

Toward Intelligent Emergency Control for Large-scale Power Systems: Convergence of Learning, Physics, Computing and Control

Qiuhua Huang^{*†}, Renke Huang[†], Tianzhixi Yin[†], Sohom Datta[†], Xueqing Sun[†], Jason Hou[†], Jie Tan[‡], Wenhao Yu[‡], Yuan Liu[‡], Xinya Li[†], Bruce Palmer[†], Ang Li[†], Xinda Ke[†], Marianna Vaiman[‡], Song Wang[¶], Yousu Chen[†]

^{*} Colorado School of Mines, Golden, CO, USA, qiuhuahuang@mines.edu

[†] Pacific Northwest National Laboratory, Richland, WA, USA, yousu.chen@pnnl.gov

[‡] Google DeepMind, Mountainview, CA, USA, jietan@google.com

[§]V&R Energy System Inc, Los Angeles, CA, USA, marvaiman@vrenergy.com

[¶]Portland General Electric, Portland, OR, USA, song.wang@pgn.com

Abstract—This paper has delved into the pressing need for intelligent emergency control in large-scale power systems, which are experiencing significant transformations and are operating closer to their limits with more uncertainties. Learning-based control methods are promising and have shown effectiveness for intelligent power system control. However, when they are applied to large-scale power systems, there are multifaceted challenges such as scalability, adaptiveness, and security posed by the complex power system landscape, which demand comprehensive solutions. The paper first proposes and instantiates a convergence framework for integrating power systems physics, machine learning, advanced computing, and grid control to realize intelligent grid control at a large scale. Our developed methods and platform based on the convergence framework have been applied to a large (more than 3000 buses) Texas power system, and tested with 56000 scenarios. Our work achieved a 26% reduction in load shedding on average and outperformed existing rule-based control in 99.7% of the test scenarios. The results demonstrated the potential of the proposed convergence framework and DRL-based intelligent control for the future grid.

Index Terms—Deep reinforcement learning, emergency control, physics-informed machine learning, high-performance computing

I. INTRODUCTION

As power systems are undergoing a significant transformation with more uncertainties, less inertia, closer to operation limits, and more frequent extreme weather events such as hurricanes and heat waves, there is an increasing risk of large outages. Existing emergency controls are either armed by human operators in the control rooms or triggered by rule-based settings. Consequently, they suffer from either slow response due to human decision-making or ineffectiveness in changing operating conditions. To address this critical issue,

researchers developed optimal power flow-based methods [1], but it is very challenging to model power system dynamic behaviors in the formulation. Optimization-based predictive control methods, e.g., model predictive control [2], have been developed to model system dynamics and derive optimal control solutions. The key challenges are computational efficiency and scalability. Another line of research includes conventional Artificial Intelligence (AI)-based methods, e.g., decision tree [3]. They can be regarded as an advanced version of the rule-based methods, but lack the sequential decision-making capability that is crucial for power system emergency control. Thus, there is an imperative need to develop new and effective solutions.

Some notable progress has been made in developing AI-based or data-driven power system emergency control solutions in recent years. Deep reinforcement learning (DRL) represents one of the latest developments in AI for sequential decision-making and has been utilized or developed for grid stability and emergency control applications [4]–[6]. Recent review papers [7]–[9] show that deep reinforcement learning (DRL) methods have been developed and applied for various grid controls, providing fast and effective solutions in power system stability and emergency control applications and overcome some limitations with traditional rule-based and optimization-based methods. Readers are referred to them for a general introduction to DRL and DRL applications in power systems. There are *multifaceted challenges such as scalability, adaptiveness, and security* posed by the complex power system landscape to achieve DRL-based intelligent wide-area or system-level grid emergency control. Existing works mostly focused on addressing one of these challenges. To tackle such grand challenges, we believe that single-domain advancements are inadequate because solutions designed for solving one challenge may not be compatible with solutions for another one. Furthermore, the scalability challenge has not been properly solved, and existing works are mostly developed for and/or tested with small transmission or distribution systems and only a few scenarios. For example, the largest system for DRL-based stability control prior to this work is

The Pacific Northwest National Laboratory (PNNL) is operated by Battelle for the U.S. Department of Energy (DOE) under Contract DE-AC05-76RL01830. This work was supported by DOE ARPA-E OPEN 2018 Program.

Qiuhua Huang, Renke Huang, Xueqing Sun, and Xinya Li were with PNNL while performing the main research work in this paper.

^{*} Corresponding author: Qiuhua Huang

the IEEE 300-bus test system with limited operation scenarios being considered [10]. To the best of our knowledge, none of them tackles critical issues such as scalability, adaptability, and security comprehensively, and is tested with realistic, large-scale power systems with diverse operation conditions. This means a large gap between existing research efforts and real-world applications in this area.

We need multi-domain advancements to address the gap, thus our key idea focuses on a framework for achieving **convergence** of the state-of-the-art reinforcement learning, physics of power systems, advanced computing and grid control methods, which is illustrated in Fig. 1 and has not been reported in the literature. We instantiate the convergence framework by developing a suite of comprehensive methods and tools, as summarized in Fig. 2. Our final solutions are comprehensively tested on a synthetic Texas power system with more than 3000 buses. We considered 14000 scenarios for training and tested the trained control policy with 56000 scenarios that are unseen in the training dataset, and compared with the existing rule-based under-voltage load-shedding (UVLS) control scheme. The results show that our approach helps reduce the load shedding by 26% on average, and is better than the baseline in 99.7% of the test scenarios in terms of the control objective value (i.e., total reward). This demonstrates great progress toward intelligent emergency control for large-scale power systems.

The main contributions of this paper include:

- 1) proposing and instantiating a convergence framework for integrating DRL, power systems physics, advanced computing and grid control to achieve intelligent grid control at a large scale.
- 2) development of a scalable and high-performance computing (HPC)-compatible grid simulation environment for training and testing large-scale learning-based control solutions and it is open-sourced.
- 3) a three-stage DRL training method including a two-stage curriculum learning method for distributed training of DRL-based control policies for each zone, followed by one-stage coordinated training of all control policies.
- 4) unprecedented large-scale training, testing of DRL-based grid emergency control with a large (more than 3000 buses) synthetic Texas system, demonstrating the potential of proposed convergence framework and DRL-based intelligent control for future grids.

The rest of the paper is organized as follows: Section II discusses the problem and our proposed framework and key methods. Section III provides more details of implementation of the convergence framework. Section IV presents the training and testing results. The paper is concluded in Section V.

II. PROBLEM STATEMENT AND OUR PROPOSED CONVERGENCE FRAMEWORK

As discussed in the introduction section, we have to address multifaceted challenges such as scalability, adaptiveness, and security to successfully develop intelligent grid stability and emergency control. Different from previous work focusing on addressing a single challenge without considering the

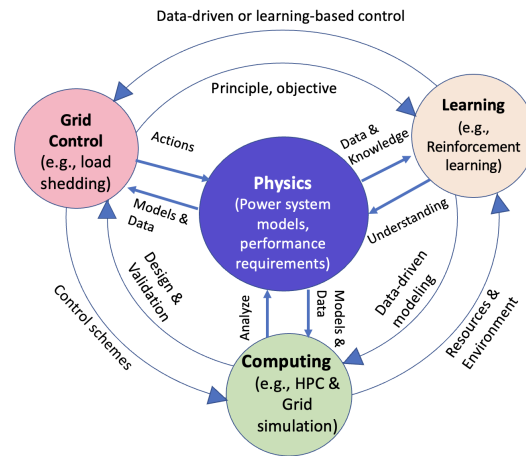


Fig. 1. Convergence of power system physics, advanced computing, machine learning and grid control for achieving intelligent control for large-scale power systems

compatibility and synergy of the underlining methods and technologies, we highlight *systematic solutions* by making technological advancements in the following key areas in a cohesive manner.

- 1) **Learning:** AI and machine learning can provide some unique learning from (big) data and decision-making capabilities that are essential for real-time monitoring, analysis, and control of modern power systems. Successful AI-based control applications require data from the physical world, existing control design principles, and computational resources and platforms.
- 2) **Computing:** Advanced computational algorithms and tools are essential for power system analysis, control design and verification, and supporting various AI applications in power systems.
- 3) **Control:** Grid control strategies are critical for maintaining grid stability and security. However, most existing controls are designed based on physics-based models for a few deterministic operation patterns and do not operate well under vastly changing operation conditions. AI-based methods can complement existing physics-based control designs.
- 4) **Physics:** Power system physics provides the essential models, data, and domain knowledge for the areas above. Advanced computational methods and AI tools help engineers gain insightful information from big data in power systems. Physics-informed machine learning is an example of combining the best of both worlds.

It can be seen that there is a strong synergy among these technologies (also see Fig. 1). However, to the best of our knowledge, few effort has been focused on the convergence of them to realize their full potential for intelligent emergency control for large-scale power systems. Therefore, we first develop a convergence framework by considering their inherent synergy in the context of power system stability and emergency control. The framework is shown in Fig. 1. In the following sections, we will discuss the details of instantiating and implementing the framework.

III. IMPLEMENTATION OF THE CONVERGENCE FRAMEWORK

With the convergence framework above as guidance, we instantiate it by developing several interdisciplinary methods to tackle the scalability, adaptability, and security challenges, as shown in Fig. 2. Below is a brief summary, with more discussions of the development of these key methods in the following subsections.

First, several key factors including computational complexity of large-scale power system dynamic simulation, high dimensional action space, large number of power system operation conditions and fault scenarios all contribute to the scalability challenge, which makes it significantly difficult to address. Therefore, we developed several complementary methods ranging from an HPC-compatible, highly scalable grid simulation environment, and a scalable DRL algorithm, to scenario reduction to overcome it.

Second, since power systems are always changing, there are a large number of grid operation conditions. We developed smart sampling techniques to identify representative scenarios for training the DRL model. In addition, we incorporated the meta-learning method into DRL to ensure it can learn from past experiences and quickly adapt to new scenarios.

Third, like other AI methods, DRL-based methods still do not have a strict performance guarantee for some corner cases. We developed a hybrid DRL-and rule-based control method to ensure the final solution can perform better than existing rule-based control solutions most of the time, and have a fallback control scheme for other extreme cases.

A. High-performance grid simulation platform and DRL algorithm

As shown in Fig. 3, DRL-based control policies are primarily first trained in a simulated environment instead of in the real-world power grid due to security concerns. Dynamic simulation of complex power systems requires solving large-scale differential algebraic equation models, which is computationally intensive. As such, for large-scale RL training and testing, a common bottleneck is environment execution, which is often the slowest part of the whole system [11]. Existing RL environments such as RLGC[5] for grid control are designed for small-scale research and demonstration purposes.

To address this performance bottleneck, we developed the first high-performance simulation and learning platform for grid stability control. The platform architecture is shown in Fig. 3. First of all, the platform is developed based on Ray [12], which is an open-sourced compute framework for scaling AI and Python workloads and supports various computing infrastructures from laptops to high-performance computing (HPC) clusters to clouds. This facilitates DRL-based control development from fast prototyping and testing to large-scale applications with the same Python codes. Secondly, we develop **GridPACK-Gym** (shown at the top of Fig. 3), a scalable RL environment based on the high-performance grid simulator GridPACK [13], [14]. We developed new functions for setting up a grid environment for DRL training and an extra layer of Application Programming Interface (APIs) on top of

GridPACK APIs to make it fully compatible with the OpenAI Gym interface. GridPACK-gym can scale up to thousands of computing cores, which is the first of its kind. It has been open-sourced along with GridPACK on GitHub in [15].

To fully leverage available computing resources and the capabilities of GridPACK-Gym, we developed a highly scalable DRL algorithm called PARS in our earlier work [10] that can easily scale to thousands of computing cores. In light of the growing uncertainties and operation envelopes in power systems, we further enhance the adaptiveness of the PARS algorithm by integrating the meta-learning method into it and create a novel deep meta-reinforcement learning algorithm [16] that is named **meta-PARS** in this paper. The incorporation of meta-learning enables the DRL-based controller (agent) to quickly learn or update a low-dimensional latent representation of the complex grid operation context to adapt the control strategy to new grid operation scenarios. We have made these novel DRL algorithms open-sourced along with the training codes on GitHub in [17].

These are combined with other techniques introduced later in this section to realize successful training of the DRL-based control agent for intelligent grid emergency control.

B. Incorporate physics into DRL agent and training process

There are significant amount of physical knowledge and best practices developed in the power system community for analyzing and operating the grid. On the other hand, recent AI research showed that it requires a massive amount of data for AI agents to learn even fundamental physical concepts or laws. Thus, a rising AI research direction is physics-informed machine learning which aims to leverage physics to jump-start the training of machine learning models and make it more robust. It is highly desirable to incorporate power system physics into DRL agent models and training processes.

First, we noticed that power system performance requirements such as voltage stability criteria [18] are good prior knowledge that can be utilized to regulate the behavior of the neural network-based control policy. For the voltage stability emergency control application that we will consider as an example in our test cases, we incorporated the system transient voltage stability and recovery performance requirements into the DRL agent model as an action mask [19]. It can help block or filter out unfavorable or unnecessary control actions, thereby enhancing both the DRL training efficiency and robustness of the DRL-based control policy. More details can be found in our recent work [19].

Second, domain decomposition is an effective approach to reducing problem complexity and allows tackling each subdomain problem individually in a more tangible way. It should also be noted that many power system stability problems have localized impacts or can be solved by controlling generators or loads near the fault that leads to stability problems. In light of both, we proposed and developed a physics-informed spatial decomposition method in our DRL training process. The basic idea is illustrated in Fig. 4 with the synthetic Texas power grid as the background. There are three main load centers in Texas that are vulnerable to fault-induced voltage stability

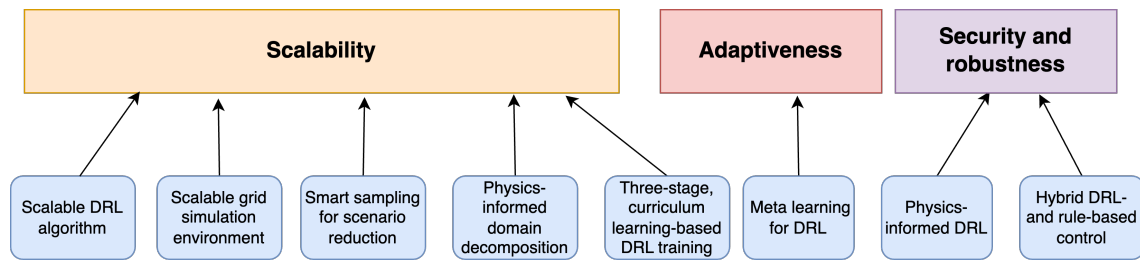


Fig. 2. Key methods based on the convergence of physics, learning, computing and control

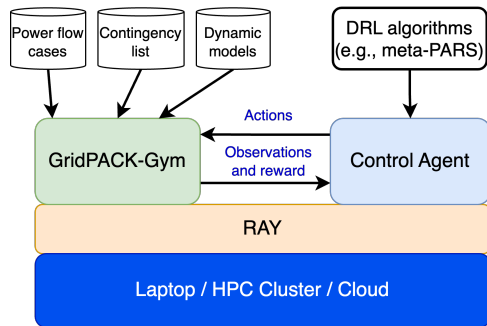


Fig. 3. A high-performance grid simulation and reinforcement learning platform

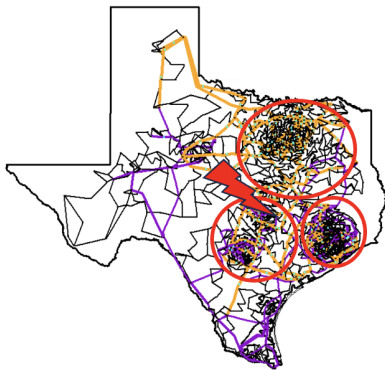


Fig. 4. Physics-informed spatial decomposition for training (Figure adapted from [20])

issues. We first train individual DRL-based controllers for three load centers, which helps reduce the size of the action and observation spaces and makes the DRL-based controllers easier to train. However, as shown in Fig. 4, faults near the boundary between two load centers could result in voltage stability problems in two or more load centers and require coordinated control actions among them. Therefore, we need to coordinate these individually trained controllers, which will be addressed in the next subsection.

C. A three-stage, curriculum learning-based RL method for overcoming difficult exploration problems

RL agents generally learn by the so-called “exploration and exploitation” technique. For RL training at the scale of hundreds of dimensional action space and observation space, tens of thousands of scenarios even after using down-sampling

techniques, it is common to have difficult exploration problems where there are some difficult-to-train cases. There is no exception to our problem.

Along with the proposed physics-informed spatial decomposition method for training, we developed a three-stage DRL training method for overcoming the abovementioned problem and achieving successful training on a large-scale power system with more than 3000 buses and over 50 GW generation and load. First, we leverage the physics-informed spatial decomposition discussed before to create and train individual control policies for each control zone individually.

Even with such a decomposition, we noticed the training can still be very difficult in some cases in each zone. Therefore, we adapt the curriculum learning concept [21] to make the training more smooth (by focusing on solving easy cases first before solving difficult cases). In the conventional setup of curriculum learning, the training starts with a subset of “easy” cases and gradually switches to “difficult” cases. However, we found it difficult to pre-define the “easy” and “difficult” categories at the beginning across very different power flow conditions and fault scenarios. Therefore, we developed a two-stage curriculum learning method. In the first stage, all power flow cases and fault scenarios are treated equally and randomly sampled. The goal for this stage is twofold: 1) training a control policy that is good enough for a majority of the cases; 2) identifying the difficult-to-train cases where the control policy failed to achieve good performance. Then, in the second stage, difficult-to-train cases will be determined dynamically by the evaluation part of the proposed PARS algorithm if the total reward value is less than a threshold. A fixed number of difficult-to-train cases will be sampled and combined with other cases to form the full training set for the second stage.

At the final stage, we perform coordinated training to combine them together in one training to enable them to learn to consider the mutual impacts of the control actions for each zone and hence coordinate their control actions.

D. Smart sampling for scenario reduction

In this paper, each grid scenario for training and testing DRL-based controllers is a combination of one power flow case for representing the grid operation condition, one contingency (defined jointly by fault location and fault duration) and one system dynamic model. We consider one system dynamic model for all studies, which is consistent with the current industry planning and operation practice. For large-scale power systems with high penetration of renewable energy resources,

there are a large number of potential grid operation conditions in the form of power system snapshots. Thus, to deal with the scenario scalability challenge, it is necessary to down-sample a relatively small but representative subset of power flow cases. However, there are challenges in handling high-dimensional data with hierarchical structures and challenges arising from the non-parametric solution space where configurations are complicated by non-Gaussian behaviors and nonlinear cross-dependence in complex dynamic systems like power grids. We developed a hierarchical Latin Hypercube Sampling (LHS) technique for smart sampling, considering free-form distributions of system load, generator commitment status, and generation levels [22]. The method can represent the probability distributions of the original high-dimensional data using fewer samples while respecting the hierarchical data-dependency among the variables. The approach is demonstrated using a large-scale Texas system and is used for scenario reduction in Section IV.

For selecting a representative contingency list, it is important to have good coverage across the system and include both normal and severe contingencies to represent the nature and distribution of contingencies in the real-world system. We adopt the critical clearing time (CCT) as a metric for ranking the contingency locations and perform system-wide CCT analysis for all high-voltage buses. Then we down-sample the contingency fault locations by considering both the CCT and location in the system. Fault duration is sampled from the range between normal fault clearing time (3-4 cycles) to delayed fault clearing time (up to 25 cycles). It is worth noting that there are other metrics for different transient stability and maybe a composite metric like the one proposed in [23] would be a better choice. This deserves further investigation and improvement.

E. DRL- and rule-based hybrid grid emergency control

Electric utilities and power system operators have widely used rule-based control methods such as under-voltage load shedding (UVLS) or under-frequency load shedding (UFLS) for grid emergency control. We should recognize that DRL-based controls have notable advantage for most of the cases and time when it is well trained, but there will always be some extreme cases or corner cases that will be very difficult, if not impossible, for them to cover. Moving forward, we should leverage their complementary strengths. In this paper, we adopt a DRL- and rule-based hybrid grid emergency control approach where the existing rule-based controls are configured with conservative settings, e.g., a relatively longer time delay for action, to become a secondary or backup control, while DRL-based control acts as the primary control. We include these existing rule-based controls in the power system simulation environment for DRL training, such that DRL-based controllers can learn to coordinate with them.

IV. TEST RESULTS

Comprehensive tests are performed on a synthetic Texas power system with more than 3000 buses and over 50 GW total peak load. It is a modified system based on the original system

developed in [20]. The main modifications include 1) adding sub-transmission and equivalent distribution low voltage buses and composite load models to better reflect the voltage stability issues in the real world; 2) adding hourly renewable generation output based on publicly available historical data. We created 1440 hourly power flow cases for 2 summer months. Details can be found in our earlier work [22].

A. DRL-based emergency voltage control

Our developed methods have been applied to different grid emergency controls including transient angular stability control via generator tripping, controlled islanding and emergency voltage control via adaptive load shedding. In this paper, our work is tested and demonstrated with the emergency voltage control via adaptive load shedding application. The goal is to develop a DRL-based closed-loop control policy for applying the load shedding at load centers including Houston areas, Dallas areas, Austin areas and San Antonio Areas to avoid the voltage stability issue and meet the voltage recovery requirements.

The possible load shedding control actions are defined for a total of 258 buses with at least 50 MW dynamic motor loads in the following load centers: zone 3 (Houston, 81 buses), zone 16 (Dallas, 76 buses), at zone 23 (Austin, 61 buses), and zone 26 (San Antonio, 50 buses). The amount of load that could be shed for each bus at each action time step (0.1 s) is a continuous variable from 0 (no load shedding) to 0.2 (shedding 20% of the initial total load at the bus). As such, *the dimension of the action space is 258*. The observations included voltage magnitudes at 115-kV and above buses in the four zones (total 468 buses) as well as the fractions of loads served at the 258 buses where load shedding could be applied. Thus, *the dimension of the observation space is 726*.

The detailed mathematical problem formulation can be found in our earlier works [10], [16]. Compared to our earlier works, we adopt the same Markovian Decision Process (MDP) formulation, but we focus on developing and testing scalable methods based on the convergence framework for grid control at a large scale in this work.

B. Dataset preparation

We employed the smart sampling method discussed in Section III-D to create scenarios for training and testing, as shown in Fig. 5. The training dataset includes 14000 different tasks (scenarios), which are combinations of 140 candidate fault buses and 100 different power flow cases. To comprehensively test the performance and generalization capability of the control policy, we sampled a larger dataset with 56000 scenarios that were unseen in the training.

C. Training results

We performed the training and testing on an HPC cluster and used 840 CPU cores for training. At each of the training iterations, 40 fault locations are sampled from the 140 fault location candidates and 30 power flow cases are sampled from the 100 power flow candidates and the fault locations and

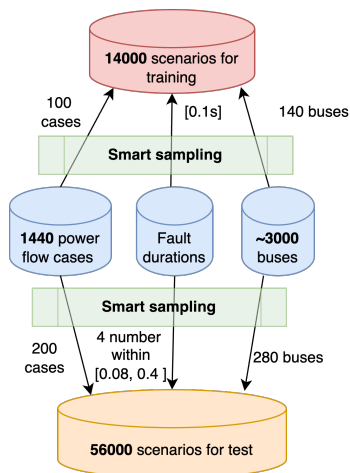


Fig. 5. Preparation of the training and testing scenarios.

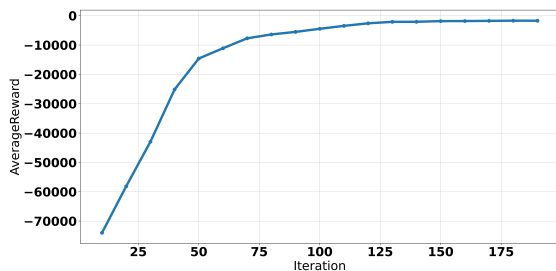


Fig. 6. Training average reward over iterations.

power flow cases are combined to create the rollout tasks (1200 tasks per training iteration). Fig. 6 shows the learning curve of the training, which indicates the training converges around 200 iterations, and it takes about 110 hours to finish the training.

We also compared the meta-PARS-based load shedding control versus the existing rule-based UVLS load-shedding scheme (readers are referred to our earlier work [5] for more details). We chose UVLS as the baseline as other existing optimal control methods such as model predictive control and DRL methods such as Proximal Policy Optimization (PPO) do not scale well to solve this complex problem using the available computing resources within a reasonable period.

Among the 14000 training scenarios, there are 2687 scenarios do not require load shedding actions, and the meta-PARS does not provide actions for all the 2687 cases. For the 11313 scenarios do require load shedding actions to bring the voltage profile back to normal. To show the comparison results, we calculated the reward differences (i.e., the reward of meta-PARS subtracts that of UVLS) for all the total of 14000 trading tasks. A positive value means that our method is better for the corresponding test scenario and vice versa. Fig. 7 shows the histogram of the rewards differences. Our method outperformed UVLS for (99.74%) of the scenarios.

Fig. 8 shows the comparison between the meta-PARS and UVLS performance for a specific training task with a three-phase fault at 500 kV bus 7037. The total rewards of the meta-PARS and UVLS relay control in this test task are -1663 and -2142, respectively, and a larger reward is better.

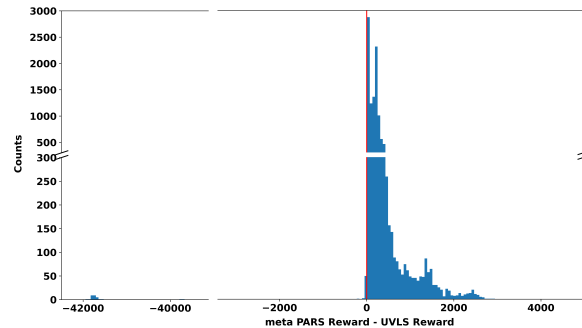


Fig. 7. Histogram of reward differences between meta-PARS and UVLS methods for the training cases.

Fig. 8 (a) and (b) show that the voltages in the system with UVLS control recover much slower than the voltages in the system with meta-PARS control. Fig. 8(c) shows that the meta-PARS-based control only shed around 1700 MW of load to bring the system voltage back to meet the standard, while the UVLS control shed more than 2200 MW load but still could not recover the system voltage to the level required by the standard. This key difference is the shedding action time instant and location.

D. Test results

Once the policy is trained, we tested the trained control policy on a set of 56000 different scenarios with the combination of 280 fault buses that are different from the training fault buses and 200 power flow cases that are different from the training power flow cases. We also compared the meta-PARS-based control versus the conventional UVLS scheme.

Among the 56000 test scenarios, there are a total of 18962 scenarios that do not require load shedding actions for system recovery, and the meta-PARS does not provide load shedding actions for 18,914 of them (99.75%). There are 37038 scenarios that require load shedding actions to meet the voltage recovery standard. In 94.147% of them, the meta-PARS control outperforms the UVLS control. To show the comparison results, we calculated the reward differences (i.e., the reward of meta-PARS subtracts that of UVLS) for all the test tasks requiring load shedding actions. A positive value means the meta-PARS method is better for the corresponding test task and vice versa. Fig. 9 shows the histogram of the rewards differences. Our meta-PARS method outperforms UVLS scheme for most of the scenarios. At the same time, the result also shows that it is still necessary to have backup controls for a small number of extreme (or corner) cases.

Fig. 10 shows the comparison of meta-PARS and UVLS performance for a specific test scenario with 0.1 s three-phase fault at the 500 kV bus 5083 in the Dallas region. The total rewards of the meta-PARS and UVLS relay control in this test task were -1439.5 and -1769.9, respectively. Fig. 10 (a) and (b) show that the voltages in the system with UVLS control recover much slower compared with the voltages in the system with meta-PARS control. Fig. 10(c) shows that meta-PARS control only shed about 1300 MW loads to bring the system

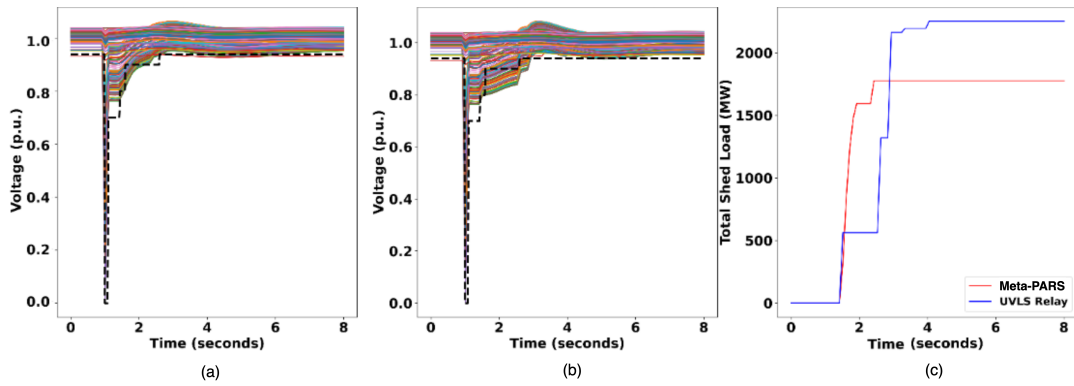


Fig. 8. Comparison of meta-PARS and UVLS-based controls with power flow case 5 and a fault of duration 0.1s at bus 7037 in the training dataset: (a) bus voltages with meta-PARS-based control; (b) bus voltages with UVLS control; (c) total load shedding amount with meta-PARS and UVLS control methods.

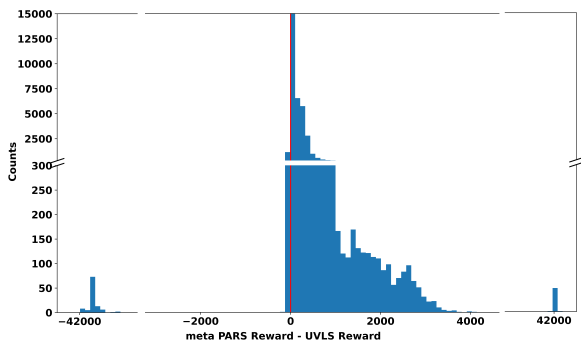


Fig. 9. Histogram of the reward differences between meta-PARS and UVLS for the testing cases.

voltage back to meet the performance requirements, while the UVLS control shed more than 1750 MW loads but still cannot recover voltages to the required level.

V. DISCUSSIONS

There are several key issues related to real-world application of the developed methods and tools.

- 1) The non-stationarity problem of the environment: The non-stationarity problem can be addressed through two complementary approaches: (i) fast adaptation to the new environment enabled by incorporating meta-learning into DRL (the resulting meta-PARS method); (ii) model update after some noticeable changes in the environment (e.g., seasonal changes). Our paper focuses on the first approach, and the second one is easy to implement by performing the training process presented in this paper for newly sampled scenarios.
- 2) The sim-to-real (moving from a simulator to real-world) problem or reality gap [24]: Methods such as improving the physical simulator and learning robust policies via domain randomization can help mitigate the reality gap.

This work can be further improved in at least two directions: First, leveraging historical emergency control events and recordings to train the DRL-based model helps make it more efficient. This is an active research topic area in the DRL

community. Imitation learning is a method that can train a DRL model using historical data. In our recent work [25], we employed this method to jump-start the DRL training before switching to conventional exploration-exploitation-based RL training, and it helped achieve notable training time reduction. Second, physics-based dynamic simulation for large-scale power systems is computationally intensive and requires many computing resources to scale it. An alternative is to leverage the fact that the same or very similar simulation tasks are repeated many times during the training stage, and we can train a machine learning-based (or data-driven) dynamic surrogate model to replace the simulator during training. In our recent work [25], we used a dynamic surrogate model in DRL training, and it achieved more than 80% training time reduction.

VI. CONCLUSIONS

The rapid evolution of the grid and the push towards decarbonization necessitate advanced solutions to ensure the stability, reliability, and resilience of electricity services. This paper has delved into the pressing need for intelligent emergency control in large-scale power systems. We considered the DRL-based methods as the backbone for achieving intelligent grid control. However, there are multifaceted challenges in DRL-based control such as scalability, adaptiveness, and security posed by the complex power system landscape. To this end, the paper proposes and instantiates a convergence framework integrating power systems physics, machine learning, advanced computing, and grid control to realize intelligent grid control at a large scale. We instantiated the framework by developing several key, interdisciplinary methods and tools including scalable grid simulation environment, highly scalable meta-PARS method and physics-informed three-stage DRL agent training process.

The proposed solutions were rigorously tested on a synthetic Texas power system with more than 3000 buses, considering 14000 scenarios for training and 56000 unseen scenarios for testing. The results were promising, showing a 26% reduction in load shedding on average compared to the existing rule-based UVLS control scheme, and outperforming the baseline in 99.7% of the test scenarios in terms of the control objective value (total reward). The comprehensive methods and tools

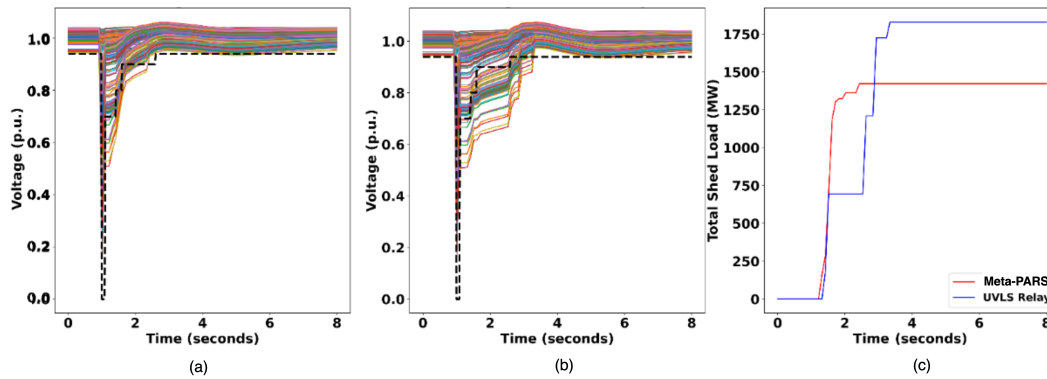


Fig. 10. Comparison of meta-PARS and UVLS methods with the power flow case 9 and a 0.1 s fault at bus 5083 in the testing dataset: (a) bus voltages with meta-PARS-based control; (b) bus voltages with UVLS control; (c) total load shedding amount with meta-PARS and UVLS control methods.

developed have demonstrated significant progress and have the potential to be a cornerstone in the evolution of intelligent grid control design.

Potential future research directions include 1) adaptation and extension for other grid control applications and 2) collaboration between human operators and AI agents for grid operation and control.

REFERENCES

- [1] S. Misra, L. Roald, M. Vuffray, and M. Chertkov, "Fast and robust determination of power system emergency control actions," *arXiv preprint arXiv:1707.07105*, 2017.
- [2] Z. Li, G. Yao, G. Geng, and Q. Jiang, "An efficient optimal control method for open-loop transient stability emergency control," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2704–2713, jul 2017.
- [3] I. Genc, R. Diao, V. Vittal, and *et al.*, "Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1611–1619, 2010.
- [4] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1653–1656, 2019.
- [5] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, "Adaptive power system emergency control using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1171–1182, 2020.
- [6] C. Chen, M. Cui, F. F. Li, S. Yin, and X. Wang, "Model-free emergency frequency control based on reinforcement learning," *IEEE Transactions on Industrial Informatics*, 2020.
- [7] M. Glavic, "(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annual Reviews in Control*, 2019.
- [8] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for selective key applications in power systems: Recent advances and future challenges," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2935–2958, 2022.
- [9] Y. Li, C. Yu, M. Shahidehpour, T. Yang, Z. Zeng, and T. Chai, "Deep reinforcement learning for smart grid operations: Algorithms, applications, and prospects," *Proceedings of the IEEE*, vol. 111, no. 9, pp. 1055–1096, 2023.
- [10] R. Huang, Y. Chen, T. Yin, X. Li, A. Li, J. Tan, W. Yu, Y. Liu, and Q. Huang, "Accelerated derivative-free deep reinforcement learning for large-scale grid emergency voltage control," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 14–25, 2022.
- [11] J. Weng, M. Lin, S. Huang, B. Liu, D. Makoviichuk, V. Makoviychuk, Z. Liu, Y. Song, T. Luo, Y. Jiang, Z. Xu, and S. Yan, "Envpool: A highly parallel reinforcement learning environment execution engine," 2022.
- [12] Anyscale, "Ray: An open-source unified framework for scaling ai and python applications," <https://www.ray.io/>, accessed: 2023-9-27.
- [13] B. Palmer, W. Perkins, Y. Chen, S. Jin, D. Callahan, K. Glass, R. Diao, M. Rice, S. Elbert, M. Vallem, and Z. Huang, "Gridpack: A framework for developing power grid simulations on high performance computing platforms," in *2014 Fourth International Workshop on Domain-Specific Languages and High-Level Frameworks for High Performance Computing*, 2014, pp. 68–77.
- [14] R. Huang, S. Jin, Y. Chen, R. Diao, B. Palmer, Q. Huang, and Z. Huang, "Faster than real-time dynamic simulation for large-size power system with detailed dynamic models using high-performance computing platform," in *2017 IEEE Power Energy Society General Meeting*, 2017, pp. 1–5.
- [15] PNNL, "GridPACK." [Online]. Available: <https://github.com/GridOPTICS/GridPACK>
- [16] R. Huang, Y. Chen, T. Yin, Q. Huang, J. Tan, W. Yu, X. Li, A. Li, and Y. Du, "Learning and fast adaptation for grid emergency control via deep meta reinforcement learning," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4168–4178, 2022.
- [17] Huang, Renke and Huang, Qihua and Yin, Tianzhixi, and Palmer, Bruce, and Li, Ang, "GridPACK." [Online]. Available: <https://github.com/pnnl/hadrec>
- [18] WECC, "Tpl-001-wecc-crt-4—transmission system planning performance," <https://www.wecc.org/Reliability/TPL-001-WECC-CRT-4.pdf>, accessed: 2023-9-28.
- [19] Y. Du, Q. Huang, R. Huang, T. Yin, J. Tan, W. Yu, and X. Li, "Physics-informed evolutionary strategy based control for mitigating delayed voltage recovery," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3516–3527, 2022.
- [20] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *The 10th Bulk Power Systems Dynamics and Control Symposium (IREP 2017)*, 2017, pp. 1–6.
- [21] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4555–4576, 2021.
- [22] X. Sun, X. Li, S. Datta, X. Ke, Q. Huang, R. Huang, and Z. J. Hou, "Smart sampling for reduced and representative power system scenario selection," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 293–302, 2021.
- [23] X. Ye and J. V. Milanović, "Composite index for comprehensive assessment of power system transient stability," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2847–2857, 2022.
- [24] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," *Robotics: Science and Systems XIV*, 2018.
- [25] R. R. Hossain, T. Yin, Y. Du, R. Huang, J. Tan, W. Yu, Y. Liu, and Q. Huang, "Efficient learning of power grid voltage control strategies via model-based deep reinforcement learning," *Machine Learning*, pp. 1–26, 2023.