Reinforcement Learning-Based Inertia and Droop Control for Wind Farm Frequency Regulation

Yanchang Liang
School of Engineering
University of Warwick
Coventry, United Kingdom
Email: Yanchang.Liang@warwick.ac.uk

Li Sun
School of Mechanical Engineering
and Automation
Harbin Institute of Technology, Shenzhen
Shenzhen, China
Email: sunli2021@hit.edu.cn

Xiaowei Zhao
School of Engineering
University of Warwick
Coventry, United Kingdom
Email: Xiaowei.Zhao@warwick.ac.uk

***Abstract*—As more and more wind turbines (WTs) are installed, there is an increasing interest in actively controlling their power output to meet power set-points and to participate in the frequency regulation for the utility grid. Conventional inertial and droop control loops use fixed gains, making it difficult to utilise the kinetic energy of WTs in a wind farm in a synergistic manner based on real-time information. In this paper, the fixed gains are modified to adaptive gains to improve frequency support performance and reduce the impact on mechanical structures. The cooperative frequency control problem for all WTs in a wind farm is modelled as a decentralised partially observable Markov decision process (Dec-POMDP) and solved using a multi-agent deep reinforcement learning (MADRL) algorithm. MATLAB/Simulink and FAST are run in connection to simulate the frequency response of a wind farm, where FAST simulates the mechanical part of WTs and Simulink simulates the electrical part. Simulation results show that the proposed method is effective in reducing frequency drops and the impact of frequency control on the mechanical structure.**

***Index Terms*—Frequency regulation, wind generation, inertia and droop control, multi-agent deep reinforcement learning.**

I. Introduction

In recent years, there has been significant growth in the penetration of offshore wind power into power systems and this trend is expected to continue in the future. Unlike conventional synchronous generators, wind turbines (WTs) do not naturally possess inertial response or participate in frequency disturbance events. The effective system inertia could be severely reduced with high penetration of wind power, resulting in high rates of change of frequency (RoCoF) and large frequency deviation after a sudden loss of generation or the connection of large loads.

Many works have investigated inertia control schemes for variable-speed WTs which temporarily release the kinetic energy stored in their rotating mass to arrest the frequency nadir. These schemes employ additional loops based on the measured frequency, i.e. inertia loop and droop loop [1]–[3]. However, in these schemes, the control gains are set to be fixed, making it difficult to adjust their kinetic energy uptake in real time based on information such as the wind speed and rotor speed of the WT. Due to the wake effect, each WT in a wind farm contains varying degrees of releasable kinetic energy. Therefore, in contrast to the constant gain control scheme, a control scheme with stable and adaptive gains is proposed in [4], [5]. The values of the two loop gains are proportional to the kinetic energy stored in the WT to exploit the releasable kinetic energy. The effect of wake effects on the inertial response of the turbine is analysed in [6]. Wu *et al.* [7] propose an advanced control strategy with time-varying gains for inertia and droop control loops. In the proposed strategy, the gains are determined according to the desired frequency-response time.

Although the above works consider that WTs should have different responses in different states, most of them ignore the synergistic operation between WTs. In [8], the primary frequency response of the WT is significantly improved by continuously adjusting its droop in response to wind velocities. However, the proposed method needs communication among the WTs in a wind farm, and the droop gain of each WT in the wind farm is somehow dependent on other WTs' performance. In order to be free from the limitations of communication, this paper will focus on the use of local information to collaboratively control the WTs in a wind farm.

When WTs are involved in frequency regulation, their output power needs to change frequently in response to changes in frequency, which adds fatigue loads to WTs. However, the inertia and droop control methods proposed in previous works do not take into account the impact on the mechanical structure. To fill the research gap, this paper will design control policies that can reduce the impact on the mechanical structure. The flexibility of WTs structure such as blades, tower, drivetrain and other components [9] cannot be neglected when the interactions between the electrical and the mechanical aspects of WTs need to be assessed. An accurate WT model must contain many degrees of freedom to capture the most important dynamic effects. Therefore, in this paper, the detailed aerodynamic and structural systems of the WT are modelled using FAST software [10]. FAST is a comprehensive aeroelastic simulator capable of predicting both the extreme and fatigue loads of three-bladed, horizontal-axis WTs.

This work was supported in part by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 861398, and in part by Research Start-Up Funding of Harbin Institute of Technology, Shenzhen, under the grant No CA45001043

Recent years witnessed tremendous success in deep reinforcement learning (DRL) in modeling computational challenging decision-making problems such as Atari [11], Go [12], and StarCraft [13]. In this paper, we develop a DRL-based controller to improve the performance of wind farm frequency regulation. Specifically, we consider each WT as an agent and model the problem of collaboratively controlling WTs for frequency regulation as a decentralized partially observable Markov decision process (Dec-POMDP) [14]. We use multi-agent deep deterministic policy gradient (MADDPG) [15], a multi-agent DRL (MADRL) algorithm, to solve this Dec-POMDP. The proposed method follows the centralised training, decentralised execution (CTDE) paradigm [16], thus avoiding the requirement for communication for online collaborative operation.

The rest of this paper is organised as follows. Section II presents the structure of the inertial and primary frequency controllers. In Section III, we model the joint frequency control problem for WTs in a wind farm as a Dec-POMDP. In Section IV, we run experiments connecting FAST to MATLAB/Simulink to verify the effectiveness of the proposed method. Finally, we conclude the paper in Section V.

II. Inertia and Primary Frequency Control

This section briefly describes a conventional fixed-gain inertial control scheme [1]–[3], which uses two additional loops: inertial and droop loops, as shown in Fig. 1. The active power reference of WT, P_{ref}, consists of three terms: P_{MPPT}, for the MPPT control; ΔP_{in}, which is the output of the inertial loop; and ΔP_{dr}, the output of the droop loop.

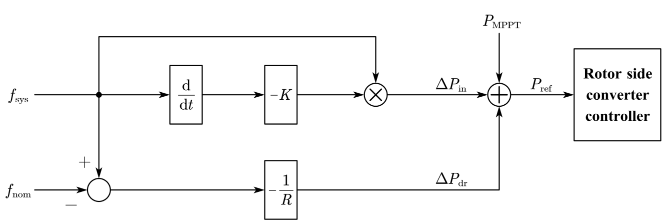

Fig. 1: Inertia and primary frequency controller.

ΔP_{in} can be expressed as

$$\Delta P_{\text{in}} = -K f_{\text{sys}} \frac{\mathrm{d}f_{\text{sys}}}{\mathrm{d}t} \tag{1}$$

where K is the inertial gain and f_{sys} denotes the measured system frequency. The function of the differentiator $\frac{\mathrm{d}}{\mathrm{d}t}$ is to obtain the RoCoF, and the inertial gain K determines the increase in the active power output when the system frequency declines.

ΔP_{dr} can be expressed as

$$\Delta P_{\text{dr}} = -\frac{1}{R}(f_{\text{sys}} - f_{\text{norm}}) \tag{2}$$

where $\frac{1}{R}$ is the droop gain and f_{norm} is the nominal frequency of power system. A high droop gain provides a large output from the droop control loop.

III. Dec-POMDP Formulation

We consider a wind farm consisting of N WTs. We tune the inertia gain and droop gain for each WT in real time at a series of discrete time $t = 1, \ldots, T$. We consider each WT as an agent, and model the joint frequency regulation problem as a Dec-POMDP, where the major components are as follows:

1) State: At each time step t, each WT n' observation o_n consists of its wind speed, rotor speed, rotor torque, generator power, pitch angle and mechanical structure information. The mechanical structure information includes blade flap-wise tip deflection $d_n^{\text{b,flap}}$, blade edge-wise tip deflection $d_n^{\text{b,edge}}$, tower fore-aft displacement $d_n^{\text{t,fore}}$, and tower side-to-side displacement $d_n^{\text{t,side}}$. The state of the entire wind farm consists of the observations of all WTs, i.e. $s = \{o_1, \ldots, o_N\}$.

2) Action: At each time step t, the action of each WT n (a_t^n) includes the inertial gain and droop gain in current time step.

3) Reward: After all agents take actions, they obtain a shared reward:

$$\begin{aligned} r_t = &- C_1(f_{\text{nom}} - f_{\text{sys}})^2 - C_2 \left| \frac{\mathrm{d}f_{\text{sys}}}{\mathrm{d}t} \right| \\ &- \frac{C_3}{N} \sum_{n=1}^{N} (|d_n^{\text{b,flap}}| + |d_n^{\text{b,edge}}| + |d_n^{\text{t,fore}}| + |d_n^{\text{t,side}}|) \end{aligned} \tag{3}$$

The three terms of equation above are used to improve the nadir of frequency, reduce the RoCoF, and reduce the displacements of mechanical quantities. C_1, C_2, C_3 are adjustable weight factors for each term. Note that we can also use other forms of penalty functions to reduce fluctuations in frequency and mechanical quantities. The mechanical quantities $d_n^{\text{b,flap}}, d_n^{\text{b,edge}}, d_n^{\text{t,fore}}, d_n^{\text{t,side}}$ that we use are normalised data. For example, if the original value of the blades flap-wise deflection is $\tilde{d}_n^{\text{b,flap}}$, the mean is $\bar{d}^{\text{b,flap}}$, and the variance is $\sigma^{\text{b,flap}}$, then the normalised deflection is

$$d_n^{\text{b,flap}} = \frac{\tilde{d}_n^{\text{b,flap}} - \bar{d}^{\text{b,flap}}}{\sigma^{\text{b,flap}}} \tag{4}$$

IV. MADRL

We use the MADRL algorithm MADDPG [15] to solve this Dec-POMDP problem. MADDPG is an actor-critic, model-free algorithm based on the deterministic policy gradient [17] that can operate in continuous state and action space. In MADDPG, each agent has a policy function and an action-value function: the policy function acts as an actor, generating actions and interacting with the environment; the action-value function acts as a critic, which evaluates the performance of the actor and guides the follow-up of the actor. Consider a collaborative operation with N WT agents with policies parameterized by $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_N\}$, and let $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_N\}$ be the set of all agent policies. Then we can write the gradient of the expected cumulative reward of agent n as:

$$\begin{aligned} \nabla_{\theta_n} J(\theta_n) = &\mathbb{E}_{s,a\sim\mathcal{D}} \big[\nabla_{\theta_n} \mu_n(a_n|o_n) \\ &\nabla_{a_n} Q_n^{\boldsymbol{\mu}}(s, a_1, \ldots, a_N)|_{a_n=\mu_n(o_n)} \big] \end{aligned} \tag{5}$$

Here $Q_n^{\boldsymbol{\mu}}(s, a_1, \ldots, a_N)$ is a centralised action-value function that takes as input the actions of all agents, $a_1, \ldots, a_N$, in addition to some state information s, and outputs the Q-value for agent n. The experience replay buffer $\mathcal{D}$ contains the tuples $(s, s', a_1, \ldots, a_N, r)$, recording experiences of all agents. The centralised action-value function $Q_n^{\boldsymbol{\mu}}$ is updated as:

$$\mathcal{L}(\theta_n) = \mathbb{E}_{s,a,r,s'}[(Q_n^{\boldsymbol{\mu}}(s, a_1, \ldots, a_N) - y)^2] \tag{6}$$

where

$$y = r + \gamma Q_n^{\boldsymbol{\mu}'}(s', a_1', \ldots, a_N')|_{a_j' = \mu_j'(o_j)} \tag{7}$$

where $\boldsymbol{\mu}' = \{\mu_{\theta_1'}, \ldots, \mu_{\theta_N'}\}$ is the set of target policies with delayed parameters θ_n'. Pseudo-code for MADDPG algorithm is shown in Alg. 1.

Algorithm 1: MADDPG for controlling WTs

for *episode* $= 1$ **to** *number of episodes* **do**
  foreach *time step* t **do**
    for each agent n, select action $a_n = \mu_{\theta_n}(o_n) + \mathcal{N}_t$
    Execute actions $a = (a_1, \ldots, a_N)$ and observe reward r and next state s'
    Store (s, a, r, s') in replay buffer $\mathcal{D}$
    $s \leftarrow s'$
    for *agent* $n = 1$ **to** N **do**
      Sample a random minibatch of S samples (s^j, a^j, r^j, s'^j) from $\mathcal{D}$
      Set $y^j = r_i^j + \gamma Q_n^{\boldsymbol{\mu}'}(s'^j, a_1', \ldots, a_N')|_{a_k' = \mu_k'(o_k^j)}$
      Update critic by minimizing the loss $\mathcal{L}(\theta_n) = \frac{1}{S}\sum_j (y^j - Q_n^{\boldsymbol{\mu}}(s^j, a_1^j, \ldots, a_N^j))^2$
      Update actor using the sampled policy gradient: $\nabla_{\theta_n} J \approx \frac{1}{S}\sum_j \nabla_{\theta_n}\mu_n(o_n^j) \cdot \nabla_{a_n} Q_n^{\boldsymbol{\mu}}(s^j, a_1^j, \ldots, a_N^j)|_{a_n = \mu_n(o_n^j)}$
    Update target network parameters for each agent n: $\theta_n' \leftarrow \tau\theta_n + (1-\tau)\theta_n'$

V. Case Study

As shown in Fig. 2, the simulation experiments are carried out on a two-area test system, which is scaled down from a two-area benchmark power system [18]. The two-area system has four synchronous generators each rated at 15 MVA and they are divided between the two areas equally. The wind farm is made up of three NREL 5 MW Baseline WTs. The major properties of the NREL 5 MW Baseline WT are shown in Table I. Detailed aerodynamic and mechanical WT model are connected to MATLAB/Simulink simulation where the electrical aspects were simulated, resulting in the grid-connected full-scale converter (FSC)-WT system [19] shown in Fig. 3.

We generated wind speeds for the three WTs using NREL TurbSim [20], which is a stochastic, full-field turbulence simulator. The wind speed of each WT is shown in Fig. 4.

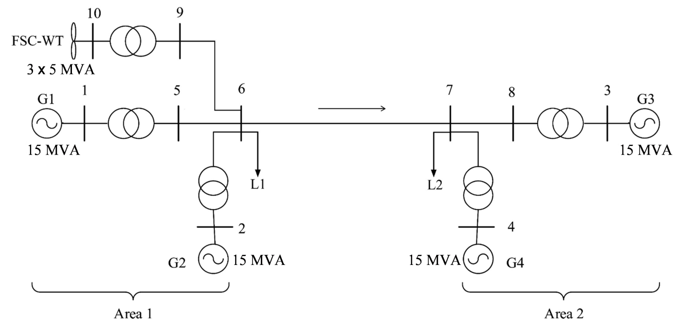

Fig. 2: Four-machine two-area test system with a wind farm.

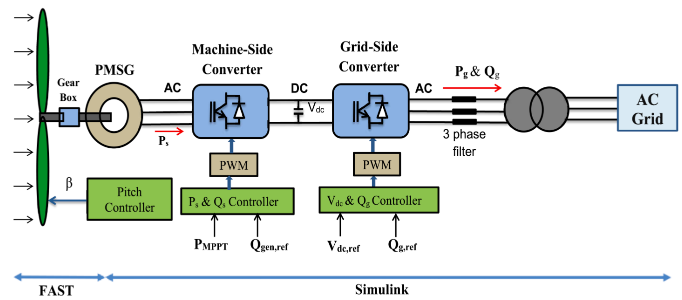

Fig. 3: Developed FSC-WT model using FAST and Simulink.

TABLE I: Properties of the NREL 5 MW Baseline WT

Rating	5 MW
Rotor orientation, configuration	UpWind, 3 blades
Control	variable speed, collective pitch
Rotor, hub diameter	126 m
Tower height	90 m
Cut-in, rated, cut-out wind speed	3 m/s, 11.4 m/s, 25 m/s
Rated generator speed	122.91 rad/s
Rotor mass	110 T
Nacelle mass	240 T
Tower mass	347.460 T

In this experiment, the length of each time step is set to 0.2 seconds, i.e., the values of inertia and droop gains are updated every 0.2 seconds. We use a sudden connection to a 12 MW load to produce a drop in frequency.

The proposed method is compared with three other methods: zero-valued gains, non-zero constant gains and parabolic gains. A value of zero for gains is equivalent to the fact that WTs do not participate in frequency regulation, which is noted as "None" in the following figures. In the constant gains setting,

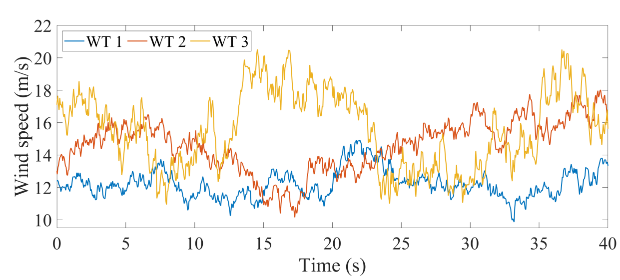

Fig. 4: Wind speed of 3 WTs.

the inertia gain is constant at 25 and the droop gain is constant at 6 [7]. Parabolic gains are proposed by [7] as a coordinated control method that combines a parabolic function for the inertia variable and a linear function for the droop variable. In this method, large gains are set to increase the power output from the WTs at the instant of frequency drop. As the time increases, the gains decreases gradually, preventing the WTs from overdecelerating.

Fig. 5 illustrates the cumulative reward variation of MADDPG algorithm during the training process. After 25 episodes of training, the cumulative reward of MADDPG algorithm converges and outperforms the other three methods. Parabolic gains perform better than constant gains, and gains of zero are the least effective.

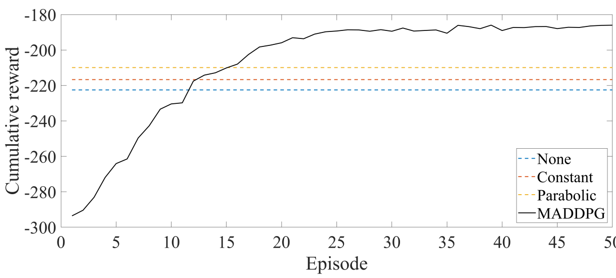

Fig. 5: Episodic average cumulative reward during training.

After training, the MADDPG-based gains of the three WTs as a function of time are shown in Fig. 6. Fig. 6 also shows the gains of the other three methods. Since the MADDPG algorithm tunes the gains based on the state of the WT, the gains are different for different WTs. The other three methods do not consider the state of the WT, so all three WTs have the same gains. It can be seen that the gains of WT 1 and WT 2 based on the MADDPG algorithm drop rapidly at the instant when the frequency drops. This is because the $|\frac{\mathrm{d}f_{\mathrm{sys}}}{\mathrm{d}t}|$ is very large at this moment, and reducing the gains appropriately will prevent the rotor from dropping too fast and causing shocks to the mechanical structure. However, WT 3 still maintains large gains due to its high wind speed and thus more energy available to be consistently taken from the wind.

The frequency variation curves for the different methods are shown in Fig. 7. It can be seen that both MADDPG-based gains and parabolic gains are effective in suppressing the drop in frequency, with frequency nadirs of 49.608 Hz and 49.631 Hz respectively. Constant gains are less effective than time-varying gains, with a frequency nadir of 49.581 Hz for non-zero constant gains and a frequency nadir of 49.439 Hz for zero gains.

The mechanical response of WT 1 for the different methods is shown in Fig. 8. It can be seen that blade flap-wise tip deflection and tower fore-aft displacement are greater with time-varying gains than with constant gains. This means that the faster the WT releases kinetic energy to the grid, the better the frequency regulation, but also the greater the mechanical structure vibrates. The MADDPG-based gains and the parabolic gains have similar frequency regulation capabilities,

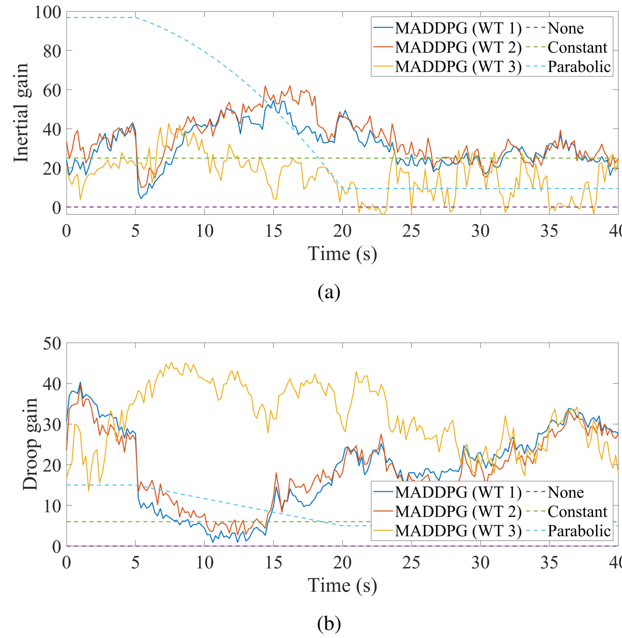

Fig. 6: Inertial and droop gains (a) inertial gains; (b) droop gains.

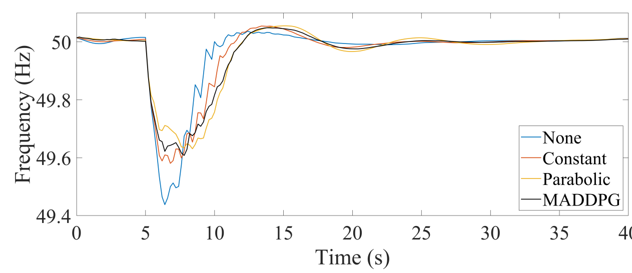

Fig. 7: System frequency variation under different frequency regulation methods.

but the mechanical structures have smaller deflections with MADDPG-based gains.

VI. Conclusion

In this paper, we model the cooperative frequency regulation problem of WTs in a wind farm as a Dec-POMDP and solve it using the MADRL algorithm. Each WT tunes its inertia gain and droop gain in real time based on its own observation. Simulation experiments based on FAST and MATLAB/Simulink verified that the proposed method is not only effective in raising the frequency nadir, but also in reducing the impact of frequency control on the mechanical structure.

References

[1] J. Morren, S. W. De Haan, W. L. 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.

[2] J. F. Conroy and R. Watson, “Frequency response capability of full converter wind turbine generators in comparison to conventional generation,” *IEEE transactions on power systems*, vol. 23, no. 2, pp. 649–656, 2008.

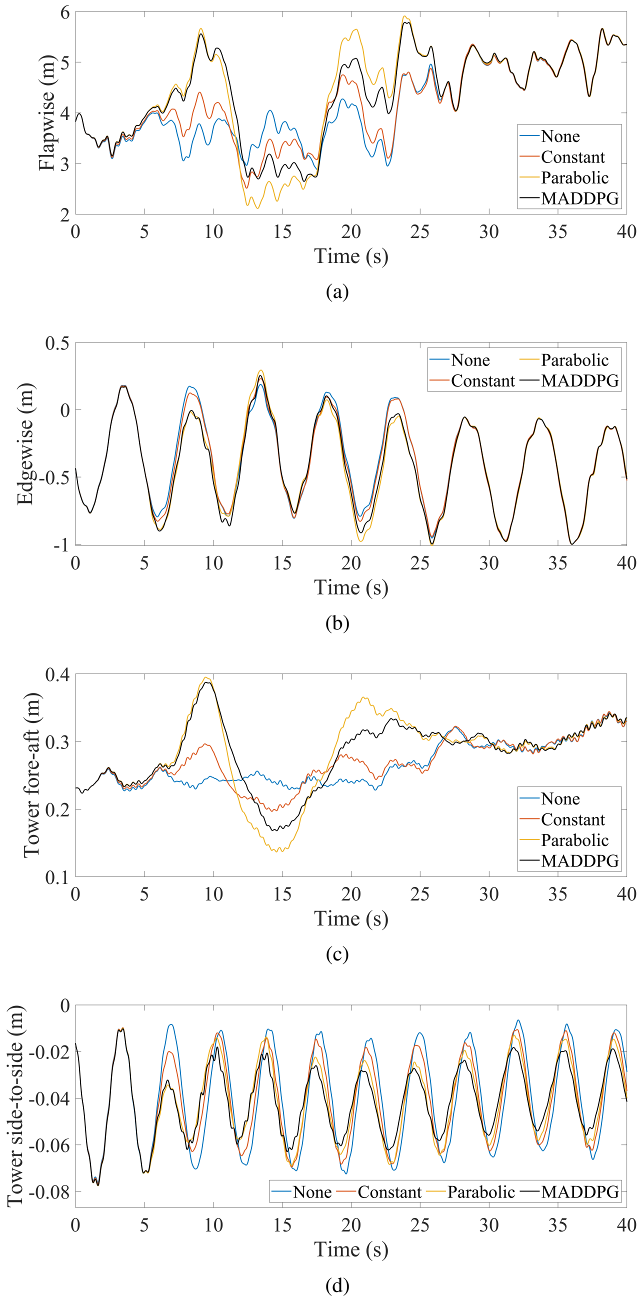

Fig. 8: Mechanical response of WT 1 (a) blades flap-wise, (b) blades edge-wise tip deflections, (c) tower fore-aft, and (d) tower side-to-side displacements.

[3] J. Van de Vyver, J. D. De Kooning, B. Meersman, L. Vandevelde, and T. L. Vandoorn, "Droop control as an alternative inertial response strategy for the synthetic inertia on wind turbines," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1129–1138, 2015.

[4] J. Lee, E. Muljadi, P. Srensen, and Y. C. Kang, "Releasable kinetic energy-based inertial control of a dfig wind power plant," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 279–288, 2015.

[5] J. Lee, G. Jang, E. Muljadi, F. Blaabjerg, Z. Chen, and Y. C. Kang, "Stable short-term frequency support using adaptive gains for a dfig-based wind power plant," *IEEE Transactions on Energy Conversion*, vol. 31, no. 3, pp. 1068–1079, 2016.

[6] S. Kuenzel, L. P. Kunjumuhammed, B. C. Pal, and I. Erlich, "Impact of wakes on wind farm inertial response," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 237–245, 2013.

[7] Y.-K. Wu, W.-H. Yang, Y.-L. Hu, and P. Q. Dzung, "Frequency regulation at a wind farm using time-varying inertia and droop controls," *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 213–224, 2018.

[8] K. Vidyanandan and N. Senroy, "Primary frequency regulation by deloaded wind turbines using variable droop," *IEEE transactions on Power Systems*, vol. 28, no. 2, pp. 837–846, 2012.

[9] G. P. Prajapat, N. Senroy, and I. N. Kar, "Wind turbine structural modeling consideration for dynamic studies of dfig based system," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 4, pp. 1463–1472, 2017.

[10] J. M. Jonkman, M. L. Buhl Jr *et al.*, "Fast user's guide," *Golden, CO: National Renewable Energy Laboratory*, vol. 365, p. 366, 2005.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[12] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.

[13] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4295–4304.

[14] F. A. Oliehoek and C. Amato, *A concise introduction to decentralized POMDPs*. Springer, 2016.

[15] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *arXiv preprint arXiv:1706.02275*, 2017.

[16] L. Kraemer and B. Banerjee, "Multi-agent reinforcement learning as a rehearsal for decentralized planning," *Neurocomputing*, vol. 190, pp. 82–94, 2016.

[17] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*. PMLR, 2014, pp. 387–395.

[18] P. Kundur, "Power system stability," *Power system stability and control*, pp. 7–1, 2007.

[19] M. Edrah, X. Zhao, W. Hung, P. Qi, B. Marshall, S. Baloch, and A. Karcanias, "Electromechanical interactions of full scale converter wind turbine with power oscillation damping and inertia control," *International Journal of Electrical Power & Energy Systems*, vol. 135, p. 107522, 2022.

[20] B. J. Jonkman, "Turbsim user's guide: Version 1.50," Tech. Rep., 2009.