A review of reinforcement learning based approaches for industrial demand response

Margi Shah[1], Yue Zhou [1*], Jianzhong Wu[1], Max Mowbray[2]

1 School of Engineering, Cardiff University, Cardiff, UK

2 The University of Manchester, Manchester, UK

(*Corresponding Author: zhouy68@cardiff.ac.uk)

ABSTRACT

Industrial demand response plays a key role in mitigating the operational challenges of smart grid brought by massive proliferation of distributed energy resources. However, industrial plants have complex and intertwined processes, which provides barriers for their participation in industrial demand response programs. This is in part due to the complexity and uncertainties of approximating systems models. More recently, reinforcement learning has emerged as a data-driven control technique for sequential decision-making under uncertainty. This emergence is strongly coupled with the abundance of data offered by advanced information technologies. The potential of applying reinforcement learning in industrial demand response is identified in this work by comparing pivotal aspects of reinforcement learning with the requirements of industrial demand response schemes.

Keywords: Reinforcement learning, industrial demand response, production process, optimisation

NONMENCLATURE

Abbreviations	
DR	Demand response
RL	Reinforcement learning
DRL	Deep reinforcement learning
ANN	Artificial neural network
MDP	Markov decision process

1. INTRODUCTION

The modernised smart grid facilitates renewable energy integration and enables electricity users to participate in demand response (DR) programs by altering or reducing their energy consumption, enhancing grid efficiency and reliability. Industrial plants account for 21% of global electricity consumption (projected to reach 30% by 2030) [1], and play a crucial role in DR adoption.

Energy-intensive production processes offer significant potential for industrial DR [2]. In [3] and [4], the chemical engineering community has presented scheduling models for industrial plants to actively take part in DR. Recent advancements in scheduling models and solution techniques have revolutionised production scheduling, with the state-of-the-art underpinned by traditional approaches using mathematical programming with heuristic decision-making algorithms also widely used (e.g. nature inspired metaheuristic optimisation) [5].

Implementing DR for industrial consumers requires a model capturing plant dynamics and operational constraints. Real-time DR is hindered by uncertain renewable power generation and electricity prices. Addressing these challenges necessitates an efficient and intelligent real-time solution for monitoring, forecasting, and decision-making.

Reinforcement learning (RL) addresses sequential decision-making in an uncertain process through learning closed-loop feedback control. Unlike traditional techniques relying on detailed plant models, RL's data-driven and model-free approach has the potential to handle complex scheduling problems, and allow for real-time implementation of DR [6]. However, as of now it's potential has been largely unrealised in practice.

In the rest of this paper, the approaches for the scheduling of industrial processes in DR programs are reviewed, with the challenges in the existing methodology and future directions identified. Then RL is introduced for DR in industrial process systems and major challenges to its implementation identified. Concluding remarks are provided thereafter.

2. INDUSTRIAL DEMAND RESPONSE

2.1 Scheduling of industrial process for demand response

In 2022, global industrial emissions reached 9.0 Gt of CO2, accounting for a quarter of global energy system CO2 emissions. The IEA's industrial report identifies major contributors as iron and steel, chemicals, cement, aluminium, and pulp and paper [7]. Industrial plants can participate in incentive-based DR for providing ancillary services to the grid and price-based DR by adjusting the consumption patterns of energy intensive processes, leading to lower energy consumption costs, reduced fossil fuel dependency, and a smaller carbon footprint [8].

Fig. 1 presents a framework that illustrates how industrial plants can participate in the energy and ancillary service markets by adjusting their production schedules in response to market signals. The day ahead market enables industrial plants to participate by submitting their daily production plans in advance, thus facilitating the optimisation of their energy consumption. The intraday market allows rescheduling of production plans in response to critical peak prices. This means that production schedules can be updated frequently to match the changing electricity prices. The ancillary service market is crucial in managing and regulating grid power flow. One of the earliest studies on modelling energy intensive process for participation in the ancillary service market was made in [9]. Approximately 50% of steel mills in Germany have pre-qualified their furnaces in the tertiary reserve market as positive capacity [10].

In [11], the authors highlight key considerations in developing production schedules for industrial plants. These include adherence to safety requirements, operational feasibility, and product delivery commitments for participating in DR. Understanding the physical characteristics, constraints, and temporal dependencies of devices involved in production processes is essential. Additionally, on-site power generation resources must be factored into scheduling considerations.

Software solutions for production scheduling in industrial plants typically employ model predictive control (MPC) and rule-based expert systems. However, rule-based systems have limited flexibility, are time-consuming to implement, and lack dynamic capabilities for improving performance beyond a narrow operating window. When implemented in a receding horizon framework with access to state feedback, MPC reduces to iteratively identifying the solution of (large) mixed integer programs at discrete points in time. In general, decision problems are formulated as mixed integer linear programs (MILP), typically providing some approximation to the complexity of industrial processes, particularly when energy balances are accounted for.

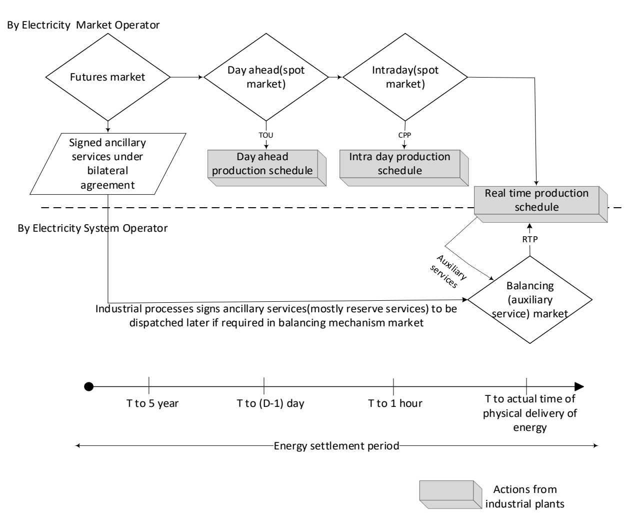

Fig. 1 Industrial process scheduling in energy and ancillary service market programs

The methods applied to solve the mathematical program in MPC broadly include exact optimisation algorithms such as branch-and-bound and nature-inspired (metaheuristic) algorithms [12]. Nature-inspired heuristic optimisation algorithms are often used to find approximate solutions to these complex problems. Some of the commonly used metaheuristic algorithms include genetic algorithm (GA), simulated annealing (SA), particle swarm optimisation (PSO), and ant colony optimisation (ACO) [12]. Although, mathematical programming techniques offer exact optimal solutions, they may have a prohibitively high computational burden online. Whereas nature-inspired algorithms can find approximate solutions with a lower computational burden.

To address uncertainties, scheduling problems often employ stochastic programming and robust optimisation. Stochastic programming explicitly models uncertainty using probability distributions, enabling decision-makers to consider different outcomes, and make informed decisions that optimise in expectation, with the opportunity for recourse decisions as the uncertainty is realised. However, considerable approximations to the decision process are made over the time horizon to ensure online tractability. Robust optimisation considers set-based descriptions of uncertainty, with focus on optimising for the worst-case scenarios, which introduces conservatism into decision-making especially when implemented online.

While these approaches can maintain operations within desired limits, they often fall short in addressing the uncertainty and complexity of industrial processes [13].

2.2 Challenges in scheduling for industrial demand response

Modelling industrial processes for DR requires a high-resolution model. This modelling effort is encompassed by a set of challenges including accurately modelling the plant dynamics and their often-considerable time dependence, and capturing exogeneous variables such as electricity prices and renewable generation [14].

Therefore, it is crucial to identify modelling approaches that integrate with the practicalities of real time scheduling.

The computational complexity of existing solution techniques poses challenges for complex processes with many coupling constraints and binary variables. Simplifying assumptions and heuristic rules may not capture system dynamics accurately. Efficient scheduling algorithms are needed to handle both short-term and long-term planning effects in a computationally feasible manner [15].

It is of interest to industrial consumers to leverage advanced information technology infrastructure for smart and intelligent DR modelling and control [16]. The major aim of RL is to identify function approximations to the optimal control policy in a data-driven manner. In doing so, RL explicitly learns about the plant's dynamics and exogenous uncertainties without the need for a mathematical model of the plant online. Instead, RL policies identify scheduling decisions conditionally to the current system state in a very short time frame through the inference processes of the function approximation. The generalisability and adaptability of RL controllers make the approach particularly compelling over alternative approaches, such as MPC. In contrast, traditional controllers are costly to upgrade and incorporate new features. However, as RL controllers are less robust, operator oversight is likely to be an inherent part of their implementation.

3. REINFORCEMENT LEARNING FRAMEWORK IN PROCESS SCHEDULING FOR INDUSTRIAL DEMAND RESPONSE

3.1 Preliminaries of Reinforcement learning

With the rapid advancement of AI, there is growing interest in utilising model-free reinforcement learning (RL) to address decision-making in the smart grid. RL offers several advantages, including being systems model-free online and identifying scheduling decisions in closed-loop. These characteristics in particular make RL a promising tool for optimising energy management in the face of dynamic factors such as fluctuating electricity prices [17]. By incorporating deep learning techniques, deep reinforcement learning (DRL) can provide scheduling decisions by leveraging artificial neural networks (ANNs) for optimal policy function approximation, providing reactive scheduling decisions cheaply through inference.

RL problems are typically represented as Markov Decision Processes (MDPs), which offer a formal mathematical framework for closed-loop decision-making in a stochastic process [18]. Fig. 2 provides the general framework of the MDP. In the following, we introduce the MDP framework and provide comment on its use within the context of scheduling.

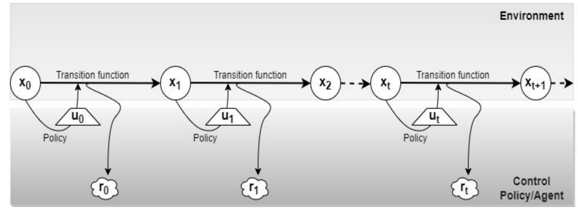

Fig. 2 Illustration of Markov Decision Process

In an MDP, decisions are made at time indices within a finite discrete time horizon of length T. At each discrete time index, $t \in \{0, 1, 2 \ldots \ldots T\}$, the state of the plant is represented by $\boldsymbol{x}_t \in \mathbb{X} \subseteq \mathbb{R}^{n_x}$ and is assumed to contain all information required for decision-making (i.e., the state is Markov). The state defines a representation of the schedule including the tasks currently being processed, indications of equipment availability, and relevant data including for example the duration and energy consumption of the processing tasks at that point in time, together with any electricity price information. At each discrete time step within the scheduling process, the decision-maker is able to observe the state of the plant, and select a control action, $\boldsymbol{u}_t \in \mathbb{U} \subseteq \mathbb{Z}^{n_u}$ as discrete integer values which represent appropriate assignment decisions on the available equipment, (i.e. an allocation of task to equipment at a given time index). These decisions are made according to a policy $\pi: \mathbb{X} \to \mathbb{U}$. Thereafter the next state $\boldsymbol{x}_{t+1}$ is generated from a transition probability function, $\mathbb{P}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)$, which may also be described as:

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{s}_t) \tag{1}$$

where $\boldsymbol{s}_t \in \mathbb{S}$ represents exogenous sources of uncertainty arising from electricity prices and onsite renewable power generation, and $f: \mathbb{X} \times \mathbb{U} \times \mathbb{S} \to \mathbb{X}$ represents the production environment dynamics. A scalar reward function, $R : \mathbb{X} \times \mathbb{U} \times \mathbb{X} \to R_{t+1}$, informs the control policy about the performance of the control action. The objective return, G is the discounted cumulative reward received over the horizon:

$$G = \sum_{t=0}^{T-1} \gamma^t R_{t+1} \tag{2}$$

$\gamma \in [0,1]$, which is a discount factor.
Solution methods for MDPs aim to identify a control policy, $\pi : \mathbb{X} \to \mathbb{U}$, that maximise the expected return from the initial system state, $\boldsymbol{x}_0$:

$$\pi^* = \underset{\pi}{argmax}\, \mathbb{E}[G|\pi, \boldsymbol{x}_0] \tag{3}$$

One can use dynamic programming to solve the MDP exactly if the transition probability function and reward function is known, and the cardinality of the state and control spaces is small. In industrial practice, this is almost never the case. This provides context for the use of RL, which is thought of as an approximate solution method for MDPs. When combined with function approximators such as neural networks, DRL instead solves:

$$\theta^* = \underset{\theta}{argmax}\ \mathbb{E}[G|\theta, \boldsymbol{x}_0] \qquad (4)$$

where $\pi^* \approx \pi(\boldsymbol{x}; \theta^*)$, and $\theta \in \mathbb{R}^{n_\theta}$ are the parameters of a parametric function.

Model-free RL underpins a set of policy learning algorithms that do not assume knowledge of the process or a model, but assume one is instead able to gain some evaluation of it, subject to the action of the policy. This is particularly beneficial in the context of industrial DR, because it enables one to parameterise an optimal control policy, without making approximations to a systems model, but instead simply by evaluating it. The optimal policy parameters can be identified offline through a process that can be viewed as simulation-based optimisation via the detailed model. Typically, the policy identification mechanism presents in the form of first order updates of θ, as evaluated through Monte Carlo simulation of the model. The policy function may then be deployed online to the real system to make fast scheduling decisions through inference.

3.2 State of the art of RL in Industrial demand response

The Energy Systems Catapult's reviews RL's broad applications in and beyond the energy sector. It compares RL with MPC and rule-based systems, highlighting RL's challenges such as limited sample efficiency, together with real-world engineering considerations, such as safety constraints. It emphasizes RL's fit for industrial and commercial control systems due to the innovations previously discussed [19]. The authors in [20] reviews RL in the context of smart grids and energy internet, highlighting its use in security, automatic generation control (AGC) and smart power generation control, voltage and reactive power control, and optimal power flow control. Furthermore, [21] offers a thorough review of RL's energy system applications such as building energy management, dispatch, energy systems in hybrid vehicles, energy markets, grid, and energy devices. Notably, a significant portion (45%) of RL research in this domain focuses on building energy management or dispatch. The review assesses problem diversity, RL techniques employed, and achieved success. Most studies show a performance boost of 10-20% over their respective benchmarks. The review highlights challenges including the absence of standardised RL benchmarking, limited reproducibility, and underutilisation of deep learning techniques. In [22], the application of RL in demand side management is highlighted, particularly in controlling domestic hot water heaters, heating ventilation air conditioning devices, residential appliances, and EV charging. Despite these applications, challenges such as addressing physical constraints, and the difficulties of centralizing information for single-agent RL in more complex problems were highlighted for further study. The authors in [23] review deep DRL approaches for smart manufacturing in industry 4.0 and 5.0 frameworks. The review emphasizes DRL's applicability in key manufacturing activities such as path planning, process control, scheduling, maintenance, and energy management. Despite DRL demonstrating as a competitive alternative to other conventional techniques in these activities, the review also highlights some significant challenges. One challenge pertains to the selection of algorithms, while another involves its limited application in real-world scenarios. This limitation is primarily attributed to industrial policies that hinder the full exploitation of the potential of DRL algorithms.

In the context of industrial DR, [24] explores the optimal control policy selection for energy storage systems (ESS) in price and incentive-based DR programs. The study assesses various RL algorithms like deep Q-network (DQN), deep recurrent Q network (DRQN), double DQN, dueling DQN and proximal policy optimisation (PPO) based on reward performance, and scalability. Ultimately, DRQN and PPO exhibited superior performance. Deep learning techniques have been employed to develop model free DRL algorithm for smart facilities. The aim was to minimise electricity costs. Results show that DRL-based DR algorithm surpassed an MPC strategy for managing energy consumption by achieving 8.29 % lower electricity cost [25]. Similar DRL methods were applied to optimise energy management in steel powder manufacturing, yielding a notable 24.12% reduction in total energy costs compared to the absence of DR implementation case [26].

Recognising limitations in traditional DQN and Deep Policy Gradient (DPG) approaches for complex multi-agent industrial contexts, Multi-agent Deep Deterministic Policy Gradient (MADDPG) was applied in managing energy consumption during a lithium-ion battery assembly process. Notably, MADDPG achieved a 9.8% reduction in total electricity costs compared to a case without DR implementation [27]. In [28], a new Multi-agent RL (MARL) based approach for energy-oriented production control integrated self-supply, batteries, and electricity trading to reduce costs. Results show 84% average performance gain over a rule-based benchmark. MARL offers computational efficiency and reactivity advantages. However, Instability in policy learning dynamics is also one of the major challenges of MARL [29].

RL has been used to schedule DR in the cement manufacturing industry [30]. Commercial solutions are also beginning to emerge, with notable reductions in energy consumption and carbon emissions reported for cement plants [31]. The authors in [32] employ a RL approach for energy-flexible production planning on job shop problem, integrating make span and electricity cost minimisation. Two distinct reward functions were developed to manage job operation and machine idle time, offering flexibility based on a weighted balance between make span and electricity cost objectives. Results demonstrate an intuitive trade-off between the two objectives, with the edge case of allocating full weight to the electricity cost objective reducing electricity costs by an average of 13%, compared to giving full weight to the make span objective. A SARSA-based composite differential evolution (DE) method has been proposed for integrated DR programs in a grid-connected industrial multi-energy microgrid framework for determining optimal schedule of power and heat energy suppliers. The objective was to reduce total energy costs and battery EES degradation costs, along with maximising renewable energy resources usage. The authors highlight combining this approach with transfer learning or deep learning to speed up the policy identification [33]. This has also been applied elsewhere [34] [35]. DRL enables load control in incentive-based DR, making it a valuable approach for industrial facilities [36]. Additionally, [37] highlighted some major deficiencies for industrial plants to participate in energy markets and described an end-to-end approach based on a platform ecosystem. In model free RL, policy and systems model are decoupled, allowing for flexibility in handling multiple goals and considering consumer preferences simultaneously which can be achieved by potential of multi-objective RL [38] and RL with human feedback [39].

Prior research demonstrates the potential of RL as a solution for addressing the complexity of industrial plants in DR. To ensure practical applicability, future research should focus on improving constraint handling techniques and conducting comprehensive benchmarking against gray box models. This will validate RL's effectiveness and suitability in industrial settings for DR [40].

3.3 Reinforcement learning framework in industrial demand response

In [41], the authors provide a framework for RL in the chemical and process industries.

Fig. 3 provides the scheme of RL for production scheduling and control for industrial process in price-based DR program. One can also view the environment model as a digital twin of the production process serving as a training environment for agents that learn in sophisticated simulation through RL [42].

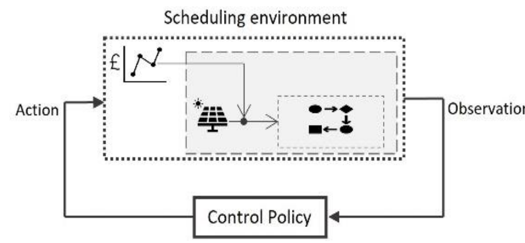

Fig. 3 RL scheme for control and decision making in an industrial plant

Scheduling for real-time DR poses computational challenges, particularly with longer time horizons and larger control state spaces. DRL addresses this by leveraging a simulation model for offline agent training. Though the training stage is computationally expensive, this will bring core benefit since an approximately optimal decision can be made quickly online through inference when the policy is deployed, provided the system model is a good approximation.

Fig. 4 illustrates the intuition of RL in scheduling for price-based industrial DR. The day ahead electricity price and generation forecast, and production data continuously collected by distributed control systems can be fed to the digital twin or simulation model of industrial plant where the detailed systems model is trained to get the optimal policy. The policy is then transferred online on real plant where it provides controls through prediction which reduces the computational burden offered by MILP. Another advantage of this approach is there is minimal requirement of approximating detailed

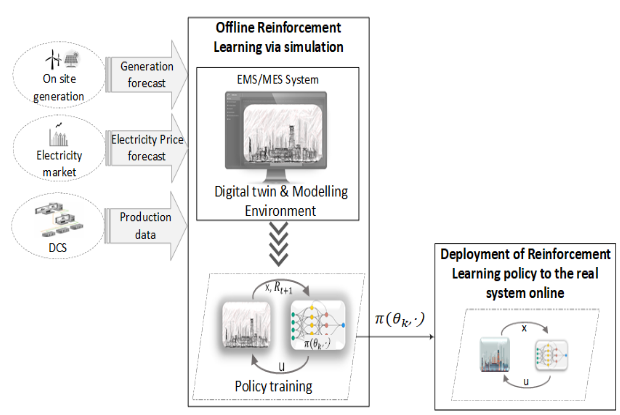

Fig. 4 Intuition of RL in price-based industrial DR

systems model due to absence of time constraints in offline training.

3.4 Challenges for Reinforcement learning

While the potential of RL is clear, its widespread adoption in industrial DR faces significant obstacles. Among algorithmic challenges, data inefficiency is prominent. RL often demands extensive data for effective learning, rendering policy identification computationally expensive. Although function approximation methods address scalability concerns, their black box nature limits their explanatory power, thus promoting safety concerns. Moreover, RL lacks robust constraint handling techniques, although efforts have been made in this direction [43]. Soft state constraints can be incorporated through the reward function design to discourage undesired policies.

The operational nuances in the context of scheduling arising from MDP framework are substantial, as industrial facilities require proactive decision-making to accommodate factors such as personnel, resources, and scheduling considerations, rather than relying on entirely reactive approaches [6].

Additional challenge stems from the decision process models assumed by RL. Creating an MDP that effectively captures decision-making information and is of fixed and finite dimension poses a significant challenge. Attention mechanisms present a promising solution for addressing this. By allowing the model to selectively focus on relevant information within the state, attention mechanisms can enhance the representation of decision-critical information without exponentially increasing dimensionality [44]. However, these approaches are not free from hyperparameters, which will require careful selection.

Industrial scheduling problems in practice are often characterized by large decision spaces, making it difficult to identify an effective policy given RL's reliance on generalised policy iteration and Monte Carlo sampling. An effective alternative is to instead construct the policy as a hyper-heuristic, which selects heuristic decision rules to identify assignment decisions, rather than identifying assignment decisions via prediction directly. This enables one to reduce the dimensionality of the control space and to handle variations in the number of production tasks over the course of the horizon. However, this is likely to introduce some inherent suboptimality into the formulation.

To handle the infinite horizon nature of the scheduling task, one is required to recursively approximate the decision problem with several fixed horizon decision problems, solved sequentially. The impact this has on the operation over the long-term is not clear, and policy retraining strategies have yet to be properly developed. As a result, this may lead to suboptimal policies and a misrepresentation of the long-term consequences of decisions. However, this is also a challenge encountered in receding horizon mathematical programming approaches.

A DRL model has the potential to be trained using forecasts for both day-ahead and intraday markets. However, it exhibits reduced robustness in the face of distributional changes, such as substantial shifts in pricing mechanisms [45]. This is significant, because if accurate forecasts cannot be made about price uncertainty, then an RL policy will likely be making extrapolative control predictions promoting safety concerns. As a result, it becomes essential to assess the decline in system performance by subjecting it to scenarios and conditions that were not encountered during the training phase. This evaluation is crucial for understanding how well the model adapts to novel states and situations, ensuring its reliability and effectiveness in real-world applications [6]. In practice, mechanisms to detect significant mismatch between the model utilised in offline simulation and the real system will be required to ensure safe operation.

3.5 Conclusions

This paper discusses the potential of reinforcement learning (RL) in scheduling of industrial processes for industrial demand response, specifically in handling uncertainties and reducing computational burden in intrinsically difficult-to-model and solve problems. While RL is not intended to replace traditional model-based methods entirely, it presents a viable alternative for scheduling complex process, which include a high degree of uncertainty and require fast decisions online. This paper discusses some of the prominent challenges arising from modelling approaches of RL in scheduling problems. Future research is required to realise the potential of RL, given the challenges identified.

ACKNOWLEDGEMENT

This work is supported by EPSRC through the projects EP/T021969/1(MC2) and SUSTAIN Feasibility project EP/S018107/1(SFSC2-203).

REFERENCE

[1] "Industry – Analysis," IEA. Accessed: Feb. 14, 2023. [Online]. Available: https://www.iea.org/reports/industry

[2] R. Heffron, M.-F. Körner, J. Wagner, M. Weibelzahl, and G. Fridgen, "Industrial demand-side flexibility:

A key element of a just energy transition and industrial development," *Appl. Energy*, vol. 269, p. 115026, Jul. 2020, doi: 10.1016/j.apenergy.2020.115026.

[3] I. Harjunkoski, "Deploying scheduling solutions in an industrial environment," *Comput. Chem. Eng.*, vol. 91, pp. 127–135, Aug. 2016, doi: 10.1016/j.compchemeng.2016.03.029.

[4] P. M. Castro, I. Harjunkoski, and I. E. Grossmann, "Optimal scheduling of continuous plants with energy constraints," *Comput. Chem. Eng.*, vol. 35, no. 2, pp. 372–387, Feb. 2011, doi: 10.1016/j.compchemeng.2010.05.008.

[5] I. Harjunkoski *et al.*, "Scope for industrial applications of production scheduling models and solution methods," *Comput. Chem. Eng.*, vol. 62, pp. 161–193, Mar. 2014, doi: 10.1016/j.compchemeng.2013.12.001.

[6] C. D. Hubbs, C. Li, N. V. Sahinidis, I. E. Grossmann, and J. M. Wassick, "A deep reinforcement learning approach for chemical production scheduling," *Comput. Chem. Eng.*, vol. 141, p. 106982, Oct. 2020, doi: 10.1016/j.compchemeng.2020.106982.

[7] "Industry - Energy System," IEA. Accessed: Aug. 10, 2023. [Online]. Available: https://www.iea.org/energy-system/industry

[8] H. Golmohamadi, "Demand-side management in industrial sector: A review of heavy industries," *Renew. Sustain. Energy Rev.*, vol. 156, p. 111963, Mar. 2022, doi: 10.1016/j.rser.2021.111963.

[9] Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, and J. M. Pinto, "Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets," *AIChE J.*, vol. 61, no. 5, pp. 1547–1558, 2015, doi: 10.1002/aic.14730.

[10] M. Paulus and F. Borggrefe, "The potential of demand-side management in energy-intensive industries for electricity markets in Germany," *Appl. Energy*, vol. 88, no. 2, pp. 432–441, Feb. 2011, doi: 10.1016/j.apenergy.2010.03.017.

[11] Q. Zhang and I. E. Grossmann, "Planning and Scheduling for Industrial Demand Side Management: Advances and Challenges," in *Alternative Energy Sources and Technologies: Process Design and Operation*, M. Martín, Ed., Cham: Springer International Publishing, 2016, pp. 383–414. doi: 10.1007/978-3-319-28752-2_14.

[12] I. Harjunkoski *et al.*, "Scope for industrial applications of production scheduling models and solution methods," *Comput. Chem. Eng.*, vol. 62, pp. 161–193, Mar. 2014, doi: 10.1016/j.compchemeng.2013.12.001.

[13] "Prospects for Reinforcement Learning," Energy Systems Catapult. Accessed: May 10, 2023. [Online]. Available: https://es.catapult.org.uk/report/prospects-for-reinforcement-learning/

[14] Q. Zhang and I. E. Grossmann, "Planning and Scheduling for Industrial Demand Side Management: Advances and Challenges," in *Alternative Energy Sources and Technologies*, M. Martín, Ed., Cham: Springer International Publishing, 2016, pp. 383–414. doi: 10.1007/978-3-319-28752-2_14.

[15] S. M. S. Siddiquee, B. Howard, K. Bruton, A. Brem, and D. T. J. O'Sullivan, "Progress in Demand Response and It's Industrial Applications," *Front. Energy Res.*, vol. 9, p. 673176, Jun. 2021, doi: 10.3389/fenrg.2021.673176.

[16] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Appl. Energy*, vol. 235, pp. 1072–1089, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.002.

[17] R. Lu, S. H. Hong, and X. Zhang, "A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Appl. Energy*, vol. 220, pp. 220–230, Jun. 2018, doi: 10.1016/j.apenergy.2018.03.072.

[18] M. R. Mowbray, C. Wu, A. W. Rogers, E. A. D. Rio-Chanona, and D. Zhang, "A reinforcement learning-based hybrid modeling framework for bioprocess kinetics identification," *Biotechnol. Bioeng.*, vol. 120, no. 1, pp. 154–168, 2023, doi: 10.1002/bit.28262.

[19] "Prospects for Reinforcement Learning," Energy Systems Catapult. Accessed: Jun. 15, 2023. [Online]. Available: https://es.catapult.org.uk/report/prospects-for-reinforcement-learning/

[20] L. Cheng and T. Yu, "A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems," *Int. J. Energy Res.*, vol. 43, no. 6, pp. 1928–1973, 2019, doi: 10.1002/er.4333.

[21] A. T. D. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renew. Sustain. Energy Rev.*, vol. 137, p. 110618, Mar. 2021, doi: 10.1016/j.rser.2020.110618.

[22] D. Cao *et al.*, "Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029–1042, Nov. 2020, doi: 10.35833/MPCE.2020.000552.

[23] A. Del Real Torres, D. S. Andreiana, Á. Ojeda Roldán, A. Hernández Bustos, and L. E. Acevedo Galicia, “A Review of Deep Reinforcement Learning Approaches for Smart Manufacturing in Industry 4.0 and 5.0 Framework,” *Appl. Sci.*, vol. 12, no. 23, p. 12377, Dec. 2022, doi: 10.3390/app122312377.
[24] S. Oh, J. Kong, Y. Yang, J. Jung, and C.-H. Lee, “A multi-use framework of energy storage systems using reinforcement learning for both price-based and incentive-based demand response programs,” *Int. J. Electr. Power Energy Syst.*, vol. 144, p. 108519, Jan. 2023, doi: 10.1016/j.ijepes.2022.108519.
[25] R. Lu, R. Bai, Z. Luo, J. Jiang, M. Sun, and H.-T. Zhang, “Deep Reinforcement Learning-Based Demand Response for Smart Facilities Energy Management,” *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8554–8565, Aug. 2022, doi: 10.1109/TIE.2021.3104596.
[26] X. Huang, S. H. Hong, M. Yu, Y. Ding, and J. Jiang, “Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach,” *IEEE Access*, vol. 7, pp. 82194–82205, 2019, doi: 10.1109/ACCESS.2019.2924030.
[27] R. Lu, Y.-C. Li, Y. Li, J. Jiang, and Y. Ding, “Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management,” *Appl. Energy*, vol. 276, p. 115473, Oct. 2020, doi: 10.1016/j.apenergy.2020.115473.
[28] M. Roesch, C. Linder, R. Zimmermann, A. Rudolf, A. Hohmann, and G. Reinhart, “Smart Grid for Industry Using Multi-Agent Reinforcement Learning,” *Appl. Sci.*, vol. 10, no. 19, Art. no. 19, Jan. 2020, doi: 10.3390/app10196900.
[29] “Multi-Agent Reinforcement Learning: Foundations and Modern Approaches.” Accessed: Aug. 10, 2023. [Online]. Available: https://www.marl-book.com/
[30] X.-Y. Ye, Z.-W. Liu, M. Chi, M.-F. Ge, and Z. Xi, “Demand Response optimization of Cement Manufacturing Industry Based on Reinforcement Learning Algorithm,” in *2022 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, Mar. 2023, pp. 402–406. doi: 10.1109/CBS55922.2023.10115387.
[31] pnpd-admin, “How AI is helping cement plant operators reduce energy consumption and carbon emissions,” Carbon Re. Accessed: Aug. 11, 2023. [Online]. Available: https://carbonre.com/how-ai-is-helping-cement-plant-operators-reduce-energy-consumption-and-carbon-emissions/
[32] M. Felder *et al.*, “Energy-Flexible Job-Shop Scheduling Using Deep Reinforcement Learning,” 2023, doi: 10.15488/13454.
[33] N. Mughees, M. H. Jaffery, A. Mughees, E. A. Ansari, and A. Mughees, “Reinforcement learning-based composite differential evolution for integrated demand response scheme in industrial microgrids,” *Appl. ENERGY*, vol. 342, Jul. 2023, doi: 10.1016/j.apenergy.2023.121150.
[34] P. Petsagkourakis, I. O. Sandoval, E. Bradford, D. Zhang, and E. A. del Rio-Chanona, “Reinforcement learning for batch bioprocess optimization,” *Comput. Chem. Eng.*, vol. 133, p. 106649, Feb. 2020, doi: 10.1016/j.compchemeng.2019.106649.
[35] M. Mowbray, R. Smith, E. A. Del Rio-Chanona, and D. Zhang, “Using process data to generate an optimal control policy via apprenticeship and reinforcement learning,” *AIChE J.*, vol. 67, no. 9, p. e17306, 2021, doi: 10.1002/aic.17306.
[36] Y. Zhao, J. H. Yoo, and C. G. Lim, “Real-Time Demand Response Management for Controlling Load Using Deep Reinforcement Learning,” *Cmc-Comput. Mater. Contin.*, vol. 73, no. 3, pp. 5671–5686, 2022, doi: 10.32604/cmc.2022.027443.
[37] M. Roesch *et al.*, “Harnessing the Full Potential of Industrial Demand-Side Flexibility: An End-to-End Approach Connecting Machines with Markets through Service-Oriented IT Platforms,” *Appl. Sci.*, vol. 9, no. 18, Art. no. 18, Jan. 2019, doi: 10.3390/app9183796.
[38] C. F. Hayes *et al.*, “A practical guide to multi-objective reinforcement learning and planning,” *Auton. Agents Multi-Agent Syst.*, vol. 36, no. 1, p. 26, Apr. 2022, doi: 10.1007/s10458-022-09552-y.
[39] “Prospects for Reinforcement Learning-Energy systems catapult,” Energy Systems Catapult. Accessed: Apr. 17, 2023. [Online]. Available: https://es.catapult.org.uk/report/prospects-for-reinforcement-learning/
[40] R. Nian, J. Liu, and B. Huang, “A review On reinforcement learning: Introduction and applications in industrial process control,” *Comput. Chem. Eng.*, vol. 139, p. 106886, Aug. 2020, doi: 10.1016/j.compchemeng.2020.106886.
[41] M. Mowbray, M. Vallerio, C. Perez-Galvan, D. Zhang, A. D. R. Chanona, and F. J. Navarro-Brull, “Industrial data science – a review of machine learning applications for chemical and process industries,” *React. Chem. Eng.*, vol. 7, no. 7, pp. 1471–1509, 2022, doi: 10.1039/D1RE00541C.
[42] “Carbon Re Talks Creativity and Innovation in the Steel Industry | ResponsibleSteel.” Accessed: May 28, 2023. [Online]. Available:

https://www.responsiblesteel.org/news/carbon-re-talks-creativity-and-innovation-in-the-steel-industry/
[43] H. Yoo, V. M. Zavala, and J. H. Lee, “A dynamic penalty approach to state constraint handling in deep reinforcement learning,” *J. Process Control*, vol. 115, pp. 157–166, Jul. 2022, doi: 10.1016/j.jprocont.2022.05.004.
[44] J. Xie, A. Ajagekar, and F. You, “Multi-Agent attention-based deep reinforcement learning for demand response in grid-responsive buildings,” *Appl. Energy*, vol. 342, p. 121162, Jul. 2023, doi: 10.1016/j.apenergy.2023.121162.
[45] “Global Energy Crisis – Topics,” IEA. Accessed: Aug. 10, 2023. [Online]. Available: https://www.iea.org/topics/global-energy-crisis