Grid-aware Online and Realtime Deep Reinforcement Learning-based Control and Optimization of Tightly Coupled Nuclear-Hydrogen System.

Temitayo O. Olowu[a], Elvan Sahin[a], Korey R. Cook[a], Jan W. Lambrechtsen[a], Nicholas J. Kane[a], Joseph McKay Barton[a], Tao Liu[a], Jeremy L. Hartvigsen[a], and Micah J. Casteel[a]

[a] Idaho National Laboratory, Idaho falls, Idaho, United States, Temitayo.Olowu@inl.gov, Elvan.Sahin@inl.gov, Korey.Cook@inl.gov, Jan.Lambrechtsen@inl.gov, Nicholas.Kane@inl.gov, Joseph.Barton@inl.gov, Tao.Liu@inl.gov, Jeremy.Hartvigsen@inl.gov, Micah.Casteel@inl.gov.

Abstract: The large-scale integration of intermittent energy resources and loads is driving the need for flexible, resilient, and economically optimized multi-energy systems capable of supporting real-time grid stability. Coupling nuclear energy systems (NESs) with high-temperature steam electrolysis (HTSE) provides a promising pathway for simultaneous reliable electricity generation and hydrogen production; however, the tightly coupled electro-thermal dynamics, operational constraints, and stochastic grid conditions present significant challenges for conventional optimization and control strategies. This paper presents a grid-aware, online, and real-time deep reinforcement learning (DRL)-based optimization and control framework for coordinated operation of integrated nuclear–hydrogen energy systems. The proposed framework employs a Twin-Delayed Deep Deterministic Policy Gradient (TD3) algorithm to learn continuous optimal dispatch policies for HTSE thermal and electrical demand while accounting for uncertain renewable generation, stochastic electricity and hydrogen market conditions, battery energy storage dynamics, and grid operational constraints. A physics-informed simulation environment is developed to capture nuclear thermal dynamics, HTSE electro-thermal behavior, inverter-based renewable generation, and nonlinear AC power flow constraints, enabling realistic closed-loop training and deployment. The framework is evaluated on a modified IEEE 9-bus system incorporating a 1000 MW nuclear power plant, multiple HTSE units, photovoltaic generation, and battery energy storage. Results demonstrate stable DRL convergence, adaptive coordination between electricity generation and hydrogen production, effective utilization of renewable resources, and sustained grid voltage regulation under highly dynamic operating conditions. The proposed approach achieves robust real-time operational flexibility while maximizing economic performance and maintaining nuclear and grid safety constraints. These findings demonstrate the potential of DRL-enabled coordinated control architectures to support scalable deployment of integrated nuclear–hydrogen systems in future low-carbon energy infrastructures.

Keywords: Nuclear Power Plant, High Temperature Steam Electrolysis, Deep Reinforcement Learning, Optimal Power Flow.

1. INTRODUCTION

Modern power systems are undergoing rapid transformation due to the increasing penetration of renewable energy resources, distributed energy systems, and flexible electrical loads [1]. The growing integration of inverter-based renewable generation, electrified industrial processes, and multi-energy infrastructures introduces significant operational complexity while simultaneously creating new opportunities for enhanced grid flexibility and resilience. In this evolving energy landscape, coordinated operation of nuclear energy systems (NESs), hydrogen production facilities, battery energy storage systems (BESSs), and other flexible demand-side assets has emerged as a promising strategy for supporting grid stability and resiliency [2].

Among emerging hydrogen production technologies, high-temperature steam electrolysis (HTSE), also referred to as high-temperature electrolysis (HTE), has attracted considerable attention due to its high thermodynamic efficiency and strong synergy with advanced nuclear reactors. HTSE utilizes both electrical and thermal energy to split steam into hydrogen and oxygen at elevated temperatures, typically between 750–850°C [3-4]. Compared with conventional low-temperature electrolysis technologies, HTSE significantly reduces electrical energy consumption by leveraging high-temperature heat, making it particularly suitable for integration with nuclear systems capable of supplying steady thermal energy.

The produced hydrogen can be stored, transported, utilized in industrial processes, or reconverted into electricity through fuel cells, thereby supporting sector coupling and long-duration energy storage [5].
The integration of HTSE with NESs provides several operational advantages. Because hydrogen production rates are directly proportional to electrical power consumption, HTSE systems can operate as highly flexible electrical and thermal loads that dynamically respond to changing grid conditions[6]. During periods of excess renewable generation or low electricity prices, HTSE facilities can increase hydrogen production to absorb surplus power and mitigate renewable curtailment. Conversely, during peak demand conditions or grid stress events, hydrogen production can be reduced to alleviate system loading and support overall grid reliability. Consequently, nuclear–hydrogen integrated systems have the potential to improve economic efficiency, enhance renewable energy utilization, and provide ancillary services such as voltage regulation, frequency support, and demand response.
Despite these benefits, tightly coupled nuclear–hydrogen systems present substantial operational and control challenges. HTSE systems exhibit strongly coupled electrochemical, thermal, electrical, and fluid-dynamic behavior with multiple interacting time scales. Simultaneously, NESs are subject to stringent operational constraints associated with reactor thermal limits, ramp-rate restrictions, and safety requirements. Coordinated operation therefore requires solving nonlinear, constrained, and stochastic optimization problems under uncertain renewable generation and time-varying load conditions. Conventional optimization and model predictive control approaches often struggle to achieve real-time scalability due to the computational burden associated with repeated nonlinear optimal power flow (OPF) calculations and high-dimensional system dynamics.
Recent advances in deep reinforcement learning (DRL) have created new opportunities for adaptive and autonomous control of complex power and energy systems [7-16]. Unlike conventional model-based optimization techniques , DRL methods can learn optimal control policies directly through interaction with the environment, enabling real-time decision making under uncertainty. Actor–critic algorithms such as Deep Deterministic Policy Gradient (DDPG) [9, 10, 14], Proximal Policy Optimization (PPO) [7, 8, 13] and Twin-Delayed Deep Deterministic Policy Gradient (TD3) [15-16] have demonstrated strong performance in continuous-control applications involving voltage regulation, frequency control, and distributed energy resource coordination. Prior studies have shown that DRL-based autonomous voltage control frameworks can effectively regulate transmission and distribution system voltages under renewable uncertainty while outperforming traditional value-based reinforcement learning methods in continuous control settings. Similarly, DRL approaches have been successfully applied to low-inertia frequency regulation, multi-area load frequency control, and decentralized ancillary service coordination.
Although DRL has shown promising capabilities for power system control, several important limitations remain. Existing frameworks primarily focus on electrical-domain objectives without explicitly considering tightly coupled multi-energy interactions or flexible electrochemical loads such as HTSE systems. Many approaches rely on offline training using simplified benchmark systems, limiting their adaptability to realistic operational environments with rapidly changing grid conditions. In addition, operational constraints and safety requirements are often incorporated only through reward penalization rather than through physics-informed formulations that explicitly account for network power flow constraints, thermal dynamics, and equipment operational limits. These challenges motivate the development of scalable, grid-aware, and real-time DRL frameworks capable of coordinating integrated nuclear–hydrogen systems under uncertainty.
To address these challenges, this paper proposes a grid-aware, online, and real-time DRL-based optimization and control framework for coordinated operation of nuclear–hydrogen integrated systems. The proposed approach employs a continuous-control Twin-Delayed Deep Deterministic Policy Gradient (TD3) algorithm to learn optimal control policies for regulating HTSE thermal and electrical demand in response to changing grid conditions. A physics-informed environment is developed to capture HTSE electro-thermal dynamics, nuclear operational constraints, uncertain renewable generation, and power flow limitations, enabling realistic closed-loop learning and deployment. By dynamically coordinating power dispatch between nuclear generation and hydrogen production, the proposed framework enhances operational flexibility, supports voltage stability, and enables ancillary grid services including demand response and voltage regulation.
The main contributions of this work are summarized as follows:

1. A physics-informed, grid-aware deep reinforcement learning (DRL) framework is developed for coordinated control and optimization of integrated nuclear–hydrogen energy systems, explicitly incorporating HTSE electro-thermal dynamics, nuclear operational constraints, renewable energy resources, battery energy storage, and AC grid power flow behavior within a unified closed-loop environment.
2. An online and real-time Twin-Delayed Deep Deterministic Policy Gradient (TD3)-based continuous-control strategy is proposed to enable adaptive coordination of hydrogen production, thermal diversion, and nuclear power dispatch under uncertain renewable generation, stochastic load demand, and dynamic electricity and hydrogen market conditions.
3. The proposed framework enables HTSE systems to operate as flexible electro-thermal loads that dynamically balance electricity generation and hydrogen production while supporting ancillary grid services including voltage regulation, demand response, and renewable energy integration.
4. Simulation studies on a modified IEEE 9-bus system demonstrate that the proposed DRL framework achieves stable learning convergence, improved operational flexibility, enhanced renewable energy utilization, robust economic performance, and sustained compliance with nuclear and grid operational constraints under highly dynamic operating conditions.

The remainder of this paper is organized as follows. Section II presents detailed mathematical formulation of the models used. Section III presents a Markov decision process formulation for the proposed TD3-based optimization and control framework. Section IV discusses the simulation environment and case studies. Section V concludes the paper and discusses future research directions.

2. SYSTEM MODELING

2.1 Nuclear Power Plant Model.

The developed model represents a reduced-order dynamic approximation of a 1000 MWe nuclear power plant. The formulation captures the dominant thermal and electromechanical dynamics required for integrated energy system analysis without introducing the complexity of full core neutronic simulations. The model is particularly suitable for studies involving thermal diversion to HTSE systems, hydrogen production, and multi-timescale energy management. The model includes the following subsystems: reactor thermal power dynamics, fuel thermal dynamics, primary coolant thermal dynamics, steam-side thermal and pressure dynamics, turbine mechanical power dynamics, generator electrical power dynamics and thermal diversion to HTSE loads.

The reactor thermal power dynamics are modelled using a first-order response with ramp-rate limitations. The reactor attempts to track a thermal power reference while respecting operational ramping constraints. The thermal power dynamics are expressed as (1)

$$\mathrm{d}P_{\mathrm{th}}/\mathrm{dt} = \left(K_p\left(P_{th,ref} - P_{th}\right)\right)/\tau_{rx} \tag{1}$$

$$-r_l/60 \le r_{rx} \le r_l/60 \tag{2}$$

where P_{th} and $P_{th,ref}$ represents reactor actual and reference thermal power, $\tau_{rx}$is the reactor time constant, K_p is the proportional controller gain, r_{rx} is the ramp rate, r_l is the limit ramp rate.

The reactor fuel and coolant thermal interactions are represented using lumped thermal capacitance models as presented in (3) – (6).

$$Q_{fc} = h_{fc}\left(T_f - T_c\right) \tag{3}$$

$$Q_{cs} = h_{cs}(T_c - T_s) \tag{4}$$

$$\mathrm{d}T_f/\mathrm{dt} = \left(Q_{rx} - Q_{fc}\right)/\left(M_f C_{pf}\right) \tag{5}$$

$$\mathrm{d}T_c/\mathrm{dt} = \left(Q_{fc} - Q_{cs}\right)/\left(M_c C_{pc}\right) \tag{6}$$

where $\mathrm{Q_{fc}}$ and $\mathrm{Q_{cs}}$ are fuel-to-coolant and coolant-to-steam heat transfer, $\mathrm{h_{fc}}$ and h_{cs} $\mathrm{T_f}$ and T_c are fuel and coolant temperatures, M_f and M_c represent the effective thermal masses of the fuel and coolant, respectively, while C_{pf} and $\mathrm{C_{pc}}$denote the corresponding specific heat capacities.

The steam-side subsystem captures thermal diversion Q_{div} to the HTSE process while maintaining thermal power (Q_{turb}) supply to the turbine. The steam mass flow rate to the turbine is constrained by both the available driving pressure and the thermal demand $\dot{\mathrm{m}}_{need}$ as expressed in (7) – (9)

$$Q_{\mathrm{turb,in}} = Q_{\mathrm{steam}} - Q_{\mathrm{div}} \tag{7}$$

$$\dot{m}_{turb} = min(\dot{m}_{cap}, \dot{m}_{need}) \quad (8)$$

$$Q_{\mathrm{turb}} = \dot{m}_{turb} h_{fg} \quad (9)$$

The steam pressure p_s is modelled as a first-order lag toward a quasi-static equilibrium pressure $p_{s,eq}$ that depends linearly on the steam temperature T_s as expressed in (10) – (11)

$$p_{s,\mathrm{eq}} = p_{s,\mathrm{ref}} + k_p(T_s - T_{s0}) \quad (10)$$

$$\mathrm{d}p_s/\mathrm{dt} = (p_{s,eq} - p_s)/\tau_{ps} \quad (11)$$

where T_{s0} is rated steam temperature, k_p is the steam temperature-pressure coupling coefficient, and τ_{ps} is the steam pressure time constant.

The turbine-generator subsystem is modelled using first-order dynamic equations. The mechanical power P_m delivered by the turbine is modelled as a first-order lag toward the product of turbine efficiency and actual thermal input as expressed in (12) – (13)

$$P_{m,\mathrm{ref}} = \eta_{\mathrm{turb}} Q_{\mathrm{turb}} \quad (12)$$

$$\mathrm{d}P_m/\mathrm{dt} = (P_{m,ref} - P_m)/\tau_{\mathrm{turb}} \quad (13)$$

$$P_{e,\mathrm{ref}} = \eta_{gen} P_m \quad (14)$$

where $P_{e,\mathrm{ref}}$ and $P_{m,ref}$ are the reference/command electrical and mechanical power respectively, η_{turb} is the turbine thermal to mechanical power conversion efficiency, τ_{turb} is the turbine time constant, and η_{gen} is the generator mechanical to electrical power conversion efficiency. The electrical power P_e output is obtained through a subsequent first-order lag representing the generator and excitation dynamics.

$$\mathrm{d}P_e/\mathrm{dt} = (P_{e,ref} - P_e)/\tau_{gen} \quad (15)$$

where τ_{gen} are turbine and generator time constants. From the above formulation, the electrical output of the nuclear power plant (NPP) which depends on the thermal diversion to the HTSE is denoted as P_e^{NPP}.

2.2 High Temperature Steam Electrolysis

The HTSE electrical power demand is modelled as a controllable load $P_{HTSE}(t)$. Where the HTSE's operating limits are defined as:

$$P_{HTSE}^{min} \le P_{HTSE}(t) \le P_{HTSE}^{max} \quad (16)$$

The HTSE ramp-rate constraints are represented by (17)

$$-\mathrm{R}_{\mathrm{HTSE}}^{\downarrow} \Delta\mathrm{t} \le \mathrm{P}_{\mathrm{HTSE}}(\mathrm{t}) - \mathrm{P}_{\mathrm{HTSE}}(\mathrm{t}-1) \le \mathrm{R}_{\mathrm{HTSE}}^{\uparrow} \Delta\mathrm{t} \quad (17)$$

where $R_{HTSE}^{\uparrow}$ and $R_{HTSE}^{\downarrow}$ denote the HTSE's electrical power demand ramp-up and ramp-down limits. The thermal energy supplied from the nuclear power plant to the HTSE process is represented by the thermal diversion variable $\mathrm{Q}_{\mathrm{div}}(\mathrm{t})$ from the nuclear power plant as expressed in (18)

$$Q_{HTSE}^{th}(t) = Q_{div}(t) \quad (18)$$

This diversion reduces the effective thermal energy available for electricity generation from the nuclear plan. The thermal diversion operating limits are defined as (19)

$$0 \le \mathrm{Q}_{\mathrm{div}}(\mathrm{t}) \le \mathrm{Q}_{\mathrm{div}}^{\mathrm{max}} \quad (19)$$

The thermal diversion ramp-rate limit is formulated as expressed in (20)

$$-\mathrm{R}_{\mathrm{div}}^{\downarrow}\Delta\mathrm{t} \le \mathrm{Q}_{\mathrm{div}}(\mathrm{t}) - \mathrm{Q}_{\mathrm{div}}(\mathrm{t}-1) \le \mathrm{R}_{\mathrm{div}}^{\uparrow}\Delta\mathrm{t} \quad (20)$$

where $\mathrm{R}_{\mathrm{div}}^{\downarrow}$ and $\mathrm{R}_{\mathrm{div}}^{\uparrow}$ are the thermal ramp-down and ramp-up limits

2.3 Battery Energy Storage

The instantaneous stored battery energy is modelled as

$$E_b(t) = SOC(t) E_b^{max} \quad (21)$$

where $E_b(t)$ is the instantaneous BESS capacity and E_b^{max} is the maximum battery energy capacity. The battery's state-of-charge SOC dynamics is modelled using (22)

$$SOC(\mathrm{t} + \Delta\mathrm{t}) = SOC(\mathrm{t}) + [\eta_{\mathrm{ch}} \mathrm{P}_{\mathrm{ch}}(\mathrm{t})\Delta\mathrm{t} / \mathrm{E}_{\mathrm{b}}^{\mathrm{max}}] - [P_{dis}(t)\Delta\mathrm{t}/(\eta_{dis} E_b^{max})] \quad (22)$$

where $P_{ch}(t)$ is the charging power, $P_{dis}(t)$ is the discharging power, and η_{ch} and η_{dis} are the charging and discharging efficiencies, respectively. The BESS operation is constrained by state-of-charge and power limits.

$$SOC^{min} \leq SOC(t) \leq SOC^{max} \quad (23)$$
$$0 \leq P_{ch}(t) \leq P_{ch}^{max} \quad (24)$$
$$0 \leq P_{dis}(t) \leq P_{dis}^{max} \quad (25)$$
$$P_{bess}(t) = \kappa_d \times P_{dis}(t) + \kappa_c \times P_{ch}(t) \quad (26)$$
$$\kappa_d + \kappa_c = 1 \quad (27)$$

Positive battery power corresponds to discharging operation, while negative battery power corresponds to charging operation. $\kappa_d \in 0,1$ and $\kappa_c \in 0,1$ are binary values that determine the state of batteries, either charging or discharging.

2.4 Inverter-based resource

The maximum DC power $P_{PV}^{dc}(t)$ generated by the solar PV array is modeled as

$$P_{PV}^{dc}(t) = P_{PV}^{rated} \frac{G(t)}{G_{STC}} [1 + \gamma_P (T_c(t) - T_{STC})] \quad (28)$$

where P_{PV}^{rated} is the rated PV capacity, where $T_c(t)$ is the PV cell temperature, T_{STC} and G_{STC} are temperature at and irradiance at standard test conditions, $G(t)$ is the actual irradiance in $\mathrm{W/m^2}$, and γ_P is the PV power temperature coefficient. The available DC power is bounded by the rated capacity as expressed in

$$0 \leq P_{PV}^{dc}(t) \leq P_{PV}^{rated} \quad (29)$$

The AC power ($P_{PV}^{ac}(t)$) dispatched to the grid by the inverter which converts DC PV power to AC power can be expressed as (30)

$$P_{PV}^{ac}(t) = \eta_{inv}(t) P_{PV}^{dc}(t) \quad (30)$$

where $\eta_{inv}(t)$ is the instantaneous inverter efficiency.

2.5 Optimal Power Flow and Optimization Formulation

The multiobjective-OPF (M-OPF) problem to optimize the production of hydrogen and provide ancillary services is formulated as a mixed-integer non-linear programming is presented as expressed in (31) – (33)

$$\text{minimize} \quad \mathrm{f}(\mathrm{x}, \boldsymbol{u}) \quad (31)$$
$$\text{st.} \quad \mathrm{h}(\mathrm{x}, \boldsymbol{u}) = \mathbf{0} \quad (32)$$
$$g(x, \boldsymbol{u}) \quad \leq 0 \quad (33)$$

where x is the vector of action variables, u is the vector of state variables, $\mathrm{f}(x, \boldsymbol{u})$ is the multiobjective function, $\mathrm{h}(x, \boldsymbol{u})$ denotes equality constraints, and $g(x, \boldsymbol{u})$ denotes inequality constraints.

The objective function f(x, u) is formulated as presented below

$$revenue = \max \sum_{t \in \mathcal{T}} \Delta \mathrm{t} \left[\pi_{\mathrm{t}}^{\mathrm{e}} \mathrm{P}_{\mathrm{e,net}}^{\mathrm{NPP}}(\mathrm{t}) + \pi_{\mathrm{t}}^{\mathrm{H2}} \left(\mathrm{m}_{\mathrm{H2}}^{\mathrm{HTSE-1}}(\mathrm{t}) + \mathrm{m}_{\mathrm{H2}}^{\mathrm{HTSE-2}}(\mathrm{t}) \right) - \frac{\pi_{\mathrm{t}}^{\mathrm{e}} \mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-2}}(\mathrm{t})}{\eta_{\mathrm{inv}}} \right] \quad (34)$$

where

$$\mathrm{m}_{\mathrm{H2}}^{\mathrm{HTSE-1}}(\mathrm{t}) = \frac{\mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-1}}(\mathrm{t})}{\eta_{\mathrm{stack}}} \quad (35)$$

and

$$\mathrm{m}_{\mathrm{H2}}^{\mathrm{HTSE-2}}(\mathrm{t}) = \frac{\mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-2}}(\mathrm{t})}{\eta_{\mathrm{stack}}} \quad (36)$$
$$P_{e,net}^{NPP} = P_e^{NPP}(t) - P_e^{HTSE-1}(t) \quad (37)$$

Δt is the optimization time-step, $\mathcal{T}$ is the optimization time horizon, π_t^e is the electricity price at time t, π_t^{H2} is hydrogen selling price at time t, $\mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-1}}$ and $\mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-2}}$ are electrical power demand of HTSE 1 and 2, while η_{stack} is the HTSE's electrolyser's stack efficiency in MWh/kg.

The power flow governing equations are set as the optimization equality constraints $\mathrm{h}(x, \boldsymbol{u})$ as expressed in (38) – (40)

$$S_i = v_i \sum_{k=1}^{N} Y_{ik}^* v_k^* \quad (38)$$
$$p_i = \Sigma_{(\mathrm{k}=1)}^{\mathrm{N}} |v_i||v_k|(G_{ik} cos\delta_{ik} + B_{ik} sin\delta_{ik}) \quad (39)$$
$$q_i = \Sigma_{(\mathrm{k}=1)}^{\mathrm{N}} |v_i||v_k|(G_{ik} sin\delta_{ik} - B_{ik} cos\delta_{ik}) \quad (40)$$

where S_i, p_i, and q_i are the apparent, real and reactive power injection at bus i $\in$ N , N is total number buses in the grid network, Y_{ik} is the element of the admittance matrix, which can be expressed as $Y_{ik} = G_{ik} + jB_{ik}$, and δ_{ik} is the voltage angle difference between bus i and k.

The inequality constraints governing the control variables which are the HTSEs electrical power demand (consequently the net power from the NPP to the grid), solar power generation and controlled BESS power are as expressed in (41) – (45)

$$\mathrm{P}_{\mathrm{e,min}}^{\mathrm{HTSE-1}} \leq \mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-1}}(\mathrm{t}) \leq \mathrm{P}_{\mathrm{e,max}}^{\mathrm{HTSE-1}} \tag{41}$$

$$\mathrm{P}_{\mathrm{e,min}}^{\mathrm{HTSE-2}} \leq \mathrm{P}_{\mathrm{e}}^{\mathrm{HTSE-2}}(\mathrm{t}) \leq \mathrm{P}_{\mathrm{e,max}}^{\mathrm{HTSE-2}} \tag{42}$$

$$\mathrm{P}_{\mathrm{min}}^{\mathrm{PV}} \leq \mathrm{P}^{\mathrm{PV}}(\mathrm{t}) \leq \mathrm{P}_{\mathrm{max}}^{\mathrm{PV}} \tag{43}$$

$$P_{e,net}^{NPP,min} \leq P_{e,net}^{NPP}(t) \leq P_{e,net}^{NPP,max} \tag{44}$$

$$P_{bess}(t) = -P_{e,net}^{NPP} - P^{PV}(t) + P_e^{HTSE-2}(t) + P^L(t) \tag{45}$$

where $P^L(t)$ is the timeseries load profile of the other loads within the network.

The steam thermal diversion constraints are as expressed in (46) – (48)

$$0 \leq Q_{div}(t) \leq Q_{div}^{max} \tag{46}$$

$$p_s^{min} \leq p_s(t) \leq p_s^{max} \tag{47}$$

$$T_s^{min} \leq T_s(t) \leq T_s^{max} \tag{48}$$

The fuel and coolant temperature constraints are as expressed in (49) - (50)

$$T_f^{min} \leq T_f(t) \leq T_f^{max} \tag{49}$$

$$T_c^{min} \leq T_c(t) \leq T_c^{max} \tag{50}$$

To ensure the proposed optimization does not violate nominal grid operating conditions, the following constraints are enforced as expressed in (51) – (52)

$$\mathrm{v}_{\mathrm{min}} \leq \mathrm{v}_{\mathrm{i}}(\mathrm{t}) \leq \mathrm{v}_{\mathrm{max}} \tag{51}$$

$$\mathrm{f}_{\mathrm{min}} \leq f(\mathrm{t}) \leq \mathrm{f}_{\mathrm{max}} \tag{52}$$

where $\mathrm{v}_{\mathrm{min}}$ and $\mathrm{v}_{\mathrm{max}}$ are the minimum and maximum allowable grid bus voltages, $\mathrm{f}_{\mathrm{min}}$ and $\mathrm{f}_{\mathrm{max}}$ are the minimum and maximum allowable grid frequency.

3. MARKOV DECISION PROBLEM FORMULATION

The optimization problem is formulated as a finite-horizon Markov Decision Process over a $\mathcal{T}$ scheduling horizon as expressed in

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \mathcal{T}) \tag{53}$$

where $\mathcal{S}$ is the state space containing all measurable system variables, $\mathcal{A}$ is the continuous action space containing controllable dispatch variables, $\mathcal{P}$ is the transition function describing how the system evolves from s_t to s_{t+1}, $\mathcal{R}$ is the reward function, and γ is the discount factor that weights future rewards relative to immediate rewards.

The state vector includes uncertain renewable generation, uncertain load demand, stochastic hydrogen prices, NPP states, and previous control actions as expressed in

$$s_t = \left[t, v_{i,t}, f_t, P_{t,u}^{PV}, P_{t,u}^{L}, \pi_{t,u}^{H2}, \pi_t^e, SOC_t, x_t^{NPP}, y_t^{NPP}, u_{t-1}\right] \tag{54}$$

where t is the current time index, $P_{t,u}^{PV}$ is uncertain PV generation, $\pi_{t,u}^{H2}$ is uncertain hydrogen selling price between \$2/kg and \$4/kg, $P_{t,u}^{L}$ is uncertain grid load demand, x_t^{NPP} is NPP's dynamic state vector, y_t^{NPP} is NPP's output parameters, and u_{t-1} is control action at $t-1$.

The control action space are HTSEs power demand and allowable PV generation as expressed in

$$u_t = \left[P_{e,t}^{HTSE-1}, P_{e,t}^{HTSE-2}, P_{e,t}^{PV}\right] \tag{55}$$

The transition state dynamics is as expressed in (55)

$$s_{t+1} = f(s_t, a_t, \omega_t) \tag{56}$$

where $s_{t+1} = \left[t, v_{i,t+1}, f_{t+1}, P_{t+1,u}^{PV}, P_{t+1,u}^{L}, \pi_{t+1,u}^{H2}, \pi_{t=1}^{e}, SOC_{t+1}, x_{t+1}^{NPP}, y_{t+1}^{NPP}, u_t\right]$ and $\omega_t = \left[P_{t,u}^{PV}, P_{t,u}^{L}, \pi_{t,u}^{H2}\right]$. The reward function is formulated as the hourly reward minus the penalties as expressed in (57)

$$r_t = \text{revenue} - \lambda_{viol}\mathcal{V}_t - \lambda_{\Delta u}\|u_t - u_{t-1}\|_2^2 \tag{57}$$

where λ_{viol} is the weight assigned to constraint violations, $\mathcal{V}_t$ is the set containing all constraint violations which include grid constraints, and NPP operational constraints, and $\lambda_{\Delta u}$ is the weight assigned to control-action movement to prevent abrupt transitions in control parameters.

The TD3 deterministic policy $u_t = \pi_\theta(s_t)$ and its objective is as expressed as (58)

$$J_{DRL} = \max_{\pi_\theta} \mathbb{E}\left[\sum_{t=1}^{\mathcal{T}} \gamma^{t-1} r_t\right] \tag{58}$$

To improve the robustness of the TD3 policy, uncertainties in power generation from the PV, load demand and hydrogen selling price is modelled. Based on a forecast base power P_t^{PV}, the uncertainty in the power generation $P_{t,u}^{PV}$ from the PV is as expressed as

$$P_{t,u}^{PV} = \max\left(0, P_t^{PV}(1 + \epsilon_t^{PV})\right) \tag{59}$$

where $\epsilon_t^{PV} \sim \mathcal{N}(0, \sigma_{PV}^2)$

Similarly, using the base load profile P_t^L, the uncertainty in grid load demand $P_{t,u}^L$ is as expressed in

$$P_{t,u}^L = \max\left(0, P_t^L(1 + \epsilon_t^L)\right) \tag{60}$$

where

$$\epsilon_t^{PV} \sim \mathcal{N}(0, \sigma_{PV}^2) \tag{61}$$

The hydrogen selling price $\pi_{t,u}^{H2}$ influences the decision to either produce hydrogen by thermal and electrical power diversion from the NPP or generate electricity by the NPP. The uncertainty in hydrogen price is captured as a variable price between a maximum price of π_{max}^{H2} and minimum price of π_{min}^{H2}. This is as expressed in (62)

$$\pi_{t,u}^{H2} \sim \mathcal{U}\left(\pi_{max}^{H2}, \pi_{min}^{H2}\right) \tag{62}$$

4. IMPLEMENTATION AND RESULTS

The proposed DRL-based optimization framework is implemented and evaluated on the IEEE 9-bus test system, as shown in Fig. 1. The network is modified to include two inverter-based resources—a photovoltaic (PV) system and a battery energy storage system (BESS)—connected at buses 5 and 6. A coupled NPP–HTSE-1 unit is integrated at bus 2, and an additional standalone HTSE unit is placed at bus 8. For model development and training, the DRL algorithm and grid model are constructed in MATLAB using MATPOWER. For real-time validation, the grid model is deployed on a real-time digital simulator using RSCAD, enabling high-fidelity assessment of the proposed control framework. The integrated NPP is rated at 1000 MW, HTSE-1 at 500 MW, HTSE-2 at 200 MW, the solar PV system at 100 MW, and the BESS at 50 MWh. Both HTSE units are constrained to operate at a minimum of 10% of their respective rated capacities.

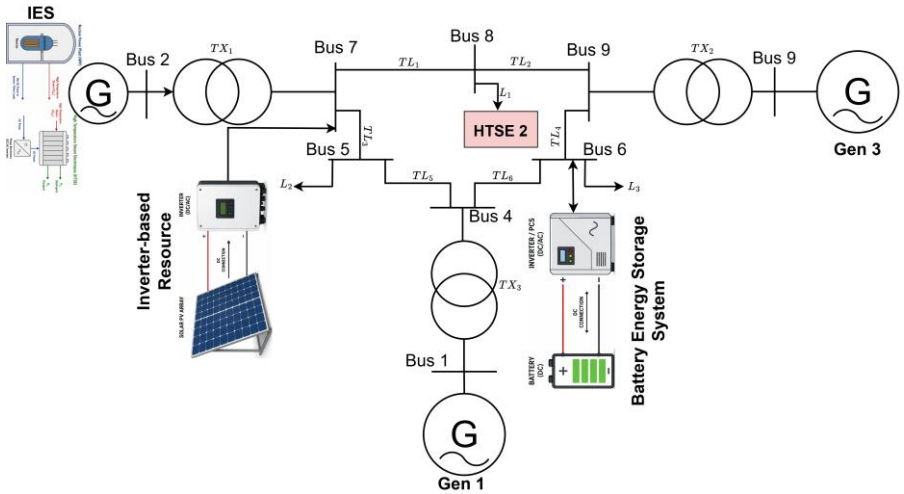

Figure 1: A modified IEEE 9-bus system for DRL-training and validation

4.1 Training and Validation Results

Figures 1 and 2 present the training convergence characteristics of the proposed robust TD3-based control framework for coordinated operation of the NPP, HTSE systems, PV generation, and grid energy exchange under uncertain renewable generation, load demand, and hydrogen pricing conditions. As shown in Fig. 2(a), the episodic reward increases rapidly during the initial training phase, indicating efficient exploration and rapid policy improvement. The reward subsequently stabilizes near 3.0×10^6, demonstrating convergence of the actor–critic learning process and stable long-horizon revenue optimization over the 48-hour scheduling horizon. Fig. 2(b) shows the per-episode reward trajectory together with the moving-average reward and variance band. The moving-

average reward converges after approximately 200 episodes, while the reduced variance during later training stages indicates improved policy consistency and robustness under randomized uncertainty realizations.

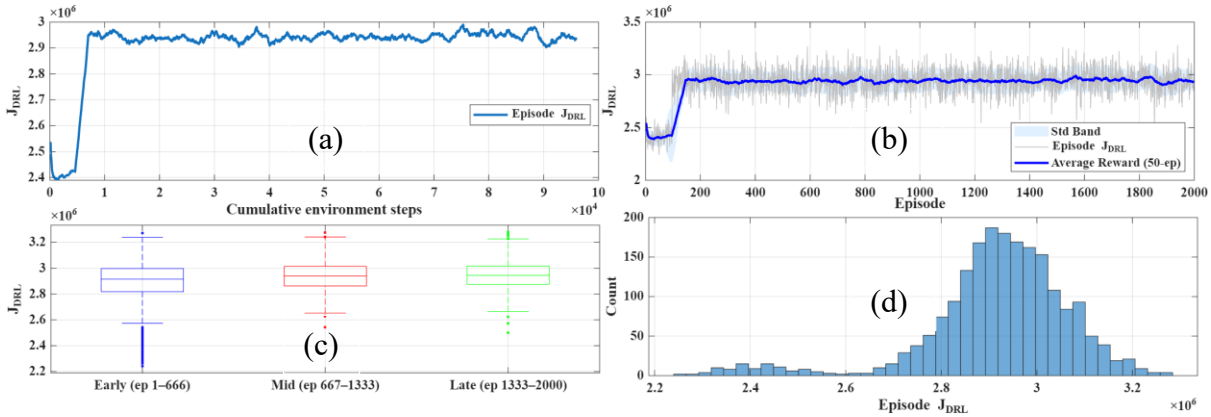

Figure 2: TD3 training performance (a) episode reward versus cumulative environment steps, (b) episode reward evolution with moving-average smoothing and variance band, (c) reward distribution across training stages (d) histogram of converged episodic rewards.

The reward distributions in Fig. 2(c) further confirm progressive policy improvement, where both the median and interquartile reward ranges shift upward from the early to late training stages. The histogram in Fig. 2(d) shows that most episodes converge around a high-reward operating region near 2.9×10^6, indicating stable and repeatable economic performance.

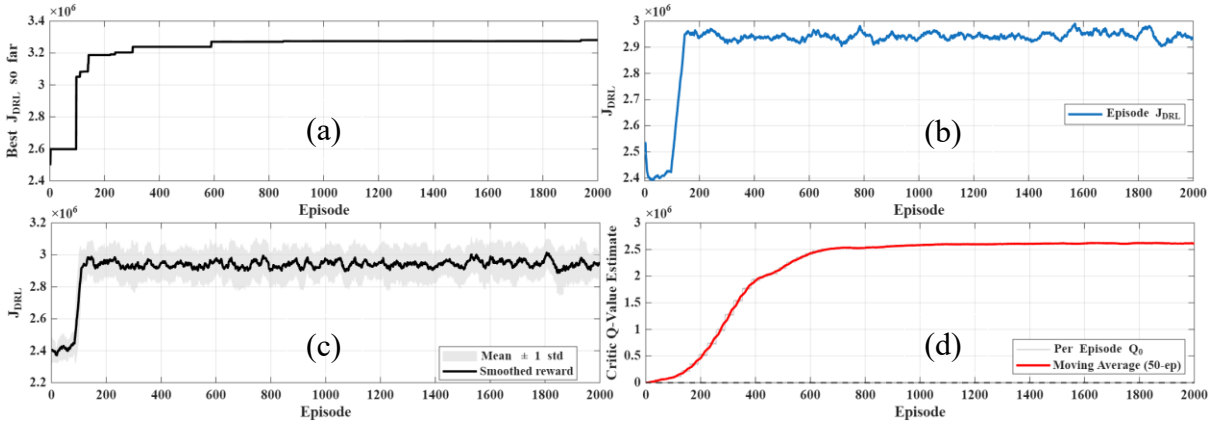

Figure 3: TD3 convergence diagnostics: (a) best episode reward evolution, (b) mean and smoothed episode reward with uncertainty bounds, (c) stabilized episodic reward trajectory after convergence, and (d) critic Q-value convergence during training.

Fig. 3(a) illustrates the monotonic improvement in the best-achieved episodic reward, demonstrating progressive discovery of increasingly profitable dispatch strategies. The mean and smoothed reward trajectories shown in Fig. 3(b) exhibit stable convergence with relatively small variance, indicating effective generalization across uncertain operating conditions. Fig. 3(c) further confirms stable learning behavior without significant oscillatory or divergent reward behavior during the later training stages.

The critic Q-value evolution shown in Fig. 3(d) demonstrates stable convergence of the learned state–action value function. The gradual saturation of the moving-average Q-value indicates effective temporal-difference learning and coordinated convergence between the actor and twin-critic networks. Overall, the results demonstrate that the proposed robust TD3 framework successfully learns economically optimal dispatch policies that balance electricity market participation and hydrogen production under stochastic operating conditions while maintaining stable training behavior and strong policy robustness.

4.2 Optimal Power Flow Results

The thermal and electrical dynamics of the coupled NPP-HTSE is as shown in Fig. 4. The temperature profiles shown in Fig. 4a indicate that the fuel, coolant, and steam temperatures remain within relatively narrow operating ranges throughout the 48-hour horizon despite continuous changes in HTSE power demand and thermal diversion.

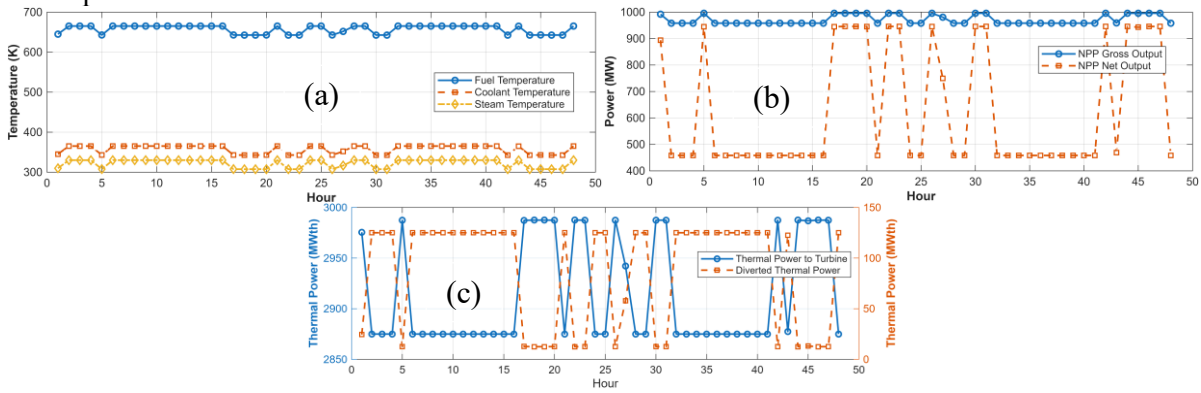

Figure 4: (a) Fuel, coolant, and steam temperature responses of the NPP (b) NPP electrical dispatch (c) thermal power delivered to the turbine and thermal power diverted for hydrogen production.

The fuel temperature remains approximately between 640–670 K, while the coolant and steam temperatures exhibit only minor fluctuations, demonstrating that the coupled NPP–HTSE operation does not introduce significant thermal instability into the nuclear plant dynamics. The electrical dispatch results shown in Fig. 4b illustrate the dynamic tradeoff between electricity market participation and hydrogen production. During periods of increased thermal diversion to the HTSE system, the net electrical output of the NPP decreases significantly relative to the gross electrical generation, indicating that a portion of the reactor thermal energy and electrical power from the NPP is redirected toward hydrogen production rather than electricity export. Conversely, when thermal diversion is reduced, the NPP net electrical output approaches the gross electrical generation level, indicating prioritization of electricity market participation. The thermal power distribution results shown in Fig. 4c shows this coordinated operational behavior. The thermal power delivered to the turbine decreases during intervals of increased diverted thermal power, while periods of reduced HTSE thermal demand allow greater thermal energy delivery to the turbine for electricity generation. This behavior demonstrates that the TD3 agent successfully learns the coupled electro-thermal tradeoff between hydrogen production and electricity revenue optimization under varying operating conditions.

The optimal control actions (variables) are as shown in Fig. 5. The HTSEs' dispatch profiles show that HTSE-1, which is directly coupled to the NPP, dynamically varies its power demand in response to system operating conditions, while the independent HTSE-2 operates near its maximum capacity for most of the scheduling horizon.

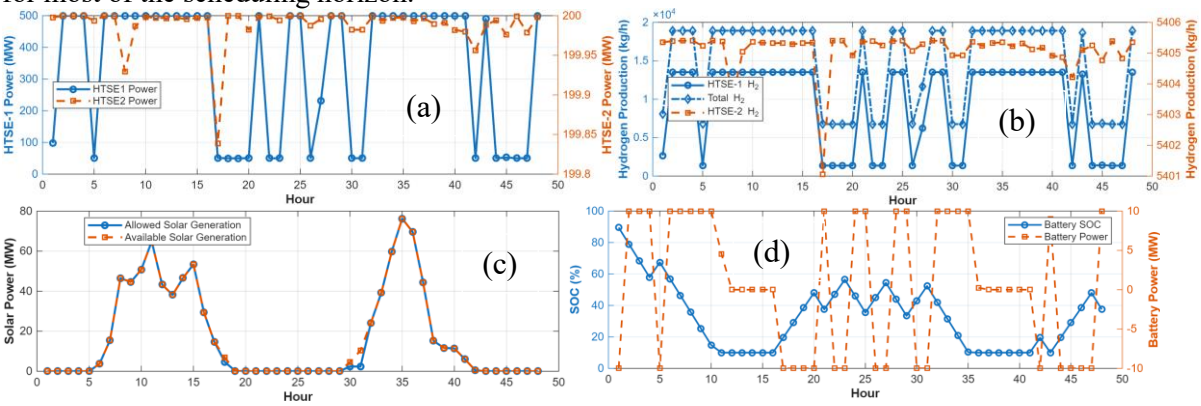

Figure 5: (a) Optimal HTSEs Power Demand (b) Hourly hydrogen production rates for HTSEs 1 & 2 (c) Optimal solar power dispatch and available power (d) Optimal BESS power and SOC

The hydrogen production results directly reflect the HTSE dispatch behavior. Variations in HTSE-1 power demand produce corresponding fluctuations in hydrogen production, whereas HTSE-2

maintains relatively stable hydrogen output due to its near-constant operating level. The total hydrogen production therefore remains comparatively stable despite dynamic power reallocation between electricity generation and hydrogen production.
The solar dispatch results indicate that the allowable solar generation closely tracks the available solar power profile with minimal curtailment throughout most of the scheduling horizon. This demonstrates effective utilization of renewable generation. The battery operation results show active charging and discharging behavior used to support system flexibility and energy balancing. The battery state of charge varies dynamically in response to renewable variability and dispatch requirements, while the battery power profile indicates alternating charging and discharging intervals that contribute to smoothing system operation and supporting coordinated multi-energy dispatch.

The hourly generated revenue based on then formulated reward function and the price signals is as shown in Fig. 6

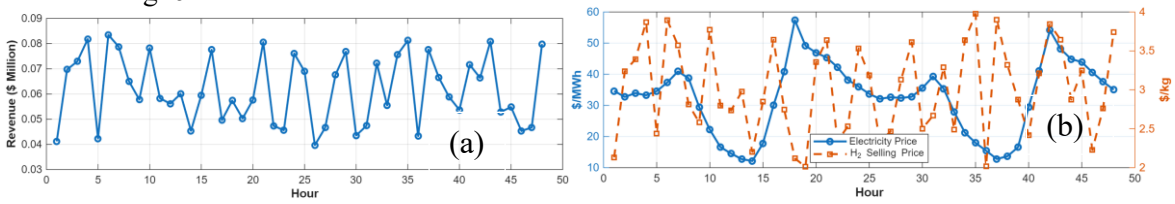

Figure 6:(a) Hourly generated revenue (b) Electricity and HTSE price variation

The revenue profile demonstrates the dynamic economic behavior learned by the proposed TD3-based control and optimization framework over the 48-hour scheduling horizon. As shown in Fig. 6a, the hourly revenue varies between approximately $0.04 million and $0.08 million depending on the prevailing electricity prices, hydrogen prices, renewable generation availability, and HTSE dispatch decisions. Higher revenue periods generally correspond to operating conditions where the agent successfully exploits favorable electricity market prices and hydrogen production opportunities simultaneously. The absence of severe revenue collapse or instability further indicates robust and economically consistent policy behavior under uncertain operating conditions. Fig. 4b shows electricity market prices and stochastic hydrogen prices during the scheduling horizon. The electricity price exhibits significant temporal variation, ranging approximately between $12/MWh and $57/MWh, while the hydrogen price fluctuates between approximately $2/kg and $4/kg due to the introduced uncertainty modeling. These varying market conditions create a continuously changing economic tradeoff between electricity export and hydrogen production.

Since the proposed DRL-based optimization framework is implemented on the IEEE 9-bus system, the resulting bus voltage profiles and the corresponding minimum and maximum voltages over the 48-hour scheduling horizon are illustrated in Fig. 7.

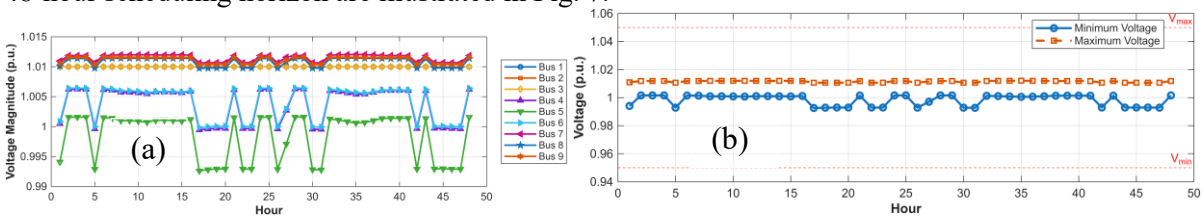

Figure 7:(a) Timeseries grid bus voltages (b) Bus maximum and minimum voltage

As shown in Fig. 7a, despite the dynamic HTSE dispatch, renewable generation variability, and evolving market-driven operating conditions, all bus voltage magnitudes remain within the prescribed voltage limits. This demonstrates that the proposed TD3-based control framework effectively maintains acceptable grid operating conditions while coordinating electricity generation and hydrogen production. Although minor voltage fluctuations occur during periods of increased HTSE demand and variations in solar output, the overall voltage trajectory remains well-regulated and dynamically stable. Furthermore, Fig. 7b confirms that the minimum and maximum bus voltages remain within the operational bounds throughout the entire scheduling horizon.

5. CONCLUSION

This paper presented a grid-aware, online, and real-time deep reinforcement learning-based optimization and control framework for coordinated operation of tightly coupled nuclear–hydrogen energy systems. The proposed framework integrates a physics-informed Twin-Delayed Deep Deterministic Policy Gradient (TD3) controller with dynamic models of nuclear power generation, high-temperature steam electrolysis (HTSE), battery energy storage, renewable energy resources, and AC power flow constraints to enable adaptive multi-energy system coordination under uncertain operating conditions.

Simulation studies conducted on a modified IEEE 9-bus system demonstrated that the proposed DRL framework successfully learns stable and economically optimal dispatch policies that dynamically coordinate electricity generation and hydrogen production in response to varying electricity prices, hydrogen market conditions, renewable generation uncertainty, and load fluctuations. The results showed stable training convergence, robust policy performance, effective renewable energy utilization, and dynamic battery coordination while maintaining all nuclear operational constraints and grid voltage limits throughout the scheduling horizon.

The developed framework also demonstrated the capability to regulate HTSE thermal and electrical demand in real time without introducing thermal instability into the nuclear plant. The coordinated electro-thermal dispatch strategy enabled flexible diversion of thermal and electrical energy between grid support and hydrogen production, thereby improving operational flexibility and supporting ancillary grid services such as voltage regulation and demand response. Furthermore, the integration of physics-informed operational constraints within the DRL environment improved policy robustness and ensured safe operation under stochastic system conditions.

Overall, the presented approach provides a scalable and adaptive control architecture for future integrated nuclear–hydrogen energy systems with high penetrations of renewable energy resources and flexible industrial loads. Future work will focus on extending the framework to larger transmission systems, incorporating detailed electrochemical degradation models for HTSE systems, integrating cybersecurity-resilient control mechanisms, and validating the proposed approach using hardware-in-the-loop and experimental test-bed implementations for practical deployment in real-world energy infrastructures.

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Nuclear Energy under DOE Idaho Operations Office Contract DE-AC07-05ID14517 and Advanced Fuels Campaign of the Nuclear Technology Research and Development program in the Office of Nuclear Energy. Accordingly, the U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript or allow others to do so, for U.S.

References

[1] T. O. Olowu, M. J. Casteel, J. W. Lambrechtsen and J. L. Hartvigsen, "Multi-Objective Control of High Temperature Electrolysis in a Microgrid," *2026 IEEE Green Technologies Conference (GreenTech)*, Boulder, CO, USA, 2026, pp. 1-6,

[2] T. O. Olowu, J. W. Lambrechtsen, J. L. Hartvigsen, and M. J. Casteel, “Dynamic Performance Analysis of High Temperature Steam Electrolysis System in an Integrated Energy Ecosystem,” in *Proc. ASME Int. Mech. Eng. Congr. Expo. (IMECE)*, vol. 6, Energy, Nov. 2025.

[3] M. Casteel, T. L. Westover, A. Shigrekar, T. Olowu, A. Ta, A. Lavernia, A. Zargari, B. Cheldelin, Ultra-high efficiency hydrogen production using a large-scale solid oxide electrolysis cell system, International Journal of Hydrogen Energy 157 (2025) 150283.

[4] H. Lange, A. Klose, W. Lippmann, L. Urbas, Technical evaluation of the flexibility of water electrolysis systems to increase energy flexibility: A review, International Journal of Hydrogen Energy 48 (42) (2023) 15771–15783.

[5] C. Bourasseau, B. Guinot, Hydrogen: a storage means for renewable energies, Hydrogen Production: Electrolysis (2015) 311–382.

[6] Y. Jiang and S. Ou, “Hydrogen storage capacity planning of nuclear-hydrogen integration under coupling of electricity and hydrogen,” *International Journal of Hydrogen Energy*, vol. 126, pp. 238–250, May 2025.

[7] Y. Zhou, B. Zhang, C. Xu, T. Lan, R. Diao, D. Shi, Z. Wang, W.-J. Lee, A data-driven method for fast ac optimal power flow solutions via deep reinforcement learning, Journal of Modern Power Systems and Clean Energy 8 (6) (2020) 1128–1139.

[8] Y. Zhou, W. Lee, R. Diao, D. Shi, Deep reinforcement learning based real-time ac optimal power flow considering uncertainties, Journal of Modern Power Systems and Clean Energy 10 (5) (2022) 1098–1109.

[9] Z. Yan, Y. Xu, Real-time optimal power flow: A lagrangian based deep reinforcement learning approach, IEEE Transactions on Power Systems 35 (4) (2020) 3270–3273.

[10] H. Nie, Y. Chen, Y. Song, S. Huang, A general real-time opf algorithm using ddpg with multiple simulation platforms, in: 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), 2019, pp. 3713–3718.

[11] A. R. Sayed, C. Wang, H. I. Anis, T. Bi, Feasibility constrained online calculation for real-time optimal power flow: A convex constrained deep reinforcement learning approach, IEEE Transactions on Power Systems 38 (6) (2023) 5215–5227.

[12] Z. Yan, Y. Xu, Real-time optimal power flow with linguistic stipulations: Integrating gpt-agent and deep reinforcement learning, IEEE Transactions on Power Systems 39 (2) (2024) 4747–4750.

[13] D. Cao, W. Hu, X. Xu, Q. Wu, Q. Huang, Z. Chen, F. Blaabjerg, Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices, Journal of Modern Power Systems and Clean Energy 9 (5) (2021) 1101–1110.

[14] J. Aldahmashi, X. Ma, Advanced machine learning approach of power flow optimization in community microgrid, in: 2022 27th International Conference on Automation and Computing (ICAC), 2022, pp. 1–6.

[15] J. Xie, W. Sun, Distributional deep reinforcement learning-based emergency frequency control, IEEE Transactions on Power Systems 37 (4) (2022) 2720–2730.

[16] A. M. Taher, S. H. A. Aleem, S. F. Al-Gahtani, Z. M. Ali, H. M. Hasanien, Modified deep reinforcement learning for frequency regulation in active distribution systems with soft open points, storage units and electric vehicles, Renewable Energy 256 (2026) 124537.