MULTI-AGENT DEEP REINFORCEMENT LEARNING FOR SOLAR-BATTERY SYSTEM TO MITIGATE SOLAR CURTAILMENT IN REAL-TIME ELECTRICITY MARKET

Jinhao Li
Department of Data Science and AI
Monash University
stephlee175@gmail.com

Changlong Wang
Department of Civil Engineering
Monash University
chang.wang@monash.edu

Hao Wang *
Department of Data Science and AI
and Monash Energy Institute
Monash University
hao.wang2@monash.edu

ABSTRACT

The increased uptake of solar energy in the energy transition towards decarbonization has caused the issue of solar photovoltaic (PV) curtailments, resulting in significant economic losses and hindering the energy transition. To overcome this issue, battery energy storage systems (BESS) can serve as onsite backup sources for solar farms. However, the backup role of the BESS significantly limits its economic value, disincentivizing the BESS deployment due to high investment costs. Hence, it is essential to effectively reduce solar curtailment while ensuring viable operations of the BESS. To better understand the synergy of a co-located solar-BESS system in the real-time electricity market, we model the cooperative bidding processes of the solar farm and the BESS as a Markov game. We use a multi-agent deep reinforcement learning (MADRL) algorithm, known as multi-agent deep deterministic policy gradient, to concurrently maximize the overall revenue from the electricity market and reduce solar curtailments. We validate our MADRL-based strategy using data from a realistic solar farm operating in the Australian electricity market. The simulation results show that our MADRL-based coordinated bidding strategy outperforms both optimization-based and DRL-based benchmarks, generating higher revenue for the BESS and reducing more solar curtailments. Our work highlights the importance of coordination between the BESS and renewable generations for both economic benefits and progress towards net-zero transitions.

1 INTRODUCTION AND BACKGROUND

As we move towards a net-zero emission future, renewable energy resources have been increasingly adopted in modern power systems for decarbonization and climate change mitigation (Lu et al. (2021)). In particular, solar photovoltaic (PV) has become the leading source of utility-scale renewables (Comello & Reichelstein (2019)) due to its declining levelized electricity costs and convenient facilitation. Solar energy has attracted about half of the global renewable investment and contributed the most net added generation capacity in the past decade, with its installed capacity increasing from 304.3 GW in 2016 to 760.4 GW in 2020 worldwide (IEA (2020)). However, the stochastic nature of solar irradiation leads to a high degree of variability in solar PV generation, subsequently causing power supply-demand imbalances and oversupply during the daytime, in particular when the solar

*Corresponding author: Hao Wang. This work has been supported in part by the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) under Grant DE230100046.

penetration in the grid grows significantly (Zahedi (2011)). Solar curtailment is often inevitable to ensure the secure operation of power systems, but leads to substantial economic losses and emerges as a new challenge for net-zero transitions (Mallapragada et al. (2020)). Flexible energy resources are needed to address the solar curtailment challenge (Frew et al. (2021)).

Battery energy storage systems (BESS) have been increasingly facilitated and co-located with solar farms as a flexible resource for reducing solar curtailments and improving the economic performance of solar farms (Hernandez et al. (2015); Chu et al. (2017); Bumpus & Comello (2017)). In a co-located solar-BESS system, the BESS is usually used as the backup source to smooth the fluctuating solar PV generation by drawing excess solar energy (otherwise to be curtailed) from the onsite solar farm and releasing stored energy in the event of solar shortages (Wang et al. (2014)). The investment cost of the BESS remains significant, despite experiencing a decrease in the past decades. The cost-effective integration of the BESS into solar farms through optimal placement and sizing has been studied in Babacan et al. (2017); Khezri et al. (2020); Hassan et al. (2022). However, when the BESS is only used for reducing solar curtailments, its flexibility has not been fully exploited, limiting the full economic potential of the BESS in the electricity market (Mallapragada et al. (2020)). The BESS can take advantage of its operational flexibility to perform energy arbitrage (i.e., buy low and sell high) in the wholesale electricity market, as a prospective revenue stream for the BESS in addition to the ancillary role of reducing solar curtailment. Hence, it is of great significance to design an effective coordination strategy for the co-located solar-BESS system to simultaneously reduce solar curtailment and improve the BESS's economic viability, benefiting not only the solar farms and the BESS, but also the power system by reducing carbon emissions and promoting climate change mitigation. However, designing such an effective coordination strategy is not trivial due to the variability of solar PV generation and the volatile electricity market prices.

BESS bidding strategies have been predominantly derived from optimization-based methods, e.g., stochastic and robust optimization. For example, related works in González-Garrido et al. (2019); Attarha et al. (2019); Aghamohamadi et al. (2021); Yang et al. (2022); Graça Gomes et al. (2023) studied the revenue maximization problems of the solar-BESS system and focused primarily on participating in the day-ahead electricity market, whose performance heavily depended on the modeling of price forecasts. In addition, these studies often neglected other tasks for the BESS, such as renewable energy curtailment mitigation. Moreover, participating in the real-time wholesale electricity market is more profitable than the day-ahead market yet more challenging due to both the high volatility of the real-time market and complex price drivers (Weron (2014)).

To bridge the research gap in the coordinated bidding strategy between a solar farm and the BESS in real-time markets, we propose a novel multi-agent deep reinforcement learning (MADRL)-based bidding strategy for the co-located solar-BESS system. Our strategy aims to concurrently manage solar curtailments and optimize the system's participation in the wholesale real-time market. The solar-BESS system is inherently decentralized, where the two assets can be managed separately. Such decentralized characteristic enables MADRL to effectively coordinate the bidding processes of the two assets and to balance the trade-off between market participation and solar curtailment reduction. Moreover, unlike the aforementioned optimization-based methods, the MADRL-based strategy can mitigate the uncertainties of solar PV generation and energy prices in a model-free manner and dynamically learn the optimal coordinated bidding strategy from historical experiences without prior knowledge or forecast results. We validate our MADRL method using the realistic electricity market and solar farm data collected from the Australian National Electricity Market (NEM), which supplies around 9 million customers with a trading value of 16.6 billion Australian dollars per annum. The coordinated bidding strategy for the co-located solar-BESS system is presented in Section 2.

2 METHODOLOGY

We consider a co-located solar-BESS system as a price-taker participating in the real-time wholesale electricity market, also known as the spot market (which is a major component of the Australian NEM). Note that our price-taker model does not consider transmission network constraints, indicating that our bids will not affect other generator bidding decisions or market clearing outcomes. The co-located solar-BESS system receives the price signal every five minutes via a dispatch process empowered by the Australian Electricity Market Operator (AEMO). The system paradigm is illus-

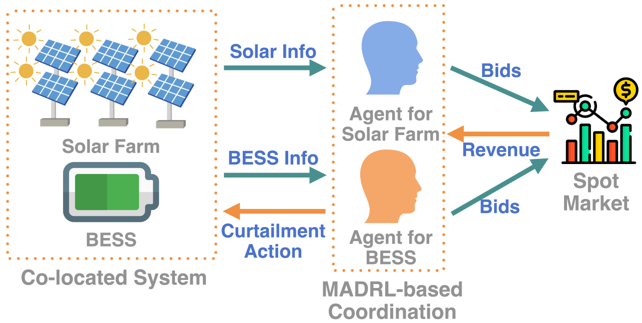


Figure 1: The coordinated bidding framework of the solar-BESS system empowered by MADRL.

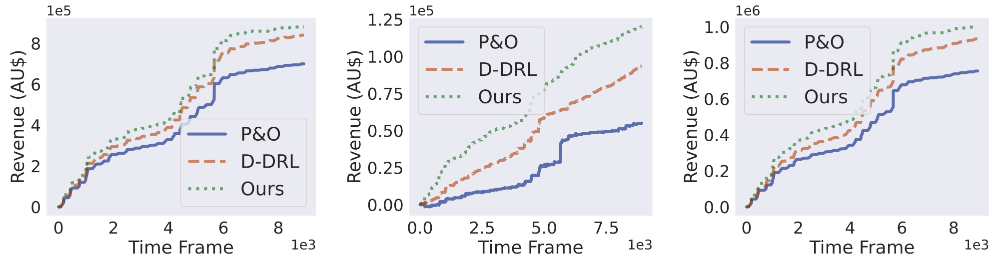
trated in Fig. 1. For the solar farm, it submits the power bid regarded as the dispatch target, denoted by p_t^{S}, at each dispatch interval in a five-minute resolution (AEMO (2017)). For the BESS, it simultaneously performs energy arbitrage in the spot market and manages solar curtailment via charging solar energy (otherwise curtailed) from the onsite solar farm. We denote the BESS's market bid and the power intended to draw from onsite solar curtailment as $p_t^{\mathrm{BESS,SM}}$ and $\hat{p}_t^{\mathrm{BESS,SC}}$, respectively.

2.1 Learning Optimal Coordinated Bidding Strategy via MADRL

The co-located solar-BESS system inherently functions in a decentralized manner, i.e., the two heterogeneous assets separately submit their bids in the spot market, whose decision-making processes can be affected by various factors, e.g., uncertain solar PV generation of the solar farm and the time-varying capacity of the BESS as a result of its previous charge/discharge actions. Such decentralized characteristic enables us to model the real-time bidding process of the co-located system as a Markov game, where the solar farm and the BESS act as two cooperative players in the Markov game to maximize the overall revenue in the spot market, while concurrently managing solar curtailments. Specifically, the BESS can take advantage of the excessive solar energy (otherwise curtailed) from the onsite solar farm which can be exported back to the grid later on through the BESS's bids.

A Markov game typically consists of five elements: agent space $\mathcal{N}$, state space $\mathcal{S}$, action space $\mathcal{A}$, probability space $\mathcal{P}$, and reward space $\mathcal{R}$. In particular, the state space $\mathcal{S}$ includes all observable information from all agents in the Markov game. Specifically, the states of the solar farm are the last clearing price of the spot market (known as the spot price) λ_{t-1}, the last solar generation $p_{t-1}^{\mathrm{S,Act}}$, and the current hour index defined as $h_t \in \{0, 1, \cdots, 23\}$. Given that solar generators sit idle when there is no solar irradiation, e.g., at night times, we introduce such an hour index to indicate the effect of solar irradiation on solar PV output profiles. The state representation of the solar farm is thus defined as $\boldsymbol{s}_t^{\mathrm{S}} = [\lambda_{t-1}, p_{t-1}^{\mathrm{S,Act}}, h_t]$. For the BESS, its states include the last spot price, the BESS's capacity e_{t-1}, and the hour index. Moreover, to achieve more effective solar curtailment management, we include the number of solar curtailment events (defined as the scenarios in which solar curtailment would occur if there were no BESS for absorbing the excessive solar energy otherwise curtailed) within the latest L dispatch intervals, denoted by f_t^{SC}, in the BESS's state. The BESS's state can be expressed as $\boldsymbol{s}_t^{\mathrm{BESS}} = [\lambda_{t-1}, e_{t-1}, f_{t-1}^{\mathrm{SC}}, h_t]$.

The action space $\mathcal{A}$ is a set including the underlying actions of all agents. For the solar farm, its action is the solar power bid in the spot market, defined as $\boldsymbol{a}_t^{\mathrm{S}} = [p_t^{\mathrm{S}}]$. The BESS's actions include its charge/discharge binary variables denoted by v_t^{Ch} and v_t^{Dch}, the power bid in the spot market $p_t^{\mathrm{BESS,SM}}$, and the power intended to draw from the onsite solar curtailment $\hat{p}_t^{\mathrm{BESS,SC}}$, which can be denoted as $\boldsymbol{a}_t^{\mathrm{BESS}} = [v_t^{\mathrm{Ch}}, v_t^{\mathrm{Dch}}, p_t^{\mathrm{BESS,SM}}, \hat{p}_t^{\mathrm{BESS,SC}}]$. A reward mechanism $\mathcal{R}$ is then developed to examine the effectiveness of agent actions under the current states. In MADRL, the optimal bidding strategy of each agent is learned via maximizing their corresponding expected cumulative rewards. Thus, designing appropriate reward mechanisms for the solar farm and the BESS plays a critical role in optimizing the formulated Markov game. The detailed reward mechanism formulation is presented in Appendix A. We introduce the multi-agent deep deterministic policy gradient (MADDPG) (Lowe et al. (2017)) for learning the coordinated bidding strategies of the solar-BESS system.



(a) Solar Farm (b) BESS (c) Solar-BESS System

Figure 2: The evaluated cumulative revenue comparisons of the P&O, D-DRL, and our method.

Table 1: The evaluated cumulative revenue, the amount of absorbed curtailed solar energy, and reduced carbon emissions derived from the P&O, D-DRL, and our MADRL.

Strategy	Solar Revenue (Unit:AU$)	BESS Revenue (Unit:AU$)	Total Revenue (Unit:AU$)	Absorbed Curtailed Energy (Unit:MWh)	Reduced Carbon Emissions (Unit:Kilos)
P&O	$699,539$	$55,031$	$754,569$	245	174
D-DRL	$839,779$	$93,490$	$933,268$	489	347
Ours	$\mathbf{880,621}$	$\mathbf{120,149}$	$\mathbf{1,000,770}$	$\mathbf{725}$	$\mathbf{514}$

3 EXPERIMENTAL RESULTS

We use real-world solar farm data collected from the Rugby Run Solar Farm located in Queensland, Australia, together with energy prices from the Queensland jurisdiction of the Australian NEM (AEMO (2020)), to train and evaluate our MADRL-based bidding strategy. Specifically, data in the first eleven months of 2020 is for training and the last month is for evaluation. One Nvidia GeForce RTX 3090 is utilized for algorithm training. Also, we develop one optimization-based benchmark and one DRL-based benchmark for comparison. The former relies on LSTM networks to produce forecast results of the solar PV generation and energy prices, which are then used by a mixed integer linear programming solver empowered by the PuLP library (Mitchell et al. (2011)) to concurrently maximize the overall revenue and minimize solar curtailments (through a weighted objective). We refer to the optimization-based method as predict-and-optimize (P&O). For the DRL-based benchmark, we decouple the bidding process of the co-located system into two Markov decision processes (MDPs) for the solar farm and the BESS, respectively. These two MDPs are then optimized by two independent DDPG Lillicrap et al. (2015) agents. The settings regarding both the MDP and the DDPG, including the state, action, reward, and algorithmic hyperparameters, remain the same for a fair comparison. We refer to it as the decoupled DRL (D-DRL) method. We present all initialized parameters of our simulations in Appendix B.

The cumulative revenues of the solar farm, the BESS, and the solar-BESS system as a whole achieved by the P&O, D-DRL, and our method are illustrated in Fig. 2, respectively. The associated statistics are presented in Table 1 for cross-comparison. The bidding outcomes reveal that our MADRL-based bidding strategy significantly outperforms both the optimization-based and DRL-based benchmarks, with substantial revenue boosts by 27% for the solar farm, 118% for the BESS, and 33% for the whole system, compared to the P&O method and by 5%, 28%, and 7% compared to the D-DRL benchmark, respectively.

Furthermore, the BESS empowered by our MADRL-based strategy tends to be more responsive to solar curtailment events and absorbs a considerable amount of solar energy (otherwise to be curtailed), as shown in Table 1. The curtailment mitigation results demonstrate the significant capability of our method in reducing solar curtailments and taking advantage of onsite curtailed solar energy to improve the economic benefits of the solar-BESS system, as depicted in Fig. 2.

Besides improving the viability of the BESS, reducing solar curtailments can reduce carbon emissions to some extent and further promote the energy transition. This is because the absorbed curtailed solar energy, i.e., renewable energy, will be exported back to the grid through the BESS's bidding, enabling more variable renewable energy transactions in the electricity market. We convert the absorbed curtailed energy (in MWh) into the reduced carbon emissions (in kilograms) in Table 1 (EPA (2022)). The associated results indicate that effective coordination between the BESS and the solar farm, i.e., managing solar curtailments while performing energy arbitrage, would have a significantly positive impact on decarbonizing the future power grids and mitigating climate change.

4 Conclusion

This paper develops an effective coordination strategy for the co-located solar-BESS system to improve the system's economic benefits in the spot market and reduce carbon emissions. We propose a model-free MADRL-based bidding strategy for the solar farm and the BESS to dynamically balance the trade-off between energy arbitrage and solar curtailment reduction. We validate the proposed MADRL-based method using realistic solar farm data in the Australian NEM. Simulation results demonstrate the effectiveness of our method, which outperforms the optimization-based and DRL-based benchmarks in achieving more economic benefits and better management of solar curtailment.

References

AEMO. *Five-minute Settlement: High Level Design*. Australian Energy Market Operator, 2017.

AEMO. *How the National Electricity Market works*. Australian Energy Market Operator, 2020.

Mehrdad Aghamohamadi, Amin Mahmoudi, John K. Ward, Mohammed H. Haque, and João P. S. Catalão. Block coordinate decent robust bidding strategy of a solar photovoltaic coupled energy storage system operating in a day-ahead market. In *2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia)*, pp. 1–6, 2021. doi: 10.1109/ISGTAsia49270.2021.9715560.

Ahmad Attarha, Nima Amjady, and Shahab Dehghan. Affinely adjustable robust bidding strategy for a solar plant paired with a battery storage. *IEEE Transactions on Smart Grid*, 10(3):2629–2640, 2019. doi: 10.1109/TSG.2018.2806403.

Oytun Babacan, William Torre, and Jan Kleissl. Siting and sizing of distributed energy storage to mitigate voltage impact by solar pv in distribution systems. *Solar Energy*, 146:199–208, 2017. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2017.02.047.

A. Bumpus and S. Comello. Emerging clean energy technology investment trends. *Nature Climate Change*, 7(6):382–385, Jun 2017. ISSN 1758-6798. doi: 10.1038/nclimate3306.

Steven Chu, Yi Cui, and Nian Liu. The path towards sustainable energy. *Nature Materials*, 16(1): 16–22, Jan 2017. ISSN 1476-4660. doi: 10.1038/nmat4834.

Stephen Comello and Stefan Reichelstein. The emergence of cost effective battery storage. *Nature Communications*, 10(1):2038, May 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09988-z.

EPA. *Greenhouse Gases Equivalencies Calculator - Calculations and References*. U.S. Environmental Protection Agency, 2022.

Bethany Frew, Brian Sergi, Paul Denholm, Wesley Cole, Nathaniel Gates, Daniel Levie, and Robert Margolis. The curtailment paradox in the transition to high solar power systems. *Joule*, 5(5): 1143–1167, 2021.

Amaia González-Garrido, Andoni Saez-de Ibarra, Haizea Gaztañaga, Aitor Milo, and Pablo Eguia. Annual optimized bidding and operation strategy in energy and secondary reserve markets for solar plants with storage systems. *IEEE Transactions on Power Systems*, 34(6):5115–5124, 2019. doi: 10.1109/TPWRS.2018.2869626.

João Graça Gomes, Juan Jiang, Cheng Tung Chong, João Telhada, Xu Zhang, Sergio Sammarchi, Shuyang Wang, Yu Lin, and Jialong Li. Hybrid solar pv-wind-battery system bidding optimisation: A case study for the iberian and italian liberalised electricity markets. *Energy*, 263:126043, 2023. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2022.126043.

Aakash Hassan, Yasir M. Al-Abdeli, Martin Masek, and Octavian Bass. Optimal sizing and energy scheduling of grid-supplemented solar pv systems with battery storage: Sensitivity of reliability and financial constraints. *Energy*, 238:121780, 2022. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2021.121780.

Rebecca R. Hernandez, Madison K. Hoffacker, and Christopher B. Field. Efficient use of land to meet sustainable energy needs. *Nature Climate Change*, 5(4):353–358, Apr 2015. ISSN 1758-6798. doi: 10.1038/nclimate2556.

IEA. *2020 Snapshot of Global PV Markets*. International Energy Agency, 2020.

Rahmat Khezri, Amin Mahmoudi, and Mohammed H. Haque. Optimal capacity of solar pv and battery storage for australian grid-connected households. *IEEE Transactions on Industry Applications*, 56(5):5319–5329, 2020. doi: 10.1109/TIA.2020.2998668.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2015. URL `https://arxiv.org/abs/1509.02971`.

Ryan Lowe, YI WU, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/68a9750337a418a86fe06c1991a1d64c-Paper.pdf`.

Xi Lu, Shi Chen, Chris P. Nielsen, Chongyu Zhang, Jiacong Li, He Xu, Ye Wu, Shuxiao Wang, Feng Song, Chu Wei, Kebin He, Michael B. McElroy, and Jiming Hao. Combined solar power and storage as cost-competitive and grid-compatible supply for china's future carbon-neutral electricity system. *Proceedings of the National Academy of Sciences*, 118(42):e2103471118, 2021. doi: 10.1073/pnas.2103471118.

Dharik S. Mallapragada, Nestor A. Sepulveda, and Jesse D. Jenkins. Long-run system value of battery energy storage in future grids with increasing wind and solar generation. *Applied Energy*, 275:115390, 2020. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2020.115390.

Stuart Mitchell, Michael OSullivan, and Iain Dunning. Pulp: a linear programming toolkit for python. *The University of Auckland, Auckland, New Zealand*, 65, 2011.

Guishi Wang, Mihai Ciobotaru, and Vassilios G. Agelidis. Power smoothing of large solar pv plant using hybrid energy storage. *IEEE Transactions on Sustainable Energy*, 5(3):834–842, 2014. doi: 10.1109/TSTE.2014.2305433.

Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2014.08.008.

Yinping Yang, Chao Qin, Yuan Zeng, and Chengshan Wang. Optimal coordinated bidding strategy of wind and solar system with energy storage in day-ahead market. *Journal of Modern Power Systems and Clean Energy*, 10(1):192–203, 2022. doi: 10.35833/MPCE.2020.000037.

A. Zahedi. Maximizing solar pv energy penetration using energy storage technology. *Renewable and Sustainable Energy Reviews*, 15(1):866–870, 2011. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2010.09.011.

A REWARD MECHANISM OF THE SOLAR-BESS SYSTEM

For the solar farm agent, to mitigate the uncertainty of solar PV generation and update accurate dispatch targets, the reward function for the solar farm can be formulated as

$$r_t^{\mathrm{S}} = -\lambda_t \left| \frac{p_t^{\mathrm{S}} - p_t^{\mathrm{S,Act}}}{p_t^{\mathrm{S,Ava}}} \right|, \tag{1}$$

where λ_t is the spot price and $p_t^{\mathrm{S,Ava}}$ is the availability, i.e., upper power limit, of the solar farm at the current dispatch interval.

For the BESS agent, to perform effective energy arbitrage, we introduce two charge/discharge indicators, denoted by $\mathbb{I}_t^{\mathrm{Ch}}/\mathbb{I}_t^{\mathrm{Dch}}$, to indicate the appropriate timing of arbitrage operations, which are defined as

$$\mathbb{I}_t^{\mathrm{Ch}} = \mathrm{sgn}\left(\bar{\lambda}_t - \lambda_t\right), \tag{2}$$

$$\mathbb{I}_t^{\mathrm{Dch}} = \mathrm{sgn}\left(\lambda_t - \bar{\lambda}_t\right), \tag{3}$$

where $\mathrm{sgn}(\cdot)$ is the sign function and $\bar{\lambda}_t$ is the exponential moving average of the spot price. The moving average price $\bar{\lambda}_t$ is expressed as

$$\bar{\lambda}_t = \tau \bar{\lambda}_{t-1} + (1 - \tau)\,\lambda_t, \tag{4}$$

where $\tau \in (0, 1)$ is a smoothing parameter. The proposed charge/discharge indicators encourage the BESS to purchase power at relatively lower spot prices, i.e., $\lambda_t < \bar{\lambda}_t$, and sell power at high prices, i.e., $\lambda_t > \bar{\lambda}_t$. If the BESS does not follow such an arbitrage guideline to bid in the spot market, the charge/discharge indicators will be negative, resulting in negative rewards. Thus, the BESS's reward function for energy arbitrage can be formulated as

$$r_t^{\mathrm{BESS,SM}} = \frac{p_t^{\mathrm{BESS,SM}}}{P_{\mathrm{max}}^{\mathrm{BESS}}} \left|\lambda_t - \bar{\lambda}_t\right| \left(\mathbb{I}_t^{\mathrm{Ch}} v_t^{\mathrm{Ch}} \frac{1}{\eta^{\mathrm{Ch}}} + \mathbb{I}_t^{\mathrm{Dch}} v_t^{\mathrm{Dch}} \eta^{\mathrm{Dch}} \right), \tag{5}$$

where $P_{\mathrm{max}}^{\mathrm{BESS}}$ is the BESS's rated power and η^{Ch}, η^{Dch} are the charging/discharging efficiencies of the BESS.

Moreover, the BESS is incentivized to store otherwise curtailed solar power from the onsite solar farm. The reward function for managing solar curtailments is designed as

$$r_t^{\mathrm{BESS,SC}} = \lambda_t \frac{p_t^{\mathrm{BESS,SC}}}{P_{\mathrm{max}}^{\mathrm{BESS}}} \frac{f_{t-1}^{\mathrm{SC}}}{L} \frac{1}{\eta^{\mathrm{Ch}}}, \tag{6}$$

where $p_t^{\mathrm{BESS,SC}}$ is the actual power drawn from the onsite solar farm.

Combining reward functions for energy arbitrage and solar curtailments, the overall reward function for the BESS can be written as

$$r_t^{\mathrm{BESS}} = r_t^{\mathrm{BESS,SM}} + \beta r_t^{\mathrm{BESS,SC}}, \tag{7}$$

where β is the incentive factor for solar curtailment reduction.

B SIMULATION PARAMETER SETUP

- Australian NEM dispatch interval: 5 minutes.
- BESS charging efficiency: 0.95.
- BESS discharging efficiency: 0.95.
- BESS rated power: 10 MW.
- BESS lower energy limit: 0.5 MWh.
- BESS upper energy limit: 9.5 MWh.
- The time range for counting the number of solar curtailment events: $L = 10$.
- Smoothing parameter for calculating the exponential moving average of the spot price: $\tau = 0.9$.
- Incentive factor for solar curtailment reduction: $\beta = 6$.
- BESS storage capacity: 10 MWh.
- BESS minimum state of charge: 5%.
- BESS maximum state of charge: 95%.
- Noise used in DDPG and MADDPG: Gaussian noise $\mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ set as 0 and 0.1, respectively.
- Structure of the actor network: Three layers all with 256 neurons.
- Structure the critic network: Three layers all with 256 neurons.
- Learning rate for the actor network in DDPG and MADDPG: 0.0008.
- Learning rate for the critic network in DDPG and MADDPG: 0.0008.
- Optimizer for neural networks: Adam.
- Batch size: 256.
- Discounted factor for reward calculation: 0.99.
- Smoothing parameter for target network update: 0.01.