 is the event starting time, t_{LS} is the scintillation time, which is governed by the LS optical properties, t_{tof} is the time of flight of photons propagating from the event vertex to PMTs, t'_{TT} is the transit time of PMTs which roughly obeys a Gaussian distribution with (μ_{TT}, σ_{TT}) , and t_d is the delay time caused by the PMT readout electronics. Both μ_{TT} and t_d are PMT dependent, and the difference between PMTs can be calibrated and absorbed by t'_h , resulting in the calibrated hit time t_h :

$$t_h = t_0 + t_{LS} + t_{tof} + t_{TT}, \quad (3)$$

where t_{TT} is described by the new Gaussian $(\overline{\mu_{TT} + t_d}, \sigma_{TT})$, and $\overline{\mu_{TT} + t_d}$ is the average value of all PMTs.

Ref. [11] defines the residual time of PMT photon hits t_r as

$$t_r = t_h - t_{tof} - t_0 = t_{LS} + t_{TT}. \quad (4)$$

To calculate the time-of-flight of the photons in the JUNO CD, a constant effective refractive index was used in Ref. [11]. However, the propagation of photons in LS is complicated and includes various optical processes, as described in Sect. 2. The time-of-flight may not necessarily be proportional to propagation distance. Given that the calibration sources can be deployed at different positions in the CD, the calibration data can be used to calibrate the time-of-flight or, equivalently, the effective refractive index n_{eff} as a function of the propagation distance. For radioactive sources, the starting time t_0 of each event is unknown. However, the precision of t_0 can reach < 0.5 ns for a laser source [21]. Thus, a laser source was chosen to calibrate n_{eff} . Although the original optical photons of the laser source have a fixed

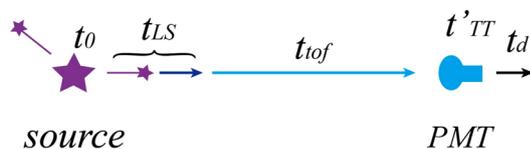


Fig. 3 Illustration of the components of photon hit time. t_0 is the event starting time, t_{LS} is the scintillation time, t_{tof} is the time of flight of photons propagating from the event vertex to PMTs, t'_{TT} is the transit time of PMTs, and t_d is the delay time caused by the PMT readout electronics

wavelength, they are rapidly absorbed and re-emitted by the LS, resulting in a wavelength spectrum similar to that of other sources [21]. By deploying the laser source at different positions along the z axis of the CD with ACU, one can plot the distribution of $t_r + t_{tof} = t_h - t_0$ for each PMT with sufficient statistics of the laser events. Because t_{LS} depends only on the LS properties, it is the same for all PMTs. In addition, t_{TT} defined in Eq. 4 follows approximately the same distribution for the same type of PMTs. Thus, the shape of the $t_h - t_0$ distribution should be the same for the same type of PMTs and the t_{tof} term merely leads to a relative shift with respect to t_r , which can be measured by the peak of the $t_h - t_0$ distribution

$$t_{\text{peak}} = t_{\text{tof}} + t_{\text{peak}}^0, \quad (5)$$

where t_{peak}^0 is the peak of the t_r distribution and is a constant with different values for each type of PMTs. We define the distance between PMT and the source position as d (Fig. 1). For PMTs of the same type and d , their t_{peak} should be aligned because t_{tof} is the same in the first-order approximation. Thus, their t_{peak} values were averaged to reduce small second-order fluctuations. Given that Dynode PMTs have much better time resolution than MCP PMTs, only Dynode PMTs are used to obtain a more precise t_h . By deploying the laser source at different positions, a large set of (t_{peak}, d) data points can be obtained, which can then be used to fit t_{tof} as a function of d . For convenience, the effective refractive index $n_{\text{eff}}(d)$ is used instead of $t_{\text{tof}}(d)$:

$$n_{\text{eff}}(d) = \frac{c \times (t_{\text{peak}}(d) - t_{\text{peak}}^0)}{d}. \quad (6)$$

t_{peak}^0 can be approximately defined as the time interval of the sharp rising edge in the $t_h - t_0$ distribution. It was estimated to be approximately 3.3 ns by extrapolating this time interval for all $t_h - t_0$ distributions with different d values. The fitting function of the effective refractive index is given by

$$n_{\text{eff}} = n_{\text{eff}}^{\text{LS}} + \frac{A}{d}, \quad (7)$$

where $n_{\text{eff}}^{\text{LS}}$ is the dominant term of the effective refractive index of liquid scintillator, and A is the coefficient of the additional term which is distance dependent. Figure 4 presents the fitting results. The best-fit value $n_{\text{eff}}^{\text{LS}} \approx 1.54$ was consistent with that in Ref. [11].

3.3 Construction of more accurate and realistic time PDF

One caveat of the residual time PDF from Ref. [11] is that it was obtained from MC simulations. Any potential discrepancy between the MC and real data will degrade

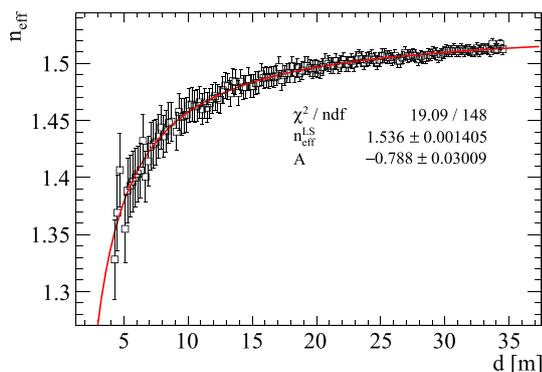


Fig. 4 (Color online) The fitting results of effective refractive index. Based on the timing error of the laser system (Fig.11 in [21]), the uncertainty of t_{peak} (Δt_{peak}) is set as 0.5 ns. The error of n_{eff} is then calculated by $c\Delta t_{\text{peak}}/d$, with negligible statistical uncertainty. For small values of d , the source is located near the north or south pole and the corresponding PMTs fall in the dark zone due to total reflection

the vertex reconstruction. Furthermore, the PDF was constructed using only the events at the detector center. This simplification did not consider the PDF dependence on the vertex and was later found to cause a large vertex bias near the detector border in Ref. [11]. Based on the calibration-data-driven construction of the nPE map in Ref. [9], the same calibration data can be used to build a realistic time PDF. Moreover, various calibration positions allow for a more precise parametrization of the time PDF.

To construct the nPE maps in Ref. [9], ^{68}Ge and laser sources were used. As previously mentioned, the optical photons of the laser source are absorbed and re-emitted by the LS. However, the other particles must first transfer energy to the LS molecules. Consequently, the photon timing profile of the laser source is different from that of the positrons. In this study, we only use ^{68}Ge to construct the time PDF of the positrons. Ideally, an electron source can be used to describe the time PDF of photons originating from the kinetic energy of positrons. Although electron sources were not available, we compared the reconstruction performance using the time PDF of ^{68}Ge to that of ^{68}Ge and electrons to check its impact. The details are presented in Sect. 4.

As shown in Eq. 4, to calculate the residual time, the information of t_0 is required for every single event. In contrast to a laser source with high precision of t_0 , this quantity of calibration events from radioactive sources is unknown. Reconstruction of t_0 for each event was attempted; however, the uncertainty was greater than 2 ns. Because the $t_h - t_{\text{tof}}$ distribution for different events with a fixed vertex and energy should have approximately the same shape, a different t_0 would merely cause a relative shift in the distribution. The peak of the $t_h - t_{\text{tof}}$ distribution t'_0 can be

used as the new reference time instead of t_0 to align the different events. Equation 4 can be modified as follows such that the distribution of the newly defined residual time t'_r always peaks at 0. For convenience, the prime symbols t'_r and t'_0 are omitted in the remainder of this paper.

$$t'_r = t_h - t_{\text{tof}} - t'_0 = t_{\text{LS}} + t_{\text{TR}} + \text{constant}. \tag{8}$$

The PDF $P_T(t_r|r, d, \mu_l, \mu_d, k)$ describes the probability of the residual time of the first photon hit falling in $[t_r, t_r + \delta t]$, where r is the radius of the event vertex, $d = |\vec{r} - \vec{r}_{\text{PMT}}|$ is the propagation distance from the previous section, μ_l and μ_d are the expected numbers of LS and dark noise photoelectrons inside the full electronic readout window, respectively, and k is the detected number of photoelectrons. The dependence of P_T on these parameters is discussed in the following subsections. For convenience, let us denote $f(t)$ as the probability density of “photoelectron hit at time t ” and $F(t)$ as the probability of “photoelectron hit after time t ” for a general case. Then, $F(t)$ can be calculated as:

$$F(t) = \int_t^{L_{\text{FADC}}} f(t')dt'. \tag{9}$$

3.3.1 Dependence on r and d

Optical processes such as absorption and re-emission or Rayleigh scattering are not negligible in an LS volume as large as JUNO and become more prominent as the photon propagation distance d increases. Meanwhile, the total reflection can significantly change the direction of the photons, which is more likely to occur for events with a larger radius r . Thus, $f(t)$ of the LS photoelectron $f_l(t)$ depends on d and r . This dependency can be addressed by deploying calibration sources at different positions with the ACU + CLS system. This data-driven construction of the time PDF does not require a comprehensive understanding of the properties of the liquid scintillator such as attenuation length and decay time.

One of the challenges in constructing a time PDF from the calibration data is that the number of calibration positions is limited. To address this challenge, 35 and 200 bins were set in the r - and d -directions, respectively. Examples of the construction results of the t_r PDF of one LS photoelectron are shown in Fig. 5. The left and right plots correspond to vertices in the central and edge regions, respectively. One can clearly observe the difference in the time PDF for different radii r . Moreover, the time PDF becomes wider as d increases, except in the total reflection region. Most of the detected photons in the total reflection region are scattered light, whose time PDF is relatively flat. Notably, the

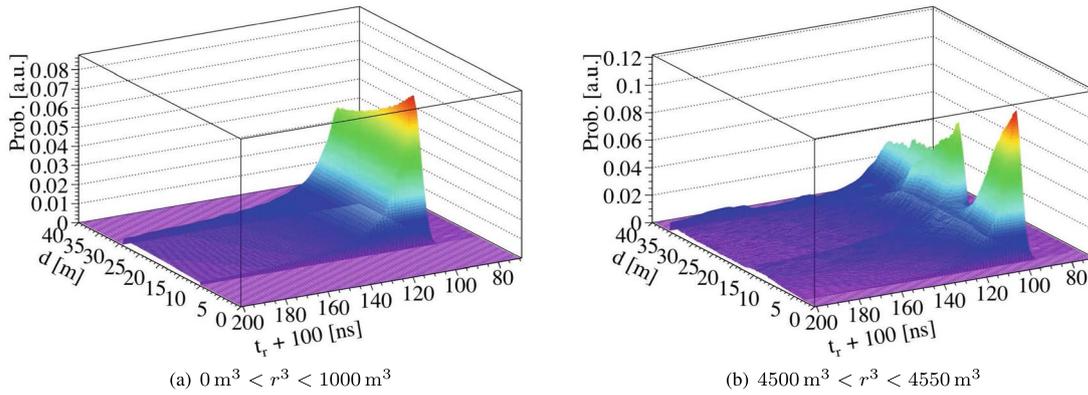


Fig. 5 (Color online) Examples of the time PDFs at different positions. (a) Central region: the time PDF becomes wider as d increases. (b) Edge region: the time PDF also becomes wider as d increases,

except in the total reflection region with $12\text{ m} < d < 16\text{ m}$. Most detected photons in the total reflection region are scattered light, whose time PDF is relatively flat

contribution from dark noise was subtracted and will be added independently in the next subsection.

Once the time PDF of one LS photoelectron is obtained, it is straightforward to calculate the time PDF of n LS photoelectrons using Eq. 10, where I_n^1 denotes the normalization coefficient.

$$P_T^1(t|k = n, r, d) = I_n^1 \times [f_1(t|r, d)F_1^{n-1}(t|r, d)] \tag{10}$$

3.3.2 Adding dark noise contribution

Photoelectrons induced by PMT DN contaminate the photoelectrons from the signal particles in the LS. Their impact on the residual-time PDF must be carefully considered. As dark-noise photoelectrons occur randomly in time, their $f(t)$ is simply $f_d(t) = 1/L_{FADC}$. The probability of a DN photoelectron falling into $[t, L_{FADC}]$ is given by Eq. 11:

$$F_d(t) = 1 - \frac{t}{L_{FADC}}. \tag{11}$$

For any given PMT with the expected nPE from the LS μ_1 , expected nPE from DN μ_d , and the detected nPE k , one can mathematically calculate the probability density of the first photon hit observed at time t . In the simplest case, where only one PE is detected by the PMT or $k = 1$, $P_T(t|\mu_1, \mu_d, k = 1)$ is given by

$$P_T(t|k = 1, r, d, \mu_1, \mu_d) = I_1 \times [P(1, \mu_1)P(0, \mu_d)f_1(t|r, d) + P(0, \mu_1)P(1, \mu_d)f_d(t)], \tag{12}$$

where I_1 is a normalization factor and $P(k_1, \mu_1)$ and $P(k_d, \mu_d)$ are the Poisson probabilities of detecting k_1 photoelectrons from the LS and k_d photoelectrons from DN, respectively, under the condition $k=k_1+k_d$. Given $k=1$ in this case, both k_1 and k_d can be either 1 or 0. Thus, one can easily see that the two terms correspond to the photoelectron emitted from LS or DN, respectively. The same method can be applied to the case where $k \geq 2$. The time PDF of the first hit of n photoelectrons is given by

$$P_T(t|k = n, r, d, \mu_1, \mu_d) \approx I_n \times [P(n, \mu_1)P(0, \mu_d)P_T^1(t|n, r, d) + P(n-1, \mu_1)P(1, \mu_d)(P_T^1(t|n-1, r, d)F_d(t) + f_d(t)F_1^{n-1}(t|r, d)) + P(n-2, \mu_1)P(2, \mu_d)(P_T^1(t|n-2, r, d)F_d^2(t) + 2f_d(t)F_d(t)F_1^{n-2}(t|r, d))], \tag{13}$$

where I_n is the normalization coefficient. The three terms correspond to the cases where $k_d=0, 1$, and 2 . Given that μ_d is rather small, terms with $k_d > 2$ will be highly suppressed by the Poisson probability $P(k_d, \mu_d)$ and thus can be safely omitted to simplify the calculation. To verify this analytical approach, the calculated results of the time PDF containing dark noise contributions were compared with those obtained using the true information of k in the MC simulation, as shown in Fig. 6. The good agreement verified the validity of Eqs. 12 and 13.

3.3.3 Charge versus nPE

As described in the above two subsections, the usage of the time PDF requires knowledge of the number of photoelectrons detected by each PMT, whereas the PMT charge is usually measured. Given that the charge information is used to estimate μ , the time PDF can be rewritten as Eq. 14:

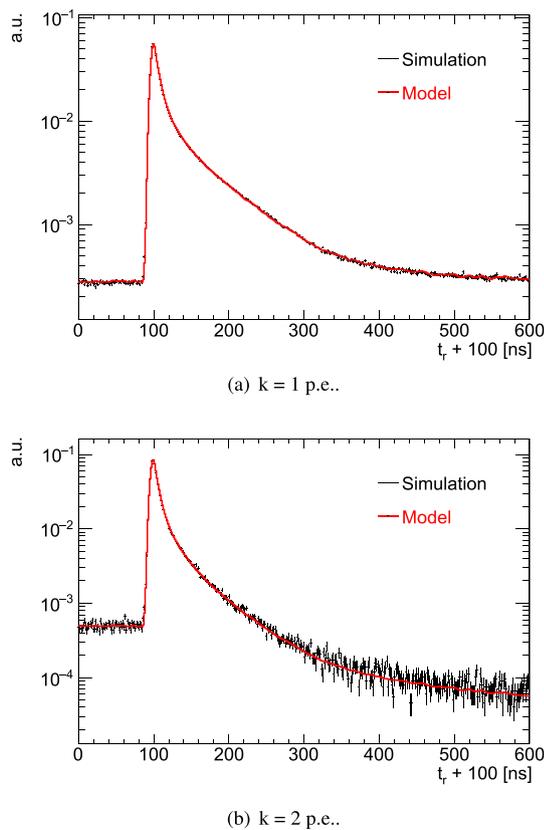


Fig. 6 (Color online) Time PDF of Dynode-PMT of central ⁶⁸Ge events considering dark noise. Red histograms are calculated by Eq. 13, while black histograms are obtained from MC simulation data

$$P'_T(t|r, d, \mu_l, \mu_d) = \frac{\sum_k P_T(t|r, d, \mu_l, \mu_d, k) \times P(k, \mu_l + \mu_d)}{\sum_k P(k, \mu_l + \mu_d)} \tag{14}$$

4 Simultaneous reconstruction of vertex and energy

In previous studies [9, 11], the reconstruction of the positron vertex and energy was decoupled for simplicity. The positron energy was reconstructed assuming that its vertex was known, and vice versa. For real data, neither the vertex nor the energy of a positron is known; both must be reconstructed. Moreover, these two variables are highly correlated. One main correlation is that the energy response for monoenergetic positrons varies at different vertices, which is also referred to as the detector energy nonuniformity. The other is that the vertex resolution depends on the energy. The higher the positron energy, the smaller the vertex resolution. This section presents the simultaneous reconstruction of the positron vertex and energy for large liquid scintillator detectors. The strong

correlation between the vertex and the energy is handled naturally. Moreover, the crucial inputs of the simultaneous reconstruction, namely the nPE map and time PDF of PMTs, could be obtained from the calibration data and did not depend on the MC simulation. In addition, with all the updates from Sect. 3, the nPE map and time PDF of PMTs were more realistic and accurate.

The reconstruction performance was evaluated in terms of radial bias, radial resolution, energy uniformity and energy resolution. The radial bias and resolution were defined as the mean and sigma of the Gaussian fit of the $r_{rec} - r_{edep}$ distribution, respectively. Energy uniformity represents the consistency of the reconstructed energy of identical particles generated at different positions, which is assessed by the deviation of the average reconstructed energies of monoenergetic positrons within different small volumes ($\sim 10\text{ m}^3$) of the detector [9]. The reconstructed energies of the monoenergetic positrons were fitted with a Gaussian function ($\bar{E}_{rec}, \sigma_{E_{rec}}$). The energy resolution is defined as $\sigma_{E_{rec}}/\bar{E}_{rec}$. The default fiducial volume condition was $r_{rec} < 17.2\text{ m}$.

4.1 Charge-based maximum likelihood estimation

Ref. [9] presented the basic strategy for energy reconstruction; a likelihood function was constructed using the expected nPE $\mu = \hat{\mu}_L \times E_{vis}$ and observed charge for each PMT. By maximizing the likelihood function, one could obtain the reconstructed energy. However, the expected nPE for each PMT depends strongly on the positron vertex, as indicated by Eq. 1. In Ref. [9], the vertex was assumed to be known; however, in real data, it must also be reconstructed. Thus, the vertex and energy can be simultaneously reconstructed using a likelihood function similar to that in Ref. [9].

This likelihood function utilizes only the charge information of PMTs and is referred to as the charge-based maximum likelihood estimation (QMLE). It is constructed using Eq. 15, which is the product of the probabilities of observing a charge q_i when the expected nPE is μ_i for the i -th PMT.

$$\mathcal{L}(q_1, q_2, \dots, q_N | \mathbf{r}, E_{vis}) = \prod_{\text{unfired}} e^{-\mu_j} \prod_{\text{fired}} \left(\sum_{k=1}^{+\infty} P_Q(q_i|k) \times P(k, \mu_i) \right), \tag{15}$$

where $\mu_i = E_{vis} \times \hat{\mu}_i^L + \mu_i^D$ and $P(k, \mu_i)$ is the Poisson probability for detecting k photoelectrons. $P_Q(q_i|k)$ is the charge PDF of k photoelectrons that can be constructed by convolving the single photoelectron charge spectrum (SPES). Indices j and i run over all "unfired" and "fired" PMTs, respectively, with a PMT firing threshold of $q > 0.1$ p.e. The index k ends when $P_Q(q_i|k) < 1.0 \times 10^{-8}$. The constraint power dramatically decreases for positrons in the central region of

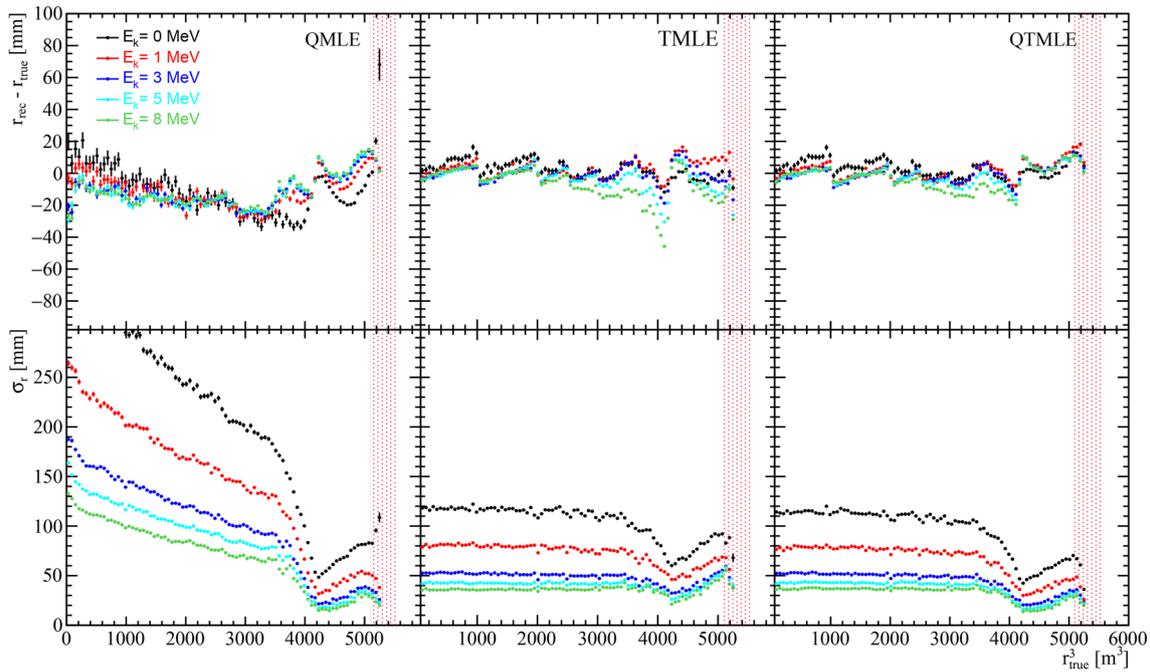


Fig. 7 (Color online) Vertex reconstruction performances. The left, middle, and right columns correspond to the QMLE, TMLE, and QTMLE methods, respectively. The top row shows the vertex bias,

and the bottom row shows the vertex resolution. The red band corresponds to the region outside the FV with $17.2 \text{ m} < r < 17.7 \text{ m}$

the CD. This was demonstrated in Ref. [11, 13] and can also be seen in the bottom-left plot of Fig. 7, where the vertex resolution in the central region is much worse than that in the border region. Moreover, the vertex bias of QMLE is large, as shown in the top-left plot of Fig. 7. An inaccurate vertex degrades the energy resolution for the simultaneous reconstruction.

4.2 Time-based maximum likelihood estimation

The event vertex is strongly constrained by the time information of the PMTs. Similar to Ref. [11], a likelihood function can be constructed using the first hit time of PMTs and a more accurate and realistic time PDF P_T from Eq. 14. This likelihood function uses only the PMT time information and is referred to as time-based maximum likelihood estimation (TMLE). It can be constructed using Eq. 16, which is the product of the probabilities of observing the residual time of $t_{i,r}$ when the expected t_{tof} is $d/(c/n_{\text{eff}})$ for the i -th PMT.

$$\mathcal{L}(t_{1,r}, t_{2,r}, \dots, t_{N,r} | \mathbf{r}, t_0) = \prod_{T\text{-valid hit}} \frac{\sum_{k=1}^K P_T(t_{i,r} | r, d_i, \mu_i^1, \mu_i^d, k) \times P(k, \mu_i^1 + \mu_i^d)}{\sum_{k=1}^K P(k, \mu_i^1 + \mu_i^d)}, \quad (16)$$

where "T-valid" hit refers to those hits satisfying $-100 < t_{i,r} < 500 \text{ ns}$ and $0.1 \text{ p.e.} < q_i^f < K = 20 \text{ p.e.}$. The

definition range of the residual-time PDF was $(-100, 500) \text{ ns}$. q^f denotes the total charge in the full electronic readout window. A cutoff value K was set for the detected nPE k to simplify the calculations. TMLE takes the reconstructed vertex and energy from the QMLE as the initial values and only updates the reconstructed vertex. Note that reference time t_0 is also a free parameter in the reconstruction. As shown in Figs. 7 and 8, the vertex bias and resolution of the TMLE are significantly improved with respect to the QMLE, especially in the central region of the CD. This is mainly because of the stronger constraint of the PMT time information.

4.3 Charge and time combined maximum likelihood estimation

Given the likelihood functions from QMLE with charge information only and TMLE with time information only, it is straightforward to construct the charge and time combined maximum likelihood estimation (QTMLE), as Eq. 17, by multiplying Eqs. 15 and 16.

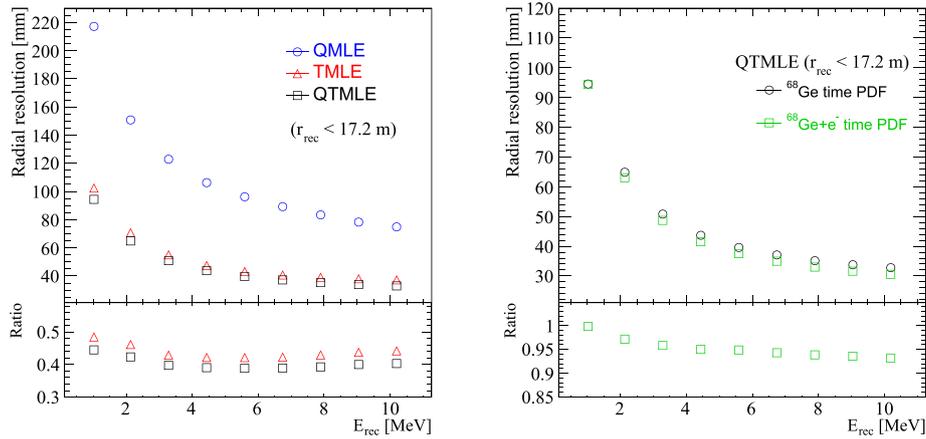


Fig. 8 (Color online) Comparison of the vertex resolution as a function of the energy. The left plot shows the comparison among the three methods QMLE (blue), TMLE (red), and QTMLE (black). The ratio is defined as $Res_R(\text{other})/Res_R(\text{QMLE})$. The QTMLE method utilizing both charge and time information of PMTs effectively

improves vertex reconstruction. The right plot shows the comparison between using $^{68}\text{Ge} + \text{electron}$ time PDF and ^{68}Ge time PDF in the QTMLE method, which will be discussed in Sect. 4.4. The ratio is defined as $Res_R(^{68}\text{Ge} + e^- \text{ time PDF})/Res_R(^{68}\text{Ge} \text{ time PDF})$

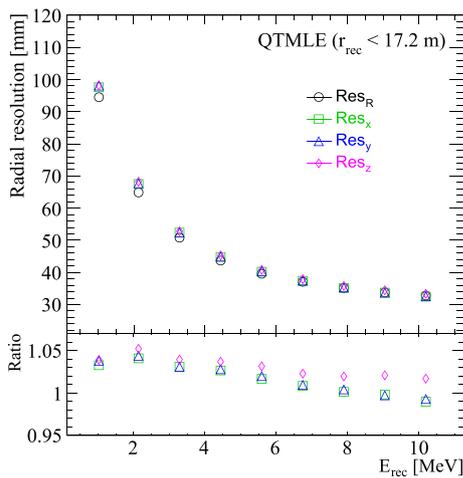


Fig. 9 (Color online) Comparison of the x , y , z , and r resolution of the QTMLE method with ^{68}Ge time PDF. The bottom panel shows the ratio to the radial resolution

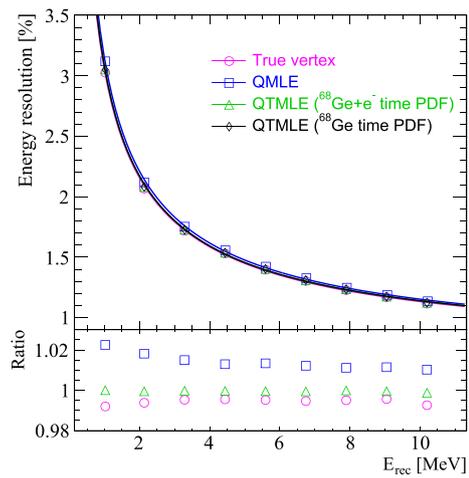


Fig. 10 (Color online) Comparison of the energy resolution with different vertex precisions. The bottom panel shows the ratio to the QTMLE method with ^{68}Ge time PDF. The blue dots correspond to the energy resolution of QMLE, which has the worst vertex resolution. The pink dots represent the energy resolution of QTMLE using true vertex, which is equivalent to an ideal vertex resolution of 0 mm. The energy resolution of QTMLE using $^{68}\text{Ge} + \text{electron}$ time PDF is also shown and will be discussed in Sect. 4.4

$$\mathcal{L}(q_1, q_2, \dots, q_N; t_{1,r}, t_{2,r}, \dots, t_{N,r} | \mathbf{r}, t_0, E_{\text{vis}}) = \prod_{\text{unfired}} e^{-\mu_j} \prod_{\text{fired}} \left(\sum_{k=1}^{+\infty} P_Q(q_i | k) \times P(k, \mu_i) \right) \prod_{T\text{-valid hit}} \left(\frac{\sum_{k=1}^K P_T(t_{i,r} | r, d_i, \mu_i^l, \mu_i^d, k) \times P(k, \mu_i^l + \mu_i^d)}{\sum_{k=1}^K P(k, \mu_i^l + \mu_i^d)} \right) \quad (17)$$

Because QTMLE uses both the charge and time information of PMTs to constrain the vertex, it yields the best vertex reconstruction performance among the three methods, which can be observed in the plot on the left side of Fig. 8. Across

the entire energy range of interest, the vertex resolutions of TMLE and QTMLE are, on average, approximately 56% and 60% better than that of QMLE, respectively. Meanwhile, a more accurate vertex also leads to a more accurate energy, given the strong correlation between them. This is shown in Fig. 10, where the energy resolution of QTMLE is significantly better than that of the QMLE. The resolutions of the x -, y -, z -, and r -components of the vertex were similar, with

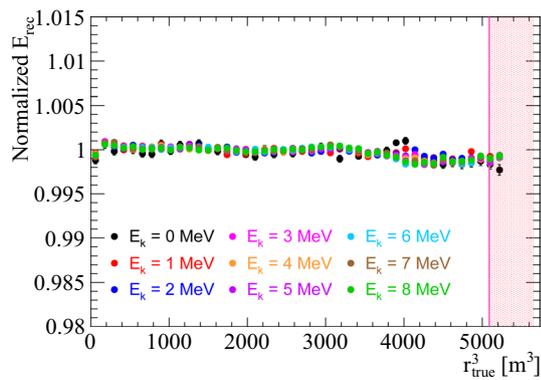


Fig. 11 (Color online) Energy uniformity of QTMLE. The red band corresponds to the region outside the FV with $17.2\text{ m} < r < 17.7\text{ m}$. The QTMLE method yields excellent energy uniformity. The residual energy nonuniformity is less than 0.23% within the FV for all the different energies

a discrepancy less than 5%, as shown in Fig. 9. Meanwhile, the energy resolution was not sensitive to the resolution of θ and ϕ components, given the approximate spherical symmetry of the CD. Thus, only the radial resolution is presented in this study.

The uniformity of the reconstructed energy using QTMLE for positrons with different energies is shown in Fig. 11. The QTMLE method yields excellent energy uniformity, and the residual energy nonuniformity is less than 0.23% inside the fiducial volume. Compared to the value of 0.17% in Ref. [9], which used a true vertex and did not include any electronic effects, one can see that the impact of vertex inaccuracy and electronic effects on energy nonuniformity is non-negligible but still under control.

4.4 Discussion

As mentioned previously, the energy deposition process of positrons in the LS usually consists of two parts, kinetic energy and annihilation with an electron emitting two gamma particles. During the construction of the expected nPE map of PMTs, laser and ^{68}Ge were used to mimic the two parts, respectively. However, because the photons from the laser contain only the fast component, they cannot be used to describe the photon timing profile of the charged particles. Although electrons can mimic the kinetic energies of positrons, monoenergetic electron sources are unavailable. Consequently, only the ^{68}Ge source was used to construct the time PDF of PMTs in all previous studies. Pseudo-electron calibration data were produced to check the impact of the accuracy of the time PDF on the vertex and energy reconstruction. The QTMLE method was used and a new set of time PDF was constructed using a weighted ^{68}Ge +

electron-time PDF. The reconstruction results are compared with those obtained using the ^{68}Ge -time PDF.

The right plot in Fig. 8 shows the comparison of the vertex resolution. The weighted ^{68}Ge + electron-time PDF was more accurate than the ^{68}Ge time PDF, and the corresponding vertex resolution was about 5% better. Figure 10 compares the energy resolutions of the two cases. Despite the better vertex resolution obtained using ^{68}Ge + electron-time PDF, the energy resolution was almost the same as that obtained using the ^{68}Ge time PDF. To verify the impact of the accuracy of the vertex on the energy resolution, two additional cases, namely, QMLE and QTMLE using the true vertex, are also shown in Fig. 10. The black dots correspond to the energy resolution of QMLE, which has the worst vertex resolution. The pink dots represent the energy resolution of QTMLE using the true vertex, which has an ideal vertex resolution of 0 mm. By comparing these cases, it is clear that better vertex resolution leads to better energy resolution. Meanwhile, comparing the default case of QTMLE using ^{68}Ge time PDF to the ideal case of QTMLE using the true vertex, the impact of vertex inaccuracy on the energy resolution is approximately 0.6%.

5 Conclusion

High-precision vertex and energy reconstruction are crucial for large liquid scintillator detectors such as JUNO, especially for the determination of neutrino mass ordering. In this study, a calibration-data-driven simultaneous vertex and energy reconstruction method was proposed. The dependence of the refractive index on the photon propagation distance was calibrated to obtain more precise PMT time information. More accurate and realistic time PDF of PMTs were constructed to consider their dependence on the vertex radius and photon propagation distance. The contribution to the time PDF from PMT dark noise was modeled using an analytical approach. With these updates, a charge and time combined likelihood function was constructed to simultaneously reconstruct the vertices and energies of the positrons. This method does not rely on MC simulations and obtains the expected PMT charge and time response directly from the calibration data. It also naturally handles the strong correlation between the vertex and energy. By combining the PMT charge and time information, the vertex resolution was improved by approximately 4% (60%) with respect to using only the time (charge) information. The vertex bias was reduced with the more accurate time PDF and was less than 2 cm. A better vertex resolution also leads to a better energy resolution. The residual energy

nonuniformity of this method was less than 0.5% within the FV. Moreover, the impact of an inaccurate vertex on energy resolution was approximately 0.6%. In the MC studies, positron samples were used to verify the method. In real data, vertex and energy resolutions can be determined using calibration sources and other natural sources such as spallation neutrons.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Gui-Hong Huang, Wu-Ming Luo, and Wei Jiang. The first draft of the manuscript was written by Wu-Ming Luo and Gui-Hong Huang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Super-Kamiokande Collaboration, Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.* **81**, 1562–1567 (1998). <https://doi.org/10.1103/PhysRevLett.81.1562>
2. SNO Collaboration, Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.* **89**, 011301 (2002). <https://doi.org/10.1103/PhysRevLett.89.011301>
3. KamLAND Collaboration, Reactor on-off antineutrino measurement with KamLAND. *Phys. Rev. D* **88**(3), 033001 (2013). <https://doi.org/10.1103/PhysRevD.88.033001>
4. SNO Collaboration, Combined analysis of all three phases of solar neutrino data from the sudbury neutrino observatory. *Phys. Rev. C* **88**, 025501 (2013). <https://doi.org/10.1103/PhysRevC.88.025501>
5. Daya Bay Collaboration, Measurement of the electron antineutrino oscillation with 1958 days of operation at Daya Bay. *Phys. Rev. Lett.* **121**, 241805 (2018). <https://doi.org/10.1103/PhysRevLett.121.241805>
6. IceCube Collaboration, Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science* **342**, 1242856 (2013). <https://doi.org/10.1126/science.1242856>
7. JUNO Collaboration, Neutrino physics with JUNO. *J. Phys. G* **43**, 030401 (2016). <https://doi.org/10.1088/0954-3899/43/3/030401>
8. W. Wu, M. He, X. Zhou et al., A new method of energy reconstruction for large spherical liquid scintillator detectors. *J. Instrum.* **14**, P03009 (2019). <https://doi.org/10.1088/1748-0221/14/03/P03009>
9. G. Huang, Y. Wang, W. Luo et al., Improving the energy uniformity for large liquid scintillator detectors. *Nucl. Instrum. Meth. A* **1001**, 165287 (2021). <https://doi.org/10.1016/j.nima.2021.165287>
10. Q. Liu, M. He, X. Ding et al., A vertex reconstruction algorithm in the central detector of JUNO. *J. Instrum.* **13**, T09005 (2018). <https://doi.org/10.1088/1748-0221/13/09/T09005>
11. Z. Li, Y. Zhang, G. Cao et al., Event vertex and time reconstruction in large-volume liquid scintillator detectors. *Nucl. Sci. Tech.* **32**, 49 (2021). <https://doi.org/10.1007/s41365-021-00885-z>
12. Z. Qian, V. Belavin, V. Bokov et al., Vertex and energy reconstruction in JUNO with machine learning methods. *Nucl. Instrum. Meth. A* **1010**, 165527 (2021). <https://doi.org/10.1016/j.nima.2021.165527>
13. Z.Y. Li, Z. Qian, J.H. He et al., Improvement of machine learning-based vertex reconstruction for large liquid scintillator detectors with multiple types of PMTs. *Nucl. Sci. Tech.* **33**, 93 (2022). <https://doi.org/10.1007/s41365-022-01078-y>
14. A. Gavrikov, Y. Malyskhin, F. Ratnikov, Energy reconstruction for large liquid scintillator detectors with machine learning techniques: aggregated features approach. *Eur. Phys. J. C* **82**, 1021 (2021). <https://doi.org/10.1140/epjc/s10052-022-11004-6>
15. JUNO Collaboration, JUNO physics and detector. *Prog. Part. Nucl. Phys.* **123**, 103927 (2022). <https://doi.org/10.1016/j.pnpnp.2021.103927>
16. JUNO and Daya Bay Collaboration, Optimization of the JUNO liquid scintillator composition using a Daya Bay antineutrino detector. *Nucl. Instrum. Meth. A* **988**, 164823 (2021). <https://doi.org/10.1016/j.nima.2020.164823>
17. JUNO Collaboration, Calibration strategy of the JUNO experiment. *JHEP* **2021**, 4 (2021). [https://doi.org/10.1007/JHEP03\(2021\)004](https://doi.org/10.1007/JHEP03(2021)004)
18. T. Lin, J. Zou, W. Li et al., The application of SNiPER to the JUNO simulation. *J. Phys. Conf. Series* **898**, 042029 (2017). [https://doi.org/10.1007/JHEP03\(2021\)004](https://doi.org/10.1007/JHEP03(2021)004)
19. Z.M. Wang, JUNO PMT system and prototyping. *J. Phys. Conf. Ser.* **888**, 012052 (2017). <https://doi.org/10.1088/1742-6596/888/1/012052>
20. JUNO Collaboration, Mass testing and characterization of 20-inch PMTs for JUNO. *arXiv:2205.08629*
21. Y. Zhang, J. Liu, M. Xiao et al., Laser calibration system in JUNO. *J. Instrum.* **14**, P01009 (2019). <https://doi.org/10.1088/1748-0221/14/01/P01009>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.