Improved primary frequency response through deep reinforcement learning

Timothy Thacker, Héctor Pulgar-Painemal
Department of Electrical Engineering and Computer Science
The University of Tennessee, Knoxville
Knoxville, TN, USA
tthacke3@vols.utk.edu, hpulgar@utk.edu

***Abstract*—This paper explores the application of deep reinforcement learning (DRL) to create a coordinating mechanism between synchronous generators (SGs) and distributed energy resources (DERs) for improved primary frequency regulation. Renewable energy sources, such as wind and solar, may be used to aid in frequency regulation of the grid. Without proper coordination between the sources, however, the participation only results in a delay of SG governor response and frequency deviation. The proposed DRL application uses a deep deterministic policy gradient (DDPG) agent to create a generalized coordinating signal for DERs. The coordinating signal communicates the degree of distributed participation to the SG governor, resolving delayed governor response and reducing system rate of change of frequency (ROCOF). The validity of the coordinating signal is presented with a single-machine finite bus system. The use of DRL for signal creation is explored in an under-frequency event. While further exploration is needed for validation in large systems, the development of this concept shows promising results towards increased power grid stabilization.**

***Index Terms*—Deep reinforcement learning, frequency excursion, distributed energy resources, renewable energy**

I. Introduction

The implementation of renewable energy sources has become increasingly relevant for the power grid in recent years. Penetration of renewable sources has increased steadily, while social and economic factors have encouraged further focus on sustainable power generation. With the advantages of decoupled control in most implementations and fast response during transient events, sources such as solar and wind are ideal candidates for participation in primary frequency control [1]. Several papers have proposed designs for large-scale energy storage to increase penetration of renewable resources [2], [3]. For wind power, it has been proposed to let variable-speed wind turbines (WTs) support primary frequency control by obtaining power from the kinetic energy stored in the rotating mass of the blades [4]. For a single turbine, this method is insufficient to support the grid. Wind farms consisting of numerous WTs, however, may be viewed as a large cache of kinetic energy. While promising for wind energy, this approach does not apply to other renewable sources such as solar. Increasing energy storage is a popular direction towards increasing renewable energy penetration, though capital cost may be a limiting factor in widespread implementation [5].

This work was supported in part by the National Science Foundation (NSF) under Grant No. 2033910, and in part by CURENT, which is an NSF Engineering Research Center funded by NSF and the Department of Energy under NSF Award EEC-1041877.

The goal of this work is to improve primary frequency control, which will consequently enhance grid stability. Primary frequency control refers to the automatic system response following a power imbalance [6]. This control is predominantly composed of SG governor response. The governors adjust the SGs power output and the frequency is balanced at a new steady-state frequency, differing from the initial value due to governor droop characteristics and load frequency dependence. The very slow response of governors, especially of conventional thermal power plants, results in significant frequency decay following a disturbance such as a sudden load increase or generator outage. Other types of control are used to return the frequency to the initial operating condition. The goal of primary frequency control is to balance the system while avoiding extreme measures to arrest frequency deviation [7]. The ideal post-disturbance frequency response would be a first-order response from the initial to final value. By reducing the frequency nadir and ROCOF, system reliability is improved considerably by decreasing the potential for under-frequency load shedding or over-frequency generation trip.

The implementation of power sources with faster dynamics such as renewable DERs can provide frequency support during the inertial response, while the generalized coordinating signal proposed in this paper holds potential for widespread application in large systems. DRL has been shown to have promising applications for power systems as increased renewable penetration introduces greater uncertainty in decision and control problems [8]. Through a reinforcement learning environment built in Simulink, a DDPG agent is trained to determine continuous actions resulting in improved primary frequency response.

The following details the organization of this paper. Section II provides a brief description of traditional frequency response and the intended impact of coordinated frequency response. Section III details the implementation of the coordinating signal to a power system through Simulink, which is used in a case study in Section IV. The concept of generalized coordination through DRL is summarized in section V.

II. Frequency Response

A. Traditional Frequency Response

Traditional frequency response to a power imbalance is the evolution of system frequency from the original steady state frequency, through inertial and governor response, to a new steady state value. Initially, as the frequency exhibits a nearly linear decay, kinetic energy from the rotating masses of the large synchronous generators is extracted to balance the power consumption. This use of generator inertia is appropriately termed the inertial response. Extreme frequency decay during inertial response may cause under-frequency load shedding and is a focus point of system instability. In a power imbalance event, participation from renewable DERs may reduce the severity of the excursion and decrease the ROCOF initially, but this supplementary control cannot be sustained for more than several seconds before the available energy is exhausted. When supplementary control ends, the typical primary frequency response is droop control, where the SGs output power is controlled by their governors. The time delay action of the droop control may cause the effects of a power imbalance event to occur regardless of participation of renewable DERs in the primary frequency control, and studies show that droop control alone is sub-optimal [9]. Thus with participation from DERs we see a delay in frequency excursion, but not a resolution to the under-frequency event.

B. Coordinated Frequency Response

By communicating to the SGs the amount of power contributed in a power imbalance event by DERs, the former can increase mechanical power output accordingly and therefore negate the possibility of a frequency excursion post-DER participation. This resolves the concern that a frequency decay will only be delayed following DER support. By modifying the frequency reference used governor, it will respond as if there is no support from DERs and governor response will begin. The droop control should take full effect when the participation from DERs has ended. This coordinated frequency response allows the slow governor response to take place during DER participation, rather than after the available energy has been used. In this proposed coordinated frequency response, the SGs retain the responsibility of maintaining grid frequency long-term, while grid stability is increased during the typical inertial response time.

Hybrid control consisting of DFIG-based WTs and conventional synchronous plants has been proposed to resolve the issue of droop control time delay by creating a coordinating signal between the wind farm and the SGs at the conventional plant [10]. The implementation of this novel coordinating mechanism to a small test system has shown decreased ROCOF during frequency excursions, in addition to a less severe frequency nadir. This approach to creating coordination begins at the DFIG-based WTs, where the coordinating signal is derived based on the power reference of the WT frequency controller. Considering this, the WT active generated power is used as an observation (input) to determine the coordinating signal and the coordination is determined to be some signal proportional to the power reference of the WT's frequency controllers. While the proposition of a coordinating mechanism between DFIG-based WTs and SGs is a promising step towards higher penetration of renewable energy sources, its implementation is limited. Considering the geophysical constraints of renewable energy sources, some renewable sources may not be viable depending on the region [11]. In addition to geophysical constraints, social constraints such as conservation may limit otherwise ideal wind power generation, while solar can be more easily distributed throughout developed areas [12].

This paper proposes proof of concept of a substantial upgrade to the coordinating mechanism presented in [10], aiming to generalize the concept of primary frequency response coordination to any DER. By beginning the approach to control coordination at the SGs rather than the DERs, a smaller set of observations can be used to determine optimal coordination for DER participation in grid frequency support. The generation of the signal would not depend on the power reference of DER's frequency controllers, but by a DRL agent determining the frequency response by observing SG speed, error, and maximizing rewards according to a well-defined reward function. The generated control signal serves two purposes - controlling the participation of DERs in a power imbalance event and communicating to the SG's governors the appropriate action. The training of a DRL agent on a multitude of power imbalance events may allow for the agent to optimally coordinate between energy sources, beginning at the SGs, to improve primary frequency response and significantly increase grid stability.

III. Signal Implementation

While DERs support grid frequency, droop control does not activate. The governor will not change mechanical power until grid frequency is above or below an acceptable value, when the DER participation has ended. By adding the coordinating signal to the frequency reference in the governor, the governor will change mechanical power by an amount proportional to the participation provided by DERs, thus immediately communicating the frequency excursion to the SG. This addition is made in a simplified version of the IEESGO governor model [13].

A. Modified IEESGO Governor Model

By setting $K_2 = K_3 = 0$, we obtain a reduced model of the governor described by the following set of differential-algebraic equations (DAEs):

$$T_1 \dot{y_1} = -y_1 + K_1 \left[\frac{\omega}{\omega_s} - (1 + \mu_c) \right] \tag{1}$$

$$T_3 \dot{y_3} = -y_3 + y_1 \tag{2}$$

$$T_4 \dot{T_m} = -T_m + P_V \tag{3}$$

$$y_2 = \left(1 - \frac{T_2}{T_3} \right) y_3 + \frac{T_2}{T_3} y_1 \tag{4}$$

where P_V is solved algebraically by:

$$P_V = \begin{cases} P_C - y_2, P_{min} \leq P_C - y_2 \leq P_{max} \\ P_{max}, P_C - y_2 \geq P_{max} \\ P_{min}, P_C - y_2 \leq P_{min} \end{cases} \quad (5)$$

Note that P_V is an intermediate variable used to implement the P_{max} and P_{min} limits algebraically to the model. The addition of the μ_c modifies the governor speed reference from ω_s to $\omega_s + \mu_c$ (Fig. 1), therefore communicating to the governor how much frequency support is provided by participating power sources.

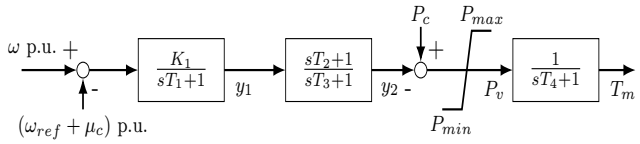

Fig. 1. Simplified IEESGO model governor with coordinating signal μ_c

B. Simulink Environment

1) Reinforcement Learning Environment: To simulate the frequency of a power system during an under-frequency event, a reinforcement learning environment was constructed in Simulink (Fig. 2). The action of the agent is to either increase or decrease the value of μ_c, the coordinating signal, where a positive value indicates the degree of DER participation in frequency response and a negative value indicates power consumption of DERs. In the environment, system frequency is compared to post-frequency event steady state value as a reference. The center of inertia frequency may be considered in a multi-machine system. Rewards and observations are generated from the system frequency and the reference and used by the agent to determine the next action. To maximize rewards, the agent generates the coordinating signal resulting in the minimum difference between system frequency and the reference value, which is the desired primary frequency response of the system. It should be noted that available DERs may not have enough energy stored to participate in grid frequency support for the full duration required by the contingency. Improper agent training may also affect performance.

2) DDPG Agent: Proper application of reinforcement learning to the power grid requires an algorithm that can learn a policy in a large, continuous action space. The Deep Deterministic Policy Gradient algorithm proposed in [14] was designed to learn policies in such an environment by using an improved actor-critic method. The actor network is composed of a deep neural network with a single input, the observation, and a single output, the action. The critic network is composed of a deep neural network with two inputs, the observation and action, and a single output (Fig. 3), the coordinating signal μ_c. The action is the change being made to the power system which is determined by the actor network, then validated by the critic network. The observation is defined as a 3x1 vector $\begin{bmatrix} \int edt & e & \omega \end{bmatrix}$ where e is the error signal, $\int edt$ is the cumulative error, and ω is the frequency. Agent training was determined to be complete when the agent received a high enough average reward over a window of 20 consecutive episodes, indicating the agent consistently produced a coordinating signal resulting in a significantly small error.

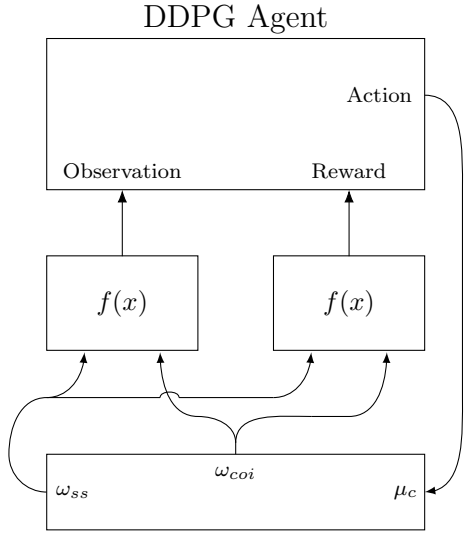

Fig. 2. Reinforcement Learning Environment Diagram

Algorithm 1 DDPG Algorithm [14]

Randomly initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer R
for Episode = 1, M **do**
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial observation state s_1
  for t = 1, T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:
    $\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$
    Update the target networks:
    $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
    $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
  end for
end for

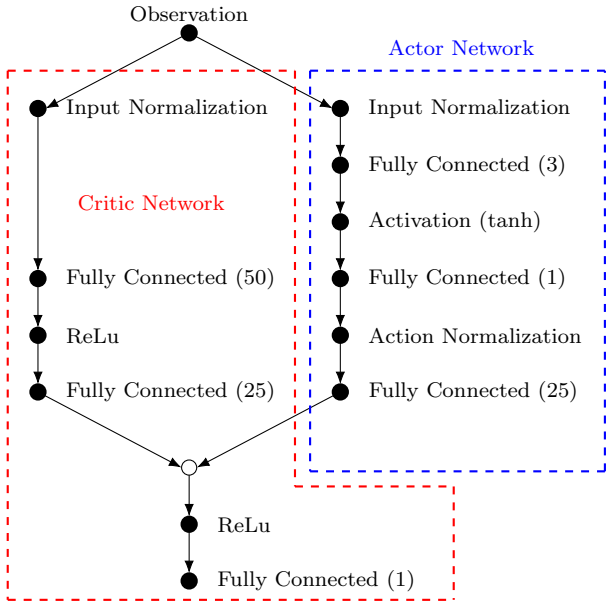

Fig. 3. Actor-Critic Network Structure

3) Semi-Explicit ODE Solver: Simulink presents a variety of ways to solve differential algebraic equations. It is unfeasible, however, to solve large systems of equations by representing each equation with an equivalent block diagram. Given semi-explicit DAEs of the form

$$\dot{u} = f_1(t, u, v) \tag{6}$$

$$0 = f_2(t, u, v) \tag{7}$$

a Simulink framework consisting of an integrator and algebraic loop can be used to solve the system of equations [15]. For an appropriate power systems application, initial conditions must be provided for both the integrator and the algebraic loop. For this application, the coordinating signal is set as the input. Simulink can solve any number of semi-explicit DAEs by expressing the differential and algebraic equations in respective interpreted MATLAB functions.

IV. Case Study

A. Single-Machine Finite Bus System

To investigate the proposed application of DRL for improved primary frequency response, we used a single-machine finite bus (SMFB) system as a proof of concept. The SG in the SMFB system is represented by the classical model with the modified IEESGO model governor presented in (1)-(5) [6]. To create a system reference, we set $E'\angle\delta = 1\angle 0$. The load was modeled as a constant impedance of 1 per unit, with a unit power factor considered. An under-frequency event was created by an instantaneous load change of 5%, providing a frequency excursion exhibiting a significant frequency nadir and notable overshoot following governor response considering relevant system parameters listed in the appendix. By beginning at the SGs for signal generation, DERs do not need to be considered in the system model and a control signal is created that can be attributed to a wide variety of DERs. Thus, no DER is considered in the SMFB system. The DDPG agent creates the generalized coordinating signal, which is then considered in the SG DAEs.

B. Training

The ideal primary frequency response is a linear decrease from the original steady state frequency of the system to the post-event steady state frequency—avoiding a significant nadir during inertial response or significant overshoot during governor response. With a step change of load at $t = 0$, the agent is trained to minimize convergence time to the post-event steady state. The error signal is defined as $e = \omega_{ss} - \omega$ where ω_{ss} is the post-event steady state frequency. The parameters for completed training were set as either the completion of 5,000 episodes or exceeding a reward of 750 for 20 consecutive episodes, where the reward function is defined as:

$$\begin{aligned} reward = {} & 10(|e| < 0.001) - 1(|e| \geq 0.001) - 4(|e| > 0.1) \\ & - 5(e < -0.005) - 10(e > 0) \\ & - 1000(\omega \leq 0.9 || \omega \geq 1.1) \end{aligned} \tag{8}$$

such that a small error magnitude results in a positive reward. This discrete action function is designed to give increasingly large negative rewards as the error increases. When the magnitude of the error is less than 0.001, the agent receives a positive reward of 10. Likewise, when the magnitude of the error is greater than or equal to 0.001, or greater than 0.1, negative rewards of -1 and -4 are given respectively. To prioritize the reduction of the frequency nadir, a reward of -10 is given if the error value indicates a frequency under the post-event steady state. To minimize overshoot, a reward of -5 is given if the frequency is more than 0.005 per unit greater than the reference. A significant negative reward is given if the system frequency exceeds the range 0.9 - 1.1 per unit, and the episode is immediately ended. These limits are not intended to represent the operating limits of a power system. They are a broad range of values that will allow the agent to learn from the entire dynamic simulation. By implementing a strict set of values representing realistic power system limits, important information is lost due to episode termination. The discrete design of the reward function results in parameters that can be understood and modified with ease, and the performance of the discrete reward function can be observed throughout training. Exploration into MATLAB-generated reward functions, consisting of both discrete and continuous parts, is encouraged to further improve upon this proof of concept. The implementation of a continuous reward function may improve agent performance and allow for training across multiple power imbalance events. With these conditions, the agent completed training after 263 episodes on a 5% load change event (Fig. 4).

The agent may not learn if proper data parameters are not set for the observations and continuous actions. To focus training in ranges of expected values, the ω value in the observation vector was limited to the range $0.9 \leq \omega \leq 1.1$, reflecting the

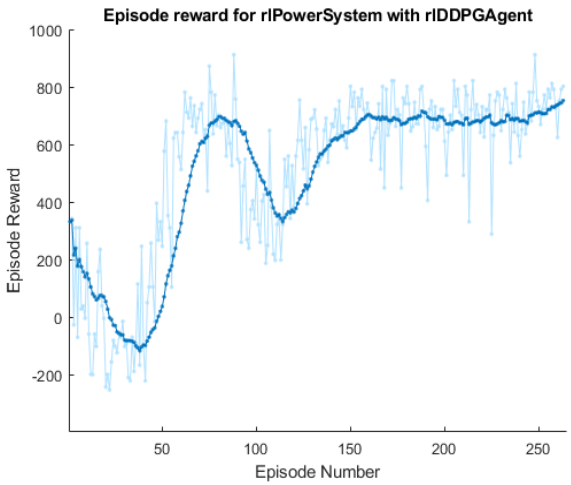

Fig. 4. DDPG Agent Training Performance

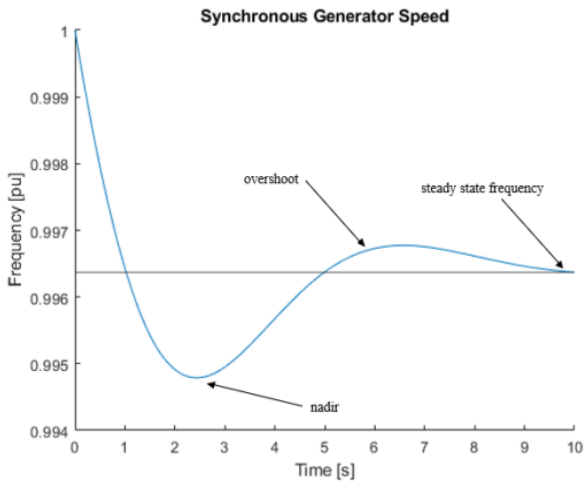

Fig. 5. Primary frequency response without control coordination

limits in the reward function. The continuous action range, the value of μ_c, was constrained to the range $-2.0 \leq \mu_c \leq 2.0$ to prevent divergence to extreme or unrealistic values. A limit was also considered such that the coordinating signal was only considered for the first two seconds of the simulation. A shorter control signal reduces agent training time, and a transient coordinating signal is more realistic as a generalized signal. DERs may not have sufficient energy storage to sustain long-term grid frequency support, limiting the potential applications for the control signal. Further accuracy can be obtained by decreasing the time step, though this significantly increases training time. To maintain realistic training time, it may be advised to extract the coordinating signal generated in the highest performing episode of the training set.

C. Simulation Results

Without coordinated participation, system simulation shows a frequency nadir of 0.9948 per unit and an overshoot due to governor response peaking at about 6.5 seconds (Fig. 5). Simulation considering two seconds of coordination, with μ_c determined by best performing episode of the training set, results in a significantly reduced frequency nadir of 0.9960 per unit and a faster governor response, with an overshoot peaking at about 4.4 seconds. The comparison in Fig. 6 visualizes the frequency nadir reduction of about 76%, and about 47.7% faster governor response time. Furthermore, ROCOF was reduced by about 26.6% after the inclusion of the coordinating signal. As illustrated in Fig. 6, a transient control effort may significantly increase grid stability and reduce risk of extreme measures such as under-frequency load shedding by considerably reducing the frequency nadir. Despite the considerable improvements, system frequency reached the new steady state value at nearly identical times for both simulations. This suggests the need for a longer duration of grid frequency support from DERs, or the implementation of further control methods, to reduce steady state convergence time.

These results reflect and improve upon the findings in [10] which consider the implementation of a coordinating signal between DFIG-based WTs and SGs in a 9-bus test system. By using an artificial neural network (ANN), a coordinating signal was created based on the power reference of the WT's frequency controllers. The application of the ANN-generated signal to the 9-bus test system showed a frequency nadir reduction of about 22% and a 29.5% improvement to system ROCOF considering a 10% load increase. Most notably, the generalized signal presented a significant improvement in frequency nadir by about 54%. While the generalized signal had a 2.9% smaller improvement of system ROCOF compared to the ANN-generated signal, the performance over a standard simulation with no coordinating signal remains promising.

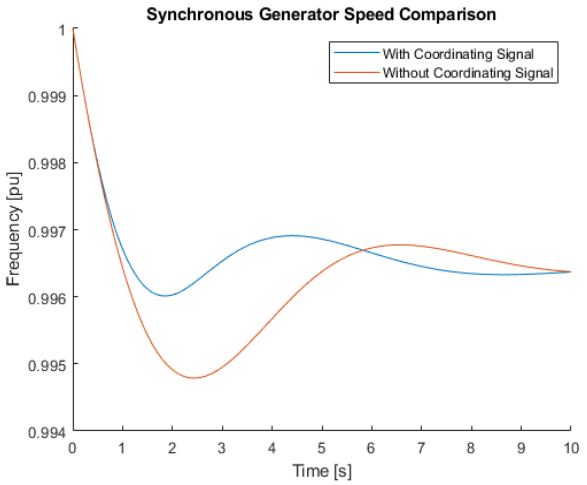

Fig. 6. Primary frequency response with and without coordinated DER participation

Fig. 7 illustrates the two-second generalized coordinating signal. The signal is comparable to the control effort determined by the ANN in [10] which was proportional to the power reference of WT frequency controllers. This indicates

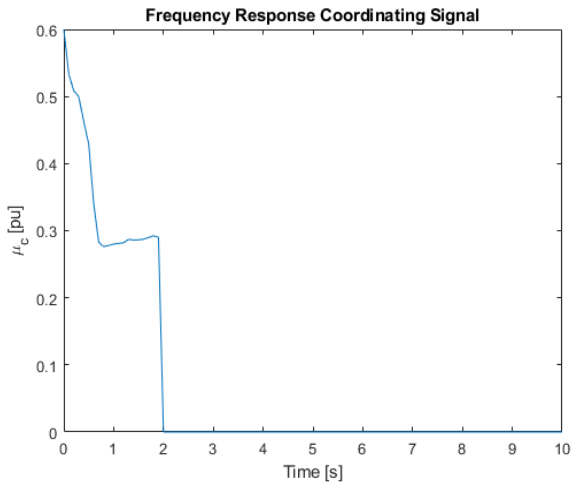

Fig. 7. DRL-Generated Signal for Primary Frequency Response Control Coordination

the potential for signal creation by DERs with limited control capability in transient time. It can be observed that the signal exhibits behavior that could lead towards discrete representation, potentially presenting control coordinating between SGs and DERs as a discrete control problem. The results show potential for significant primary frequency response improvement, and the generalized coordinating signal generated may serve as an improved control signal for DER-based grid frequency support.

Contrasting previously proposed methods of coordinating signal generation, the coordinating signal created by the trained DDPG agent considers no specific origin of grid frequency participation and signal derivation begins at the SG. This introduces the possibility that any DER with appropriately fast dynamics has the potential to participate in coordinated grid frequency support using an appropriate μ_c. With the varying implementation of DERs due to geophysical and geopolitical constraints, a generalized coordination presents a practical improvement for widespread power grid improvement amid strong pushes for the implementation of renewable DERs. Further work is needed to explore the application to a large system and performance over multiple contingencies.

V. Conclusion

In this paper, a DRL-generated coordinating mechanism for distributed energy resources is presented. As the demand for renewable penetration increases, so too does the potential for the implementation of this coordination in the power grid. The DDPG agent used in the application of DRL to this simple power system is designed to scale to larger networks while learning using low-dimension observations - showing promise for expansion of this proof of concept to large power systems. When applied to a SMFB example with a 5% load change, simulation showed a decrease of about 76% in frequency nadir and 26.6% in ROCOF. Further training is needed to adapt this to a variety of power imbalance events. Continued development shows promise for improved primary frequency response and power grid stability.

Appendix

Synchronous Generator: $H = 5, K_D = 5, X_d = 0.1$
Modified Governor: $T_1 = 1, T_2 = 1, T_3 = 1, T_4 = 1, K_1 = 8$

References

[1] M. Liserre, T. Sauter, and J. Y. Hung, "Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics," *IEEE Industrial Electronics Magazine*, vol. 4, no. 1, pp. 18–37, 2010.

[2] A. Solomon, D. M. Kammen, and D. Callaway, "The role of large-scale energy storage design and dispatch in the power grid: A study of very high grid penetration of variable renewable resources," *Applied Energy*, vol. 134, pp. 75–89, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261914007867

[3] E. Rodrigues, R. Godina, S. Santos, A. Bizuayehu, J. Contreras, and J. Catalão, "Energy storage systems supporting increased penetration of renewables in islanded systems," *Energy*, vol. 75, pp. 265–280, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544214008949

[4] J. Morren, S. de Haan, W. Kling, and J. Ferreira, "Wind turbines emulating inertia and supporting primary frequency control," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 433–434, 2006.

[5] E. Hittinger, J. Whitacre, and J. Apt, "What properties of grid energy storage are most valuable?" *Journal of Power Sources*, vol. 206, pp. 436–449, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775311024220

[6] P. Sauer and M. Pai, "Power system dynamics and stability, prentice-hall," *New Jersey*, 1998.

[7] *EPRI Power Systems Dynamics Tutorial.*, EPRI, Palo Alto, CA: 2009. 1016042.

[8] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, 2020.

[9] H. Mahmood, D. Michaelson, and J. Jiang, "Reactive power sharing in islanded microgrids using adaptive voltage droop control," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3052–3060, 2015.

[10] S. Morovati and H. Pulgar-Painemal, "Control coordination between dfig-based wind turbines and synchronous generators for optimal primary frequency response," in *2020 52nd North American Power Symposium (NAPS)*, 2021, pp. 1–6.

[11] M. R. Shaner, S. J. Davis, N. S. Lewis, and K. Caldeira, "Geophysical constraints on the reliability of solar and wind power in the united states," *Energy Environ. Sci.*, vol. 11, pp. 914–925, 2018. [Online]. Available: http://dx.doi.org/10.1039/C7EE03029K

[12] A. N. Arnette and C. W. Zobel, "Spatial analysis of renewable energy potential in the greater southern appalachian mountains," *Renewable Energy*, vol. 36, no. 11, pp. 2785–2798, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148111001960

[13] I. C. Report, "Dynamic models for steam and hydro turbines in power system studies," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-92, no. 6, pp. 1904–1915, 1973.

[14] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015. [Online]. Available: https://arxiv.org/abs/1509.02971

[15] L. F. Shampine, M. W. Reichelt, and J. A. Kierzenka, "Solving index-1 daes in matlab and simulink," *SIAM Review*, vol. 41, no. 3, pp. 538–552, 1999. [Online]. Available: https://doi.org/10.1137/S003614459933425X