



Extraction of fissile isotope antineutrino spectra using feedforward neural network

Jian Chen¹ · Jun Wang² · Wei Wang^{2,3} · Yue-Huan Wei²

Received: 29 July 2024 / Revised: 30 October 2024 / Accepted: 6 December 2024 / Published online: 16 July 2025

© 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 2025

Abstract

The precise measurement of the antineutrino spectra produced by isotope fission in reactors is of great significance for studying neutrino oscillations, refining nuclear databases, and addressing the reactor antineutrino anomaly. In this paper, we report a method that utilizes a feedforward neural network (FNN) model to decompose the prompt energy spectrum observed in a short-baseline reactor neutrino experiment and extract the antineutrino spectra produced by the fission of major isotopes such as ^{235}U , ^{238}U , ^{239}Pu , and ^{241}Pu in the nuclear reactor. We present two training strategies for the model and compare them with the traditional χ^2 minimization method by applying them to the same set of pseudo-data corresponding to a total exposure of $(2.9 \times 5 \times 1800) \text{ GW}_{\text{th}} \cdot \text{tons} \cdot \text{days}$. The results show that the FNN model not only converges faster and better during the fitting process but also achieves relative errors of less than 1% in the 2 – 8 MeV range in the extracted spectra, outperforming the χ^2 minimization method. The feasibility and superiority of this method were validated in the study.

Keywords Reactor neutrinos · Isotope antineutrino spectra · Feedforward neural network

1 Introduction

Since the direct discovery of neutrinos by Cowan and Reines at the Savannah River reactor power plant in 1956 [1], reactor neutrino experiments have played a pivotal role in the

advancement of neutrino physics. Reactor neutrinos are also known as reactor antineutrinos because they are composed exclusively of electron antineutrinos ($\bar{\nu}_e$). In commercial pressurized water reactors (PWRs), more than 99.7% of the reactor neutrinos are emitted from the beta decay branches of neutron-rich fission products generated by four isotopes: ^{235}U , ^{238}U , ^{239}Pu , and ^{241}Pu . In research reactors utilizing 93% ^{235}U -enriched fuel, 99.3% of the reactor neutrinos result from the fission of ^{235}U . Each isotope releases approximately six $\bar{\nu}_e$ per fission along with a corresponding antineutrino flux and spectrum. Precise fissile isotope antineutrino spectra are required for reactor monitoring and safeguarding applications [2–4] and serve as valuable inputs to reactor neutrino experiments utilizing the inverse beta decay (IBD) reaction [5–7] or coherent elastic neutrino-nucleus scattering (CEvNS) [8, 9].

Fissile isotope antineutrino spectra and fluxes have been evaluated several times in the past decades. The methodologies employed can be classified into three major categories comprising summation [10], conversion [11, 12], and extraction methods [13–15]. The summation method, i.e., the *ab initio* approach, utilizes information on fission products and decays from nuclear databases to calculate and sum the contributions of all possible beta decay chains to $\bar{\nu}_e$ [16].

This work was supported by the China Postdoctoral Science Foundation (No. 2024M753715), Fundamental Research Funds for the Central Universities, Sun Yat-sen University (Nos. 24qnpy125 and 22lgjl11), and Guangdong Basic and Applied Basic Research Foundation (No. 2023B1515120030).

✉ Jun Wang
wangj933@mail.sysu.edu.cn

✉ Wei Wang
wangw223@mail.sysu.edu.cn

✉ Yue-Huan Wei
weiyh29@mail.sysu.edu.cn

¹ School of Physics and Astronomy, Sun Yat-sen University, Zhuhai 519082, China

² Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-sen University, Zhuhai 519082, China

³ School of Physics, Sun Yat-sen University, Guangzhou 510275, China

However, the presence of incomplete or inaccurate information in nuclear databases introduces complexities and challenges in constructing reliable spectral models, ultimately leading to potentially large and unknown uncertainties in model predictions. The conversion method relies on the measured beta spectra of uranium and plutonium. The beta spectra for thermal-neutron-induced fissions of ^{235}U , ^{239}Pu , and ^{241}Pu have been measured at the Institut Laue–Langevin High Flux Reactor in the 1980 s [11, 17–19], while those for the fast-neutron-induced fission of ^{238}U were measured at the Heinz Maier-Leibnitz (FRM II) research neutron source in 2013 [20]. The measured beta spectra for each isotope are fitted by a set of virtual beta decay branches based on the allowed beta decay transitions, which are then converted into antineutrino branches and summed to the corresponding isotope antineutrino spectra [16, 21]. Although the reduced dependence on nuclear databases in this method provides spectral shapes with typical relative uncertainties of a few percent, the fine structural information in the spectral shapes is not as rich as that obtained using the summation method. To address these shortcomings, several antineutrino spectrum models have been developed based on the conversion method or a combination of both methods. One example is the Huber–Mueller model [16, 21], which provides predictions that roughly agree with earlier experimental data and is widely accepted in reactor neutrino experiments. However, measurements from short-baseline reactor neutrino experiments such as Double Chooz [22], RENO [23], Daya Bay [13], and NEOS [6] confirmed a $\sim 6\%$ deficit in the measured reactor antineutrino flux and an excess in the 4–6 MeV prompt energy range compared to the predictions of the Huber–Mueller model. These discrepancies, which are, respectively, known as the “reactor antineutrino anomaly (RAA)” [12] and “5 MeV excess” or “5 MeV bump” [24, 25], cannot be ignored in the era of precise measurements. The extraction method, in which the fission isotope antineutrino spectrum is inferred from the reconstructed prompt energy spectrum measured by the detector and independent of nuclear databases, has become a common approach for testing various RAA formation hypotheses, including explanations of sterile neutrinos. Using this method, the Daya Bay experiment [13, 14] extracted the ^{235}U and ^{239}Pu antineutrino spectra from PWRs, while the PROSPECT [15, 26] and STEREO [27, 28] experiments extracted the ^{235}U antineutrino spectrum from highly enriched uranium research reactors. Moreover, it was revealed that the flux deficit was primarily carried by ^{235}U , and the 5 MeV bump had shared contributions from uranium and plutonium. However, the extraction of the ^{238}U and ^{241}Pu antineutrino spectra was not satisfactory owing to statistical limitations [14].

The current general practice in experiments for extracting fissile isotope antineutrino spectra involves first unfolding the reconstructed prompt energy spectrum to obtain an

antineutrino energy spectrum weighted by the IBD cross section, and then further fitting the unfolded spectrum with the χ^2 minimization method to extract individual or combined isotope antineutrino spectra [14, 15, 26, 27]. Unfolding is a common technique used in high-energy physics (HEP) to disentangle detector effects, correct migration effects, suppress fluctuations, and reconstruct approximate distributions of quantities. Common methods for unfolding include singular value decomposition (SVD) [29], Wiener SVD [30], and Bayesian iterations [31]. In the Daya Bay experiment, these methods were used to yield consistent extraction results. Although the Wiener-SVD method produces the smallest unfolded spectrum mean square error (MSE) within the energy range of 3–6 MeV, it does not perform as well as the other methods outside this energy range because of the large statistical fluctuations in the intrinsic neutrino energy spectrum [14]. To obtain more precise solutions, the number of bins for the unfolded spectrum in experiments is typically limited to that of the intrinsic spectrum [32]. Although this simplifies the subsequent fitting process for extracting the specific fission isotope antineutrino spectrum, it also suppresses the fine structure of the spectrum shape.

In our previous study [33], we proposed a machine learning method in which a convolutional neural network (CNN) model is employed to extract fission isotope antineutrino spectra from the unfolded prompt energy spectrum in a virtual short-baseline reactor neutrino experiment. The analysis results demonstrate that the proposed CNN model can achieve subpercentage uncertainties in the extracted ^{235}U and ^{239}Pu antineutrino spectra whereas the ^{238}U and ^{241}Pu antineutrino spectra need to be constrained via prior knowledge during the fitting process. In this study, we extend the method and establish a feedforward neural network (FNN) model to resolve this extraction problem. This new method is designed to directly extract the antineutrino spectra of the four fission isotopes from the reconstructed prompt energy spectrum without highlighting the unfolding process or any constraints on the spectra while better preserving the fine structure of the extracted spectra.

The remainder of this paper is organized as follows: In Sect. 2, we present the antineutrino spectra of the IBD reactions and the generation of the simulation dataset for this study. In Sect. 3, we introduce the conceptual and technical details of the proposed FNN model and its training strategies. In Sect. 4, we compare the performance of this new method in extracting fission isotope antineutrino spectra with that of the benchmark traditional method, that is, the χ^2 minimization method, and discuss the obtained results. Finally, a summary and future outlook are presented in Sect. 5.

2 Dataset generation for FNN model

In this study, we constructed a virtual reactor neutrino experiment in a layout comprising a PWR and a detector. To verify the feasibility of the virtual experiment, we referred to the Daya Bay [14] and Taishan Antineutrino Observatory (TAO, also known as JUNO-TAO) [34, 35] experiments, and made the following assumptions about the experimental parameters: The reactor is operated for 1800 days at a full thermal power of 2.9 GW_{th} with an initial uranium fuel mass of 72 tons. The detector is loaded with 5 tons of liquid scintillator (LS) with 12% hydrogen by mass, has an energy resolution of 8% at 1 MeV and detection efficiency of 50%, and is situated at a baseline distance of 30 m. We adopted the Huber–Mueller model as the foundational theory for the phenomenological prediction of the IBD yield to generate the simulated sample dataset for this study. The model selection did not significantly affect the analysis. We disregarded the contributions of the spent nuclear fuel and the non-equilibrium effect on the IBD yield [16, 32].

2.1 IBD yield prediction

The Huber–Mueller model is a theoretical framework for predicting the antineutrino spectra produced by the fission reactions of four isotopes in reactors. Each of these isotopic antineutrino spectrum can be parameterized using the exponent of a fifth-order polynomial as follows:

$$s_l(E_\nu) = \exp\left(\sum_{p=1}^6 \alpha_{lp} E_\nu^{p-1}\right), \quad (1)$$

where $l = \{^{235}\text{U}, ^{238}\text{U}, ^{239}\text{Pu}, ^{241}\text{Pu}\}$, E_ν is the $\bar{\nu}_e$ energy, and the α_{lp} s are polynomial coefficients for the isotope l . The α_{lp} coefficients for ^{235}U , ^{239}Pu , and ^{241}Pu were derived using the conversion method by Huber [21], whereas the α_{lp} coefficients for ^{238}U were obtained using the summation method of Mueller et al. [16]. To incorporate the RAA in this study, we modified the isotopic antineutrino spectrum in Eq. (1) as follows:

$$S_l(E_\nu) = s_l(E_\nu) r_{\text{RAA}}(E_\nu), \quad (2)$$

where $r_{\text{RAA}}(E_\nu)$ is the ratio of the RAA between the spectra measured in the Daya Bay experiment [32] and the Huber–Mueller model prediction. To evaluate $r_{\text{RAA}}(E_\nu)$, we performed cubic spline interpolation within the provided energy range of 1.8–8 MeV and set it uniformly to 1 for energy values above 8 MeV.

The antineutrino yield per fission can be expressed as

$$\phi(E_\nu, t) = \sum_l f_l(t) S_l(E_\nu), \quad (3)$$

where the fission fraction $f_l(t)$ represents the relative contribution of the isotope l to the fission reaction at time t . The event rate of antineutrinos emitted from the reactor core can be calculated as

$$\frac{dN}{dE_\nu} = \frac{W(t)}{\sum_l f_l(t) \epsilon_l} \phi(E_\nu, t), \quad (4)$$

where $W(t)$ is the thermal power of the reactor at time t , ϵ_l is the mean energy released per fission of the isotope l , and the values for ϵ_l were obtained from Ref. [36].

In the standard three-flavor neutrino oscillation framework, the survival probability P_{ee} of $\bar{\nu}_e$ after propagating a distance L is given by [37]

$$\begin{aligned} P_{ee}(L, E_\nu) &= P(\bar{\nu}_e \rightarrow \bar{\nu}_e; L, E_\nu) \\ &= 1 - \cos^4 \theta_{13} \sin^2(2\theta_{12}) \sin^2(\Delta_{21}) \\ &\quad - \cos^2 \theta_{12} \sin^2(2\theta_{13}) \sin^2(\Delta_{31}) \\ &\quad - \sin^2 \theta_{12} \sin^2(2\theta_{13}) \sin^2(\Delta_{32}), \end{aligned} \quad (5)$$

where the θ_{ij} s represent the neutrino mixing angles. The oscillation phases Δ_{ij} are given by

$$\Delta_{ij} = \frac{\Delta m_{ij}^2 L}{4E_\nu} \simeq \frac{1.267 \Delta m_{ij}^2 [\text{eV}^2] L [\text{m}]}{E_\nu [\text{MeV}]}, \quad (6)$$

where Δm_{ij}^2 denotes the mass-squared difference between the two mass eigenstates m_i and m_j , i.e., $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$.

For short-baseline reactor neutrino experiments, considering that the term involving Δ_{21} is negligible and $\Delta m_{31}^2 \approx \Delta m_{32}^2$, Eq. (5) can be simplified to

$$P_{ee}(L, E_\nu) \approx 1 - \sin^2(2\theta_{13}) \sin^2\left(\frac{\Delta m_{31}^2 L}{4E_\nu}\right). \quad (7)$$

Unless otherwise specified, $\sin^2 \theta_{13} = (2.20 \pm 0.07) \times 10^{-2}$, $\Delta m_{32}^2 = (2.437 \pm 0.033) \times 10^{-3} \text{ eV}^2$, and $\Delta m_{21}^2 = (7.53 \pm 0.18) \times 10^{-5} \text{ eV}^2$ in this study based on the values from the Particle Data Group (PDG) 2022 [37].

As the $\bar{\nu}_e$ emitted by the reactor propagate to the LS detector, some of them engage in IBD reactions with the free target protons in the LS, which are denoted as $\bar{\nu}_e + p \rightarrow e^+ + n$. In this process, the positron e^+ rapidly deposits its energy and annihilates the surrounding electron e^- to form two 0.511 MeV gammas, generating a prompt signal. The neutron n scatters within the detector until it is thermalized and subsequently captured by hydrogen (99%) or carbon (1%) within $\sim 200 \mu\text{s}$, thereby releasing a 2.22 or 4.95 MeV gamma, respectively, and yielding a delayed signal [38]. An IBD event is identified by the prompt-delayed signal pair during such a brief interval. The measured IBD event number M_k in the k -th reconstructed prompt energy E_{rec} bin

observed at a detector within the data acquisition time T_{DAQ} is therefore given by [5]

$$M_k = \frac{N_p \epsilon}{4\pi L^2} \int_{E_{\text{rec}}^k}^{E_{\text{rec}}^{k+1}} dE_{\text{rec}} \int_{T_{\text{DAQ}}} dt \int_{E_{\text{thr}}} dE_\nu \times \frac{dN}{dE_\nu} P_{\text{ec}}(L, E_\nu) \sigma_{\text{IBD}}(E_\nu) G(E_\nu, E_{\text{rec}}) \quad (8)$$

where N_p is the number of free target protons in the LS, ϵ is the detection efficiency of the detector, the IBD threshold energy $E_{\text{thr}} \sim m_n - m_p + m_e \sim 1.8$ MeV, $\sigma_{\text{IBD}}(E_\nu)$ is the cross section of the IBD taken from Ref. [39], and $G(E_{\text{rec}}, E_\nu)$ is a normalized Gaussian smearing function, which includes the energy resolution effect.

To simplify the calculation, we assumed that the detector has no energy leakage or LS nonlinearity [32]. Thus, the prompt energy $E_{\text{pro}} \sim E_\nu - 0.78$ MeV, and E_{rec} is expected to obey the distribution $G(E_{\text{rec}}, E_\nu)$ defined as follows [5]:

$$G(E_\nu, E_{\text{rec}}) \simeq \frac{1}{\sqrt{2\pi}\delta_{E_{\text{pro}}}} \exp \left\{ -\frac{(E_{\text{pro}} - E_{\text{rec}})^2}{2(\delta_{E_{\text{pro}}})^2} \right\}. \quad (9)$$

The energy resolution $\delta_{E_{\text{pro}}}$ is parameterized as

$$\frac{\delta_{E_{\text{pro}}}}{E_{\text{pro}}} = \sqrt{\left(\frac{p_0}{\sqrt{E_{\text{pro}}}} \right)^2 + p_1^2 + \left(\frac{p_2}{E_{\text{pro}}} \right)^2}, \quad (10)$$

where p_0 quantifies the statistical fluctuations in the photons detected by the detector, p_1 is predominantly influenced by residual effects resulting from the spatial nonuniformity and temporal instability correction of the detector, and p_2 quantifies the effects associated with the photomultiplier tube (PMT), notably the PMT dark noise [5, 38].

For simplicity, we set $p_0 = 0.08$, $p_1 = 0$, and $p_2 = 0$ in this study for an energy resolution of 8% at 1 MeV in the detector. Therefore, under full reactor power and classical fission fractions conditions [32], the detector observes the energy spectrum of the IBD events (i.e., the reconstructed prompt energy spectrum) distorted by the RAA in one day, as shown in Fig. 1, and approximately 7473 IBD events are recorded.

2.2 Simulated samples and targets in dataset

Considering the significant computational resources and time required for the integral terms in Eq. (8), Eq. (8) is typically converted to a discrete summation or matrix multiplication equivalent form in practical computations. In this study, the integral form of the reconstructed prompt energy spectrum is rewritten as an element of the row matrix

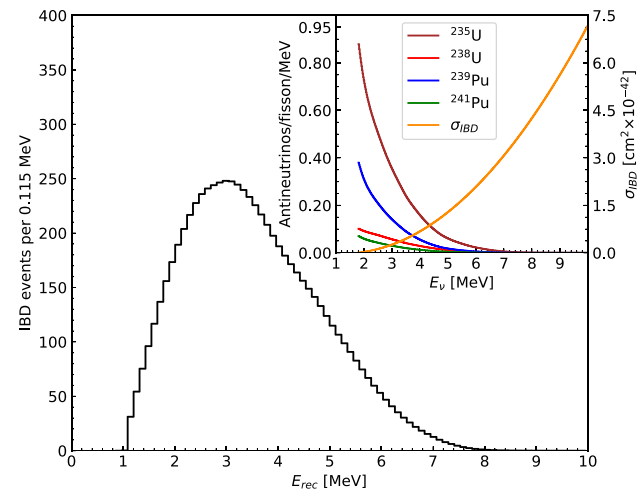


Fig. 1 (Color online) Virtual detector observes the reconstructed prompt energy spectrum by the RAA with an exposure of $(2.9 \times 10^7 \text{ s})$. The inset shows the cross section of the IBD for four isotope antineutrino spectra obtained by the Fueller model according to Eq. (2)

Thus, the prompt energy spectrum is represented by the matrix $M_{1 \times N_{E_{\text{rec}}}}$. Each element of the matrix represents the reconstructed IBD event number in the corresponding energy bin.

$$M_{1 \times N_{E_{\text{rec}}}} = X_{1 \times 4} \cdot S_{4 \times N_{E_\nu}} \cdot P_{N_{E_\nu} \times N_{E_\nu}} \cdot \sigma_{N_{E_\nu} \times N_{E_\nu}} \cdot R_{N_{E_\nu} \times N_{E_{\text{rec}}}} \\ = X_{1 \times 4} \cdot S_{4 \times N_{E_\nu}} \cdot P \sigma R_{N_{E_\nu} \times N_{E_{\text{rec}}}}, \quad (11)$$

where the subscripts denote the dimensions of the corresponding matrices and the number 4 indicates the four isotopes. $N_{E_{\text{rec}}}$ is the number of bins in the reconstructed prompt energy spectrum while N_{E_ν} is the number of terms in the discretized sum for integration over E_ν , which is also the number of bins in the extracted isotopic antineutrino spectrum. The ranges of E_ν and E_{rec} are 1.8 – 10 MeV and 0.8 – 10 MeV, respectively. In this study, $N_{E_{\text{rec}}}$ was set to 80 based on the limit of the virtual detector energy resolution whereas N_{E_ν} was optimized to 401 after balancing model performance and computational cost. The element X_i in $X_{1 \times 4}$ can be expressed as

$$X_i = \sum_u \frac{N_p \epsilon W(t^u) f_i(t^u)}{4\pi L^2 \sum_l f_l(t^u) \epsilon_l} \Delta t \Delta E_\nu \Delta E_{\text{rec}}, \quad (12)$$

where T_{DAQ} in Eq. (8) is divided into N_t time units of Δt , u is the time unit index, and ΔE_ν and ΔE_{rec} are the bin widths of the extracted isotopic antineutrino spectrum and reconstructed prompt energy spectrum, respectively. In Eq. (12), W , f_l , and ϵ_l are reactor-related parameters. W and f_l vary as the reactor evolves while ϵ_l and the remaining parameters are

constants. X_l is therefore referred to as the reactor dynamic evolution information.

Each row of $S_{4 \times N_{E_v}}$ represents the binned antineutrino spectrum for isotope l , as given by $S_l(E_v)$. Both $P_{N_{E_v} \times N_{E_v}}$ and $\sigma_{N_{E_v} \times N_{E_v}}$ are diagonal matrices whose diagonal elements are given by $P_{ce}(L, E_v)$ and $\sigma_{IBD}(E_v)$, respectively. The role of $R_{N_{E_v} \times N_{E_{rec}}}$ is to map each E_v to a spectrum of E_{rec} . $R_{N_{E_v} \times N_{E_{rec}}}$ is therefore also referred to as the detector response matrix [R_{qk}], which is defined as follows:

$$\begin{aligned} R_{qk} &= R(E_v^q, E_{rec}^k) = G\left(E_v^q, \frac{E_{rec}^k + E_{rec}^{k+1}}{2}\right) \\ &= G\left(E_v^q, E_{rec}^k + \frac{\Delta E_{rec}}{2}\right), \end{aligned} \quad (13)$$

where q is the index for binning E_v , $q \in [1, 2, \dots, N_{E_v}]$, and $k \in [1, 2, \dots, N_{E_{rec}}]$. In contexts that do not involve the oscillation parameters or unfolding, $P_{N_{E_v} \times N_{E_v}}$, $\sigma_{N_{E_v} \times N_{E_v}}$, and $R_{N_{E_v} \times N_{E_{rec}}}$ can be pre-multiplied to obtain the matrix $P\sigma R_{N_{E_v} \times N_{E_{rec}}}$.

The matrix multiplication relation in Eq. (11) provides the mathematical foundation for constructing the FNN architecture presented in Sect. 3. Furthermore, $X_{1 \times 4}$ and $M_{1 \times N_{E_{rec}}}$, respectively, constitute a sample and its associated target in our dataset, which serve as a feature-label pair for supervised learning in the FNN model implemented in this study.

As described in Eq. (12), the fission fraction varies dynamically with burn-up as the reactor operates. In each reactor core refueling cycle, the cycle burn-up can be calculated as [32]

$$\text{Burn-up} = \frac{W \cdot D}{M_{Uini}}, \quad (14)$$

where W , D , and M_{Uini} represent the total thermal power of the reactor, the number of days since the refueling cycle started, and the mass of the initial uranium fuel loaded into the reactor, respectively. The unit for burn-up is $\text{GW}_{th} \cdot \text{day} \cdot \text{ton}^{-1}$. Given that the real-time power output of the reactor is dynamic and cannot exceed its maximum capacity of 2.9 GW_{th} for safe operation, we used a random number generator for a normal distribution with a mean of 2.9 GW_{th} and downward fluctuation of 0.5% to determine the daily average power output of the virtual reactor [33]. By incorporating the fission fraction evolution data of the isotopes during a complete burn-up cycle from Ref. [32], we obtained the evolution of the fission fractions for the four main isotopes as a function of the operation day, as shown in Fig. 2.

Under the assumption that the thermal power and fissile fractions for the four main isotopes of the reactor are constant within each day, we accumulated the exposure over

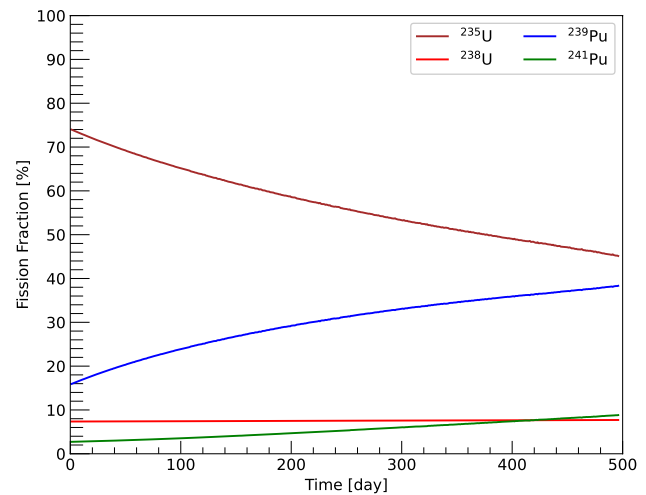


Fig. 2 (Color online) Evolution of fissile fractions for the four main isotopes in the reactor core as a function of operation day, which includes one complete refueling cycle [32]. The cumulative fission fraction of the four main isotopes used in this experiment is normalized to unity, and other isotopes contributing less than 0.3% are excluded from our analysis

each 3-day interval as a sample to create a dataset of 600 simulated samples and their corresponding targets for subsequent analysis.

3 Implementation of FNN model

Machine learning algorithms such as neural network (NN) models have attracted increasing attention from high-energy and nuclear physics researchers [33, 40–44]. However, most of these applications are characterized by black-box models in which the meaning of the model parameters are challenging to understand or interpret. In this section, we present a FNN-based white-box model where each layer and parameter has a clear physical or mathematical meaning, thereby ensuring the interpretability of the model.

3.1 Mathematical foundations of FNN model

The NN is a powerful machine learning model that has been widely explored and applied across various fields. The universal approximation theorem [45, 46] implies that any continuous function can be approximated with arbitrary precision using an appropriate NN, even if the NN is an FNN with only one hidden layer containing a sufficient number of neurons. However, the internal structure and parameters of the NN in such scenarios often lack physical meaning or interpretability. This results in black-box models, which are not fully trusted by high-energy physicists. Therefore, we designed and implemented a white-box NN model in this

study for converting the mathematical mapping function in Eq. (11) to a FNN model.

An FNN is typically composed of one to several single-layer perceptrons, which are considered the fundamental building units of the FNN and play a vital role in its overall functionality [47]. Each perceptron in the FNN follows the computational flow shown in Fig. 3 to process data. Forward and backward propagation are two phases in the NN training process that interact to optimize network performance.

During the forward propagation phase, the perceptron performs computation by computing the dot product of the input vector $\vec{x} = [x_1, x_2, \dots, x_N]^T$ with the weight coefficient vector $\vec{w} = [w_1, w_2, \dots, w_N]^T$, adding the bias b , and applying the activation function ψ to yield the activation result y as the output. The discrepancy between the output y and target \hat{y} is then calculated using the loss function $L(y, \hat{y})$. Forward propagation provides the foundation for evaluating network performance. Backward propagation in turn determines how the network parameters (weights and bias) are updated to reduce loss. It can be described as

$$\omega'_\mu = \omega_\mu - \eta \times [\nabla_{\omega_\mu} L(y, \hat{y}) + \lambda \omega_\mu], \quad (15)$$

$$b' = b - \eta \times \nabla_b L(y, \hat{y}), \quad (16)$$

where ω_μ and ω'_μ represent the μ -th weight coefficient of the current and subsequent steps, respectively; b and b' the biases of the current and subsequent steps, respectively; and η and λ are the learning rate and weight decay rate, respectively. This iterative update process of the parameters based on the computed gradients allows the NN to learn and improve its predictions over time.

To allow matrix multiplication in the perceptrons, the bias b must be eliminated, i.e., set to zero. The absence of negative values in our data flow justifies the use of the default rectified linear unit (ReLU) activation function, which is defined as $\psi(z) = \max\{0, z\}$. This setup also permits the perceptrons to be chained to perform successive matrix dot product operations, which is integral to the development of our FNN model.

As shown in Fig. 4, the architecture of the FNN model consists of three layers comprising, from left to right, the input, hidden, and output layers with four, N_{Ev} , and N_{Rec}

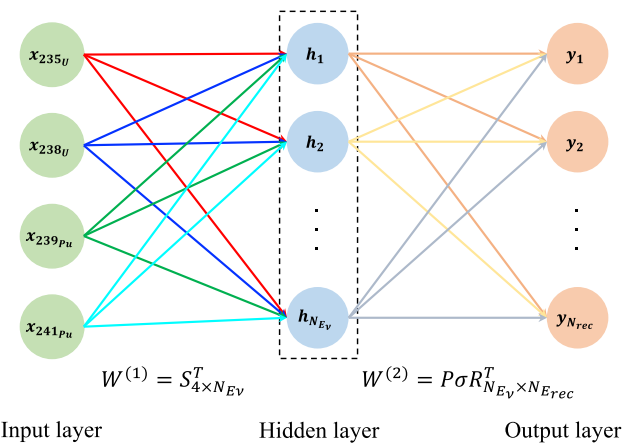


Fig. 4 (Color online) FNN is a white-box model that describes the mapping relation between the reactor dynamic evolution information and reconstructed prompt energy spectrum. The architecture of the FNN model includes an input layer, hidden layer, output layer, and two sets of weight coefficient matrices $W^{(1)}$ and $W^{(2)}$. The weight values between neurons associated with connections of the same color form the rows of the weight coefficient matrix

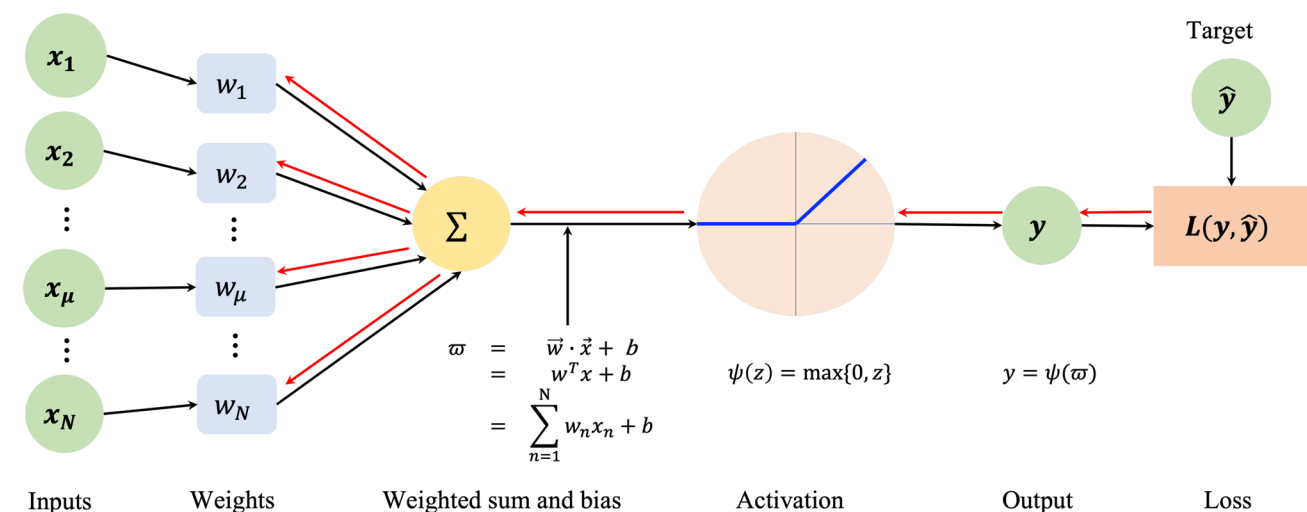


Fig. 3 (Color online) An example illustration of the structure of a single-layer perceptron along with forward (black flow arrows) and backward (red flow arrows) propagation

neurons, respectively. The neurons in adjacent layers are connected using a fully connected approach; that is, each neuron in one layer is connected to every neuron in the subsequent layer with no connections between neurons within the same layer. The training process of the model starts from the input layer, at which each neuron receives the reactor dynamic evolution information corresponding to its fission isotope. The output of the hidden layer is the scaled total spectrum $[h_q]$ of antineutrinos emitted by the reactor. The output layer then provides a predicted reconstructed prompt energy spectrum $[y_k]$. The two weight coefficient matrices $W^{(1)}$ and $W^{(2)}$ correspond to the transposes of the matrices $S_{4 \times N_{Ev}}$ and $P\sigma R_{N_{Ev} \times N_{Erec}}$, respectively. The matrix $W^{(1)}$ contains the fission isotope antineutrino spectra to be extracted, which are learned during training. In contrast, the matrix $W^{(2)}$ is fixed as $P\sigma R_{N_{Ev} \times N_{Erec}}^T$ because it is assumed to be a constant matrix without uncertainties in this study. The FNN is therefore a supervised learning model that iteratively refines $W^{(1)}$ to minimize the discrepancies between its outputs and corresponding targets.

3.2 Training strategy

All the samples generated in Sect. 2.2 were utilized solely to train the FNN model. The validation and testing processes were omitted. This approach was chosen because our aim is to minimize the loss function during the training process to determine the optimal $W^{(1)}$ for extracting the four main isotopic antineutrino spectra. Our focus is on optimizing spectra extraction performance rather than evaluating model performance across various datasets, as

well as on simplifying the process and aligning with our primary research objective.

The loss function is a fundamental component in deep learning models. It serves as the criterion for evaluating how well the model predictions match the actual outcomes and provides a numerical indicator of model accuracy. The Combined Neyman–Pearson (CNP) chi square model is a statistical model frequently employed in HEP experiments to quantify the error between predicted and measured values [48]. Based on this model, we define the loss function for the FNN model as

$$\chi_{\text{CNP}}^2 = \sum_{k=1}^{N_{\text{Erec}}} \frac{[M_k - y_k(W^{(1)})]^2}{3 \left/ \left[\frac{1}{M_k} + \frac{2}{y_k(W^{(1)})} \right] \right.}, \quad (17)$$

where M_k is the IBD event number in the k -th bin for the measured reconstructed prompt neutrino energy spectrum given by Eq. (8) and y_k is the corresponding predicted value output of the model. We used this loss function to guide the optimization process of $W^{(1)}$ during the training process so that the FNN was driven toward increasingly precise predictions.

After defining the loss function, it is essential to select a suitable optimizer, learning rate schedule, batch size, and epoch, among other hyperparameters. Following hyperparameter tuning using the Optuna framework [49] and extensive testing, we developed two training strategies denoted as the short and long epoch strategies to investigate the performance of the FNN model in extracting the antineutrino spectra of the four fission isotopes from the reconstructed prompt energy spectrum [50]. As shown in Table 1, a critical commonality between these two

Table 1 Configurations of the two training strategies for FNN model

Strategy	Short epoch	Long epoch
Epoch	2×10^3	2×10^6
Optimizer	AdamW	Adam
Hidden layer partitions	[1], (1, 180], (180, 225], (225, 303], (303, 401]	
Learning rates for hidden layer	$[3.4892 \times 10^{-4}, 9.9485 \times 10^{-4}, 2.754 \times 10^{-4}, 1.8272 \times 10^{-4}, 0]$	
Weight decay rates for hidden layer	$[7.418 \times 10^{-3}, 7.748 \times 10^{-3}, 4.155 \times 10^{-3}, 9.999 \times 10^{-3}, 0]$	[0, 0, 0, 0, 0]
Learning rate for output layer	0	
Weight decay for output layer	0	
Learning rate scheduler	ReduceLROnPlateau (factor = 0.32, patience = 1×10^2)	ReduceLROnPlateau (factor = 0.32, patience = 1×10^4) & epoch $\geq 2 \times 10^5$
Batch size	30	

The configurations were derived based on our empirical knowledge and optimized using Optuna [49]. The partition numbers correspond to neuron indices in the hidden layer of the FNN model. ReduceLROnPlateau is a Python class that dynamically adjusts the learning rate during deep learning model training to improve convergence speed and performance [50]

strategies is the segmentation of the hidden layer in the FNN model into multiple partitions or parallel hidden layers. This setup allows distinct learning and weight decay rates to be assigned to each partition to facilitate differential performance outcomes. Because the focus in this study is not on the isotope antineutrino spectra above 8 MeV, i.e., in the (303, 401] partition or the matrix $P\sigma R_{N_{E_\nu} \times N_{E_{\text{rec}}}}$, we fixed their learning and weight decay rates to zero and disabled the gradient calculations for the corresponding weight coefficients. Additionally, we set the initial value of $W^{(1)}$ based on the Huber–Mueller model.

As indicated by their names, the main distinction between the short and long epoch strategies lies in the epochs. The short epoch strategy leverages the AdamW [51] optimizer with non-zero weight decay rates for faster loss reduction. In contrast, in the long epoch strategy, the Adam [52] optimizer is applied without weight decay, i.e., the weight decay rates are set to zero. Superior convergence results were obtained using the long epoch strategy. The results are presented and discussed in Sect. 4. As illustrated in Table 1, these circumstances also led to minor differences in the configurations of the learning rate schedulers. Nonetheless, the same metric, i.e., the sum of the losses for all samples denoted as $\chi^2_{\Sigma \text{CNP}}$, was monitored in both schedulers.

We also extracted the antineutrino spectra of the four fission isotopes using the χ^2 minimization method to provide a comparison and benchmark for the FNN model. We employed the Minuit2 minimization library from ROOT [53] to implement this method. $\chi^2_{\Sigma \text{CNP}}$ was used as the objective function to be minimized to find the best fit. The same dataset as that for the FNN model was used as the measured value in this fitting process. In contrast, the predicted value was derived from Eq. (11) where the $S_{4 \times N_{E_\nu}}$ matrix elements corresponding to ≤ 8 MeV are the parameters to be fitted and the remaining elements considered as fixed parameters in the fitting procedure. We adopted the “Combined” minimizer algorithm to minimize the objective function with initial fitting values from the Huber–Mueller model and fitting step sizes of 1% of the order of magnitude of these values. We set the tolerance for the fitting procedure to 1×10^{-30} . The fitting stopped automatically only when the improvement in the $\chi^2_{\Sigma \text{CNP}}$ value between consecutive iterations fell below this threshold.

The FNN model was implemented using PyTorch [54], a Python-based deep learning library that supports both CPU and GPU platforms and is one of the mainstream tools for developing and training NN models. A NVIDIA GeForce RTX 3060 Ti GPU platform was used to deploy the FNN model, whereas tasks involving Optuna and ROOT were performed on two identical servers, each of which was equipped with two 28-core Intel(R) Xeon(R) Gold 6330 CPUs @ 2.00 GHz.

4 Results and discussions

To facilitate the discussion and comparative analysis of the short and long epoch strategies of our FNN model and the χ^2 minimization method, we first consider their performance in fitting all the samples and reducing the losses. As shown in Fig. 5, did the loss $\chi^2_{\Sigma \text{CNP}}$ decreased more rapidly in both FNN strategies, and lower ultimate $\chi^2_{\Sigma \text{CNP}}$ values were obtained compared to those obtained by the χ^2 minimization method. The $\chi^2_{\Sigma \text{CNP}}$ values at the conclusion of the epochs are 5.51×10^{-6} , 5.42×10^{-10} , and 9.34×10^{-6} , respectively.

The short epoch strategy can rapidly reduce the loss in the early stages of training mainly because of the regularization effects and optimization efficiency due to the combination of non-zero weight decay rates and the AdamW optimizer. However, in the later stages of training, the model must be able to respond to small changes in the loss function for fine adjustments of the parameters. Weight decay may interfere with this process and make it challenging for the model to determine the optimal solution within regions of small loss function gradients.

Figure 6 shows a comparison of the performance in extracting the antineutrino spectra of the four isotopes using these three approaches. The extraction performance decreases in the order of the long epoch strategy, short epoch strategy, and χ^2 minimization method. The FNN model accurately extracted the antineutrino spectra of ^{235}U , ^{239}Pu , and ^{241}Pu in the energy range of 2–5 MeV. The FNN model with the short epoch strategy achieved relative errors of less

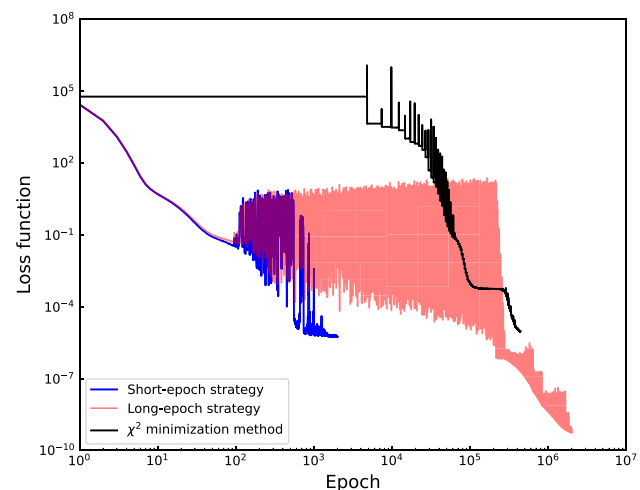


Fig. 5 (Color online) Evolution of loss function across epochs for the short and long epoch strategies and the χ^2 minimization method. The epochs of the first two were manually specified to be 2×10^3 and 2×10^6 , respectively, while that of the χ^2 minimization method was automatically determined as approximately 4.39×10^5

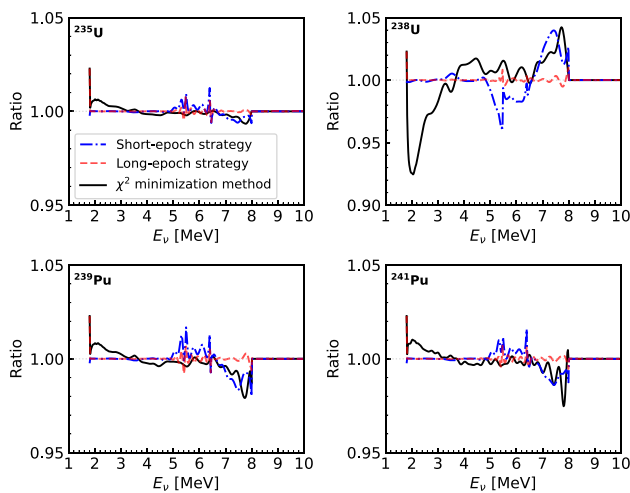


Fig. 6 (Color online) Comparison of the ratios between the four isotope antineutrino spectra extracted using the short and long epoch strategies in our FNN model and the χ^2 minimization method, and the assumed true spectra described by Eq. (2)

than 2% in the 5–8 MeV range, which decreased to less than 1% with the long epoch strategy. In comparison, the χ^2 minimization method achieved relative extraction errors of less than 2% and 3% for these three isotopes in the respective energy ranges. For the isotope ^{238}U , both the short epoch strategy and χ^2 minimization method showed relatively poor extraction performance compared to that for the other isotopes. The maximum extraction relative errors in the 2–8 MeV range are approximately 4% and 8%, respectively, whereas only the long epoch strategy maintained relative errors of less than 1%.

It is worth noting that although ^{241}Pu has a lower average fission fraction throughout the entire refueling cycle compared to ^{238}U , the extraction performance for the former is better in all the extraction approaches. This indicates that in addition to large fission fractions, significant variations are also crucial for extracting isotopic antineutrino spectra accurately. Greater variations produce better extraction results. This is further confirmed by the extraction performance for the ^{235}U and ^{239}Pu antineutrino spectra. Therefore, such long epochs are employed in the long epoch strategy primarily to enhance the extraction performance for ^{238}U . Overall, regardless of the extraction approach used, the extraction performance for the isotopic antineutrino spectra in descending order is as follows: ^{235}U , ^{239}Pu , ^{241}Pu , and ^{238}U .

The above results and discussion reveal that because of the exceptional capability of NNs in optimizing large-scale parameters, the FNN model achieved faster and more effective convergence than the traditional χ^2 minimization method. Based on PyTorch's extensive array of optimization algorithms [55], various model training strategies can be designed to satisfy the practical requirements for extracting

isotope antineutrino spectra. Moreover, executing spectrum extraction algorithms on GPU platforms can significantly increase the inference speed of the process, thereby improving extraction efficiency.

5 Summary and outlook

In this study, we presented an FNN model designed to infer and extract the corresponding antineutrino spectra generated by the fission of ^{235}U , ^{238}U , ^{239}Pu , and ^{241}Pu from the reconstructed prompt energy spectrum measured by the detector in a reactor neutrino experiment. Using a simulated short-baseline reactor neutrino experiment with an exposure of $(2.9 \times 5 \times 1800) \text{ GW}_{\text{th}} \cdot \text{tons} \cdot \text{days}$, we demonstrated how this FNN model establishes a mapping from reactor evolution information to the reconstructed prompt energy spectrum and enables the extraction of antineutrino spectra for the four isotopes through its training process.

By comparing the extraction effects of the short and long epoch training strategies for our FNN model with the traditional χ^2 minimization method, as shown in Fig. 6, we found that the FNN model converged faster and better, and the performance of the three approaches for extracting the isotope antineutrino spectra in descending order is as follows: long epoch strategy, short epoch strategy, and χ^2 minimization method. Furthermore, the relative extraction errors of the antineutrino spectra for the four isotopes are reduced to less than 1% in the 2–8 MeV range of interest by the FNN model with the long epoch strategy, which is better than the error of 8% or less obtained using the χ^2 minimization method in the control group. These results show that the FNN model has considerable potential for extracting fission isotope antineutrino spectra.

In the near future, TAO will serve as a satellite experiment of JUNO and achieve an energy resolution exceeding 2% at 1 MeV in measuring reactor antineutrinos [34]. Its primary physics goals include constraining the fine structures of isotope antineutrino spectra and providing a model-independent reference spectrum for JUNO and a benchmark measurement to test nuclear databases. Employing the FNN model in high-precision experiments such as TAO would therefore be an excellent match. In addition, depending on the research objectives, new NN models can be developed using the methodologies outlined in this study to further investigate a broader range of physics topics such as unfolding, neutrino oscillation parameter measurements, sterile neutrino searches, and reactor monitoring. For example, the unfolded neutrino energy spectrum is represented by the output of the hidden layers in our FNN model, which can achieve a relative error of less than 1% in the 2–8 MeV range.

Acknowledgements We thank Yuda Zeng for the fruitful discussions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jian Chen and Jun Wang. The first draft of the manuscript was written by Jian Chen and Jun Wang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at <https://cstr.cn/31253.11.scienicedb.j00186.00689> and <https://doi.org/10.57760/sciencedb.j00186.00689>.

Declarations

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

References

1. F. Reines, C.L. Cowan, F.B. Harrison et al., Detection of the free antineutrino. *Phys. Rev.* **117**, 159–173 (1960). <https://doi.org/10.1103/PhysRev.117.159>
2. A. Bernstein, Y.F. Wang, G. Gratta et al., Nuclear reactor safeguards and monitoring with anti-neutrino detectors. *J. Appl. Phys.* **91**, 4672 (2002). <https://doi.org/10.1063/1.1452775>
3. M. Fallot, B. Littlejohn P. Dimitriou, Antineutrino spectra and their applications. Tech. rep., International Atomic Energy Agency. <https://doi.org/10.61092/iaea.e4zk-7ryk>
4. N.S. Bowden, J.M. Link, W. Wang, Report of the Topical Group on Neutrino Applications for Snowmass 2021. in: Snowmass 2021, (2022). <https://doi.org/10.48550/arXiv.2209.07483>
5. JUNO collaboration, Neutrino physics with JUNO. *J. Phys. G* **43**, 030401 (2016). <https://doi.org/10.1088/0954-3899/43/3/030401>
6. NEOS collaboration, Sterile neutrino search at the NEOS experiment. *Phys. Rev. Lett.* **118**, 121802 (2017). <https://doi.org/10.1103/PhysRevLett.118.121802>
7. PROSPECT collaboration, Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR. *Phys. Rev. D* **103**, 032001 (2021). <https://doi.org/10.1103/PhysRevD.103.032001>
8. CONUS collaboration, Constraints on elastic neutrino nucleus scattering in the fully coherent regime from the CONUS experiment. *Phys. Rev. Lett.* **126**, 041804 (2021). <https://doi.org/10.1103/PhysRevLett.126.041804>
9. RELICS collaboration, Reactor neutrino liquid xenon coherent elastic scattering experiment. *Phys. Rev. D* **110**, 072011 (2024). <https://doi.org/10.1103/PhysRevD.110.072011>
10. M. Fallot, S. Cormon, M. Estienne et al., New antineutrino energy spectra predictions from the summation of beta decay branches of the fission products. *Phys. Rev. Lett.* **109**, 202504 (2012). <https://doi.org/10.1103/PhysRevLett.109.202504>
11. K. Schreckenbach, H.R. Faust, F. von Feilitzsch et al., Absolute measurement of the beta spectrum from ^{235}U fission as a basis for reactor antineutrino experiments. *Phys. Lett. B* **99**, 251–256 (1981). [https://doi.org/10.1016/0370-2693\(81\)91120-5](https://doi.org/10.1016/0370-2693(81)91120-5)
12. G. Mention, M. Fechner, T. Lasserre et al., The reactor antineutrino anomaly. *Phys. Rev. D* **83**, 073006 (2011). <https://doi.org/10.1103/PhysRevD.83.073006>
13. Daya Bay collaboration, Extraction of the ^{235}U and ^{239}Pu antineutrino spectra at Daya Bay. *Phys. Rev. Lett.* **123**, 111801 (2019). <https://doi.org/10.1103/PhysRevLett.123.111801>
14. Daya Bay collaboration, Antineutrino energy spectrum unfolding based on the Daya Bay measurement and its applications. *Chin. Phys. C* **45**, 073001 (2021). <https://doi.org/10.1088/1674-1137/abfc38>
15. Daya Bay and PROSPECT collaborations, Joint determination of reactor antineutrino spectra from ^{235}U and ^{239}Pu fission by Daya Bay and PROSPECT. *Phys. Rev. Lett.* **128**, 081801 (2022). <https://doi.org/10.1103/PhysRevLett.128.081801>
16. T.A. Mueller, D. Lhuillier, M. Fallot et al., Improved predictions of reactor antineutrino spectra. *Phys. Rev. C* **83**, 054615 (2011). <https://doi.org/10.1103/PhysRevC.83.054615>
17. F. Von Feilitzsch, A.A. Hahn, K. Schreckenbach, Experimental beta-spectra from ^{239}Pu and ^{235}U thermal neutron fission products and their correlated antineutrino spectra. *Phys. Lett. B* **118**, 162–166 (1982). [https://doi.org/10.1016/0370-2693\(82\)90622-0](https://doi.org/10.1016/0370-2693(82)90622-0)
18. K. Schreckenbach, G. Colvin, W. Gelletly et al., Determination of the antineutrino spectrum from ^{235}U thermal neutron fission products up to 9.5 MeV. *Phys. Lett. B* **160**, 325–330 (1985). [https://doi.org/10.1016/0370-2693\(85\)91337-1](https://doi.org/10.1016/0370-2693(85)91337-1)
19. A.A. Hahn, K. Schreckenbach, G. Colvin et al., Anti-neutrino Spectra From ^{241}Pu and ^{239}Pu thermal neutron fission products. *Phys. Lett. B* **218**, 365–368 (1989). [https://doi.org/10.1016/0370-2693\(89\)91598-0](https://doi.org/10.1016/0370-2693(89)91598-0)
20. N. Haag, A. Gütlein, M. Hofmann et al., Experimental determination of the antineutrino spectrum of the fission products of ^{238}U . *Phys. Rev. Lett.* **112**, 122501 (2014). <https://doi.org/10.1103/PhysRevLett.112.122501>
21. P. Huber, Determination of antineutrino spectra from nuclear reactors. *Phys. Rev. C* **84**, 024617 (2011). <https://doi.org/10.1103/PhysRevC.84.024617>
22. Double Chooz collaboration, Double chooz θ_{13} measurement via total neutron capture detection. *Nature Phys.* **16**, 558–564 (2020). <https://doi.org/10.1038/s41567-020-0831-y>
23. RENO collaboration, Measurement of reactor antineutrino oscillation amplitude and frequency at RENO. *Phys. Rev. Lett.* **121**, 201801 (2018). <https://doi.org/10.1103/PhysRevLett.121.201801>
24. RENO collaboration, New results from RENO and the 5 MeV excess. *AIP Conf. Proc.* **1666**, 080002 (2015). <https://doi.org/10.1063/1.4915563>
25. P. Huber, NEOS data and the origin of the 5 MeV bump in the reactor antineutrino spectrum. *Phys. Rev. Lett.* **118**, 042502 (2017). <https://doi.org/10.1103/PhysRevLett.118.042502>
26. PROSPECT collaboration, Final measurement of the ^{235}U antineutrino energy spectrum with the PROSPECT-I detector at HFIR. *Phys. Rev. Lett.* **131**, 021802 (2023). <https://doi.org/10.1103/PhysRevLett.131.021802>
27. STEREO and PROSPECT collaborations, Joint measurement of the ^{235}U antineutrino spectrum by PROSPECT and STEREO. *Phys. Rev. Lett.* **128**, 081802 (2022). <https://doi.org/10.1103/PhysRevLett.128.081802>
28. STEREO collaboration, STEREO neutrino spectrum of ^{235}U fission rejects sterile neutrino hypothesis, *Nature* **613**, 257–261 (2023). <https://doi.org/10.1038/s41586-022-05568-2>
29. A. Hocker, V. Kartvelishvili, SVD approach to data unfolding. *Nucl. Instrum. Meth. A* **372**, 469–481 (1996). [https://doi.org/10.1016/0168-9002\(95\)01478-0](https://doi.org/10.1016/0168-9002(95)01478-0)
30. W. Tang, X. Li, X. Qian et al., Data unfolding with Wiener-SVD method. *JINST* **12**, P10002 (2017). <https://doi.org/10.1088/1748-0221/12/10/P10002>
31. G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth. A* **362**, 487–498 (1995). [https://doi.org/10.1016/0168-9002\(95\)00274-X](https://doi.org/10.1016/0168-9002(95)00274-X)
32. Daya Bay collaboration, Improved measurement of the reactor antineutrino flux and spectrum at Daya Bay. *Chin. Phys. C* **41**, 013002 (2017). <https://doi.org/10.1088/1674-1137/41/1/013002>

33. Y.D. Zeng, J. Wang, R. Zhao et al., Decomposition of fissile isotope antineutrino spectra using convolutional neural network. *Nucl. Sci. Tech.* **34**, 79 (2023). <https://doi.org/10.1007/s41365-023-01229-9>
34. JUNO collaboration, TAO conceptual design report: a precision measurement of the reactor antineutrino spectrum with sub-percent energy resolution. <https://doi.org/10.48550/arXiv.2005.08745>
35. G. Luo, Y. Hor, P. Lu et al., Design optimization of JUNO-TAO plastic scintillator with WLS-fiber and SIPM readout. *Nucl. Sci. Tech.* **34**, 99 (2023). <https://doi.org/10.1007/s41365-023-01263-7>
36. X.B. Ma, W.L. Zhong, L.Z. Wang et al., Improved calculation of the energy release in neutron-induced fission. *Phys. Rev. C* **88**, 014605 (2013). <https://doi.org/10.1103/PhysRevC.88.014605>
37. Particle Data Group collaboration, Review of particle physics. *PTEP* **2022**, 083C01 (2022). <https://doi.org/10.1093/ptep/ptac097>
38. JUNO collaboration, Calibration strategy of the JUNO experiment. *JHEP* **03**, 004 (2021). [https://doi.org/10.1007/JHEP03\(2021\)004](https://doi.org/10.1007/JHEP03(2021)004)
39. A. Strumia, F. Vissani, Precise quasielastic neutrino/nucleon cross-section. *Phys. Lett. B* **564**, 42–54 (2003). [https://doi.org/10.1016/S0370-2693\(03\)00616-6](https://doi.org/10.1016/S0370-2693(03)00616-6)
40. 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>
41. 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>
42. T.S. Shang, J. Li, Z.M. Niu, Prediction of nuclear charge density distribution with feedback neural network. *Nucl. Sci. Tech.* **33**, 153 (2022). <https://doi.org/10.1007/s41365-022-01140-9>
43. W.B. He, Y.G. Ma, L.G. Pang et al., High-energy nuclear physics meets machine learning. *Nucl. Sci. Tech.* **34**, 88 (2023). <https://doi.org/10.1007/s41365-023-01233-z>
44. X.C. Ming, H.F. Zhang, R.R. Xu et al., Nuclear mass based on the multi-task learning neural network method. *Nucl. Sci. Tech.* **33**, 48 (2022). <https://doi.org/10.1007/s41365-022-01031-z>
45. G. Cybenko, Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2**, 303–314 (1989). <https://doi.org/10.1007/BF02551274>
46. K. Hornik, Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991). [https://doi.org/10.1016/0893-6080\(91\)90009-t](https://doi.org/10.1016/0893-6080(91)90009-t)
47. I. Goodfellow, Y. Bengio, A. Courville, Deep learning. MIT press, (2016). <http://www.deeplearningbook.org>
48. X. Ji, W. Gu, X. Qian et al., Combined Neyman-Pearson chi-square: an improved approximation to the Poisson-likelihood chi-square. *Nucl. Instrum. Meth. A* **961**, 163677 (2020). <https://doi.org/10.1016/j.nima.2020.163677>
49. T. Akiba, S. Sano, T. Yanase et al., Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, (2019), pp. 2623–2631. <https://doi.org/10.1145/3292500.3330701>
50. PyTorch, ReduceLROnPlateau class. https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html Accessed 30 June (2024)
51. I. Loshchilov F. Hutter, Decoupled weight decay regularization, In: International Conference on Learning Representations, (2017). <https://doi.org/10.48550/arXiv.1711.05101>
52. D.P. Kingma, J. Ba, Adam: a method for stochastic optimization. (2014). <https://doi.org/10.48550/arXiv.1412.6980>
53. R. Brun, F. Rademakers, ROOT: an object oriented data analysis framework. *Nucl. Instrum. Meth. A* **389**, 81–86 (1997). [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X)
54. A. Paszke, S. Gross, F. Massa et al., PyTorch: an imperative style, high-performance deep learning library. <https://doi.org/10.48550/arXiv.1912.01703>
55. H. Kim, pytorch_optimizer: optimizer & lr scheduler & loss function collections in PyTorch. https://github.com/kozistr/pytorch_optimizer Accessed 30 June (2024)

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.