Towards Understanding the Challenges of Applying Reinforcement Learning to the Power Grid

Matthew Schlegel[1], **Martha White**[2,3], **Matthew E. Taylor**[2, 3, †], **Mostafa Farrokhabadi**[1,†]
{matthew.schlegel, mostafa.farrokhabadi}@ucalgary.ca,
{whitem, mtaylor3}@ualberta.ca

[1]**Schulich School of Engineering, University of Calgary, Canada**
[2]**Department of Computing Science, University of Alberta, Canada**
[3]**Alberta Machine Intelligence Institute (Amii)**

[†] Indicates equal supervisory roles

Abstract

This paper presents a high-level overview of the challenges facing the application of reinforcement learning (RL) to power system control. As power systems evolve with increasing complexity and uncertainty, RL offers promising approaches for adaptive control. Building on the framework of Dulac-Arnold et al. (2021), we explore how power systems embody and expand on these challenges. We also introduce two additional challenges that arise from the power systems domain: (1) learning policies over multi-timescale action spaces, and (2) fostering effective collaboration between RL agents and human operators. By outlining these challenges for the power systems domain, this work aims to enable future research and collaborative efforts between the power systems and RL communities.

1 Introduction

The widespread adoption of renewable and distributed energy resources challenges current modeling and control tools in power systems (Pfenninger et al., 2014; Lopion et al., 2018). Reliable, scalable, and economical control of the grid with these new energy resources is growing in complexity, uncertainty, and volatility (Chen et al., 2022). With these rising challenges come new opportunities to develop paradigms and algorithms for a more sustainable future.

To enable stable decarbonization of the power grid, traditional operating paradigms and control strategies need to be rethought (Ilic & Jaddivada, 2022). The current control framework for power systems control is strictly hierarchical, with clear spatiotemporal boundaries between different control hierarchies. However, the increasing penetration of untraditional resources has challenged the adequacy of conventional control methods inter- and intra-hierarchies (Farrokhabadi et al., 2020). Thus, control strategies derived from the traditional control perspective may result in suboptimal and, in extreme cases, unstable power grid operation (Cardell & Ilic, 2004; Keyhani & Chatterjee, 2012). Moreover, the use of legacy controllers is one reason for the increasing need for operational reserves, which are often emissions-producing generation. Consequently, operators keep the penetration of these untraditional resources below a certain limit (Yang et al., 2020b).

Reinforcement learning (RL) is a potential candidate to address emerging challenges in power system control (Sutton & Barto, 2018; Chen et al., 2022; Bertozzi et al., 2024; Yu et al., 2024). Due to the flexibility of RL, it has been successfully applied to several complex control systems where classical approaches have failed or performed suboptimally (Luo et al., 2022; Degrave et al., 2022).

A large body of work has proposed various forms of RL for a diverse range of applications in power systems (Li & Yu, 2020; Chen et al., 2022; Yu et al., 2024). Although previous RL and machine learning methods have shown promising attributes in power system modeling and control, they have not yet been adopted by industry leaders (Chatzivasileiadis et al., 2022; Chen et al., 2022; Bertozzi et al., 2024; Yu et al., 2024).

Previous review articles have considered the application of RL to specific problem settings in power system control, such as in Chen et al. (2022) and Zhang et al. (2020), or on a specific design principle such as safety in RL (Yu et al., 2024). In contrast, this paper focuses on how the challenges proposed by Dulac-Arnold et al. (2021) apply to power systems. We hope that this framing will enable new areas of collaboration and research while targeting key areas in which RL is currently deficient for power system control.

2 Background

This section reviews RL and power grid control. Our goal is to provide familiarity with the ideas and jargon in both fields to contextualize the discussion in Section 3.

2.1 Reinforcement Learning

The dynamics of the RL problem setting is most often described as a Markov decision process (MDP) (Sutton & Barto, 2018), where an agent (in our case a control algorithm) interacts with an environment (the power grid). Given a state $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^n$ and an action $a \in \mathcal{A} \subset \mathbb{R}^b$, the environment transitions to a new state $\mathbf{s}' \in \mathcal{S}$ according to the transition probabilities $\mathrm{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, \infty)$[1]. A reward is received for each transition according to the (human-designed) reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. The goal of an RL agent is to maximize the sum of cumulative (discounted) rewards, that is, the discounted return $G_t = \mathbb{E}[\sum_{i=t}^{T} \gamma^{i-t} r_i]$ with discount $\gamma \in [0, 1)$[2] and (potentially infinite) horizon $T \in [t, \infty]$ (the time that the agent acts in the environment). The agent learns a policy $\mu : \mathcal{S} \times \mathcal{A} \to [0, \infty)$[1] that describes the probability distribution of actions for a given state. The agent samples an action from this distribution for every state.

2.2 Power System Control

The power grid is an interconnected set of electronic devices that can generate, transfer, distribute, and consume power. To effectively control and effect the dynamics of a power system, the control surface is decomposed into several axes (Schweppe & Mitter, 1972; Palizban & Kauhaniemi, 2015): modes of operation, the temporal horizon of decisions, physical hierarchies, and functionalities. In this section, we provide an overview of these axes.

- **Modes of operation**: During the operation of the power grid, due to various uncontrollable events, the dynamics of the system can demand radically different plans of action (Schweppe & Wildes, 1970). These events (or contingencies) include exogenous disturbances to the generation-consumption equilibrium and the failure of any component of the grid. Examples of exogenous disturbances include weather events that affect generation output, cyber-security events, and normal or sudden changes in electricity consumption. Examples of component failures include the loss of a generation unit or transmission line. Instead of designing a controller that manages all possible conditions, the specification of four discrete modes (normal, preventative, emergency, and restorative) allows controllers to be designed for specific needs (Schweppe & Mitter, 1972).
- **Temporal Horizons for Decision Making**: Several temporal horizons for decision making naturally emerge from the system. The temporal horizons can range from microseconds to minutes, hours, or even days and weeks, depending on the specific problem being addressed (Hatziargyriou

[1] When considering discrete states and actions, the distributions will be probability mass functions as opposed to probability distribution functions described here.

[2] The discount can be 1 if the horizon of the control problem is finite.

et al., 2021). At the smallest horizon of microseconds are the controllers responsible for maintaining operational stability in view of ultra-fast dynamic phenomena referred to as electromagnetic transients. At longer horizons, up to hours, is set-point dispatch for power generation derived from the optimal power flow problem (Wood et al., 2013; Ela et al., 2010).

- **Size of grid**: Although power grids of any kind have decompositions analogous to those discussed this far, the size and purpose of a grid can change physical dynamics and thus the available control mechanisms, responsibilities, fail-safes, and interactions with an operator (Olivares et al., 2014; Farrokhabadi et al., 2020). At the largest size, there are synchronous interconnections, such as the Western Interconnection, spanning large areas, often across political borders. Within each interconnection, there could be several regional operators responsible for various sections of the grid. For example, the Alberta Electric System Operator (AESO) is responsible for centralized market-based dispatch and real-time reliability responsibilities of the Alberta Interconnected Electric System. At the smallest scale are microgrids, a cluster of electricity demand and generation resources at medium to low voltage levels. By definition, a microgrid can operate in synchronous (i.e., the frequencies are matched with the larger grid) or islanded (i.e., the grid is operated independently of the main grid) modes (Farrokhabadi et al., 2020).
- **Layers of Control**: Traditionally, the control framework is separated into three strict layers. The decomposition of these layers (or levels) is primarily done based on the temporal characteristics of various controls (Schweppe & Mitter, 1972). At the fastest level, the *Primary* control happens closest to the device, e.g., autonomous regulation of a local generation output. Higher control layers pertain to managing optimal device set-points for improved economics, restoration of voltage and frequency to their nominal values, as well as coordinated operation of interconnected sub-grids for enhanced reliability and economics. Specific control functionalities belong to each control layer depending on the extent of their impact on the system and their temporal horizon. Without loss of generality, some examples of such functionalities that may benefit from an RL-based control are provided below (Chen et al., 2022):
  - *Frequency regulation* refers to the control actions necessary to maintain the system frequency within acceptable operational limits. Conventionally, this is done by balancing power generation and demand, primarily through regulating the intake fuel of traditional generators.
  - *Voltage Control* is the maintenance of voltages across the network within an acceptable interval. This is important to maintain the health of the electrical system equipment as well as the stability of power flow in the system. Conventional voltage regulators include synchronous generators, tap-changing transformers (devices that can change the ratio of voltages across their input and output), and capacitor banks.
  - *Energy management* covers a broad range of constrained optimization applications to enhance the stability, reliability, sustainability, and economics of energy delivery. This can be at the device level, for example, minimizing the total cost of charging an electric vehicle, or at the system level, e.g., optimal coordination of multiple energy resources for a reduced carbon footprint (Chen et al., 2022).

3 Challenges

In this section, we outline how the challenges presented by Dulac-Arnold et al. (2021) apply to the power system control problems described above.

3.1 Main Challenges

1) Being able to learn on live systems from limited samples: In power systems control, it is unlikely that an RL agent will be deployed to learn online without prior training. First, there are stringent requirements for the safety and reliability of the power supply, resulting in considerably conservative control implementation practices. Second, simulation models are relatively pervasive in the power system domain and can be leveraged for training RL agents. Although the challenge

of learning from limited samples is reduced by the use of such simulators, the challenge is not fully overcome. High-fidelity simulators may be slow to generate enough data for data-hungry RL algorithms. In this context, sample-efficient RL algorithms are of high importance, specifically for control functionalities with high execution rates such as those belonging to the primary control layer. There have been several directions to address this challenge, including: physics-informed RL (Cao et al., 2024; Du et al., 2022; Hossain et al., 2023; Li et al., 2024; She et al., 2024; Wang et al., 2024; Wu et al., 2025), reward shaping (Lu et al., 2023), and model-based RL (Hossain et al., 2024). Another interpretation of this challenge is to overcome the sim-to-real gap (Peng et al., 2018; Zhao et al., 2020), which has been underexplored in the power system control literature.

2) System Delays: Delays can occur in all control hierarchies and in both sensing and acting mechanisms. Telemetry delays can occur when there are communication delays or unreported device faults. This can cause old observational data or default values to be displayed instead of the real values. Computational bottlenecks due to outdated infrastructure can also cause communication delays. Operational constraints can also delay when a decision made by an RL agent is fully realized. For example, a generator or power plant can take hours to meet a power set-point due to physical limitations or organizational inertia (Schweppe & Mitter, 1972). Delays in sensing and acting can also be caused by physical limitations of a device. For example, tap-change transformers will take several seconds to react to a command partially due to the physical mechanism. Finally, there could be operational conditions for which the causal effects of control actions or disturbances can take a significant amount of time to manifest in sensor observations. For example, certain control actions may trigger inter-area oscillations that would take a significant amount of time to grow and manifest in the system.

3) Learning and acting in high-dimensional state and action spaces: The development of new sensor devices and emerging smart electronics has spurred a new generation of data-driven approaches to power system control (Bertozzi et al., 2024). Although new sensors and devices with more complex control mechanisms mean greater flexibility in control, RL faces uphill battles in learning policies from large state and action spaces (Dulac-Arnold et al., 2021), but may be better suited than currently deployed control systems. In particular, the action spaces of many power system control functionalities are hybrid discrete-continuous with many dimensions (Chen et al., 2022). For instance, in voltage control, an RL agent might need to select discrete control modes (e.g., switching capacitor banks or inverter modes), while simultaneously adjusting continuous control parameters (e.g., power outputs of inverter-based distributed energy resources). The different devices may also be controlled on different timescales. Only a small number of studies have applied RL to hybrid action spaces in power systems (Yang et al., 2020a; Gao et al., 2022), and more work is needed to improve RL in hybrid action settings.

4) Reasoning about system constraints that should never or rarely be violated: Ensuring the reliability, stability, and general health of the system is critical to approaching a power system control problem. The power grid is responsible for continually delivering energy to consumers. This could cause rolling, or, in worst-case scenarios, total power supply interruptions (Pourbeik et al., 2006; Chen et al., 2022; Chatzivasileiadis et al., 2022; Yu et al., 2024). There are many types of safety constraints in power systems. In frequency regulation, the frequency is indirectly controlled through the power generation of each generator. The agent's actions are explicitly constrained through the minimum and maximum bounds around the generator's power output, while the agent cannot explore frequencies outside nominal values (making frequencies implicitly constrained) (Gu et al., 2022; Yu et al., 2024). On the other hand, in voltage control, the main instantaneous constraint is often to ensure the voltage is within a reasonable range (that is, an instantaneous implicit constraint), while other cumulative costs associated with certain actions are included in the reward function (Gu et al., 2022; Yu et al., 2024). We recommend the work of Yu et al. (2024) for a detailed review of safe RL in power systems (including frequency regulation, voltage control, and energy management), and Gu et al. (2022) for an in-depth discussion of safe RL in general.

5) Partially observability, non-stationary, and stochasticity: An emerging concern in power system control is the increasing stochasticity of untraditional energy sources. The stochasticity of

renewable devices can cause major voltage disturbances (Sun et al., 2019). In addition to the challenges of new devices, power systems are constantly changing. In normal operation of the power grid, the topology is changing due to operational decisions, line faults, or long-term planning. These topology changes can be overcome through the inclusion of topology as a state variable into value functions and policies, or through transfer learning. Finally, it is unlikely that the state space available to an agent is Markov even when there are no sensor delays (Challenge 2) or topology changes.

6) Learning from multiobjective or poorly specified reward functions: When applying RL to different areas of power system control, considerable effort should go into developing a reward function. The objective function of a controller is dictated by several axes in power system control. Chen et al. (2022) discusses some of the main features of the reward function for several control problems in power systems (voltage control, frequency regulation, and energy management), and below we discuss some of the other axes to consider. These sources are not comprehensive.

- **Multiple objectives**: RL agents at every level, mode, and control horizon will have to balance many objectives when learning a policy. This is especially true for the optimal power flow (OPF) problem, where the objective can include economic concerns, minimizing power loss, maximizing power quality, environmental impacts, and many more (Frank et al., 2012). In frequency regulation at the secondary level (automatic generation control (AGC)), the reward can simply discourage the amount of required change in area generation (referred to as area control error) by giving a negative reward whenever this value goes beyond a fixed limit (Imthias Ahamed et al., 2002) or including the cost of power generation and keeping frequencies within the nominal range (Li & Yu, 2020).
- **Different modes of operation**: As discussed in Section 2, there are multiple modes in which the grid operates depending on the state of the grid. In the different modes, the reward function could be radically different. For example, in preventive or emergency mode, the economic and environmental objectives will likely be ignored in favor of ensuring the stability and health of the system (Schweppe & Mitter, 1972; Palizban & Kauhaniemi, 2015; Frank et al., 2012). Recovering from blackout also requires a different set of objectives (Wu et al., 2024).
- **Multiple objectives when straddling different control hierarchies**: Although traditional control approaches have only taken responsibility in a single layer, more flexible approaches, such as RL, may allow a controller to implement better policies by straddling multiple hierarchies. Building a reward function that balances the different responsibilities is challenging (Chaturvedi et al., 2024). We believe that this is an underexplored but impactful research area in power system control.
- **Operator influence and shaping**: One key factor in power systems (and any critical infrastructure or utilities project) is incorporating operator influence into the agent's behavior. Operators can apply reward shaping to guide the agent towards desirable behavior (Ibrahim et al., 2024). The main challenge is not only to numerically describe the operator's requirements but also to design the reward function so that it does not cause unintended behavior (Knox et al., 2023).

7) Being able to provide actions quickly, especially for systems requiring low latencies: In many of the control problems in power systems, decisions occurring at the secondary and tertiary control levels happen at relatively long temporal horizons (i.e., multiple seconds to hours). As discussed in Section 2.2, there are problems that require microsecond-level precision at the primary level. In some cases, a deep RL agent may not be fast enough to handle such a control frequency, but there may be opportunities to use RL or other ML techniques to tune classic controllers that operate at the required speeds (Bevrani & Shokoohi, 2013; Chaturvedi et al., 2024).

8) Training offline from the fixed logs of an external behavior policy: Offline RL has been applied to several areas and problem settings within power system control (Chen et al., 2022) but it is often presented as a solution to another of the above challenges, not as a particular challenge that needs to be overcome. For example, offline RL has been used to learn safe policies (Yu et al., 2024) following expert trajectories with the incorporation of imitation learning in frequency regulation (Lesage-Landry & Callaway, 2022). Another example is to learn a surrogate model and warm start policies in emergency voltage control (Hossain et al., 2024). Although there have been other

uses of offline RL in power systems (Chen et al., 2022; Yu et al., 2024), these approaches often use simulators to generate data. In the future, we expect this challenge to become more prevalent in relation to Challenge 1 (the sim-to-real gap) and a growing ecosystem of open simulators and datasets as data-driven research becomes more prevalent in power system control problems (Wiese et al., 2019; Zheng et al., 2022; Lara et al., 2024).

9) Providing system operators with explainable policies: This is a key challenge in the deployment of RL agents in a real-world power system. Due to the major consequences of mismanagement of the grid, operators are ultimately responsible for the decisions that are made. For any new autonomous system, either the dynamics of the controller should be so well analyzed that they will never produce surprising output (e.g., a droop controller for frequency regulation), or the operators are able to monitor the controller's decisions and intervene when necessary. Automatic generation control is an example of the latter, in that operators monitor the decisions made by the controller and make adjustments and interventions as necessary (Bevrani & Hiyama, 2011). Considerable effort needs to go into designing explainability approaches for RL agents in power system control, but also in studying their usefulness and reliability with operators.

3.2 New Emerging Challenges in Power System control

Although the set of challenges developed by Dulac-Arnold et al. (2021) is robust for many real-world problems, the unique nature of the way the power grid is organized, constructed, and controlled causes unique problems. In this section, we describe two additional challenges.

10) Action spaces with different temporal horizons of use: This challenge is partially covered under challenge 3 as there are control problems that have different action frequencies. For example, in voltage control, the activation of OLTCs, capacitor banks, and voltage regulators must be slowly controlled due to physical constraints and minimization of wear and tear on the physical device (Maschinenfabrik Reinhausen GmbH, n.d.; Chen et al., 2022). Devices that can be controlled on faster time scales include distributed energy resources based on inverters (such as a photovoltaic cell, wind turbine, or battery) can have their voltages controlled nearly instantaneously (Chen et al., 2022). Some approaches that can tackle a problem with actions that have different execution lengths have been studied in the power system control setting (Yang et al., 2020a; Liu & Wu, 2021), but more work is required. Separating this challenge from the discrete-continuous issues in challenge 3 can provide a new set of research directions for foundational research.

11) Collaboration with operators: Although briefly discussed in challenges 5, 6, and 9, mechanisms for operator-agent collaboration is a broader topic. As pointed out in challenges 5 and 6, there are some collaborations that are best expressed as rewards or new features, but there is no clear distinction between the two, especially for those who are not experts in RL. There may even be more avenues of collaboration (Bicho et al., 2011; Pilarski et al., 2013; 2019; Retzlaff et al., 2024).

4 Future Perspectives

This work provides a high-level overview of the challenges encountered when applying RL to power system control. The organization followed the challenges proposed by Dulac-Arnold et al. (2021), which was able to cover a wide range of problems in applying RL to the power grid. In addition to Dulac-Arnold et al. (2021)'s challenges, there are two additional challenges that require investigation in power system control. The main limitation of this work is the lack of detail on techniques used when applying RL to power systems, instead focusing on areas where RL will face hurdles. In the future, we will add more detail around the literature applying RL to power systems through the above lens, developing an overview of how each challenge has been addressed in the current literature, enabling researchers to more easily deploy new RL techniques.

References

Otavio Bertozzi, Harold R. Chamorro, Edgar O. Gomez-Diaz, Michelle S. Chong, and Shehab Ahmed. Application of data-driven methods in power systems analysis and control. *IET Energy Systems Integration*, 6(3):197–212, 2024. ISSN 2516-8401. DOI: 10.1049/esi2.12122.

Hassan Bevrani and Takashi Hiyama. Automatic Generation Control (AGC): Fundamentals and Concepts. In *Intelligent Automatic Generation Control*. CRC Press, 2011. ISBN 978-1-315-21740-6.

Hassan Bevrani and Shoresh Shokoohi. An Intelligent Droop Control for Simultaneous Voltage and Frequency Regulation in Islanded Microgrids. *IEEE Transactions on Smart Grid*, 4(3):1505–1513, September 2013. ISSN 1949-3061. DOI: 10.1109/TSG.2013.2258947.

Estela Bicho, Wolfram Erlhagen, Luis Louro, and Eliana Costa e Silva. Neuro-cognitive mechanisms of decision making in joint action: A human-robot interaction study. *Human Movement Science*, 30(5):846–868, October 2011. ISSN 1872-7646. DOI: 10.1016/j.humov.2010.08.012.

Di Cao, Junbo Zhao, Jiaxiang Hu, Yansong Pei, Qi Huang, Zhe Chen, and Weihao Hu. Physics-Informed Graphical Representation-Enabled Deep Reinforcement Learning for Robust Distribution System Voltage Control. *IEEE Transactions on Smart Grid*, 15(1):233–246, January 2024. ISSN 1949-3061. DOI: 10.1109/TSG.2023.3267069.

J. Cardell and M. Ilic. Maintaining stability with distributed generation in a restructured industry. In *Proceedings of the IEEE Power and Energy Society General Meeting*, 2004.

Shivam Chaturvedi, Van-Hai Bui, Wencong Su, and Mengqi Wang. Reinforcement Learning-Based Integrated Control to Improve the Efficiency of DC Microgrids. *IEEE Transactions on Smart Grid*, 15(1):149–159, January 2024. ISSN 1949-3061. DOI: 10.1109/TSG.2023.3286801.

Spyros Chatzivasileiadis, Andreas Venzke, Jochen Stiasny, and Georgios Misyris. Machine Learning in Power Systems: Is It Time to Trust It? *IEEE Power and Energy Magazine*, 20(3):32–41, May 2022. ISSN 1558-4216. DOI: 10.1109/MPE.2022.3150810.

Xin Chen, Guannan Qu, Yujie Tang, Steven Low, and Na Li. Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges. *IEEE Transactions on Smart Grid*, 13(4):2935–2958, July 2022. ISSN 1949-3061. DOI: 10.1109/TSG.2022.3154718.

Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis, and Martin Riedmiller. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897):414–419, February 2022.

Yan Du, Qiuhua Huang, Renke Huang, Tianzhixi Yin, Jie Tan, Wenhao Yu, and Xinya Li. Physics-Informed Evolutionary Strategy Based Control for Mitigating Delayed Voltage Recovery. *IEEE Transactions on Power Systems*, 37(5):3516–3527, September 2022. ISSN 1558-0679. DOI: 10.1109/TPWRS.2021.3132328.

Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. *Machine Learning*, 110(9):2419–2468, September 2021. ISSN 1573-0565. DOI: 10.1007/s10994-021-05961-4.

Erik Ela, Brendan Kirby, Eamonn Lannoye, Michael Milligan, Damian Flynn, Bob Zavadil, and Mark O'Malley. Evolution of operating reserve determination in wind power integration studies. In *IEEE PES General Meeting*, pp. 1–8, July 2010. DOI: 10.1109/PES.2010.5589272.

Mostafa Farrokhabadi, Claudio A. Cañizares, John W. Simpson-Porco, Ehsan Nasr, Lingling Fan, Patricio A. Mendoza-Araya, Reinaldo Tonkoski, Ujjwol Tamrakar, Nikos Hatziargyriou, Dimitris Lagos, Richard W. Wies, Mario Paolone, Marco Liserre, Lasantha Meegahapola, Mahmoud Kabalan, Amir H. Hajimiragha, Dario Peralta, Marcelo A. Elizondo, Kevin P. Schneider, Francis K. Tuffner, and Jim Reilly. Microgrid Stability Definitions, Analysis, and Examples. *IEEE Transactions on Power Systems*, 35(1):13–29, January 2020. ISSN 1558-0679. DOI: 10.1109/TPWRS.2019.2925703.

Stephen Frank, Ingrida Steponavice, and Steffen Rebennack. Optimal power flow: A bibliographic survey I. *Energy Systems*, 3(3):221–258, September 2012. ISSN 1868-3975. DOI: 10.1007/s12667-012-0056-y.

Yuan Gao, Yuki Matsunami, Shohei Miyata, and Yasunori Akashi. Multi-agent reinforcement learning dealing with hybrid action spaces: A case study for off-grid oriented renewable building energy system. *Applied Energy*, 326:120021, November 2022. ISSN 0306-2619. DOI: 10.1016/j.apenergy.2022.120021.

Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, and Alois Knoll. A Review of Safe Reinforcement Learning: Methods, Theory and Applications. https://arxiv.org/abs/2205.10330v5, May 2022.

Nikos Hatziargyriou, Jovica Milanovic, Claudia Rahmann, Venkataramana Ajjarapu, Claudio Canizares, Istvan Erlich, David Hill, Ian Hiskens, Innocent Kamwa, Bikash Pal, Pouyan Pourbeik, Juan Sanchez-Gasca, Aleksandar Stankovic, Thierry Van Cutsem, Vijay Vittal, and Costas Vournas. Definition and Classification of Power System Stability – Revisited & Extended. *IEEE Transactions on Power Systems*, 36(4):3271–3281, July 2021. ISSN 1558-0679. DOI: 10.1109/TPWRS.2020.3041774.

Ramij R. Hossain, Kaveri Mahapatra, Qiuhua Huang, and Renke Huang. Physics-informed Deep Reinforcement Learning-based Adaptive Generator Out-of-step Protection for Power Systems. In *2023 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, July 2023. DOI: 10.1109/PESGM52003.2023.10252299.

Ramij Raja Hossain, Tianzhixi Yin, Yan Du, Renke Huang, Jie Tan, Wenhao Yu, Yuan Liu, and Qiuhua Huang. Efficient learning of power grid voltage control strategies via model-based deep reinforcement learning. *Machine Learning*, 113(5):2675–2700, May 2024. ISSN 1573-0565. DOI: 10.1007/s10994-023-06422-w.

Sinan Ibrahim, Mostafa Mostafa, Ali Jnadi, Hadi Salloum, and Pavel Osinenko. Comprehensive Overview of Reward Engineering and Shaping in Advancing Reinforcement Learning Applications. *IEEE Access*, 12:175473–175500, 2024. ISSN 2169-3536. DOI: 10.1109/ACCESS.2024.3504735.

Marija Ilic and Rupamathi Jaddivada. Modeling and control of multi-energy dynamical systems: Hidden paths to decarbonization. In *Proceedings of the 11th Bulk Power Systems Dynamics and Control Symposium*, 2022.

T. P Imthias Ahamed, P. S Nagendra Rao, and P. S Sastry. A reinforcement learning approach to automatic generation control. *Electric Power Systems Research*, 63(1):9–26, August 2002. ISSN 0378-7796. DOI: 10.1016/S0378-7796(02)00088-3.

Ali Keyhani and Abir Chatterjee. Automatic Generation Control Structure for Smart Power Grids. *IEEE Transactions on Smart Grid*, 3(3):1310–1316, September 2012. ISSN 1949-3061. DOI: 10.1109/TSG.2012.2194794.

W. Bradley Knox, Alessandro Allievi, Holger Banzhaf, Felix Schmitt, and Peter Stone. Reward (Mis)design for autonomous driving. *Artificial Intelligence*, 316:103829, March 2023. ISSN 0004-3702. DOI: 10.1016/j.artint.2022.103829.

Jose Daniel Lara, Rodrigo Henriquez-Auba, Deepak Ramasubramanian, Sairaj Dhople, Duncan S. Callaway, and Seth Sanders. Revisiting Power Systems Time-Domain Simulation Methods and Models. *IEEE Transactions on Power Systems*, 39(2):2421–2437, March 2024. ISSN 1558-0679. DOI: 10.1109/TPWRS.2023.3303291.

Antoine Lesage-Landry and Duncan S. Callaway. Batch reinforcement learning for network-safe demand response in unknown electric grids. *Electric Power Systems Research*, 212:108375, November 2022. ISSN 0378-7796. DOI: 10.1016/j.epsr.2022.108375.

Jiawen Li and Tao Yu. Deep Reinforcement Learning Based Multi-Objective Integrated Automatic Generation Control for Multiple Continuous Power Disturbances. *IEEE Access*, 8:156839–156850, 2020. ISSN 2169-3536. DOI: 10.1109/ACCESS.2020.3019535.

Yuanzheng Li, Shangyang He, Yang Li, Yang Shi, and Zhigang Zeng. Federated Multiagent Deep Reinforcement Learning Approach via Physics-Informed Reward for Multimicrogrid Energy Management. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):5902–5914, May 2024. ISSN 2162-2388. DOI: 10.1109/TNNLS.2022.3232630.

Haotian Liu and Wenchuan Wu. Bi-level Off-policy Reinforcement Learning for Volt/VAR Control Involving Continuous and Discrete Devices, April 2021.

Peter Lopion, Peter Markewitz, Martin Robinius, and Detlef Stolten. A review of current challenges and trends in energy systems modeling. *Renewable and Sustainable Energy Reviews*, 96:156–166, November 2018. ISSN 1364-0321. DOI: 10.1016/j.rser.2018.07.045.

Renzhi Lu, Zhenyu Jiang, Huaming Wu, Yuemin Ding, Dong Wang, and Hai-Tao Zhang. Reward Shaping-Based Actor–Critic Deep Reinforcement Learning for Residential Energy Management. *IEEE Transactions on Industrial Informatics*, 19(3):2662–2673, March 2023. ISSN 1941-0050. DOI: 10.1109/TII.2022.3183802.

Jerry Luo, Cosmin Paduraru, Octavian Voicu, Yuri Chervonyi, Scott Munns, Jerry Li, Crystal Qian, Praneet Dutta, Jared Quincy Davis, Ningjia Wu, et al. Controlling commercial cooling systems using reinforcement learning. *arXiv preprint arXiv:2211.07357*, 2022.

Maschinenfabrik Reinhausen GmbH. *On-Load Tap-Changers for Power Transformers: Basic Principles and Practical Applications*. Maschinenfabrik Reinhausen GmbH, 2 edition, n.d. URL https://www.reinhausen.com/fileadmin/downloadcenter/company/publikationen/f0126405_on-load_tap-changers_for_power_transformers.pdf. Accessed: 2025-05-19.

Daniel E. Olivares, Ali Mehrizi-Sani, Amir H. Etemadi, Claudio A. Cañizares, Reza Iravani, Mehrdad Kazerani, Amir H. Hajimiragha, Oriol Gomis-Bellmunt, Maryam Saeedifard, Rodrigo Palma-Behnke, Guillermo A. Jiménez-Estévez, and Nikos D. Hatziargyriou. Trends in Microgrid Control. *IEEE Transactions on Smart Grid*, 5(4):1905–1919, July 2014. ISSN 1949-3061. DOI: 10.1109/TSG.2013.2295514.

Omid Palizban and Kimmo Kauhaniemi. Hierarchical control structure in microgrids with distributed generation: Island and grid-connected mode. *Renewable and Sustainable Energy Reviews*, 44:797–813, April 2015. ISSN 1364-0321. DOI: 10.1016/j.rser.2015.01.008.

Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3803–3810, May 2018. DOI: 10.1109/ICRA.2018.8460528.

Stefan Pfenninger, Adam Hawkes, and James Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, May 2014. ISSN 1364-0321. DOI: 10.1016/j.rser.2014.02.003.

Patrick M. Pilarski, Travis B. Dick, and Richard S. Sutton. Real-time prediction learning for the simultaneous actuation of multiple prosthetic joints. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pp. 1–8, June 2013. DOI: 10.1109/ICORR.2013.6650435.

Patrick M. Pilarski, Andrew Butcher, Michael Johanson, Matthew M. Botvinick, Andrew Bolt, and Adam S. R. Parker. Learned human-agent decision-making, communication and joint action in a virtual reality environment, May 2019.

P. Pourbeik, P.S. Kundur, and C.W. Taylor. The anatomy of a power grid blackout - Root causes and dynamics of recent major blackouts. *IEEE Power and Energy Magazine*, 4(5):22–29, September 2006. ISSN 1558-4216. DOI: 10.1109/MPAE.2006.1687814.

Carl Orge Retzlaff, Srijita Das, Christabel Wayllace, Payam Mousavi, Mohammad Afshari, Tianpei Yang, Anna Saranti, Alessa Angerschmid, Matthew E. Taylor, and Andreas Holzinger. Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. *Journal of Artificial Intelligence Research*, 79:359–415, January 2024. ISSN 1076-9757. DOI: 10.1613/jair.1.15348.

Fred C Schweppe and Sanjoy K Mitter. Hierarchical system theory and electric power systems. In *Proceedings Symposium on Real Time Control of Electric Power Systems*. Elsevier Publishing Company, 1972.

Fred C. Schweppe and J. Wildes. Power System Static-State Estimation, Part I: Exact Model. *IEEE Transactions on Power Apparatus and Systems*, PAS-89(1):120–125, January 1970. ISSN 0018-9510. DOI: 10.1109/TPAS.1970.292678.

Buxin She, Fangxing Li, Hantao Cui, Hang Shuai, Oroghene Oboreh-Snapps, Rui Bo, Nattapat Praisuwanna, Jingxin Wang, and Leon M. Tolbert. Inverter PQ Control With Trajectory Tracking Capability for Microgrids Based on Physics-Informed Reinforcement Learning. *IEEE Transactions on Smart Grid*, 15(1):99–112, January 2024. ISSN 1949-3061. DOI: 10.1109/TSG.2023.3277330.

Hongbin Sun, Qinglai Guo, Junjian Qi, Venkataramana Ajjarapu, Richard Bravo, Joe Chow, Zhengshuo Li, Rohit Moghe, Ehsan Nasr-Azadani, Ujjwol Tamrakar, Glauco N. Taranto, Reinaldo Tonkoski, Gustavo Valverde, Qiuwei Wu, and Guangya Yang. Review of Challenges and Research Opportunities for Voltage Control in Smart Grids. *IEEE Transactions on Power Systems*, 34(4):2790–2801, July 2019. ISSN 1558-0679. DOI: 10.1109/TPWRS.2019.2897948.

Richard Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, Massachusetts, second edition edition, 2018. ISBN 978-0-262-03924-6.

Yi Wang, Dawei Qiu, Xiaotian Sun, Zhaohong Bie, and Goran Strbac. Coordinating Multi-Energy Microgrids for Integrated Energy System Resilience: A Multi-Task Learning Approach. *IEEE Transactions on Sustainable Energy*, 15(2):920–937, April 2024. ISSN 1949-3037. DOI: 10.1109/TSTE.2023.3317133.

Frauke Wiese, Ingmar Schlecht, Wolf-Dieter Bunke, Clemens Gerbaulet, Lion Hirth, Martin Jahn, Friedrich Kunz, Casimir Lorenz, Jonathan Mühlenpfordt, Juliane Reimann, and Wolf-Peter Schill. Open Power System Data – Frictionless data for electricity system modelling. *Applied Energy*, 236:401–409, February 2019. ISSN 0306-2619. DOI: 10.1016/j.apenergy.2018.11.097.

Allen J Wood, Bruce F Wollenberg, and Gerald B Sheblé. *Power generation, operation, and control*. John wiley & sons, 2013.

Zhuorui Wu, Meng Zhang, Song Gao, Zheng-Guang Wu, and Xiaohong Guan. Physics-Informed Reinforcement Learning for Real-Time Optimal Power Flow With Renewable Energy Resources. *IEEE Transactions on Sustainable Energy*, 16(1):216–226, January 2025. ISSN 1949-3037. DOI: 10.1109/TSTE.2024.3452489.

Zirui Wu, Changcheng Li, and Ling He. A novel reinforcement learning method for the plan of generator start-up after blackout. *Electric Power Systems Research*, 228:110068, 2024. ISSN 0378-7796. DOI: 10.1016/j.epsr.2023.110068.

Qiuling Yang, Gang Wang, Alireza Sadeghi, Georgios B. Giannakis, and Jian Sun. Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning. *IEEE Transactions on Smart Grid*, 11(3):2313–2323, May 2020a. ISSN 1949-3061. DOI: 10.1109/TSG.2019.2951769.

Y. Yang, M. Bao, Y. Ding, H. Jia, Z. Lin, and Y. Xue. Impact of down spinning reserve on operation reliability of power systems. *Journal of Modern Power Systems and Clean Energy*, 8(4):709–718, 2020b.

Peipei Yu, Zhenyi Wang, Hongcai Zhang, and Yonghua Song. Safe Reinforcement Learning for Power System Control: A Review, June 2024.

Zidong Zhang, Dongxia Zhang, and Robert C. Qiu. Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems*, 6(1):213–225, March 2020. ISSN 2096-0042. DOI: 10.17775/CSEEJPES.2019.00920.

Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 737–744, December 2020. DOI: 10.1109/SSCI47803.2020.9308468.

Xiangtian Zheng, Nan Xu, Loc Trinh, Dongqi Wu, Tong Huang, S. Sivaranjani, Yan Liu, and Le Xie. A multi-scale time-series dataset with benchmark for machine learning in decarbonized energy grids. *Scientific Data*, 9(1):359, June 2022. ISSN 2052-4463. DOI: 10.1038/s41597-022-01455-7.