Developing a Robust Multi-agent Reinforcement Learning Framework for Autonomous SMR Abnormal Operations using Abstraction Hierarchy

Gwanwoo Kim [a], Hee-Jae Lee [a], Jonghyun Kim [a*]
[a] *Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*
[*] *Corresponding author: : jonghyun.kim@kaist.ac.kr*

***Keywords :** Nuclear Power Plant, Operator Support, Abstraction Hierarchy

1. Introduction

There is a growing consensus that advanced automation and autonomous control systems are essential to support operators and mitigate human errors in small modular reactors (SMRs). To secure economic viability, SMRs are typically designed for multi-module operation, where a single control room manages multiple reactor units simultaneously. This operational shift inevitably increases the complexity of the system and the cognitive workload imposed on human operators, particularly given the reduced staffing levels compared to large-scale commercial nuclear power plants [1].

While rule-based automation has proven effective for normal operations [2], abnormal operations present a significant challenge. Unlike Design Basis Accidents (DBAs), which have well-defined mitigation strategies, abnormal states often arise from stochastic failures or unforeseen combinations of system malfunctions for which specific operating procedures may not exist. In such "undefined" scenarios, human operators must rely on their diagnostic capabilities to stabilize the plant, a task that becomes increasingly prone to error under high stress and time pressure [3]. Therefore, an intelligent agent capable of autonomous decision-making in diverse and unpredicted scenarios is required to ensure safety and continuity of operation.

Reinforcement Learning (RL) has demonstrated remarkable potential in controlling complex, non-linear systems without requiring explicit physical models [4]. Recent studies have applied RL to autonomous startup and power control of nuclear reactors, showing that Artificial Intelligence (AI) agents can adapt to various operational transients [5]. Directly training RL policies for abnormal operations in safety-critical plants is notoriously difficult because the control problem is knowledge-intensive: many actuators are available, yet only a small subset is context-relevant, and naive exploration wastes data and destabilizes learning.

To improve learnability, this paper proposes a quantitative Abstraction Hierarchy (AH)–guided hierarchical multi-agent RL (H-MARL) framework for an integral Pressurized Water Reactor (iPWR) simulator that operationalizes domain knowledge as explicit guidance. The framework reduces control complexity by decomposing decision-making into a meta-controller and localized agent teams that act on context-specific actuator subsets, and it further injects knowledge through two mechanisms; AH-aligned hierarchical reward shaping and a safety-margin-triggered, rule-based action masking, so the multi-agent system learns with a structured mental map while avoiding physically implausible exploration and maintaining execution-level safety.

2. Work Domain Analysis and AH modeling

Work Domain Analysis (WDA) is used to describe the iPWR simulator in terms of invariant plant purposes, governing physical principles, and means-ends relations, so that the control logic is anchored to what must be maintained (safety-relevant functions) rather than to a fixed procedure library [6]. For abnormal SMR operation, the principal challenge is not only control optimization but control validity under uncertainty: the system must remain within physically and operationally admissible regions even when the initiating event is unclear or compounded. This motivates a constraint-based formulation of the control problem.

Based on WDA, this study constructs an AH of the iPWR simulator with five levels: Functional Purpose, Abstract Function, Generalized Function, Physical Function, and Physical Form. Fig. 1 presents the overall AH diagram of the iPWR simulator.

At the top level, the functional purpose is defined as Maintain Energy Balance. The three abstract function branches include Energy Generation, Energy Conversion, and Energy Dispatch. Energy Generation covers primary-side thermal-hydraulic processes, including core heat removal, power generation, and coolant control. Energy Conversion addresses secondary-side steam generation and feedwater heat exchange. Energy Dispatch encompasses electricity generation and ultimate heat rejection. Each branch is further decomposed through the Generalized Function and Physical Function levels down to Physical Form, where individual actuators (e.g., valves, pumps, and control rod assemblies) are identified as control endpoints.

This decomposition is used to map high-level control objectives to process-level functions and ultimately to component-level actuation pathways [7]. The AH therefore provides the structural scaffold for both state interpretation and action admissibility.

The main methodological step of this study is the quantitative operationalization of the AH. For each

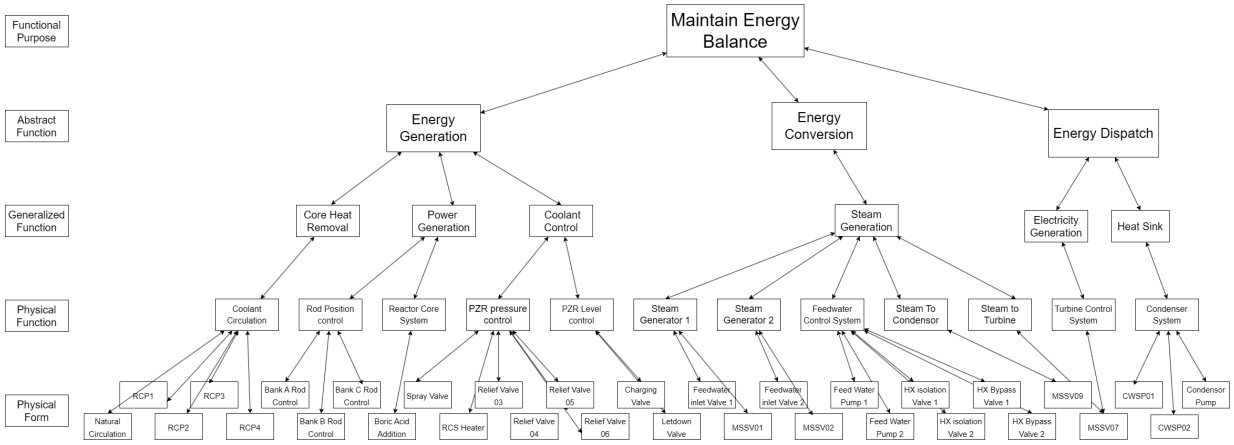

Fig. 1. AH diagram of iPWR simulator. The ultimate goal of this diagram is to maintain energy balance.

selected AH node, we define a computable functional indicator and a graded margin metric with respect to operating thresholds. Table I presents representative examples of these indicators. These AH-derived quantities are incorporated into the RL formulation to shape rewards and to restrict the effective action space in a convergence-oriented manner, providing structured learning signals and reducing unproductive exploration. In this way, the AH is used as an executable supervisory constraint layer on top of the RL policy.

Table I: Example of mathematical formulation of the abstraction hierarchy

Abstraction Level	Function	Mathematical Formulation
Functional Purpose	Maintain Energy balance	$\left\|\dot{m_{pri}}\Delta h_{pri} - \dot{m_{sec}}\Delta h_{sec}\right\| \le \epsilon_{EC}$
Generalized Function	Core Heat Removal	$\left\|(T_{hot} - T_{cold}) - \Delta T_{ref}\right\| \le \epsilon_{\Delta T}$
Physical Function	PZR pressure control	$\left\|P_{PZR} - P_{ref}\right\| \le \epsilon_P$
Physical Function	PZR Level control	$\left\|L_{PZR} - L_{ref}\right\| \le \epsilon_L$

The variables in Table I are defined as follows: $\dot{Q_{core}}$ and $\dot{Q_{SG}}$ are the core thermal power and the steam generator heat transfer rate. $\dot{W_{out}}$ and $\dot{Q_{loss}}$ represent the secondary system power output and heat loss. $\dot{m_{pri}}$ and $\dot{m_{sec}}$ denote the primary and secondary mass flow rates, while Δh_{pri} and Δh_{sec} are the enthalpy changes across the steam generator. T_{hot} and T_{cold} indicate the primary hot and cold leg temperatures, with ΔT_{ref} as the reference temperature difference. P_{PZR} and L_{PZR} are the operating pressurizer pressure and level, corresponding to their reference setpoints P_{ref} and L_{ref}. The ϵ terms define the tolerance margins for steady-state verification.

3. Design of AH-based H-MARL Framework

This section proposes a practical method for integrating the quantitative AH into an H-MARL framework [8] for the autonomous control of a SMR, specifically utilizing an iPWR simulator. This study proposes an implementable methodological template consisting of two core elements. The first element is AH-informed modular task decomposition, and the second element is a two-stage hierarchical control architecture reinforced by execution-level action masking.

3.1. RL Problem Formulation

Global observation includes raw simulator measurements and AH-derived functional indicators from Section 2. Raw measurements include temperatures, pressures, flow rates, and component states.

Instead of representing the action as a single monolithic continuous vector for all actuators, this study decomposes the decision-making process into two levels [9] as shown in Fig.2. The high-level meta-controller's action is defined as a discrete selection of a specific AH-defined functional module. The low-level actions are continuous vectors generated by a localized multi-agent group assigned to that specific module. Each dimension in the low-level action maps to an actuator command at the Physical Form layer in AH (e.g., valve opening).

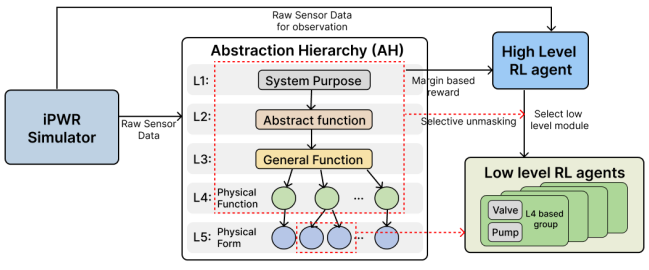

Fig. 2. Framework of AH-based H-MARL.

Rewards are also hierarchical. Both the high-level meta-controller and the low-level multi-agent groups utilize the tolerance margins of the AH nodes to

formulate their reward structures, with the margin initially set to 3%.

3.2. AH-Based Hierarchical Architecture

Unconstrained exploration across over all actuators simultaneously can be unsafe and computationally infeasible due to the exponential explosion of the joint action space. This study proposes an AH-based H-MARL architecture to compute state-dependent admissible subsets. The method first groups causally related actuators into independent multi-agent modules by traversing AH means-ends relations from the functional nodes to lower levels.

To stabilize learning, this study employ a two-stage training paradigm. In the first stage, each low-level multi-agent group is trained independently to resolve specific localized scenarios using cooperative MARL algorithms. In the second stage, the high-level meta-controller is trained to orchestrate these pre-trained modules.

When violations occur, the meta-controller selectively expands feasibility by activating only the multi-agent module on the AH branches connected to the violated function. To guarantee absolute physical safety at the execution level, this study retains a dynamic action masking mechanism. This mechanism generates a dynamic binary mask vector based on AH physical constraints to restrict the action space of the activated low-level agents. Masked dimensions are suppressed before reaching the physical actuators, ensuring that the low-level policies explore and execute only within the AH-consistent feasible subset.

4. Case Study

This section reports a prototype experiment that validates a lower-layer building block of the proposed framework. This study focuses on the pressurizer level control path within the Coolant Control branch of the AH. The experiment is designed to demonstrate that a RL agent can regulate a selected Physical Function by manipulating a linked Physical Form actuator under abnormal conditions. The purpose is not to validate the full AH-informed masking mechanism. The purpose is to confirm the feasibility of the actuator-level control layer that will later be placed under AH-constrained supervision. Fig. 3 presents the prototype-specific AH mapping, where only the nodes and links relevant to pressurizer level recovery are extracted from the full iPWR AH.

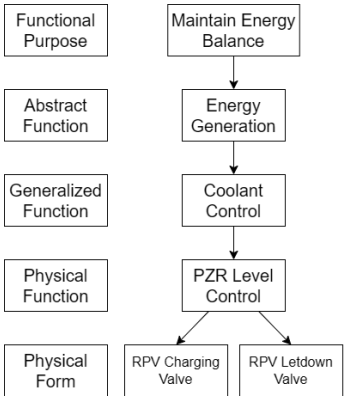

Fig. 3. Extracted AH path for the prototype experiment

4.1. Experimental Configuration

We isolated a limited AH path associated with pressurizer level control and implemented the experiment using an iPWR simulator representing an integral pressurized water reactor. The target abnormal scenario assumes a letdown valve that is stuck at a random open position between 50% and 100%. In addition, the PID controllers for the spray and charging valves are assumed unavailable. Under this condition recovery must be achieved through charging-valve actuation. If proper recovery actions are not taken, the pressurizer level continuously decreases, eventually triggering a reactor trip, as shown in Fig. 4.

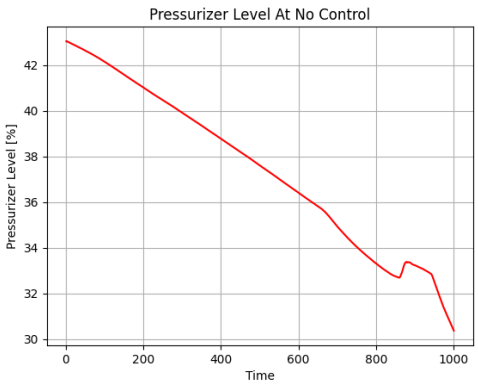

Fig. 4. Pressurizer level variation under no-control conditions.

In the prototype, the control objective is to maintain the Pressurizer Level node at the Physical Function level near the nominal target of 43%. The control input is restricted to the Charging Valve RCSV07 at the Physical Form level. This defines a single Physical Function to Physical Form linkage within the broader Coolant Control branch. We train a RL agent based on Proximal Policy Optimization (PPO) to learn the control policy. The reward is dominated by the pressurizer level tracking error relative to the 43% target. We also include episode-level bonuses and penalties to reflect sustained recovery and out-of-range termination. To limit unnecessary

control motion, one RL decision is applied every 5 s, and the valve command is constrained within the allowable operating range.

4.2. Evaluation of Physical Function Control

The prototype results indicate that the actuator-level control layer is feasible. In the single-module abnormal test, the agent learned to compensate for the persistent disturbance caused by the stuck letdown valve and kept the pressurizer level near the target region. This was achieved without relying on an explicit abnormal-operation procedure. Fig. 5 and Fig. 6 summarizes the validation results.

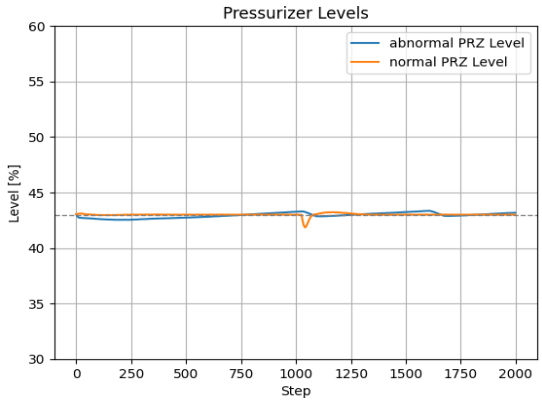

Fig. 5. Comparison of the pressurizer level between the normal state and the RL-controlled abnormal state.

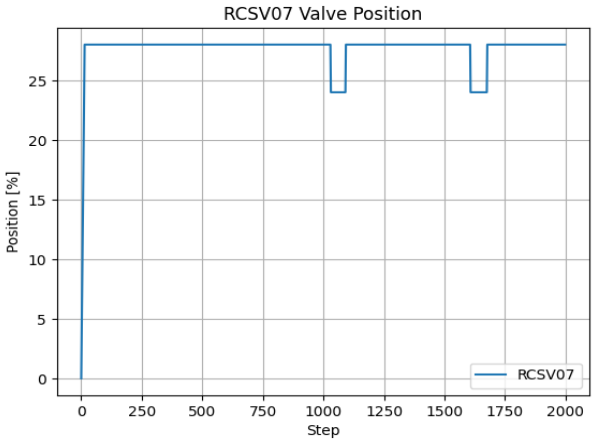

Fig. 6. Control action history of the RL agent modulating the charging valve to compensate for the coolant loss.

The learned control pattern is consistent with physical intuition. The agent tends to hold the charging-valve position when the level deviation is small. The agent applies incremental adjustments when the deviation grows. This suggests that the policy captures the dynamic relationship between the chosen Physical Function and Physical Form pair rather than memorizing a fixed sequence of actions.

5. Conclusion

This study proposed a safety-aware autonomous control framework for iPWR-based SMRs, integrating quantitative AH with H-MARL architecture. Using WDA, we organized the plant's physical and functional constraints to define modular task decomposition and execution-level action masking. The main contribution is a structured control design that safely guides the exploration of a high-level meta-controller and localized multi-agent groups using AH-derived constraints, rather than relying only on end-to-end reward penalties.

The case study demonstrated the feasibility of the lower-layer control concept on the pressurizer level control path within the Coolant Control branch. In the prototype environment, a PPO agent learned to regulate the pressurizer level by manipulating the charging valve under an abnormal disturbance scenario, showing that the Physical Function–Physical Form linkage can be learned without an explicit abnormal-operation procedure. However, this result should be interpreted as a proof-of-concept for a foundational actuator-level control unit, since the full AH-informed dynamic masking mechanism and higher-level functional constraints were not yet integrated.

ACKNOWLEDGEMENT

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIT) (No. GTL24031-111).

REFERENCES

[1] J. Hartmann, J. Hyvärinen, and V. Rintala, "The operator and the seven small modular reactors — An estimate of the number of reactors that a single reactor operator can safely operate," Nucl. Eng. Des., vol. 418, p. 112929, Mar. 2024

[2] J. Kim, S. Lee, and P. H. Seong, Autonomous Nuclear Power Plants with Artificial Intelligence, vol. 94. in Lecture Notes in Energy, vol. 94. Cham: Springer International Publishing, 2023

[3] A. D. Swain and H. E. Guttmann, "Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report," NUREG/CR-1278, SAND-80-0200, 5752058, Aug. 1983.

[4] A. Gong, Y. Chen, J. Zhang, and X. Li, "Possibilities of reinforcement learning for nuclear power plants: Evidence on current applications and beyond," Nucl. Eng. Technol., vol. 56, no. 6, pp. 1959–1974, Jun. 2024

[5] D. Lee, A. M. Arigi, and J. Kim, "Algorithm for Autonomous Power-Increase Operation Using Deep Reinforcement Learning and a Rule-Based System," IEEE Access, vol. 8, pp. 196727–196746, 2020

[6] H. J. Lee, D. Lee, and J. Kim, "Anomaly Recovery Algorithm Based on Robust AI Concept for Nuclear Power Plants," in Proceedings of 13th Nuclear Plant Instrumentation, Control and Human-Machine Interface

Technologies, NPIC and HMIT 2023, American Nuclear Society, 2023, pp. 1346–1355.
[7] M. Lind, “Making sense of the abstraction hierarchy in the power plant domain,” Cogn. Technol. Work, vol. 5, no. 2, pp. 67–81, Jun. 2003
[8] M. Ghavamzadeh, S. Mahadevan, and R. Makar, “Hierarchical Multi-Agent Reinforcement Learning”.
[9] H. M. S. Ahmad et al., “Hierarchical Multi-Agent Reinforcement Learning with Control Barrier Functions for Safety-Critical Autonomous Systems,” Aug. 18, 2025, arXiv