Deep Q-learning for autonomous optimization of neutron thermalization devices for PGNAA applications

Authors: Abdelnour, Miss Marina, Liu, Juntao, Wajid, Mr. Muneeb, Yuan, Mr. Chao, Heng, Mr. Tian, Ms. Li Wenxin, Liu, Prof. Zhiyi, Liu, Mr. Juntao

Date: 2025-11-28T00:00:00+00:00

Abstract

Prompt Gamma Neutron Activation Analysis (PGNAA) uses thermal neutron capture to execute isotopic analysis using characteristic gamma emissions. However, practical neutron sources mostly emit fast neutrons, and efficient thermalization is required to enhance signal quality and analytical accuracy. The complex multiparameter optimization of material compositions and geometries makes the design of thermalization devices computationally challenging. In this study, we introduce a Deep Q-Learning (DQL) framework that combines reinforcement learning with Monte Carlo N-Particle Code to autonomously optimize neutron thermalization device design. By defining the optimization as a Markov Decision Process, the DQL agent successfully explores over seven million possible design configurations over 1,500 training episodes. By episode 520, the agent had reached an ideal configuration, increasing thermalization efficiency by 1.42 times and reducing computing costs by 23% compared to standard genetic algorithm. This work demonstrates that transport simulations can be utilized as dynamic reinforcement learning settings, offering a scalable approach to intelligent, self-adaptive nuclear system design for complex analytical applications.

Full Text

Preamble

Deep Q-learning for autonomous optimization of neutron thermalization devices for PGNAA applications

Marina R. Abdelnour[1,2,3], Juntao Liu[1,2*], A. M. Wajid[1,2], Chao Yuan[4], Tian Heng[1,2], Wenxin Li[1,2], Zhiyi Liu[1,2†]

[1] Frontiers Science Center for Rare Isotopes, Lanzhou University, Lanzhou, Gansu, 730000, China
[2] School of Nuclear Science and Technology, Lanzhou University, Lanzhou, Gansu, 730000, China
[3] Department of Physics, Faculty of Women for Arts, Science, and Education, Ain Shams University, Cairo, Egypt

*Corresponding author: ljt@lzu.edu.cn
†Corresponding author: zhiyil@lzu.edu.cn

Abstract

Prompt Gamma Neutron Activation Analysis (PGNAA) uses thermal neutron capture to execute isotopic analysis using characteristic gamma emissions. However, practical neutron sources mostly emit fast neutrons, and efficient thermalization is required to enhance signal quality and analytical accuracy. The complex multiparameter optimization of material compositions and geometries makes the design of thermalization devices computationally challenging. In this study, we introduce a Deep Q-Learning (DQL) framework that combines reinforcement learning with Monte Carlo N-Particle Code to autonomously optimize neutron thermalization device design. By defining the optimization as a Markov Decision Process, the DQL agent successfully explores over seven million possible design configurations across 1,500 training episodes. By episode 520, the agent had reached an ideal configuration, increasing thermalization efficiency by 1.42 times and reducing computing costs by 23% compared to standard genetic algorithm. This work demonstrates that transport simulations can be utilized as dynamic reinforcement learning settings, offering a scalable approach to intelligent, self-adaptive nuclear system design for complex analytical applications.

Keywords: Neutron thermalization, PGNAA, deep Q-learning, reinforcement learning, Monte Carlo simulation, nuclear system optimization

Introduction

Prompt gamma-ray neutron activation analysis (PGNAA) has been extensively studied due to its high potential and broad range of applications as a quantitative isotope identification method. The main setup components of a PGNAA system include a neutron source, an optimized moderator, a collimator, the sample under investigation, and a detector array. PGNAA systems widely use modern neutron generators, which offer advantages in terms of compactness and operational efficiency compared to conventional neutron sources. These generators use deuterium-deuterium (D-D) and deuterium-tritium (D-T) fusion processes to create neutrons at 2.5 or 14.1 MeV. Effective thermalization is necessary to optimize these high-energy neutrons to thermal energies appropriate

for analytical use [1].

The analytical method requires bombarding the target material with thermal neutrons that interact with target nuclei to induce distinct gamma emissions, resulting in energy spectra with discrete peaks that serve as isotopic fingerprints for elemental identification. One of the main advantages of PGNAA in material characterization is the simultaneous irradiation and detection technique, which shows significantly high efficiency in real-time analysis. Given this potential, PGNAA has become an essential tool in many applications [1], including advanced material analysis in nuclear technology, medicine, and environmental monitoring; geophysical tracking for oil and coal exploration; security screening at airports and borders; and the detection of buried explosive devices [2].

Although PGNAA systems have seen significant advancements, optimizing neutron thermalization remains a major challenge. Insufficiently thermalized neutrons affect gamma spectra, which can lead to inaccurate determination of isotopic composition. Designing effective neutron thermalization devices (NTDs) is challenging due to the complex interplay of moderator geometry, material choice, and surrounding system components. Even small adjustments in these parameters can significantly affect thermal neutron flux and spectrum quality, making traditional optimization methods often insufficient [3] (Fig. 1 [Figure 1: see original paper]).

Monte Carlo (MC) simulations provide highly accurate neutron transport modeling, but their computational demands make practical optimization of neutron thermalization devices (NTDs) challenging. It often requires thousands of calculations through gradual layer adjustments or exhaustive parameter sweeps. Monte Carlo N-Particle transport (MCNP) based PGNAA designs have achieved success [3–7], yet the lengthy design cycles highlight the need for more efficient strategies to navigate complex, multi-dimensional parameter spaces. Automated approaches, including simulated annealing [8], particle swarm optimization [9], differential evolution, and genetic algorithms (GA) [10,11], reduce computational effort compared to manual tuning. However, these methods remain constrained by fixed parameter spaces, sensitivity to initial conditions [12], extensive hyperparameter adjustments [13], and the inability to leverage insights from previous optimization steps. While recent applications of GA in the design of neutron thermalization devices are promising, they are limited by local optima, subjective weighting, and high MC costs [14]. Their static nature prevents adaptation, highlighting the need for self-learning strategies that dynamically respond to simulation feedback.

In recent years, nuclear physics has made extensive use of artificial intelligence (AI) methods such as machine learning and neural networks to enhance nuclear systems [15]. AI has transformed nuclear design by enabling the automation of complex processes that were previously too computationally taxing. Supervised learning techniques have been successfully used to optimize reactor fuel loading and beam parameters [16], and recent research has demonstrated the successful integration of MC simulations and neural networks for beam shaping and isotope

detection [17, 18]. However, supervised approaches are vulnerable to Monte Carlo noise, require large pre-computed datasets, and suffer from criticality limitations.

Reinforcement learning (RL), which learns optimal policies by direct interaction with the environment, offers a good alternative. RL agents are well-suited for complex optimization problems [19]. By receiving rewards or punishments based on their behavior, they independently discover optimal strategies and gradually improve performance through trial-and-error research. Recent PGNAA applications of Q-learning have shown promise in optimizing collimator geometries and moderator dimensions [20], although these implementations are still limited to discrete action spaces, simpler configurations, and have not fully utilized RL' s potential for continuous parameter optimization and autonomous learning capabilities.

In this study, we present the first application of deep Q-learning (DQL) for autonomous optimization of thermal neutron assemblies in PGNAA systems. The DQL agent autonomously identifies optimal configurations by exploring the parameter space and interacting with MCNP simulations (Fig. 2 [Figure 2: see original paper]), leveraging learned value functions to capture complex dependencies between six optimization parameters: shielding, collimator diameter, multiplier, and moderator thickness/material. Morris' s sensitivity analysis identified the most crucial elements for guiding research [21]. With a reward function that increases thermal flux and penalizes epithermal/fast neutrons, the method was trained over 1,500 episodes and achieves optimal configurations with 23% lower computational cost compared to genetic algorithm.

---

2 Materials and Methods

2.1 Computational Environment and Setup

We optimized PGNAA setups using deep Q-learning (DQL) and MCNP6 [22] radiation transport simulations, comparing the outcomes to a genetic algorithm (GA) commonly used in neutron activation studies [2, 10, 11, 14]. The GA used standard hyperparameters (mutation rate 0.001, crossover rate 0.7, ranking selection, and simulated binary crossover with 0.1 precision [10]) with a population of 50 across 30 generations (1,500 evaluations) and a random seed of 42 for repeatability (see supplementary materials). Testing both strategies with the same computing budget allowed for direct performance comparison.

The baseline geometry included a 14.1 MeV DT neutron source and a multilayer assembly consisting of 45 cm beryllium oxide (BeO) primary moderator (S1), 31 cm polyethylene (PE) secondary moderator (S2), 10 cm lead (Pd) neutron multiplier (S3), 6 cm lead gamma shield (S4), and 5 cm beryllium (Be) collimator (S5). While the use of such materials is common in PGNAA designs [1, 20, 23], the distinguishing aspect of this work lies in the method used to

determine their optimal arrangement and thickness. Rather than relying on expert knowledge or predefined rules, the DQL agent autonomously explored the design space, interacting directly with the MCNP simulation environment to discover high-performance configurations. The optimization process included iterative simulation, learning, and validation stages.

2.2 Sensitivity Analysis and Design Framework

We used the Morris method, a computationally effective one-at-a-time (OAT) global sensitivity strategy appropriate for high-dimensional models [24], to prioritize important design variables for DQL optimization (Fig. 3 Figure 3: see original paper), with 600 assessments produced with 6 levels and 100 trajectories. This technique calculates the standard deviation (σ) to evaluate parameter interactions and nonlinearities and the mean absolute elementary effects (*) to measure overall sensitivity.

Five primary parameters (moderator layer thicknesses, collimator diameter, shielding, and neutron multiplier thickness) were chosen because they had significant impact on thermal neutron flux, although secondary factors like climatic variations and neutron source stability were considered insignificant in the controlled simulation context [1, 25, 26].

The results demonstrated considerable parameter coupling in the PGNAA system (Fig. 3(b)). The primary moderator layer exhibited the highest sensitivity ($* = 112.04 \pm 12.52$), indicating its critical involvement in thermal neutron flux enhancement. The large standard deviation ($\sigma = 63.96$) indicates strong parameter interactions, where the moderator' s effectiveness is strongly dependent on other component configurations. The same interaction pattern was observed in the side moderator S2 ($\sigma = 26.02$) and Pd multiplier ($\sigma = 30.10$). For each of these three parameters, $\sigma > 20$ indicated strong nonlinear coupling between design variables.

These significant σ values demonstrate that enhancing any single parameter independently will lead to suboptimal results, since the impact of changing one parameter depends on the settings of others. In contrast, detector shielding showed low sensitivity ($* = 0.19$) and limited interactions ($\sigma = 0.22$), suggesting its impact on neutron thermalization is essentially independent of other parameters. Its low σ value indicates it can be tuned independently without concern for coupling effects, although it remains crucial for improving the signal-to-noise ratio (SNR) through background radiation attenuation [23].

The interaction patterns show that sequential parameter adjustment cannot capture the coupled dynamics predicted by large σ values. This motivated the use of Deep Q-Learning, which naturally accounts for parameter interactions and explores the entire joint parameter space. DQL' s exploration allows it to find synergistic parameter combinations that the Morris analysis reveals, in contrast to gradient-based or grid-search methodologies that may fall into local optima in coupled systems.

2.3 State Space Definition

The state space was defined by geometrical properties and materials of the moderator, collimator, neutron multiplier, and shielding components. To reduce computational overhead, symmetric components were optimized on one side only, with results automatically mirrored to preserve symmetry. State variables and ranges are summarized in Table 1 .

2.4 Action Space and Material Selection

Actions defined discrete adjustments to component geometry and moderator materials within predefined ranges. Each parameter had two actions: increase toward the upper bound or decrease toward the lower bound. Based on sensitivity analysis results, six candidate materials were selected for the primary moderator layer using key nuclear property criteria: optimal neutron slowing-down power through elastic scattering with light nuclei (hydrogen, carbon, beryllium, and oxygen) and low thermal neutron absorption cross-sections to minimize neutron loss [3]. The material selection prioritized solid moderators to meet compact PGNAA system requirements. While water exhibits excellent neutron moderation properties, solid moderator materials provide superior mechanical stability, eliminate containment complexities, and maintain effective neutron thermalization capabilities for portable applications [27]. These materials enable the DQL agent to explore trade-offs between geometry, shielding synergy, and moderation efficiency.

A unique ID between 1 and 6 was assigned to each potential material (Table 2). The DQL algorithm' s actions 10 and 11 allow the agent to explicitly choose from all six candidate materials by changing the primary moderator layer' s material ID. Table 3 contains complete definitions of actions.

2.5 Reward Function

We aim to obtain optimal system geometry via simultaneous maximization of thermal neutron flux (Φ_{th}) and minimization of fast neutron contributions to improve SNR for PGNAA applications. The reward function R guides the optimization process by incorporating three key performance metrics as defined in Eq. (1) [4, 30]:

$$R = w_1\Phi_{th} + w_2\left(\frac{\Phi_{th}}{\Phi_{fast}}\right) + w_3\left(\frac{\Phi_{th}^2}{\Phi_{total}}\right) \quad [\text{neutron cm}^{-2}\text{s}^{-1}]$$

where Φ_{th}, Φ_{fast}, and Φ_{total} are thermal, fast, and total neutron fluxes, respectively. The weights w_1 = 0.3, w_2 = 0.3, and w_3 = 0.4 balance direct thermal flux contribution, thermal-to-fast neutron ratio, and thermalization efficiency. Larger R values correspond to higher thermal neutron flux and more efficient neutron beams for PGNAA applications. To evaluate neutron flux

components, the MCNP F4 tally was applied across thermal ($< 10^{-7}$ MeV), epithermal (10^{-7} to 10^{-2} MeV), and fast ($> 10^{-2}$ MeV) energy ranges with relative errors less than 0.5%.

2.6 Deep Q-Learning Architecture

We implement deep Q-learning to optimize PGNAA system parameters through reinforcement learning. The end-to-end framework consists of three main stages: initial state, training process, and terminal state (Fig. 4 [Figure 4: see original paper]).

The framework optimizes the action-value function Q(s, a), representing the expected cumulative reward for action a in state s [21, 31]:

$$Q(s,a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_t = s, a_t = a\right]$$

where r_t is the reward at time step t, and γ is the discount factor ($0 \leq \gamma \leq 1$). The reward function incorporates thermal neutron flux maximization to guide convergence toward optimal configurations. The complete DQL training integrating all components is presented in Algorithm 1 .

We ensure stable learning through experience replay and target networks. Experience replay stores agent interactions (s, a, r, s) in memory D, breaking data correlations during training. The target network with parameters $^{-}$ provides stable targets:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim D}\left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^{-}) - Q(s, a; \theta_i)\right)^2\right]$$

where _i represents primary network parameters.

2.7 Training and Exploration Strategy

The model uses dual neural networks: a primary network for action prediction and a target network for weight updates. Experience replay stabilizes learning from past interactions, while random weights enable comprehensive exploration of the configuration space. To balance exploration and exploitation, we use an epsilon-greedy approach. During training, epsilon declines from 0.98 to 0.01 while maintaining 1% random exploration. At each step, action selection is determined by random number generation; values below epsilon initiate random actions (exploration), while values above epsilon choose the optimal action (exploitation).

MPI parallelization was used to run MCNP6 simulations on a high-performance computing cluster (Intel® Xeon® Platinum 8358, 503 GB RAM, 128 cores).

All simulations used 128 cores via the command `mpirun -np 128 mcnp6.mpi i=<input_{file}> o=<output_{file}>`. This setup enabled efficient parallel execution utilizing nearly all available cores. Over 1,500 training episodes, the agent gradually improved its choices while reducing loss. Every geometric and material parameter is included in the state representation, and the policy is updated at each stage based on simulation feedback. The epsilon-greedy decay curve (Fig. 5 Figure 5: see original paper) demonstrates the transition from exploration-dominated early training to exploitation-focused later episodes. Training performance metrics (Fig. 5(b)) show total reward, average reward, and training loss evolution, indicating successful convergence with increasing reward stability over episodes. Hyperparameters are listed in Table 4 and were selected within established reinforcement learning ranges [32, 33].

2.8 Model Validation and Reproducibility

To verify the robustness of our DQL technique, we performed 30 independent training runs with various random initializations to evaluate learning consistency and policy stability. The stability of the optimization process was assessed as each experiment began with distinct baseline environmental factors and randomly initialized network weights. The model showed remarkable optimization consistency with a mean end reward of 16.41 ± 2.32 across all runs (CV = 14.13%) [34].

Action 9, which increased the thickness of the S5 collimator by 0.2 cm, consistently yielded the most significant rewards (mean = 17.18, 95% CI [14.59, 19.77], n = 6, CV = 14.4%). In contrast, Action 1 (increasing S1 moderator thickness toward -26 cm) performed consistently poorly (mean = 5.56, 95% CI [4.20, 6.92], n = 6) (Fig. 7 [Figure 7: see original paper]). Welch' s t-test [35] ($t(10) = 10.18$, $p < 0.001$) and Cohen' s d = 5.93 both revealed a highly significant performance difference. The significant effect size and strong statistical power (0.99) [36] validate this result' s stability. Complete normality tests [36] and Mann-Whitney U values [37, 38] are provided in Table 5 , which compares high- and low-reward actions from 30 DQL runs.

---

3 Results and Discussion

The DQL-MCNP framework found the optimal solution at episode 520 after 1,035 training steps and 2.7 hours, exploring a design space of 7×10^6 possible configurations. This represents a 23% reduction in computational time compared to the GA baseline. Table 6 lists the five best-performing episodes together with relevant geometric parameters and complete thermal metrics. BeO was identified as the optimal primary moderator at 44 cm thickness (Fig. 8 Figure 8: see original paper), illustrating systematic material exploration. The probability distributions (Fig. 8(b)) confirm that BeO outperforms other materials in performance and reliability. Panels (c)-(f) in Figure 8 illustrate secondary

component optimization: 26 cm S2, 8.5 cm S3, 4 cm S4, and 3.2 cm S5.

Our DQL-MCNP approach outperforms the baseline algorithm across several important criteria. The primary goal of maximizing thermal neutron production is demonstrated by a 1.81× improvement factor for thermal neutron flux (Φ_{th}) over baseline. Thermalization efficiency also improved with a DQL-to-baseline factor of 1.42×, indicating more efficient neutron moderation. The SNR decreased to 0.7× relative to baseline; however, since the reward function does not target SNR, this change reflects shielding and detector effects and should be considered an additional benefit rather than a direct optimization outcome. The SNR in PGNAA systems is calculated by comparing the net gamma signal from sample elements to background contributions from surrounding structures [23].

The FMESH neutron flux tally is displayed in Fig. 9 [Figure 9: see original paper]. The thermal neutron flux inside the sample region is comparatively smaller in the baseline arrangement (Fig. 9(a)). In contrast, thermal and total flux inside the sample region are significantly higher in the DQL-optimized design (Fig. 9(b)). The DQL-derived design improves thermalization efficiency by optimizing the reward term Φ_{th}^{2}/Φ_{total} (Eq. 1).

In the optimal configuration, the SNR is 2, the thermal flux improves from 1.58 $\times$ 10^{-5} to 2.86 $\times$ 10^{-5} n/cm^2/s (a 1.81× enhancement), and the thermalization efficiency reaches 1.65 $\times$ 10^{-5}.

Spectrum performance analysis shows considerable enhancement of peak neutron flux at thermal energies. With similar energy resolution (FWHM = 0.600 for both configurations), the DQL-optimized design achieves peak intensity 1.8× higher (1.89 $\times$ 10^{-7} n/cm^2/s) at approximately 10^{-7} MeV than the baseline (1.04 $\times$ 10^{-7} n/cm^2/s) (Fig. 10 Figure 10: see original paper). Figure 10(b) shows improvement factors for the DQL-optimized design across the energy spectrum, ranging from 2.3× at lower indices to 7.4× at higher indices. Spectral fitting analysis (Fig. 10(c,d)) indicates Lorentzian profiles provide the best fit for both configurations (R^2 = 0.891 baseline, R^2 = 0.829 optimized). This increase in peak flux without compromising energy selectivity validates the DQL optimization approach.

We chose Aluminum (Al) to validate our proposed DQL approach due to its industrial relevance and fundamental importance as the third most abundant crustal element, with annual global production of approximately 62 million tons [39]. With characteristic gamma signatures at 7.724 MeV and 4.133 MeV (σ_γ = 0.0493 and 0.0149 barns, respectively), the DQL modification effectively improved Al detection. Comparing three geometries reveals important performance trade-offs. Geometry 1 (Episode 240) represents an early-stage DQL solution with mediocre performance. Geometry 2 (Episode 780) trades absolute gamma-ray flux intensity to achieve maximum SNR (2.19) with good noise suppression. The DQL-optimized configuration achieves enhanced thermalization efficiency (1.65 $\times$ 10^{-5}) while maintaining balanced spectral performance.

Gamma spectrum analysis demonstrates improvements in Al characteristic peaks. At the primary 7.724 MeV line, the DQL-optimized configuration provides a significant increase, enhancing intensity by 25% from 4.0×10^{-8} to 5.0×10^{-8} photons/cm$^2 \cdot$s. At 4.133 MeV, the baseline setup shows slightly greater intensity (1.18×10^{-8} versus 1.14×10^{-8} photons/cm$^2 \cdot$s).

To evaluate generalizability, we used sodium chloride (NaCl), a compound widely studied in neutron activation investigations [40, 41]. The DQL optimization increased prompt gamma-ray emissions from both sodium and chlorine by improving neutron thermalization efficiency in the NaCl compound. This approach reliably increases thermal neutron flux and detection sensitivity, as demonstrated by nearly identical enhancement factors (31% for Na and 30% for Cl) (Fig. 11 [Figure 11: see original paper]).

3.1 Highlighting Differences Between Deep Q-Learning-Based Monte Carlo and Genetic Algorithm Optimization

The DQL-MCNP and standard GA approaches differ in how they explore the optimization space and handle parameter tuning. The GA approach uses fixed, well-established hyperparameters for mutation, population size, and crossover rates [2, 13]. In contrast, DQL adaptively adjusts its policy and action-selection strategy based on feedback from previous simulation outcomes.

DQL demonstrates online convergence because it uses real-time reward feedback to modify its policy at each step, enabling identification of optimal solution areas during training and real-time discovery of performance plateaus. In contrast, GA requires extensive generational analysis; actual convergence can only be confirmed after all generations complete, as population-based search makes it difficult to distinguish true convergence from early plateaus during training [21, 42].

DQL and GA examined comparable numbers of evaluations (DQL: 1,500 episodes $\times$ 10 steps 15,000 MCNP; GA: 30 generations $\times$ 50 population = 1,500 MCNP evaluations); however, their approaches to achieving optimal solutions differed significantly. The GA achieved its best fitness at generation 21 after 3.56 hours of training, while DQL achieved its best reward at episode 520, equivalent to 2.74 hours of training. This indicates DQL found its optimal configuration approximately 23% faster (Fig. 12 [Figure 12: see original paper]). Since DQL continuously modifies its strategy and identifies high-reward regions throughout training, it converges earlier. In contrast, GA may become trapped in local optima. The ideal GA configuration for this run comprised material set to BeO and parameters at extreme or boundary values (S1, S2, S3, S4, and S5) (Table 7 , Supplementary Figure S1). This suggests that while the approach generated the highest reward GA could find, it might not fully explore the search space and could overlook better-performing configurations.

Compared with other studies, our approach surpasses the multi-objective automated GA of Ma et al. [11] in thermal flux and offers superior thermalization.

Single-objective automated methods such as those reported by Cheng et al. [10] achieve higher SNR (2.56×). However, as SNR was not included in our reward function, the observed values are comparable. In terms of thermal flux and thermalization efficiency, real-time adaptive optimization performs better than pre-trained or manual techniques like hybrid MLP + Q-learning and manual tuning [20,23,43]. Although the reported improvement factors (Table 8) are relative due to variations in neutron sources, system designs, and simulation parameters, they demonstrate how well the DQL approach adapts to novel geometries in real-time MCNP simulations. While this study focuses on a 14.1 MeV D-T neutron source, the framework is inherently flexible and can be extended to various neutron energies and PGNAA system scales by modifying the state representation and reward function. Larger state spaces or more complex geometries may increase computational costs, though the technique is expected to maintain effective performance.

4 Conclusions

We conclude that our deep Q-learning approach yields superior multi-parameter optimization for PGNAA system design when coupled with Monte Carlo simulations. The framework incorporated sensitivity analysis utilizing the Morris method to reduce computational cost and identify key elements. The optimization approach for 14.1 MeV neutron energy was divided into three stages: initialization (random state and network setup), training (iterative state observation, action selection, reward evaluation, and Q-network updates), and termination (convergence criteria fulfilled). The reward function includes thermalization efficiency, thermal-to-fast flux ratio, and thermal flux related to geometric properties of moderator, collimator, multiplier, and shielding components. The DQL agent efficiently explored 7 million states and performed multi-objective optimization, reducing computation time by 23% without requiring expert intervention. Optimization effectiveness was demonstrated by notable improvement in distinctive gamma signals observed during materials validation. The DQL algorithm demonstrates broad applicability to different neutron thermalization device designs and material compositions. Its text-based input processing allows direct optimization from Monte Carlo simulation results without requiring pre-trained data, offering autonomous, data-driven nuclear system development with demonstrable advantages over traditional methods. Although this study focuses on a 14.1 MeV D–T neutron source, the framework is automatically flexible and could be extended to different neutron energies, system scales, and reward functions, highlighting its potential generalizability to a wide range of PGNAA configurations. We believe that automated hyperparameter optimization and real-time experimental data integration can significantly improve the proposed pipeline efficiency. Future work may include transfer learning methodologies for broader nuclear engineering applications and experimental verification to validate simulation-based results.

Data Availability

The complete reproducibility analysis across 30 independent DQL runs is provided in the supplementary materials. The pseudocode and key results of the GA baseline are also included for reference. Additional implementation details and simulation scripts are available from the corresponding author upon reasonable request. Due to proprietary formats and large file sizes, the full MCNP simulation datasets are not publicly available but can be shared upon request for research purposes.

References

[1] M. R. Abdelnour, J. Liu, K. Hossny, A. Wajid, W. Li, and Z. Liu, "Prompt gamma neutron activation analysis: A review of applications, design, analytics, challenges, and prospects," *Radiation Physics and Chemistry*, p. 112693, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969806X25001859

[2] Z. U. Koreshi and H. Khan, "Optimization of Moderator Design for Explosive Detection by Thermal Neutron Activation Using a Genetic Algorithm," *Journal of Nuclear Engineering and Radiation Science*, vol. 2, no. 3, p. 031018, 06 2016. [Online]. Available: https://doi.org/10.1115/1.4032702

[3] M. Zolfaghari, S. F. Masoudi, and F. Rahmani, "Optimization of linac-based neutron source for thermal neutron activation analysis," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 317, pp. 1477–1483, 2018.

[4] M. Vatani, M. Hassanvand, J. Mokhtari, and M. Choopan Dastjerdi, "Design of an in-tank thermal neutron beam for pgnaa application at isfahan mnsr," *Nuclear Engineering and Design*, vol. 412, p. 112451, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002954932300300X

[5] R. Uhlář, M. Kadulová, P. Alexa, and J. Pištora, "A new reflector structure for facility thermalizing d-t neutrons," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 300, pp. 809–818, 2014.

[6] A. H. Hegazy, V. Skoy, and K. Hossny, "Optimization of shielding-collimator parameters for ing-27 neutron generator using mcnp5," *EPJ Web of Conferences*, vol. 177, p. 02003, 2018.

[7] C. Cheng, Z. Wei, D. Hei, W. Jia, A. Sun, J. Li, P. Cai, D. Zhao, Q. Shan, and Y. Ling, "Design of a pgnaa facility using d-t neutron generator for bulk samples analysis," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 452, pp. 30–35, 2019.

[8] Z. Wang, Y. Wang, H. Xu, and H. Xie, "Application of simulated annealing algorithm in core flow distribution optimization," *Energies*, vol. 15, no. 21,

p. 8242, 2022.

[9] Y.-H. Lin, M.-T. Lee, and Y.-H. Hung, "A thermal management control using particle swarm optimization for hybrid electric energy system of electric vehicles," *Results in Engineering*, vol. 21, p. 101717, 2024.

[10] C. Cheng, Y. Xie, X. Xia, J. Gu, P. Wang, L. Xing, M. Wang, D. Hei, H. Lei, and J. Wenbao, "Neutron collimator optimization for 14.1 mev dt neutrons using monte carlo and genetic algorithms," *Applied Radiation and Isotopes*, vol. 198, p. 110838, 2023.

[11] B. Ma, M. Yan, X. Li, Q. Jiang, S. Wang, and Z. Liu, "Optimization design for moderator, reflector, and shielding of deuterium-deuterium neutron source," *Journal of Instrumentation*, vol. 19, no. 05, p. P05076, 2024.

[12] B. Kazimipour, X. Li, and A. K. Qin, "A review of population initialization techniques for evolutionary algorithms,"in *2014 IEEE Congress on Evolutionary Computation (CEC)*, 2014, pp. 2585–2592.

[13] H. Alibrahim and S. A. Ludwig, "Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021, pp. 1551–1559.

[14] Y. Ge, Y. Zhong, I. Murata, S. Tamaki, N. Yuan, Y. Sun, W. Ma, L. Zou, Z. Yang, and L. Lu, "Efficient optimization of an accelerator neutron source for neutron capture therapy using genetic algorithms," *Medical Physics*, vol. 51, no. 9, pp. 6445–6457, 2024.

[15] V. Sobes, B. Hiscox, E. Popov, R. Archibald, C. Hauck, B. Betzler, and K. Terrani, "Ai-based design of a nuclear reactor core," *Scientific Reports*, vol. 11, no. 1, p. 19646, 2021.

[16] A. Erdoğan and M. Geçkinli, "A pwr reload optimisation code (xcore) using artificial neural networks and genetic algorithms," *Annals of Nuclear Energy*, vol. 30, no. 1, pp. 35–53, 2003.

[17] M. Kamuda and C. J. Sullivan, "An automated isotope identification and quantification algorithm for isotope mixtures in low-resolution gamma-ray spectra," *Radiation Physics and Chemistry*, vol. 155, pp. 281–286, 2019.

[18] L.-F. Chen, "Machine learning-assisted optimization of modular neutron shielding based on monte carlo simulations," arXiv preprint arXiv:2504.17319, 2025.

[19] Y. Liu, B. Wang, S. Tan, T. Li, W. Lv, Z. Niu, J. Li, P. Gao, and R. Tian, "Applications of deep reinforcement learning in nuclear energy: A review," *Nuclear Engineering and Design*, vol. 429, p. 113655, 2024.

[20] M. Zolfaghari, S. F. Masoudi, F. Rahmani, and A. Fathi, "Thermal neutron beam optimization for pgnaa applications using q-learning algorithm and neural network," *Scientific Reports*, vol. 12, no. 1, p. 8635, 2022.

[21] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning,"*Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[22] X.-M. C. Team, "Mcnp –a general monte carlo n-particle transport code, version 6," Los Alamos National Laboratory, Tech. Rep. LA-UR-13-22934, 2013.

[23] J. Li, W. Jia, D. Hei, Z. Yao, and C. Cheng, "Research on the optimization method for pgnaa system design based on signal-to-noise ratio evaluation," *Nuclear Engineering and Technology*, vol. 54, no. 6, pp. 2221–2229, 2022.

[24] C. Wang, M. Peng, and G. Xia, "Sensitivity analysis based on morris method of passive system performance under ocean conditions," *Annals of Nuclear Energy*, vol. 137, p. 107067, 2020.

[25] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008.

[26] Y. Wang, Y. Song, W. Yin, H. Li, J. Lv, A.-J. Wang, and H.-C. Wang, "Modeling processes and sensitivity analysis of machine learning methods for environmental data," in *Water Security: Big Data-Driven Risk Identification, Assessment and Control of Emerging Contaminants*. Elsevier, 2024, pp. 511–522.

[27] L. L. Snead, D. Sprouster, B. Cheng, N. Brown, C. Ang, E. M. Duchnowski, X. Hu, and J. Trelewicz, "Development and potential of composite moderators for elevated temperature nuclear applications," *Journal of Asian Ceramic Societies*, vol. 10, no. 1, pp. 9–32, 2022.

[28] R. S. Detwiler, R. J. McConn, T. F. Grimes, S. A. Upton, and E. J. Engel, "Compendium of material composition data for radiation transport modeling," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Tech. Rep., 2021.

[29] R. G. Williams, C. J. Gesh, and R. T. Pagh, "Compendium of material composition data for radiation transport modeling," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Tech. Rep., 2006.

[30] R. Uhlář, P. Alexa, and J. Pištora, "A system of materials composition and geometry arrangement for fast neutron beam thermalization: An mcnp study," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 298, pp. 81–85, 2013.

[31] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.

[32] T. Eimer, M. Lindauer, and R. Raileanu, "Hyperparameters in reinforcement learning and how to tune them," in *International Conference on Machine Learning*. PMLR, 2023, pp. 1–12.

[33] X. Dong, J. Shen, W. Wang, L. Shao, H. Ling, and F. Porikli, "Dynamical hyperparameter optimization via deep reinforcement learning in tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1515–1529, 2021.

[34] S. Aronhime, C. Calcagno, G. H. Jajamovich, H. A. Dyvorne, P. Robson, D. Dieterich, M. Isabel Fiel, V. Martel-Laferriere, M. Chatterji, H. Rusinek, and B. Taouli, "Dce-mri of the liver: Effect of linear and nonlinear conversions on hepatic perfusion quantification and reproducibility," *Journal of Magnetic Resonance Imaging*, vol. 40, no. 1, pp. 90–98, 2014.

[35] M. Delacre, D. Lakens, and C. Leys, "Why psychologists should by default use welch' s t-test instead of student' s t-test," *International Review of Social Psychology*, vol. 30, no. 1, pp. 92–101, 2017.

[36] D. S. Quintana, "Statistical considerations for reporting and planning heart rate variability case-control studies," *Psychophysiology*, vol. 54, no. 3, pp. 344–349, 2017.

[37] P. E. McKnight and J. Najab, "Mann-whitney u test," *The Corsini Encyclopedia of Psychology*, pp. 1–1, 2010.

[38] Z. Birnbaum, "On a use of the mann-whitney statistic," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. 3. University of California Press, 1956, pp. 13–18.

[39] H. Folz, J. Henjes, A. Heuer, J. Lahl, P. Olfert, B. Seen, S. Stabenau, K. Krycki, M. Lange-Hegermann, and H. Shayan, "Pgnaa spectral classification of aluminium and copper alloys with machine learning," arXiv preprint arXiv:2404.14107, 2024.

[40] T. Czakoj, M. Košťal, E. Novák, E. Losa, J. Šimon, M. Schulc, F. Mravec, F. Cvachovec, J. Rataj, and Z. Matěj, "Measurement of prompt neutron capture gamma coming from iron and chlorine," *Annals of Nuclear Energy*, vol. 198, p. 110317, 2024.

[41] N. A. Elsheikh, "Characterization of (252cf-zrh2) monte carlo model for detection of nitrogen and chlorine by thermal neutron-capture pgnaa," *Radiation Physics and Chemistry*, vol. 188, p. 109591, 2021.

[42] M. Sarfraz and S. A. Raza, "Visualization of data using genetic algorithm," in *Soft Computing and Industry*. Springer, London, 2002, pp. 403–410.

[43] R. Uhlar, M. Kadulova, P. Alexa, and J. Pistora, "A new reflector structure for facility thermalizing d-t neutrons," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 300, no. 2, pp. 809–818, 2014.

Acknowledgments

The authors acknowledge support from: the Fundamental Research Funds for Central Universities at Lanzhou University (lzujbky-2023-ct05, lzujbky-2023-stlt01); the Central Government' s Guidance Funds for Local Science and Technology Development (24ZYQA045, YDZX20216200001297); the Ling Chuang Research Project of China National Nuclear Corporation (CNNC-LCKY-2024-080); the Special Funds from Gansu Nuclear Industry Research Institute; the National Key Research and Development Program of China (2023YFF1303501); the Lanzhou University Talent Cooperation Research Funds sponsored by Lanzhou City (561121203); and the National Natural Science Foundation of China (11975115).

Author contributions: All authors contributed to the research and manuscript preparation.
Competing interests: The authors declare no competing interests.

Supplementary Material

Deep Q-learning for autonomous optimization of neutron thermalization devices for PGNAA applications

Supplementary Tables

Table S1: Reproducibility over 30 independent DQL runs. The optimal geometry parameters (cm) are shown by M1–M6. Actions with the highest and lowest rewards are shown by Action_{high} and Action_{low}, demonstrating consistent learning behavior.

Run	S1 (BeO)	S2	S3	S4	S5	Action_{high}	Reward ($\times 10^{\{-5\})\|Action_{low}\|Reward(\times 10^{\{-5\}})$		
M1	44	26	8.5	4	3.2	Action 9	17.18	Action 1	5.56
...	...	...	...	...	...	...	...	...	...
Mean ± SD	43.80 ± 0.99	24.17 ± 1.33	8.42 ± 0.85	4.15 ± 0.19	3.18 ± 0.24	-	16.41 ± 2.32	-	5.30 ± 1.54
CV (%)	2.25	5.50	10.1	4.58	7.55	-	14.13	-	29.1

High-reward actions: Action 5 (26.7%), Action 9 (20.0%), Action 0 (13.3%), Action 2 (13.3%)
Low-reward actions: Action 1 (16.7%), Action 6 (16.7%), Action 8 (16.7%),

Action 4 (13.3%)
Performance gap: Mean reward_{high} = 16.41 vs. Mean reward_{low} = 5.30 (67.7% lower, $p < 0.001$)

Supplementary Figures

Algorithm 2: Genetic Algorithm for PGNAA Optimization

```
Require: Population size P, generations G, mutation rate  , crossover rate c, elitism e, pre
Parameters: P = 50, G = 30,   = 0.001, c = 0.7, e = 2, p = 0.1
Input: MCNP base input file, parameter ranges, material options
Initialize random seed
Create initial population P $\leftarrow$ { create_{individual}() }^P
best_{ind} $\leftarrow$  , best_{fitness} $\leftarrow$ -$\infty$

for generation g $\leftarrow$ 1 to G do
    Evaluate fitness for each individual i in P:
        - create modified MCNP input from individual's parameters
        - run MCNP and extract counts (thermal, fast, total)
        - compute reward: R = 0.3 · (thermal/fast) · 100 + 0.4 · (thermal$^{2}$/total) · 100 + (

    if reward > best_{fitness} then
        update best_{fitness} and best_{ind}
    end if

    Record statistics (best, average)
    Select parents p1, p2 by ranking selection (rank-based probabilities)

    if rand() < c then
        (c1, c2) $\leftarrow$ SBX(p1, p2,  )  // simulated binary crossover
        c1 $\leftarrow$ mutate(c1,  ); c2 $\leftarrow$ mutate(c2,  )
        append c1 (and c2 if space) to P'
    end if

    Create new population P' $\leftarrow$ [elites] (top e by fitness)
    while |P'| < P do
        append mutate(p1,  ) to P'
    end while
    P $\leftarrow$ P'[1:P]
end for

Save results, plots and best configuration
Return: best_{ind}
```

Figure S1: Fitness evolution in genetic algorithms: (a) 50 populations over 30 generations, (b) 50 populations over 50 generations, and (c) 30 populations over 100 generations.