

# Real-Time Load Frequency Control for PV-Integrated Power Systems

A Deep Reinforcement Learning Approach for Enhanced Grid Stability

Prof. Pankaj Chaudhari  
Electrical Engineering  
Department  
Gyanmanjari Innovative  
University  
Bhavnagar, India

Prof. Disha M. Siddhpura  
Electrical Engineering  
Department  
Gyanmanjari Innovative  
University  
Bhavnagar, India

Prof. Jaydip Ranva  
Electrical Engineering  
Department  
Dr. Shubhash University  
Junagadh, India

Ms. Vibhuti Varu  
Independent Researcher  
Junagadh, India

Mr. Abhyuday Katariya  
Independent Researcher  
Surat, India

**Abstract** - The global shift toward sustainable energy has led to the extensive integration of photovoltaic systems into power grids. This paradigm shift, however, causes a significant reduction in system inertia, making grids highly vulnerable to frequency disturbances and a high Rate of Change of Frequency. Traditional control methods, such as fixed-gain controllers, are ill-equipped to manage the severe nonlinearities and high operational volatility inherent in these low-inertia, renewable-heavy systems. This research proposes an adaptive control solution utilizing Deep Reinforcement Learning (DRL) to address the Load Frequency Control challenge. By formulating the control problem as a Markov Decision Process, an intelligent agent learns an optimal, model-free policy that maximizes frequency stability and operational efficiency. The DRL framework inherently handles the uncertainty and computational complexity of the system. The approach is designed for scalability using multi-Agent DRL in interconnected systems, enhancing both performance and resilience. Validation metrics focus on minimizing frequency deviations and the Rate of Change of Frequency, aligning with regulatory standards like NERC's Control Performance Standards. The findings underscore that DRL is an essential, highly adaptable paradigm for securing the stability and security of future low-inertia power systems.

**Keywords** - Deep Reinforcement Learning; Load Frequency Control; Photovoltaic Systems; Grid Stability; Low-Inertia Systems; Multi-Agent Systems; Continuous Control

## I. INTRODUCTION

Background and Motivation: The Grid Inertia Crisis:

The global imperative for energy transition toward sustainable and clean sources has accelerated the large-scale integration of Renewable Energy Sources (RES), particularly photovoltaic (PV) systems, into the existing power infrastructure [1]. While

PV systems offer substantial environmental benefits, their intensive penetration fundamentally alters the operational dynamics of the power network. A critical technical challenge arising from this paradigm shift is the displacement of traditional synchronous generators (SGs) by inverter-based generation (IBG) [2]. Since PV systems interface with the alternating current (AC) grid via power electronic components, they lack the intrinsic rotational inertia and damping mechanisms traditionally provided by SGs [3].

This resulting inertia deficit renders modern power grids highly vulnerable to disturbances. Any power imbalance, whether caused by load fluctuations or unpredictable changes in PV output, translates directly into a higher Rate of Change of Frequency (ROCOF) and larger transient frequency deviations [3]. The intermittency inherent to solar output, such as rapid changes due to passing clouds, exacerbates these power mismatches, creating severe frequency disturbances, particularly in regions near the equatorial zone [4]. Consequently, maintaining frequency stability - a crucial indicator of reliable and secure power system operation - has become increasingly complex [5].

Conventional Load Frequency Control (LFC) methods, which often rely on fixed-gain Proportional-Integral-Derivative (PID) controllers or simplified system models, are frequently unable to maintain stability under these severe nonlinearities and rapid, stochastic dynamics [6]. The control framework must simultaneously manage the structural deficiency of low inertia and the operational volatility introduced by PV intermittency. This dual challenge requires an advanced, adaptive, and model-free control paradigm capable of handling the computational complexity, high dimensionality, and rapid decision-making necessary for future low-inertia grids [6].

The Promise of Deep Reinforcement Learning (DRL) for LFC:

Deep Reinforcement Learning (DRL) presents a viable solution to the challenges facing LFC in high-penetration PV systems. DRL combines the decision-making framework of Reinforcement Learning (RL), where an intelligent agent learns optimal actions through trial and error to maximize a cumulative reward signal, with the powerful feature extraction capabilities of Deep Neural Networks (DNNs) [8].

Unlike conventional control methods that require an explicit, exact mathematical model of the system, DRL operates in a

II. CHALLENGES AND LIMITATIONS OF CONVENTIONAL CONTROL

Dynamic Impacts of Photovoltaic Integration:

The widespread integration of PV systems introduces several interconnected stability challenges that strain traditional LFC mechanisms. The most immediate impact is the degradation of frequency stability, resulting in notable frequency disturbances [13]. This is directly attributable to the inertia deficit, causing systems to experience larger frequency deviations and severe ROCOF, which increases the likelihood of unintentional tripping of distributed PV and generators [3].

Beyond frequency, angle stability is also adversely affected as increasing PV penetration reduces system inertia and effectively increases generator reactance, negatively impacting generator transient stability [13]. Furthermore, large-scale PV integration can introduce voltage instability and exacerbate overvoltage and overload issues, often requiring proactive measures such as active power curtailment or careful management of reactive power control via advanced inverters [1].

Limitations of Conventional Frequency Support Methods:

Conventional methods struggle because they lack the necessary adaptability. Fixed-gain controllers, such as PIDs, cannot adapt

model-free manner, learning an optimal policy  $\pi$  (a state-action mapping) through interaction with the environment (the power system simulator) [10]. This capability allows the DRL agent to autonomously discover complex, non-linear control strategies necessary to stabilize the grid under the highly stochastic and uncertain conditions introduced by PV integration [11]. Furthermore, by framing the LFC problem as an optimal control task maximizing long-term rewards, DRL inherently shifts the control philosophy from merely correcting known frequency deviations to implementing preventive stability strategies, continuously adjusting resources to avoid critical frequency breaches [12].

to the rapid structural and operational changes characteristic of low-inertia grids, requiring constant and complex parameter adjustment [6].

Another widely adopted solution is Virtual Synchronous Generator (VSG) control technology, which attempts to replicate the mechanical characteristics (inertia and damping) of traditional SGs using power electronics [3]. While effective in theory, the practical implementation of VSG is complicated by the large number of parameters that must be accurately tuned and adjusted, limiting operational flexibility and responsiveness to diverse real-time grid conditions. This issue is so significant that researchers have proposed DRL, specifically algorithms like TD3, to adaptively replace the conventional VSG control module, thereby using the learning capability of DRL to autonomously manage the complex parameter optimization inherent in inertia emulation.

Model Predictive Control (MPC) offers robust control based on optimization; however, it requires a computationally expensive, precise mathematical model of the power system dynamics. Comparative studies indicate that while MPC performs well when modeling errors are minimal, its performance degrades significantly when system complexity introduces large modeling errors—a frequent occurrence in highly non-linear, stochastic PV-integrated systems [14]. The model-free nature of DRL circumvents this dependency on precise system knowledge.

TABLE I. Comparison Of Control Paradigms For LFC In Low-Inertia Grids

Control Method	Model Requirement	Adaptability to Uncertainty	Scalability/Computational Speed	Primary Limitation
PID (Fixed Gain)	Low	Low (Requires retraining/tuning)	High (Fast Execution)	Highly sensitive to system parameter changes [6]
VSG Control	Medium (Requires accurate emulation model)	Medium (Complex parameter adjustment)	Medium	Operational complexity, many parameters to tune [10]
MPC (Model Predictive Control)	High (Requires accurate dynamic model)	Medium	High (If model is simple)	Performance degrades significantly with large modeling errors [14]
DRL (Model-Free)	Low (Only requires high-fidelity simulator for training)	High (Learns optimal policy under stochastic dynamics)	High (Fast Policy Inference)	Requires extensive, expensive offline training [23]

Accurate PV System Modeling for DRL Training:

A crucial requirement for developing a reliable DRL-based LFC is the fidelity of the simulation environment used for

training. Although DRL produces a model-free control policy, the quality of the learned policy is entirely dependent on the accuracy of the "experience" gathered during the training phase.

Traditional LFC studies often rely on simplified first-order or second-order PV models, which fail to capture the dynamic and steady-state complexity of PV power plants connected to the grid [15]. For rigorous LFC studies, the modeling approach must accurately mimic the real power generation profiles and, critically, account for the effect of dynamic frequency deviation on the PV output power through inverter synchronization [15].

Advanced techniques, such as Artificial Neural Networks using Radial Basis Function (ANN-RBF), have been adopted to model the highly non-linear behavior of PV power plants (see Figure 1). These models take inputs such as temperature, irradiation level, humidity, and the dynamic frequency deviation, producing an output power prediction superior to simplified models [15]. The use of such physics-informed, high-fidelity modeling ensures that the DRL policy, once trained, can generalize and perform robustly when interacting with the real complexity of the PV-integrated grid [11].

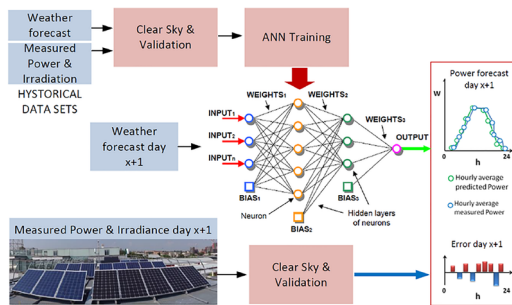


Fig. 1. High-Fidelity Photovoltaic (PV) Power Plant Modeling using Artificial Neural Network-Radial Basis Function (ANN-RBF) for Load Frequency Control (LFC) Studies.

### III. DEEP REINFORCEMENT LEARNING FRAMEWORK

Formulation as a Markov Decision Process (MDP):

To apply Deep Reinforcement Learning, the Load Frequency Control problem must be framed as a Markov Decision Process (MDP) (as conceptually illustrated in Figure 2) [5]. An MDP provides a mathematical framework for modeling sequential decision-making under uncertain outcomes, defined by the tuple  $(\mathbf{S}, \mathbf{A}, P, R)$ , representing the State, Action, Transition Probability, and Reward [9]. The objective of the DRL agent is to learn an optimal control policy  $\pi(a|s)$  that maximizes the expected cumulative discounted reward  $R_t$  over time [8].

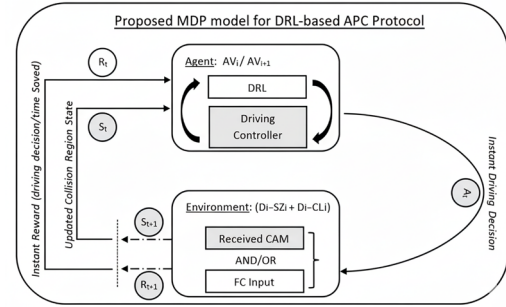


Fig. 2. Deep Reinforcement Learning (DRL) Framework for Load Frequency Control (LFC) formulated as a Markov Decision Process (MDP).

Definition of State Space ( $\mathbf{S}_t$ ):

The state space  $\mathbf{S}_t$  must encapsulate all necessary information at time  $t$  to satisfy the Markov property, allowing the agent to make optimal decisions based solely on the current observations. For LFC in a multi-area system, the state vector must be comprehensive:

Frequency Dynamics: Local frequency deviation ( $\Delta f$ ) and the Rate of Change of Frequency (ROCOF).

Inter-Area Exchange: Area Control Error (ACE) and tie-line power exchange deviations ( $\Delta P_{tie}$ ) [7].

Generation Status: Operational status and capacity limits of conventional generators.

Renewable and Storage Status: Current PV power output, possibly short-term forecasts, and the State of Charge (SOC) of any integrated Battery Energy Storage Systems (BESS) [18].

The complexity and high dimensionality of this state vector necessitate the use of Deep Neural Networks for efficient feature extraction and policy mapping.

Definition of Action Space ( $\mathbf{A}_t$ ):

The action space  $\mathbf{A}_t$  consists of the control signals that the DRL agent can issue to influence the power system frequency. Since LFC involves continuously adjusting mechanical governors and modulating inverter active power output, the action space is typically continuous [19].

The DRL agent's output represents normalized analog control signals, such as adjustments to the governor setpoints of committed synchronous units or active power adjustment commands directed to the Grid-Supporting Inverters within PV or BESS units [2]. The continuous nature of the action space dictates the use of advanced DRL algorithms designed specifically for continuous control tasks, such as Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), or Twin Delayed DDPG (TD3) [24].

#### Multi-Objective Reward Function Design ( $R_t$ ):

The reward function  $R$  is the most critical component of the MDP formulation, as it mathematically encodes the performance objectives and guides the agent's learning toward stability and efficiency [9]. For LFC, the reward must reflect the dual objectives of maximizing frequency stability while minimizing operational costs and control effort.

A standard multi-objective reward function  $r_t$  is constructed as a weighted penalty of undesirable outcomes [23]:

$$r = -\mu_1(\Delta f)^2 - \mu_2 \sum_{i=1}^n C_i + P_p$$

The critical element for ensuring system security is the Penalty Function ( $SP\_P$ ), which imposes severe negative rewards for transient frequency deviations that breach hard operational limits. For instance, regulatory practice requires frequency to be maintained within narrow bounds (e.g.,  $\pm 0.05$  Hz) [23]. A penalty function ensures that when this threshold is violated, a massive negative reward is applied :

$$P_p = \begin{cases} 0 & |\Delta f| < 0.05 \text{ Hz} \\ -K_{penalty} & |\Delta f| \geq 0.05 \text{ Hz} \end{cases}$$

Where  $K_{penalty}$  is a large negative value (e.g., -3 standard units, as proposed in some microgrid contexts) [23].

The design of the reward function allows the DRL agent to discover non-intuitive, cost-effective actions by continuously weighing the marginal cost of a control action against the expected future penalty associated with a trajectory toward instability. This effectively achieves multi-objective optimization [22]. Furthermore, advanced reward shaping, such as incorporating Lyapunov-based criteria into the reward function, serves a crucial role. This method integrates classical stability constraints directly into the model-free learning objective, offering a path to enhance DRL's robustness and bridge the gap toward obtaining formal stability guarantees required in safety-critical power system applications [9].

#### IV. DRL ALGORITHM SELECTION AND EFFICIENCY

##### Comparative Analysis of Continuous DRL Algorithms:

The selection of the appropriate DRL algorithm is paramount for achieving reliable, real-time LFC performance. Since the LFC action space is continuous, the choice typically narrows to continuous action Actor-Critic methods.

Proximal Policy Optimization (PPO) is an on-policy algorithm known for its high training stability and robustness [23]. However, PPO suffers from low sample efficiency, meaning it requires a large volume of data (many interactions with the simulator) to converge [21]. For environments that are computationally expensive to sample from, such as high-fidelity power system simulations, this can lead to excessively long training times [25].

Off-Policy Methods (DDPG, TD3, SAC), which utilize experience replay buffers, are significantly more sample efficient. This efficiency is essential when simulation requires significant computation, minimizing the total wall-clock time required for training [21].

Soft Actor-Critic (SAC): A highly stable algorithm that incorporates an entropy term to promote broad exploration, yielding strong performance in continuous control [20]. However, managing the associated temperature parameter, which controls policy stochasticity, can be complex and non-trivial [21].

Twin Delayed DDPG (TD3): An advancement over DDPG, TD3 mitigates the tendency of DDPG to overestimate Q-values, thus enhancing stability and performance [26]. TD3 generally offers performance comparable to SAC but provides parameters (like noise injection) that are easier to visualize and tune than SAC's temperature, making it a preferred choice for controlling physical systems where robust, predictable policy output is necessary.

For LFC in high-fidelity PV-integrated systems, which represent an expensive environment to simulate, the high sample efficiency of off-policy algorithms is crucial. Therefore, TD3 or SAC are strongly recommended [21].

TABLE II. DRL Algorithm Suitability for Real-Time LFC

Algorithm	Policy Type	Sample Efficiency	Stability in Continuous Control	Training/Deployment Trade-off
SAC	Off-Policy	High	Very High (Entropy regularization)	Excellent performance but complex tuning of temperature parameter [19].
PPO	On-Policy	Low	High (Robust)	Long training time due to sample inefficiency; simple implementation [20].
DDPG	Off-Policy	High	Medium (Prone to Q-value overestimation)	Shorter training time; better for expensive environments [20].
TD3	Off-Policy	High	Very High (Twin delayed updates)	Preferred for physical systems; balances efficiency and robustness [20].

Enhancing Training Efficiency: Model-Based and Informed DRL:

The primary computational challenge in scaling DRL control to large power systems is the substantial time required for policy



learning, which can take several hours even for state-of-the-art algorithms [25]. This training bottleneck is imposed mainly by the conventional power system simulators themselves.

To overcome this, novel model-based DRL frameworks are employed. This involves replacing the conventional simulator with a trained Deep Neural Network (DNN)-based surrogate model [25]. This surrogate model approximates the power system dynamics, allowing the DRL agent to learn the control policy much faster. Studies have demonstrated that this model-based DRL approach can reduce the necessary training time by a substantial margin (e.g., 87.7% reduction observed in voltage control applications) compared to model-free counterparts [25].

In addition to speed, ensuring the policy's generalization capability is vital. To mitigate the risk of performance degradation when the DRL policy is tested against conditions outside its training data range [16], methods like physics-informed DRL integrate domain knowledge and critical physical constraints directly into the network architecture or loss function [11].

## V. MULTI-AGENT CONTROL FOR SCALABILITY AND RESILIENCE

### Necessity of Decentralized/Distributed Control:

Load Frequency Control is typically executed across interconnected areas in bulk power systems. When scaling control to large systems, such as the IEEE 39-bus or 118-bus benchmarks, a centralized LFC approach rapidly becomes computationally prohibitive. Centralized control suffers from poor scalability and unmanageable computational dimensionality as the system size and complexity increase. The conventional solution, Multi-Area LFC, addresses this by coordinating control among localized units, primarily aiming to dampen inter-area frequency oscillations by adjusting governor references in each area [7].

### Cooperative Multi-Agent Deep Reinforcement Learning (MA-DRL):

Multi-Agent Deep Reinforcement Learning (MA-DRL) provides a sophisticated and inherently scalable framework for

decentralized LFC (refer to Figure 3 for architecture) [6]. MA-DRL decomposes the complex control problem into smaller, locally manageable MDPs, where multiple localized agents interact cooperatively with the shared electrical environment.

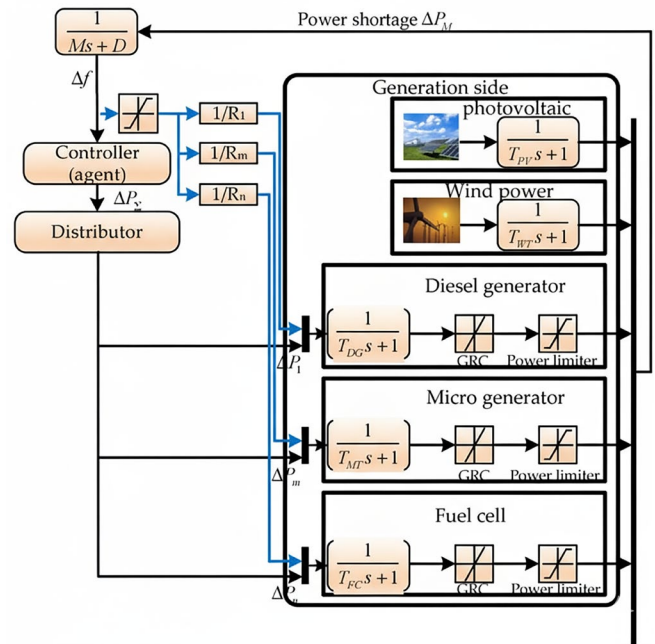


Fig. 3. Multi-Agent Deep Reinforcement Learning (MA-DRL) Architecture for Decentralized Load Frequency Control (LFC) in Interconnected Power Systems.

In this framework, each DRL controller operates autonomously within its area, relying only on local measurements such as frequency deviation ( $\Delta f$ ), Area Control Error (ACE), and tie-line power flow [6]. The agents are trained cooperatively - often utilizing algorithms like Multi-Agent DDPG (MA-DDPG) or Multi-Agent Actor-Attention-Critic (MAAC) - to minimize system-wide control errors caused by fluctuations in renewable generation and load. This approach has been proven effective in minimizing control errors in non-linear multi-area systems, including the complex New-England 39-bus system [24].

The intrinsic decomposition of the control task into local decisions effectively solves the centralized computational dimensionality issue [24]. Furthermore, a decentralized MA-DRL strategy significantly enhances system resilience by minimizing the reliance on continuous, high-bandwidth centralized communication, thereby reducing the impact of potential communication failures or targeted cyberattacks [6].

Advanced applications of MA-DRL extend beyond control signal generation, supporting Dynamic Algorithm Selection (DAS), which can be formulated as an MDP. The DRL agent can learn to dynamically switch between different underlying optimization or control algorithms to maximize overall performance.

VI. VALIDATION AND BENCHMARKING

System Model and Simulation Environment:

Rigorous validation of the DRL-based LFC strategy necessitates the use of standard, high-fidelity power system models. Recommended benchmarks for multi-area LFC research include the IEEE 39-bus (New-England) and IEEE 118-bus systems [5].

The standard system model must be structurally modified to account for high PV penetration, requiring the incorporation of

realistic PV generation models (such as ANN-RBF-based non-linear models [17]) and typically fast-response ancillary service providers like Battery Energy Storage Systems (BESS). The resulting comprehensive system model serves as the DRL environment, handling the complex physics (power flow, governor dynamics) and providing the state transitions and reward signals necessary for training the agent [8].

Key Performance Indicators (KPIs) and Stability Metrics:

TABLE III. Key Performance Indicators (KPIs) for DRL LFC Validation

KPI Category	Metric	Relevance to PV Integration	Regulatory Context
Transient Stability	Maximum Frequency Deviation ( $\Delta f_{\max}$ )	Indicates severity of power mismatch response.	Must remain within defined regulatory limits [3].
Transient Stability	Rate of Change of Frequency (ROCOF)	Directly linked to low system inertia. DRL must minimize peak $df/dt$ .	High ROCOF can trip protection relays [3].
Operational Reliability	NERC CPS1 and CPS2	Measure continuous frequency and tie-line error regulation accuracy.	Essential for meeting grid code and interconnection standards [20].
System Efficiency	Control Cost ( $\sum C_i$ )	Measures economic viability of the control solution.	Minimized via the multi-objective reward function [22].

Evaluation must focus on performance metrics that align with both transient stability requirements and operational regulatory compliance.

Transient Stability Metrics:

Maximum Frequency Deviation ( $\Delta f_{\max}$ ): Measures the peak transient error following a disturbance [3].

Rate of Change of Frequency (ROCOF): Directly measures the effectiveness of the control in countering the effects of low system inertia and must be minimized to prevent relay trips [3].

Settling Time ( $T_s$ ): The time required for frequency deviation to return and remain within acceptable bounds.

Regulatory Compliance and Operational Metrics:

Performance must be benchmarked against industry standards, most notably the North American Electric Reliability Council (NERC)'s Control Performance Standards (CPS1 and CPS2) [17]. Furthermore, the economic viability is assessed through the cumulative operational cost ( $\sum C_i$ ), which is intrinsically linked to the reward function design [23].

Comparative Performance Analysis:

The DRL controller must be benchmarked against established methods under identical stress conditions.

Classical Controllers (PID/Fuzzy): DRL-based LFC is expected to deliver superior transient performance, exhibiting

reduced overshoot (e.g., 25% less compared to classical PID) and a lower Integral Time Absolute Error (ITAE) due to its adaptive nature [23].

VSG Control: DRL policies demonstrate greater robustness against system uncertainties and the need for complex, manual parameter adjustment inherent in VSG implementations.

Model Predictive Control (MPC): DRL generally performs better than MPC when the power system exhibits high degrees of non-linearity and uncertainty, meaning the modeling errors are large [14].

VII. Practical Implementation and Robustness

Real-Time Deployment Constraints and Computational Latency:

While DRL demonstrates significant efficacy in simulated environments, its successful translation to real-world power system operation faces practical hurdles, including ensuring real-time adaptability and reliability during disturbances.

Real-time LFC requires extremely low latency for issuing control signals. Although the computational cost of the DRL policy during deployment (inference) is significantly lower than the cost of training [27], the decision-making cycle must be nearly instantaneous. This necessitates implementing the trained policy network (DNN) on dedicated, low-latency computational hardware, such as Field-Programmable Gate Arrays (FPGAs) or high-performance edge computing units, to effectively manage CPU usage and tail latency [28].

### Hardware-in-the-Loop (HIL) Validation Methodology:

To ensure that the theoretical advantages of DRL translate into practical feasibility, validation must proceed through a Hardware-in-the-Loop (HIL) methodology [29].

HIL testing involves implementing the DRL controller on the actual physical hardware intended for deployment (e.g., a dedicated controller box) and connecting it in a closed-loop with a Real-Time Digital Simulator (RTDS) that models the power grid dynamics [30]. HIL allows for the rigorous evaluation of control performance under real-world operational constraints, including realistic communication delays, sensor noise, and the latency imposed by the control hardware [29]. Successful HIL tests, demonstrating efficient frequency arrest and stability under simulated frequency events, confirm that the DRL control scheme is robust enough for field implementation [30].

### Robustness against Uncertainty and Cybersecurity Threats:

The success of DRL hinges on its ability to generalize, meaning the policy must reliably handle extreme PV power output fluctuations or stochastic load changes that may not have been explicitly encountered during offline training [31]. Furthermore, modern LFC systems are Cyber-Physical Systems (CPS) and are highly susceptible to security breaches, such as False Data Injection (FDI) attacks [16]. DRL techniques can be integrated to enhance system resilience, achieving high detection accuracy (e.g., over 99%) and low latency in identifying frequency characteristic attacks, thereby serving simultaneously as the control mechanism and the security layer [31].

## VIII. CONCLUSION

The large-scale integration of photovoltaic systems presents a fundamental challenge to grid stability, primarily through a drastic reduction in system inertia, which results in severe frequency deviations and high ROCOF during load and PV intermittency events. Conventional control methods are incapable of reliably managing this complexity due to their reliance on accurate models and static parameters.

The Deep Reinforcement Learning paradigm, leveraging continuous control algorithms such as TD3 or SAC, offers a powerful, model-free alternative. By formulating LFC as a Markov Decision Process with a multi-objective reward function that explicitly balances frequency stability ( $\Delta f$ ) against economic cost ( $C_i$ ), DRL agents learn highly adaptive and efficient control policies. Scalability to multi-area systems is achieved through the multi-Agent DRL framework, which allows for decentralized, local control execution while maintaining cooperative, system-wide optimization.

The practical deployment of these DRL policies demands rigorous validation. HIL testing is mandatory to confirm the DRL controller's capability to operate within real-time computational and communication constraints. Furthermore, robustness must be guaranteed not only against physical

uncertainties but also against cyber threats. The evolution toward DRL-driven LFC is essential for maintaining the stability and security of low-inertia, high-penetration PV power systems of the future.

## ACKNOWLEDGMENT

The authors sincerely thank Gyanmanjari Innovative University and the Electrical Engineering Department for their support. We also appreciate the guidance, encouragement, and valuable feedback from our colleagues and faculty members, which greatly contributed to the completion of this work.

TABLE IV. List of Abbreviations

Abbreviation	Description
ACE	Area Control Error
ANN-RBF	Artificial Neural Network using Radial Basis Function
BESS	Battery Energy Storage Systems
CPS	Control Performance Standards
DDPG	Deep Deterministic Policy Gradient
DNN	Deep Neural Networks
DRL	Deep Reinforcement Learning
FPGA	Field-Programmable Gate Array
HIL	Hardware-in-the-Loop
IBG	Inverter-Based Generation
KPI	Key Performance Indicator
LFC	Load Frequency Control
MA-DRL	Multi-Agent Deep Reinforcement Learning
MDP	Markov Decision Process
MPC	Model Predictive Control
NERC	North American Electric Reliability Council
PID	Proportional-Integral-Derivative
PV	Photovoltaic
RES	Renewable Energy Sources
ROCOF	Rate of Change of Frequency
SAC	Soft Actor-Critic
SGs	Synchronous Generators
SOC	State of Charge
TD3	Twin Delayed DDPG
VSG	Virtual Synchronous Generator

## REFERENCES

- [1] M. Dreidy, H. Mokhlis, and S. Mekhilef, "Inertia response and frequency control techniques for renewable energy sources: A review," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 144–155, 2017.
- [2] J. Morren, S. W. H. de Haan, W. L. Kling, and J. Ferreira, "Wind turbines emulating inertia and supporting primary frequency control," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 433–434, Feb. 2006.

- [3] F. Gonzalez-Longatt, "Impact of synthetic inertia from wind power on the protection of low inertia power systems," in *Proc. IEEE PowerTech*, Grenoble, France, 2013.
- [4] X. Zhang and G. Hug, "Real-time robust power system frequency control via deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3397–3407, Jul. 2021.
- [5] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York, NY, USA: Springer, 2008.
- [6] R. K. Sahu, S. P. Ghosh, and S. Panda, "A novel hybrid LFC technique for multi-area systems with renewable generation," *IET Gener. Transm. Distrib.*, vol. 10, no. 9, pp. 2172–2184, 2016.
- [7] H. Bevrani, *Robust Power System Frequency Control*. New York, NY, USA: Springer, 2014.
- [8] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [9] S. Das, N. R. Chaudhuri, and R. G. Kavasseri, "Lyapunov-based reinforcement learning for power system frequency regulation," *Electr. Power Syst. Res.*, vol. 212, p. 108423, 2022.
- [10] T. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," in *Proc. ICLR*, 2016.
- [11] A. St-Germaine, A. Kargarian, and D. Nguyen, "Physics-informed reinforcement learning for power system stability," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 631–641, Jan. 2022.
- [12] D. Silver *et al.*, "Deterministic policy gradient algorithms," in *Proc. ICML*, 2014.
- [13] N. Hatziaargyriou *et al.*, "Stability definition and classification of power system stability," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3271–3281, Jul. 2021.
- [14] E. Camacho and C. Bordons, *Model Predictive Control*, 2nd ed. London, U.K.: Springer, 2007.
- [15] M. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using ANN-RBF models," *Renew. Energy*, vol. 33, no. 7, pp. 1481–1487, 2008.
- [16] Y. Liu *et al.*, "False data injection attacks against power grid: A survey," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 1, pp. 594–630, 2017.
- [17] NERC, *Balancing and Frequency Control*, N. Amer. Electr. Rel. Corp., Atlanta, GA, USA, Tech. Rep., 2011.
- [18] A. Oudalov, D. Chartouni, and C. Ohler, "Optimizing a battery energy storage system for primary frequency control," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1259–1266, Aug. 2007.
- [19] H. K. Khalil, *Nonlinear Control*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.
- [20] T. Haarnoja *et al.*, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. ICML*, 2018.
- [21] Y. Duan *et al.*, "Benchmarking deep reinforcement learning for continuous control," in *Proc. ICML*, 2016.
- [22] Y. Zhou, Z. Hu, and Y. Min, "Reinforcement learning-based multi-objective optimal frequency control in hybrid renewable grids," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2176–2187, May 2022.
- [23] S. Deb *et al.*, "Load frequency control of multi-area systems using deep reinforcement learning," *Int. J. Electr. Power Energy Syst.*, vol. 135, p. 107563, 2022.
- [24] M. Chen and S. H. Low, "Multi-agent DDPG for decentralized power system control," in *Proc. IEEE CDC*, 2020.
- [25] Z. Wang *et al.*, "Model-based deep reinforcement learning for power system voltage control," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1474–1486, Mar. 2022.
- [26] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. ICML*, 2018. (TD3)
- [27] X. Chen *et al.*, "Real-time DRL-based power system emergency control," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4862–4873, Nov. 2020.
- [28] NVIDIA, *Real-Time Inference Optimization for Edge AI*, NVIDIA White Paper, 2021.
- [29] OPAL-RT Technologies, *Hardware-in-the-Loop (HIL) Testing Guide*, Montreal, Canada, 2020.
- [30] A. M. Annaswamy *et al.*, "Real-time simulation for power system control using HIL frameworks," *Proc. IEEE*, vol. 108, no. 9, pp. 1477–1495, Sep. 2020.
- [31] J. Wang *et al.*, "Reinforcement-learning-enhanced intrusion detection in cyber-physical power systems," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4230–4242, Sept. 2021.