ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202603.00136

From Sparse Measurements to Dense DDX: Bayesian Tensor Modeling of Proton-^{nat}Pb for ADS Source Terms

Authors: LIU, Miss Huizi, HUANG, Mr. Yingge, Su, Dr. Jun, WANG, Miss Hui, XIAO, Mr. Erxi, HUANG, Miss Jiali, FENG, Miss Yujie, GU, Mr. Fuchang, CHEN, Mr. Qiafeng, Su, Dr. Jun

Date: 2026-03-24T12:03:48+00:00

Abstract

Accurate spallation-neutron source terms are essential for accelerator-driven systems (ADS), yet double-differential cross-section (DDX) measurements remain sparse—particularly for proton-^{nat}Pb, a canonical ADS target. We present a data-driven pathway from sparse measurements to dense DDX using a Bayesian tensor model together with a physics-consistent interpolation scheme tailored for ADS source-term generation. A total of 1{,}727 DDX points for proton-^{nat}Pb, compiled from EXFOR and literature, spanning eight incident energies ($10_{\text{MeV-3}}$GeV) and seven angles (7.5°-150°), are used in full to fit the tensor model. Under the optimal configuration, the model shows strong in-sample reproduction on a log scale: 96.3% of points lie within ±20% of experiment, 80.6% within ±10%, and 52.46% within ±5%. For out-of-measurement checks, we compare predictions with the Bertini intranuclear-cascade model in Geant4 (and BERT_{HP} where available) on common discrete grids: agreement is close below 10~MeV; in the 20-100~MeV band, where BERT/BERT_{HP} are known to underestimate data, our predictions remain physically plausible. To deliver application-ready inputs, we construct a high-resolution database via bilinear interpolation in (E_p, θ) on self-similar energy slices $\kappa = \ln(E_n/E_p)$ (default $\Delta\kappa = 0.2$), evaluated on a regular grid with 1~MeV spacing in E_p and 0.5° in θ, with a no-extrapolation policy. The interpolants preserve evaporation-like low-energy behavior and forward-peaked high-energy emission and are consistent with Geant4 trends. The resulting proton-^{nat}Pb DDX database, covering the CiADS design point (500~MeV) and its neighborhood, is ready to be coupled to transport codes (e.g., OpenMC) for anisotropic source-term calculations and can be extended to other targets and reactions.

ChinaRxiv

Full Text

Preamble

From Sparse Measurements to Dense DDX: Bayesian Tensor Modeling of Proton–Pb for ADS Source Terms Huizi Liu, Yingge Huang, Hui Wang, Erxi Xiao, Jiali Huang, Yujie Feng, Fuchang Gu, Qiafeng Chen, and Jun Su 1, 2, 1 Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-sen University, Zhuhai 519082, China China Nuclear Data Center, China Institute of Atomic Energy, Beijing 102413, China Accurate spallation-neutron source terms are essential for accelerator-driven systems (ADS), yet dou- ble-differential cross-section (DDX) measurements remain sparse—particularly for proton–Pb, a canonical ADS target. We present a data-driven pathway from sparse measurements to dense DDX using a Bayesian tensor model together with a physics-consistent interpolation scheme tailored for ADS source-term generation.

A total of 1,727 DDX points for proton–Pb, compiled from EXFOR and literature, spanning eight incident energies (10 MeV–3 GeV) and seven angles (7.5), are used in full to fit the tensor model. Under the optimal configuration, the model shows strong in-sample reproduction on a log scale. For out-of-measurement checks, we compare predictions with the Bertini intranuclear-cascade model in Geant4 (and BERT HP where available) on common discrete grids: agreement is close below 10 MeV; in the 20-100 MeV band, where BERT/BERT HP tend to underestimate data, our predictions remain physically plausible. To deliver applica- tion-ready inputs, we construct a high-resolution database via bilinear interpolation in on self-similar energy slices = ln(, evaluated on a regular grid with 1 MeV spacing in and 0.5 , with a no-extrapolation policy. The interpolants preserve evaporation-like low-energy behavior and forward-peaked high-energy emission and are consistent with Geant4 trends. The resulting proton–Pb DDX database, cov- ering the CiADS design point (500 MeV) and its neighborhood, is ready to be coupled to transport codes (e.g., OpenMC) for anisotropic source-term calculations and can be extended to other targets and reactions.

Keywords

Spallation neutrons; Double-differential cross sections (DDX); Bayesian tensor model; Proton–Pb; ADS source terms

INTRODUCTION

Nuclear energy offers notable advantages, including high energy density and stable operational performance; however, the management of long-lived high-level radioactive waste remains a major challenge [ ]. The accelerator-driven system (ADS) has emerged as a promising route by enabling the transmutation of long-lived radionuclides into short-lived or stable isotopes, thus reducing both radioactivity and volume [ ]. Compared with conventional critical reactors, ADS can exhibit favorable neutron economy and transmutation performance in appropriate designs [ADS utilizes high-energy protons to bombard a spallation target, generating fast neutrons that drive the subcritical core to sustain transmutation reactions [ ]. Consequently, the accuracy of the external neutron source term is of critical importance.

Extensive studies have demonstrated that the neutron source term significantly influences both the computational accuracy and operational performance of the ADS reactor. For instance, Jiang Xiao-feng et al. [ ] employed the ANIANDOT4.2-ORIGEN2 code to analyze the IAEA ADS benchmark problem and concluded that the external neutron intensity must be dynamically adjusted in synchronization with fuel burnup to maintain the reactor' s rated power output. Similarly, Cao Jian et al. [ ] utilized the MCNP code to calculate transmutation reaction rates for nine types of minor actinides (MA) and long-lived fission products (LLFP), revealing that higher external neutron energies lead to increased transmutation rates across all nuclides. Furthermore, Pan Dong-mei et al. [ ] conducted neutron physics simulations of multi-region ADS cores using the VisualBUS4.2 code and found that both the number and spatial configuration of external neutron sources markedly affect the effective multiplication factor ( ) and the core' s power distribution. Collectively, these findings underscore that the precision of the neutron source Supported by the National Natural Science Foundation of China under Grant No. 12475136, and the Guangdong Basic and Applied Basic Research Foundation No. 2024A1515012310.

term is fundamentally linked to key reactor characteristics, including core neutronics behavior, power regulation, and safety evaluation.

In practice, source terms in target–core coupling simulations are often simplified. For example, the OECD/NEA ADS bench- mark adopts ring-shaped volume sources, a broad HETC-based spectrum, and isotropic emission [ ]; the optimization design of an ADS transmutation reactor sometimes use Maxwellian fission spectra with isotropy [ ]. Such approximations intro- duce two uncertainties. First, the fission neutron spectrum differs substantially from the actual energy distribution of spallation neutrons; spallation reactions can generate high-energy neutrons with energies approaching that of the incident protons [Second, spallation neutrons exhibit pronounced forward-peaked angular distribution, with high-energy neutron yield decreasing significantly as the emission angle increases [ ]. Zhao Zijia et al. [ ] further demonstrated in their neutron physics analysis of an ADS spallation target that the assumption of isotropy underestimates the contribution of forward-directed neutrons, thereby affecting predictions of and core power distribution. Therefore, fission-spectrum and isotropic assumptions can lead to biased results and are not fully justified for accurate source modeling.

A key reason for these approximations is the scarcity of double-differential cross-section (DDX) measurements for spallation neutrons. Although EXFOR includes DDX for proton-induced spallation on about 14 target materials (from light to heavy nuclei) at incident energies between 63 MeV and 3 GeV and angles between , the coverage is incomplete and lacks systematic consistency [ ]. In particular, for natural lead, which is a key ADS target material, available measurements exist only at a few discrete values [ ]. Major evaluated libraries (ENDF/B, JENDL, JEFF) provide only partial isotope-specific data at select energies (e.g., 150, 200, 1000, 3000 MeV), without a comprehensive evaluation for natural lead.

Given that the China Initiative ADS (CiADS) plans a 500 MeV, 5 mA superconducting linac [ ], filling these data gaps is important for design and optimization.

Parallel to data efforts, spallation models have been benchmarked extensively. In 2008, the International Atomic Energy Agency (IAEA) launched a global spallation-model benchmark campaign, which compared predictions of mainstream models (e.g., CEM, INCL++, Geant4-Bertini, PHITS, etc.) [ ]. The theoretical foundations of these models have been contin- uously developed over the past two decades. The Li‘ege intranuclear cascade (INCL) model, comprehensively presented by Boudard et al. [ ] and later extended to INCL4.6 with improved cluster production and low-energy capabilities [ ], has be- come one of the widely used spallation models. For quantum molecular dynamics approaches, the JQMD model implemented in PHITS and its improved variants [ ] have been applied to proton-induced spallation reactions, with studies indicating that the choice of effective interaction can influence the low-energy part of neutron spectra. Recent developments have also pursued microscopic improvements, such as incorporating in-medium nucleon–nucleon scattering cross sections based on Brueckner–Hartree–Fock calculations [The first results of the IAEA benchmark were released in 2009, followed by comprehensive analyses by David et al. [that compared model predictions for neutron DDX. Despite progress, it has been reported that noticeable discrepancies can remain in the intermediate-energy range (e.g., 20–100 MeV) and in forward-angle high-energy emissions. Subsequent exper- iments further revealed model-specific deficiencies and database inconsistencies. For example, Iwamoto et al. [ ] measured neutron DDX for 103 MeV protons on natPb and Bi and compared with PHITS-, MCNP-, and INCL++-based calcula- tions and JENDL-5; while some settings showed good overall agreement, deficiencies were identified for specific models and kinematics. Sun et al. [ ], motivated by CiADS design requirements, used Geant4 (BERT HP, BIC HP, INCLXX HP) and FLUKA to compute neutron DDX for 256 MeV protons on various targets, showing that model performance remains strongly energy- and angle-dependent. These limitations motivate higher-fidelity approaches to complement existing models and data.

Recently, data-driven methods have shown promise in nuclear data applications [ ]. In the specific context of spallation- neutron spectra, Schnabel [ ] employed a sparse Gaussian process—a non-parametric Bayesian method—to estimate model bias in INCL/ABLA predictions of double-differential neutron spectra over a broad set of nuclides. Within this context, our group has adapted the Bayesian Gaussian CP (BGCP) tensor model to neutron-induced fission product yields [ ], (n,2n) cross sections of LLFPs [ ], and elastic proton-scattering differential cross sections [ ], providing a foundation to extend tensor-based modeling to spallation-neutron DDX.

This work is part of a broader project to build a high-energy nucleon–nucleus database for CiADS reactor analysis. The DDX of neutrons from proton–reactions constitutes a necessary subset for source-term construction. In this study, we focus on proton-induced spallation on natural lead. Using the tensor model, we predict and reconstruct DDX across eight

incident energies (10 MeV–3 GeV) and seven emission angles ( ), based on 1,727 experimental points from EXFOR and literature [ ]. To provide application-ready inputs, we further construct a high-resolution database by performing interpolation in on fixed self-similar energy slices = ln(. For external checks where measurements are absent, we compare with BERT as implemented in Geant4 (and BERT HP where available), using a common binning and the same DDX conversion from counts.

THEORETICAL FRAMEWORK Tensor model We adopt a Bayesian Gaussian CP (BGCP) tensor model to complete and predict multiway nuclear data with missing en- tries. The Bayesian approach to matrix factorization was initially developed by Salakhutdinov and Mnih [ ], who proposed a Bayesian probabilistic matrix factorization model solved via Markov chain Monte Carlo. Subsequently, Xiong et al. [extended this framework by incorporating Bayesian inference into tensor decomposition while accounting for temporal depen- dencies. Building upon these foundations, Chen et al. [ ] further improved the methodology by considering spatial structures, developing the Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) tensor decomposition model for spatiotemporal traffic data imputation. Wang et al. [ ] extended the BGCP approach to the nuclear data domain by constructing weighted evolu- tion networks to investigate uncertainty issues in datasets and data points. Within our research group, this tensor modeling framework has been successfully applied to evaluate fission product yields [ ] and to predict (n,2n) reaction cross-sections of long-lived fission products [ ], as well as DDX for elastic proton scattering [ ]. Building upon these methodological foundations, the present work further develops the tensor model for predicting and completing the DDX of spallation neutrons produced by proton-induced reactions on natural lead targets.

Tensorization of DDX data. The DDX for spallation neutrons produced by proton-induced reactions on natural lead depends on three independent variables: incident proton energy, emission angle, and outgoing neutron energy. To reduce scale variation across energy decades, the outgoing neutron energy is logarithmically transformed before binning. The resulting data are arranged into a third-order tensor , where the element corresponds to incident energy bin , angle bin , and outgoing neutron-energy bin . Measured entries from EXFOR and the literature provide a subset of observations indexed by the set i, j, k, p , where enumerates replicate measurements (e.g., different experiments) for the same i, j, k CP factorization. be the CP rank and be the factor matrices. The reconstructed tensor is

S =

so that the model mean for an entry is

ˆ σ ijk =

Here "" denotes the outer product. Likelihood. For each observed replicate i, j, k, p , we assume a Gaussian noise model where denotes a Gaussian distribution, is the observation precision. This formulation assumes that the experimental errors are uncorrelated and homoscedastic (i.e., constant variance across all measurements), which is a simplifying assumption of the current approach. If reported experimental uncertainties are available, the model can use entry-specific precisions place of a global

Priors. Rows of the factor matrices are given independent Gaussian priors with shared hyperparameters where each is a precision matrix. Conjugate Gamma priors are placed on the precisions, , and Nor- mal–Wishart hyperpriors may be used for Gibbs sampling.

Thanks to conjugacy, the full conditionals are Gaussian or Gamma in closed form. For example,

conditioning on D , E , τ ϵ , the posterior for z i is 109

Λ (z) i = Λ (z) + τ ϵ

j,k,p

µ (z) i = Λ (z) i − 1

j,k,p Here "" is the Hadamard product, is 1 for index (i,j,k,p) with measurement and 0 for the case without measurement.

Analogous expressions hold for by cycling the tensor modes. Similarly, using the likelihood with the Gaussian form and the prior term of with the Gamma form, it will give the posterior with the Gamma distribution. Let

(i,j,k,p) I [(i, j, k, p) Ω] . The conditional posterior of τ ϵ is 114

i,j,k,p We alternate sampling of ; after burn-in, posterior means (or medians) of are reported together with credible intervals.

Implementation notes. We provide the bin definitions for , θ, E , the CP rank selection procedure, prior hyper- parameters, number of chains/iterations/burn-in, random seeds, and units (e.g. mb sr) to ensure full reproducibility.

Interpolation method The tensor model provides discrete DDX values on a grid indexed by (incident proton energy (emission angle), and (outgoing neutron energy sampled on the log grid). For applications requiring continuous queries, we construct a three-dimensional linear interpolation scheme with an energy-axis reparametrization that respects the kinematic scaling.

Energy-axis reparametrization. For each incident-energy index , we map the energy grid to a self- similar coordinate This stretch/compression along the outgoing-energy axis removes the artificial under-coverage that occurs if one interpolates directly on varies widely.

Stage 1: 1D linear interpolation along For each fixed pair , the discrete points define a piecewise-linear function . Given a query , compute = ln(. Let be such that

$\kappa * \leq \kappa$ ik 1 . Define 132

σ ij ($\kappa *$) = (1 − α i) σ ijk 0 + α i σ ijk 1 ,

= 113 MeV and emission angle , plotted against the neutron energy . (b) The same data plotted against the self-similar variable = ln(. The transformation aligns the spectra on a common support, enabling interpolation with physically correct energy bounds.

Stage 2: 1D linear interpolation over incident energy.

Locate and set) = (1 Stage 3: 1D linear interpolation over angle.

For fixed , locate and set) = (1 This sequential construction is equivalent to a tri-linear interpolant in , θ, κ , with = ln(. Note that the cross section itself is never log-transformed.

By reparametrizing the energy axis with the self-similar coordinate = ln((i.e., = ln( ), spectra at different are aligned on a common support so that the query is simultaneously bracketed by both neighboring datasets at the same . This alignment enables a straightforward tri-linear interpolation in , θ, κ without extrapolation. Figure illustrates why we perform interpolation in the self-similarity coordinate , using the incident-energy dimension as an example.

Shown are the tensor-completed double-differential cross sections at an emission angle of for three adjacent proton energies (= 113 , 256, and 597 MeV), which form the discrete basis for subsequent interpolation. In the original neutron- energy coordinate , each spectrum has a different kinematic endpoint (increases with). As a result, a direct linear interpolation between two incident energies must be restricted to their overlapping range; for example, interpolating between 256 and 597 MeV would truncate the interpolated spectrum at MeV, thereby underestimating the physically allowed high-energy tail at intermediate . By contrast, transforming to = ln(maps spectra at different a common support and improves their alignment, enabling physically consistent interpolation while preserving the correct neutron-energy range. The interpolation itself uses standard geometric linear weights (bilinear in at fixed) and does not introduce additional empirical weighting. In the current work, uncertainty from the tensor model is not propagated through the interpolation step; the interpolated database is reported as point estimates.

Geant4 baseline with BERT We use Geant4 as an independent, physics-based baseline to benchmark the tensor-model predictions. The standalone source code of the Bertini-style intranuclear cascade (BERT) is not publicly available to us; therefore, we run the reference implementation shipped with Geant4 via a BERT-enabled hadronic physics list and denote the outcomes as "BERT (Geant4)" [BERT (Geant4) describes proton–nucleus spallation through an intranuclear cascade followed by pre-equilibrium, evaporation and fission de-excitation (if relevant) [ ]. Elementary reaction cross sections and scattering channels are provided by the Geant4 hadronic package, which allows simulations up to the few-GeV regime. No tuning to experimental data is performed in this work; default model parameters are used.

The scattering neutrons measured in the experiment exhibit specific angular and energy distributions, which are primarily characterized and recorded using DDX. The DDX quantifies the probability that neutrons are emitted into a unit solid angle and within a unit energy interval following the interaction of protons with a target nucleus at a given incident energy. It is a normalized physical quantity representing the average neutron production cross-section per incident proton interacting with a single target nucleus. In contrast, GEANT4 simulation calculations yield the energy and angular information of neutrons emitted from the surface of the scattering target. To derive the DDX of these scattering neutrons, additional post-processing calculations must be performed according to the following formula.

d 2 σ d E n dΩ= 10 24 $\cdot$ N n N p $\cdot$ N $\cdot$ x $\cdot$ Δ E n $\cdot$ $\Delta\Omega$

dΩ= sin(Among them, the left side of the equation represents the DDX in units of b/Sr/MeV, while the right side is derived from cal- culations that convert the number of emitted spallation neutrons into the corresponding DDX. The factor 10 is introduced for unit conversion, transforming the calculated cross section from cm (obtained directly from the neutron count in the simulation) to barns (1 barn = 10 denotes the number of incident protons,) is the number density of target nuclei, (cm) is the thickness of the thin target, and represents the number of neutrons emitted within a solid angle interval in a given scattering direction and within an energy interval (MeV) around a specific neutron energy.

Where experimental DDX data exist, we compare tensor-model predictions and BERT (Geant4) to measurements. In en- ergy–angle regions without measurements, BERT (Geant4) serves as an out-of-sample physics reference to assess the plausi- bility of tensor-model predictions and interpolation within the admissible domain.

RESULTS AND DISCUSSION Selection of tensor-model parameters We examine three hyperparameters of the tensor model: the CP rank , the number of Gibbs iterations , and the logarithmic energy-bin width used to discretize the outgoing-neutron energy after the transformation (unless otherwise noted).

A smaller yields more energy bins and denser spectra but also increases sparsity of the tensor and computational cost.

Model accuracy is assessed by the root-mean-square error on a logarithmic scale,

RMSE log =

where is the number of experimental data points, avoids taking the log of zero (not active for our datasets), and all observed points are equally weighted. RMSE is computed over the available experimental points by comparing predictions with measurements. vs. rank with = 100 decreases with and saturates for Increasing from 30 to 50 only improves marginally (0.1418 0.1406) while roughly doubling runtime; therefore we

adopt r = 30 as a practical compromise between fidelity and cost. 191

vs. the total number of Gibbs iterations (with). RMSE drops rapidly and stabilizes after . We use = 100 to ensure stable posterior means; burn-in is included in increases with . When , the improvement is modest, but for the degradation accelerates. Reducing can effectively lower the RMSE, but the tensor model will require more scattered

$d(\ln E_n) = 0.1, N = 100$

$d(\ln E_n) = 0.1, r = 30$

neutron energy points to be predicted and completed, and the prediction curve is prone to overfitting. If is too large, it will lead to a large root mean square error of the predicted values, and the reproduction effect of the prediction curve on the experimental values will deteriorate. Therefore, we need to compare and analyze the prediction curves of the tensor model at different values to find the value that can better reproduce the experimental values without overfitting. = 256 MeV under = 100 . The uncertainty of the tensor-model predictions is quantified by 95% credible intervals derived from posterior samples (shown as error bars where applicable). the curve shows oscillatory behavior in regions without measurements (both low- and high-energy tails), indicating sensitivity to sparse bins despite a slightly smaller the spectrum is over-smoothed and the high-energy peak near the beam-energy region is severely attenuated. best reproduces both the global trend and the peak shape while keeping the computation moderate.

Balancing accuracy (Fig.), spectral-shape fidelity (Fig.), and runtime, we adopt = 100 , and as the default configuration for subsequent analyses.

$E_p = 256$ MeV, $= 7.5$,

(a) $d(\ln E_n) = 0.05$

= 256 MeV protons on natPb at the emission angle under different discretization intervals in the self-similarity coordinate. The BGCP tensor-model parameters are = 100 . Panels (a)-(d) correspond to , respectively. Error bars (where shown) indicate the 95% credible interval estimated from posterior samples of the BGCP tensor model (computed in log-space and back-transformed by exponentiation). , (b) incident proton energies , and (c) neutron energies Reproduction of experimental DDX with the tensor model In this study, 1 727 experimental DDX data points for spallation neutrons from proton-induced reactions on a natural lead target were used in full to fit the tensor model. We did not split the dataset into training/validation subsets due to the sparsity and tensor-completion nature of the problem. Accordingly, Figures mainly quantify the model' s reproduction on measured kinematics, while physics-based credibility checks in data-sparse kinematics are provided through comparisons with indepen- dent spallation-model calculations on common discrete grids (Figs.). The data span eight incident proton energies from

E p = 10 to 3 000 MeV and seven emission angles from 7 . 5 ∘ to 150 ∘ . 215

, (b) incident energy , and (c) neutron energy . A pronounced concentration appears at , with = 256 MeV also well sampled; along the energy axis, most measurements lie in MeV, whereas the high-energy tail accounts for only 16.2% of all points (Fig. (c)). These coverage patterns guide expectations for the reproduction quality: regions with more measurements should be better constrained.

E p = 5 9 7 M e V , 1 0

(a) = 7 . 5

E p = 2 5 6 M e V , 1 0

E p = 1 1 3 M e V , 1 0

E p = 1 0 3 M e V , 1 0

E p = 1 0 M e V , 1 0

E p = 5 9 7 M e V , 1 0

(d) = 9 0

E p = 2 5 6 M e V , 1 0

E p = 1 1 3 M e V , 1 0

E p = 1 0 M e V , 1 0

E p = 1 0 3 M e V , 1 0

absence of explicit physical equations.

E p = 5 9 7 M e V , 1 0

(c) = 6 0

E p = 2 5 6 M e V , 1 0

E p = 1 1 3 M e V , 1 0

E p = 1 0 3 M e V , 1 0

E p = 1 0 M e V , 1 0

E p = 5 9 7 M e V , 1 0

(f) = 1 5 0

E p = 2 5 6 M e V , 1 0

E p = 1 1 3 M e V , 1 0

E p = 1 0 3 M e V , 1 0

E p = 1 0 M e V , 1 0

MeV. Tensor-model (TM) parameters: = 100 , and (logarithmic step in the self-similarity coordinate). Panels (a)-(f) correspond to , respectively.

Solid lines: kinematics with experimental data; dashed lines: without data. Error bars (where shown) indicate the 95% credible interval estimated from posterior samples of the BGCP tensor model (computed in log-space and back-transformed by exponentiation).

With the optimal hyperparameters fixed from the previous section (, and the logarithmic energy step), Figure shows tensor-model DDX curves for natural lead at multiple (10–597 MeV) and). Solid lines indicate kine- matics where experimental measurements exist; dashed lines denote settings without corresponding data.

Overall, the model reproduces measurements well. The uncertainty of TM predictions is quantified by 95% credible intervals derived from posterior samples (shown as error bars where applicable). For example in Fig. (c) (= 256 MeV), the predicted curves are smooth and closely follow both the magnitude and shape of the data across the full range. However, when the experimental data points are very sparse, the smoothness of the tensor model prediction curve decreases. In Fig. = 103 MeV, the high-energy tail fluctuates and the mid-energy segment does not fully pass through the data—consistent with the lower local sampling at these energies/angle. Even when the local measurements at MeV are sparse, the prediction at = 120 is smoother and better centered on the data (Fig. (e)). This is attributable to the collaborative filtering mechanism: the overall measurement coverage at (Fig. (a)) is more than twice that at so information shared across within the same angle stabilizes the reconstruction. This demonstrates that the collaborative filtering algorithm can effectively leverage relevant experimental information to capture intrinsic physical trends, even in the Across Fig. (c)-(f), as increases and coverage improves (Fig. (b)), both smoothness and shape fidelity are noticeably enhanced. This further confirms the positive correlation between model performance and the size of the available sample.

To assess agreement uniformly across orders of magnitude, we analyze the prediction-to-experiment ratio tensor mb sr where prevents division by zero in vanishing tails; points with are masked in ratio plots. , (b) , and (c) over all 1 727 points. Most ratios cluster around 1 and lie within A ratio greater than 1 indicates overestimation by the model, while a ratio below 1 indicates underestimation. The spread tends

tensor versus (a) , (b) , and (c) , aggregated over 1 727 experimental points (eight values from 10–3 000 MeV and seven angles Protocol: a small mb sr is added for numerical stability; points with are masked. to be smaller at angles with richer coverage (e.g.,), and broader where the coverage is poorer when combined with high-energy tails. Larger deviations concentrate at MeV (Fig. (b)), consistent with Fig. (c) where only 16.2% of points lie in this range. At well-sampled (e.g., 256, 597 MeV) the ratios are tightly distributed, while at sparsely sampled (e.g., 103, 113 MeV) the spread increases—matching the behavior seen in Fig. (within), 80.6% within), and 52.46% within). The distribution centers near with modest tails; the outlying bins are mainly associated with the sparsely sampled high-energy region MeV noted in Fig. (c). These statistics, together with the qualitative overlays in Fig. , indicate strong in-sample agreement and support the reliability of the tensor model in reproducing DDX across diverse kinematics.

Analysis of the prediction effect of the tensor model on the cross-section To date, the discussion has primarily focused on the model' s ability to reproduce experimental data and its associated per- formance. However, a comprehensive evaluation of the model's effectiveness requires further assessment of its accuracy in predicting DDX in the absence of corresponding experimental data—a key aspect for evaluating its generalization and predic- tive capabilities.

Protocol: same as Fig. ; mask Typically, model predictions are validated by comparison with data from nuclear databases. However, in this study, for the DDX predicted by the tensor model in the absence of corresponding experimental measurements, no relevant evaluated data were available in the five major evaluated nuclear data libraries (ENDF/B, JEFF, ROSFOND, JENDL, and CENDL). Therefore, we employ the Bertini intranuclear cascade model as implemented in Geant4 to simulate the spallation reaction, using its results as a physics-based reference for comparative analysis. Where applicable, we also refer to results from the high-precision variant reported in the literature [BERT HP, Ref. [ ]] for additional cross-checks.

Firstly, it is necessary to ensure that the Geant4 program is used reasonably and reliably for simulating proton-induced spallation reactions. To verify the reliability of the Geant4 setup, this work utilized BERT (Geant4) to calculate the DDX of spallation neutrons at multiple emission angles for proton bombardment of a thin natural-lead target at incident energies of 113 MeV, 256 MeV, and 597 MeV. These results were compared with the corresponding experimental data and with the HP results in the literature [ ]. The number of incident protons in each Geant4 run was (to control statistical fluctuations); the thin-target parameters are given in Table . The conversion from neutron counts to DDX follows Eq. (and PHITS/JQMD calculations from relevant literature [ ], as well as the predicted values of the tensor model at the , θ, E points. As analyzed in the previous section, the tensor-model curves closely match the experimental data.

The overall trends of all physics-based models (BERT, BERT HP, FLUKA, PHITS/INCL4.6, PHITS/JQMD) are generally consistent with the measurements, but their reproduction can differ noticeably from measurements in the

20–100 MeV band and at some backward angles. Since the tensor model is directly constrained by experimental points, it provides a data-anchored reference on measured kinematics, while the physics-based models serve as independent baselines for cross-checks in data- Specifically, for multiple angles at MeV the BERT (Geant4) calculations agree well with experiments and reproduce the small-angle peak (e.g.,). In the energy range MeV, where experimental data are relatively sparse, the various models exhibit considerable scatter. BERT (Geant4) tends to underestimate the cross section. For example, at = 113 MeV and , the deficit relative to data can reach . As the incident energy increases, the deviation

] and various spallation model calculations at (a) = 113 MeV, (b) = 256 MeV, and (c) = 597 MeV. The results shown include: BERT (Geant4) from this work; BERT HP from Ref. [ ]; FLUKA from Ref. [ ] (at angles = 256 MeV) and from Ref. [ ] (at angles = 597 MeV); PHITS/INCL4.6 and PHITS/JQMD from Ref. [ ] (at angles = 256 MeV).

Protocol: common discrete , θ, E grids; no linear interpolation; Geant4 counts converted to DDX via Eq. (generally decreases. Furthermore, other physics-based models show different behaviors. For instance, at = 256 (Fig. (b)), the FLUKA and PHITS/INCL4.6 results tend to be higher than the tensor model predictions, while BERT and HP are lower. At , FLUKA shows better agreement with experimental data in some cases compared to the BERT models. This inter-model spread underscores the challenge of accurately modeling spallation reactions in this energy regime and highlights the value of data-driven approaches as complementary tools.

For the high-energy tail (MeV), where measurements are particularly sparse and model dependence is expected to increase, Fig. shows that our tensor-model predictions remain physically plausible and are consistent with the envelope spanned by the independent spallation calculations (BERT/BERT HP, FLUKA, and PHITS where available) on the same dis- crete grids. In particular, the forward-peaked behavior at small angles and the comparatively reduced yield at backward angles are preserved. We therefore use these cross-model comparisons as a practical physics-based credibility check for the data-sparse tail, while noting that additional dedicated measurements in this region would be valuable to further constrain both data-driven

d (l n E n) = 0 . 1 r = 3 0 N = 1 0 0

d (l n E n) = 0 . 1 r = 3 0 N = 1 0 0

d (l n E n) = 0 . 1 r = 3 0 N = 1 0 0

= 113 Protocol: same discrete grids as Fig. ; no interpolation; Geant4 DDX conversion via Eq. (and physics-based approaches.

Moreover, as shown in Fig. (b), the BERT (Geant4) results in this work and the BERT HP results in Ref. [ ] are in good overall agreement. Below 20 MeV, their reproduction of the experimental data is similar. At very large angles (e.g.,), both models underestimate the 20–100 MeV region, with BERT HP slightly closer to data. Taken together, these checks support the use of BERT (Geant4) as a reasonable physics baseline for out-of-measurement comparisons, while recognizing that it is not a substitute for experimental truth.

When the incident proton energies are 113, 256, and 597 MeV, experimental data are missing at some angles (e.g.,). The tensor model predicts and completes the missing DDX on the same discrete grids. Figure compares the tensor-model predictions (no corresponding experimental points) with BERT (Geant4). Across incident energies and angles, the overall trends are consistent. In the low-energy region MeV the two are often very close; in MeV moderate deviations appear. For instance, at = 113 MeV the tensor model is lower than Geant4 in MeV and slightly higher in MeV, with the spectral falloff occurring earlier. As increases, these discrepancies decrease. This behavior is consistent with the three-way comparison in Fig. where experimental data exist. Overall, these results suggest that the tensor-model predictions in data-sparse regions are physically plausible relative to an independent baseline.

d (l n (E n / E p)) = 0 . 2

they agree with available measurements at the same kinematics. peak weakens progressively.

(c) = 3 0

(f) = 9 0

(i) = 1 5 0

on fixed = ln(slices with . Panels (a)–(i) show representative angles . Common -grid ( ); bilinear in ; no extrapolation; when compared on an grid, mapping at each with interpolation in Interpolation results and analysis This section uses the tensor-model DDX for spallation neutrons on natural Pb at the first five incident energies (10–597 MeV) and seven emission angles (7.5) as inputs. A interpolation is then applied over proton incident energy and emission angle on fixed energy slices defined by the self-similar coordinate = ln(. Unless otherwise noted, the -grid spacing is . The resulting high-resolution database is evaluated on a regular grid with step of 1 MeV and step of ; along the energy dimension, values are obtained at prescribed (and, when reporting on a uniform grid, by mapping at the target and linear interpolation in). No extrapolation is performed outside the convex hull of at fixed . Only nine representative cases are displayed for illustration.

Solid segments indicate kinematics with experimental constraints; dashed segments denote interpolated values where no mea- surements exist. Overall, the curves are smooth across without non-physical oscillations or negative cross sections, and The interpolants also follow expected physics trends. In the low-energy region (MeV), all curves show evaporation- like spectra with rapidly decreasing DDX. At higher , forward-peaked features emerge at small angles (e.g., Fig. (a)), consistent with anisotropic emission of high-energy neutrons in spallation. As increases, the high-energy tail flattens and the Because the energy axis is parameterized by = ln(with a finite , the minimum covered after mapping back depends on . Consequently, some curves may not reach exactly MeV (typical gap MeV). Given that most experiments report from MeV, the impact is limited. If needed for downstream use, the gap can be removed by (i)

d (l n (E n / E p)) = 0 . 2

d (l n (E n / E p)) = 0 . 2

d (l n (E n / E p)) = 0 . 2

= 113 MeV for six angles (10 extending the -grid toward lower values during preprocessing, or (ii) linearly (or exponentially) extending the spectrum from the first bin above 1 MeV down to 1 MeV, with the filled segment flagged.

The DDX database produced by the tensor model plus bilinear interpolation provides a fine resolution (1 MeV in , 0.5 ; and either or a user-specified grid obtained via -linear evaluation). As the five major evaluated libraries lack comparable DDX at the same kinematics/resolution, we also use the Geant4 BERT model as an independent physics reference.) to 140) for proton incident energies of 113 MeV, 256 MeV, and 597 MeV comparing with the calculations from the BERT (Geant4) model. It is shown that the interpolated curves and BERT (Geant4) results share consistent global trends: high DDX at low with rapid falloff, and a diminishing cascade tail and peak with increasing . However, at specific angles and within certain energy ranges, the results obtained from the interpolation method exhibit notable discrepancies compared to those calculated by the BERT (Geant4) model. At MeV, agreement is generally excellent across angles. At MeV, BERT (Geant4) tends to be lower than the interpolation, with the difference growing toward larger and diminishing at higher . This is consistent with Fig.

Ref. [ ], where BERT (Geant4) and BERT HP systematically underestimate data in the 20–100 MeV region. In the absence of experimental DDX at these kinematics, the interpolated cross sections remain physically plausible and show consistent trends

with Bertini(Geant4) calculations. 338

SUMMARY

The ADS system is one of the key facilities for managing high-level radioactive nuclear waste. The neutron physics design of its core depends critically on accurate neutron source term data. However, experimental data on the DDX of spallation neutrons remain severely limited, leading to the widespread use of isotropic and fission spectrum approximations in simula-

tions—approximations that may introduce significant errors. 343

To address this issue, this work employs a Bayesian Gaussian CP (BGCP) tensor model, leveraging data from the EXFOR database and experimental measurements in the literature [ ], to systematically predict and complete the DDX of spallation neutrons for a natural lead target over = 10 MeV

3 GeV

. Under optimal hyperparameters for the tensor model—CP rank , total iterations = 100 , and logarithmic outgoing-energy discretization —the model demonstrates excellent in-sample reproduction: 96.3% of points are within of experiment, 80.6% within , and 52.46% within . Where experimental DDX are unavailable, the predictions are checked against an independent physics baseline, BERT as implemented in Geant4. The trends are consistent, especially at

10 MeV

; in the 20–100 MeV range, Fig. shows that BERT (Geant4) tends to underestimate the measured points, while our predictions remain physically plausible.

Building on the tensor model, we construct a high-resolution DDX database via interpolation over on fixed self- similar energy slices = ln((default). The database is provided on a regular grid with step of

1 MeV

step of ; along the energy dimension the values are evaluated at prescribed and, when needed, mapped back to grid at the target . No extrapolation is performed outside the convex hull in , which avoids non-physical oscillations and negative cross sections. Because the minimum after mapping depends on , a small incomplete segment (

0 MeV

) may appear below

1 MeV

; this has limited practical impact since most experiments report from

E n ≥

1 MeV

. 359

In summary, this study successfully applies the tensor-model approach to predict spallation-neutron DDX and delivers a validated, high-resolution database for natural lead covering proton energies from

597 MeV

(including the CiADS design point at = 500 MeV) and emission angles from . The method is extensible to other targets and reaction types.

For deployment in transport simulations, the database can be directly coupled to Monte Carlo codes (e.g., OpenMC) to provide

anisotropic, energy-angle-resolved source terms for subcritical assemblies. 364

Future work will: (i) incorporate reported experimental uncertainties as entry-wise precisions to account for heteroscedastic noise; (ii) propagate the tensor-model posterior through the interpolation to provide credible intervals on the interpolated DDX; (iii) extend target materials and incident-energy coverage; and (iv) integrate the database into CiADS analyses to quantify impacts on reactivity, power distribution, and transmutation efficiency.

BIBLIOGRAPHY S. Guo, Optimization design of top shield for China Initiative Accelerator Driven System based on genetic algorithm. Master thesis, Institute of Modern Physics, Chinese Academy of Sciences (2023).

H. A¨ıt Abderrahim, D. De Bruyn, G. Van den Eynde, R. Fernandez, Accelerator driven subcritical systems. In:

Encyclopedia of Nuclear Energy: Reference Module in Earth Systems and Environmental Sciences , 1st ed., pp. 1–21, Elsevier (2021).

A. Stanculescu, Accelerator driven systems (ADSs) for nuclear transmutation.

Ann. Nucl. Energy , 607–612 (2013). H. A¨ıt Abderrahim, D. De Bruyn, G. Van den Eynde, S. Michiels, Transmutation of high-level nuclear waste by means of accelerator driven system (ADS).

Wiley Interdiscip. Rev.: Energy Environ. (2015).

Q. Zhao, Design of the monitoring system for the temperature and neutrons in high power spallation for an ADS system. Doctoral dissertation, Institute of Modern Physics, Chinese Academy of Sciences (2017).

X.-F. Jiang, Z.-S. Xie, Transport-burnup code system and its application to IAEA ADS benchmark.

Nucl. Power Eng. (5), 6 (2004). J. Cao, Y.-Q. Shi, P. Xia et al., ADS transmutation research based on Venus 1#.

At. Energy Sci. Technol. (10), 1185–1188 (2012). 10.7538/yzk.2012.46.10.1185shu D. Pan, Study on the performance of neutron source of multiple beam ADS and fusion plasma. Doctoral dissertation, University of Science and Technology of China (2019).

A. M. A. Burner, Comparison calculations for an accelerator-driven minor actinide burner.

Nucl. Sci. Eng. (2), 186–200 (2002). Q.-L. Wei, A.-X. Wang, Y.-B. Liu et al., Monte Carlo simulation of ADS transmutation reactor coolant and fuel optimal arrangement.

Nucl. Electron. Detect. Technol. (8), 981–984 (2013).

N. V. Shetty, Study of particle transport in a high power spallation target for an accelerator-driven transmutation system. Doctoral dissertation, RWTH Aachen University (2013).

L. Song, Neutronic simulation of the ADS target-reactor coupled system using Geant4. Master' s thesis, Institute of Modern Physics, Chinese Academy of Sciences (2018).

Z.-J. Zhao, Neutronic analysis and design optimization for spallation target of ADS. Doctoral dissertation, University of Science and Technology of China (2014).

Ikeda, Nuclear relevant accelerator driven system.

Nucl. Technol. (2011). 10.1080/00223131.2002.10875029 N. Otuka, E. Dupont, V. Semkova et al., Towards a more complete and accurate experimental nuclear reaction data library (EX- FOR): international collaboration between nuclear reaction data centers (NRDC).

Nucl. Data Sheets , 272–276 (2014). M. M. Meier, W. B. Amian, C. A. Goulding et al., Differential neutron production cross sections for 256-MeV protons.

Nucl. Sci. Eng. , 289–298 (1992). H. Iwamoto et al., Neutron-production double-differential cross sections of natPb and 209Bi in proton-induced reactions near 100 MeV.

Nucl. Instrum. Methods B , 165107 (2023). W. B. Amian, R. C. Byrd, D. A. Clark et al., Differential neutron production cross sections for 597-MeV protons.

Nucl. Sci. Eng. 1–12 (1993). H. Xu, W.-L. Zhan, G.-Q. Xiao et al., HIAF and CiADS national major science and technology infrastructure: progress and prospect.

Nucl. Phys. Rev. (1), 60–66 (2024). D. Filges et al., Joint ICTP-IAEA advanced workshop on model codes for spallation reactions.

Proc. ICTP-IAEA Workshop , Trieste, R. Michel, Benchmark of spallation models 2009: production of residual nuclides.

Report GSI-2009-1 , GSI Darmstadt (2009). A. Boudard et al., Intranuclear cascade model for a comprehensive description of spallation reaction data.

Phys. Rev. C , 044615 (2002). A. Boudard et al., New potentialities of the Li‘ege intranuclear cascade model for reactions induced by nucleons and light charged particles.

Phys. Rev. C , 014606 (2013). L. Ou et al., Analysis of intermediate energy proton-induced spallation reactions by an improved quantum molecular dynamics plus statistical decay model.

J. Phys. G: Nucl. Part. Phys. , 827 (2007). Medium Effects of Nucleon-Nucleon Cross Section in Spallation Reactions Nucl. Phys. Rev. (3), 414-419 (2025). clPhysRev.42.2024084 J.-C. David et al., Benchmark of spallation models.

Prog. Nucl. Sci. Technol. , 942-947 (2011). S. Leray et al., Results from the IAEA benchmark of spallation models.

J. Korean Phys. Soc. (2), 791-796 (2011). 10.3938/jkps.59.791 J.-C. David, Spallation reactions: a successful interplay between modeling and applications.

Eur. Phys. J. A (6), 68 (2015). 10.1140/epja/i2015-15068-1 Q. Sun, Z.-Q. Chen, R. Han et al., GEANT4 and FLUKA calculations of 256 MeV proton-induced spallation neutron spectra.

Nucl. Phys. Rev. (1), 118-123 (2019). X. Dong and L. Geng, Machine learning method in study of nuclear charge radius.

At. Energy Sci. Technol. , 679 (2023). J. Li, T. Zhao, N. Ma et al., Application of neural network approach in nuclear mass.

At. Energy Sci. Technol. , 696 (2023). 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.

(2022). Z.-P. Gao, Y.-J. Wang, H.-L. L¨u et al., Machine learning the nuclear mass.

Nucl. Sci. Tech. , 109 (2021). Y. Tian, R. Xu, X. Tao et al., Analyzing fission cross section of U isotope by neural network method.

At. Energy Sci. Technol. , 805 (2023). C.-W. Ma, X.-B. Wei, X.-X. Chen et al., Precise machine learning models for fragment production in projectile fragmentation reactions using Bayesian neural networks.

Chin. Phys. C , 074104 (2022). Q.-F. Song, L. Zhu, J. Su, Target dependence of isotopic cross sections in the spallation reactions 238U + p, d and 9Be at 1 A GeV.

Chin. Phys. C , 074108 (2022). Y. Wang and Q. Li, Machine learning transforms the inference of the nuclear equation of state.

Front. Phys. , 64402 (2023). 10.1007/s11467-023-1313-3 Y. Wang, F. Li, Q. Li et al., Finding signatures of the nuclear symmetry energy in heavy-ion collisions with deep learning.

Phys. Lett. B , 136669 (2021). Y. Wang, Z. Gao, H. L¨u et al., Decoding the nuclear symmetry energy event-by-event in heavy-ion collisions with machine

ChinaRxiv 中X

learning.

Phys. Lett. B , 137508 (2022). P. Vicente-Valdez, L. Bernstein, M. Fratoni, Nuclear data evaluation augmented by machine learning.

Ann. Nucl. Energy , 108596 (2021). D. Neudecker, M. Grosskopf, M. Herman et al., Enhancing nuclear data validation analysis by using machine learning.

Nucl. Data Sheets , 36–60 (2020). W. He, Q. Li, Y. Ma et al., Machine learning in nuclear physics at low and intermediate energies.

Sci. China Phys. Mech. Astron. 282001 (2023). G. Schnabel, Estimating model bias over the complete nuclide chart with sparse Gaussian process at the example of INCL/ABLA and double-differential neutron spectra.

EPJ Nucl. Sci. Technol. , 33 (2018). Q. Song, L. Zhu, B. Cai et al., Image processing of isotope yield in neutron-induced fission.

Phys. Rev. C , 044609 (2023). 10.1103/PhysRevC.107.044609 Q.-F. Song, L. Zhu, H. Guo et al., Verification of neutron-induced fission product yields evaluated by a tensor decomposition model in transport-burnup simulations.

Nucl. Sci. Tech. , 32 (2023). J.-L. Huang et al., Prediction of (n, 2n) reaction cross-sections of long-lived fission products based on tensor model.

Nucl. Sci. Tech. 184 (2024). H. Wang, J.-L. Huang, J. Su, Studying differential cross section for elastic proton scattering by a tensor model.

Prog. Nucl. Energy 104891 (2023). R. Salakhutdinov and A. Mnih, Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In Proceedings of the 25th International Conference on Machine Learning (ICML' 08) , pp. 880–887, Helsinki, Finland (2008).

L. Xiong, X. Chen, T.-K. Huang, J. G. Schneider, and J. G. Carbonell, Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In Proceedings of the 2010 SIAM International Conference on Data Mining (SDM' 10) , pp. 211–222, SIAM (2010). 10.1137/1.9781611972801.19 X. Chen, Z. He, and L. Sun, A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation.

Transp. Res. Part C Emerg. Technol. , 73–84 (2019).

X. Wang, L. Zhu, and J. Su, Modeling complex networks of nuclear reaction data for probing their discovery processes.

Chin. Phys. C (12), 124103 (2021). L. Song, Z. Chen, R. Han, Q. Sun, G. Tian, and P. Ren, FLUKA and Geant4 simulation of spallation neutrons from lead and tungsten targets bombarded with 400–1500 MeV protons.

Nucl. Phys. Rev. (1), 100–104 (2018).

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.