Energy Conversion and Management 258 (2022) 115401

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

AI agents envisioning the future: Forecast-based operation of renewable energy storage systems using hydrogen with Deep Reinforcement Learning

Alexander Dreher [a,*], Thomas Bexten [b], Tobias Sieker [b], Malte Lehna [a], Jonathan Schütt [a], Christoph Scholz [a], Manfred Wirsum [b]

[a] Fraunhofer Institute for Energy Economics and Energy System Technology, Department Energy Informatics and Information Systems, Joseph-Beuys-Straße 8, 34117 Kassel, Germany

[b] RWTH Aachen University, Institute of Power Plant Technology, Steam and Gas Turbines, Mathieustraße 9, 52074 Aachen, Germany

ARTICLE INFO

Keywords:
Hydrogen
Renewable energy storage
Energy management
Deep reinforcement learning
Dynamic programming

ABSTRACT

Hydrogen-based energy storage has the potential to compensate for the volatility of renewable power generation in energy systems with a high renewable penetration. The operation of these storage facilities can be optimized using automated energy management systems. This work presents a Reinforcement Learning-based energy management approach in the context of CO_2-neutral hydrogen production and storage for an industrial combined heat and power application. The economic performance of the presented approach is compared to a rule-based energy management strategy as a lower benchmark and a Dynamic Programming-based unit commitment as an upper benchmark. The comparative analysis highlights both the potential benefits and drawbacks of the implemented Reinforcement Learning approach. The simulation results indicate a promising potential of Reinforcement Learning-based algorithms for hydrogen production planning, outperforming the lower benchmark. Furthermore, a novel approach in the scientific literature demonstrates that including energy and price forecasts in the Reinforcement Learning observation space significantly improves optimization results and allows the algorithm to take variable prices into account. An unresolved challenge, however, is balancing multiple conflicting objectives in a setting with few degrees of freedom. As a result, no parameterization of the reward function could be found that fully satisfied all predefined targets, highlighting one of the major challenges for Reinforcement Learning -based energy management algorithms to overcome.

1. Introduction

In most industrialized countries, the energy sector is responsible for a major share of total green house gas (GHG) emissions [1]. Therefore, the transition of energy sectors towards GHG neutrality is key to successful mitigating global warming [2]. The comprehensive deployment of renewable power generation (RPG) capacity is considered to be the most promising way to achieve GHG-neutral energy sectors [3]. However, the inherent volatility of wind- and solar-based power generation poses various challenges to electricity grid stability [4]. Energy storage systems (ESS) based on hydrogen have the potential to compensate for the volatility of RPG on time scales ranging from hours to months [5]. Morevoer, hydrogen can be utilized to substitute conventional energy carriers in other sectors where the direct use of renewable electricity is not feasible [6]. As a result, European governance has identified hydrogen as a key element for the transition towards a sustainable and renewable energy supply. This is particularly reflected in the Green Deal, the European Commission's policy initiatives for achieving climate neutrality in 2050 [7,8]. National efforts to study the potential and applications of hydrogen and accelerate technology development are already underway, driven by European norms and directives [9]. In general, automated energy management systems (EMSs) are able to facilitate the operation of energy conversion and storage systems [10–15]. Due to the interaction with volatile RPG, the application of automated EMSs is particularly relevant for hydrogen-based ESS. For making operational decisions, rule-based (RB) approaches as well as discrete mathematical optimization (DMO) models are commonly used [16]. By solving the Unit Commitment Problem (UCP) and the economic dispatch problem, DMO methods, such as Linear, Nonlinear, and Dynamic Programming (DP), can be beneficial for the economic optimization of complex systems [16,17]. Of late, energy management approaches based on Machine Learning (ML) algorithms have resurfaced as a research subject [17]. Reinforcement Learning (RL) as one of the

* Corresponding author.
E-mail address: alexander.dreher@iee.fraunhofer.de (A. Dreher).

https://doi.org/10.1016/j.enconman.2022.115401
Received 5 October 2021; Received in revised form 15 February 2022; Accepted 16 February 2022
Available online 12 March 2022

Nomenclature

General optimization framework

Δ_t	time interval [s]
G	final state cost
H	current state cost
J	total cost
t	time [s] / time step
u	control signal
x	state variable

Reinforcement Learning framework

$\mathscr{S}$	state space
$\mathscr{P}$	transition probabilities
$\mathscr{R}$	reward function
$\mathscr{A}$	action space
$\pi_\theta(a\|s)$	agent's policy
a	agent's action
s	state of the environment
θ	policy parameter vector
τ	final time step of episode
b^{St}	binary variable for storage target fulfilment
b^{GT}	binary variable for hydrogen usage target of GT
$r_t^{(1)}$	instantaneous reward incentivizing price orientation
$r^{(2)}$	rewards for partial target fulfilment
$R^{(1)}, R^{(2)}$	rewards for target fulfilment
Φ_t	electricity spot prices

Energy system and economics

$\dot{Q}_{th}^{(k)}$	heat generation or consumption of unit k [W]
$\dot{m}_{H_2}^{(k)}$	hydrogen mass flow of unit k [kg/s]
$\dot{m}_{EG}^{GT}$	exhaust gas mass flow of GT [kg/s]
$\dot{m}_{NG}^{GT}$	natural gas mass flow of GT [kg/s]
$\dot{m}_{Steam}^{HRSG}$	steam mass flow of HRSG [kg/s]
$m_{H_2}^{ST}$	state of charge of storage vessels [kg]
$m_{H_2}^{GT}$	Hydrogen consumption of GT [kg]
$P_{el}^{(k)}$	electric power generation or consumption of unit k
p^{St}	pressure level of storage vessels [bar]
$R_{feed-in}$	grid feed-in revenues [EUR]
R_{H_2}	specific gas constant [J/kg K]
T^{St}	storage temperature [K]
V_{geo}^{St}	geometric volume of storage vessels [m^3]
Z_{H_2}	compressibility factor
$\eta_{el,LHV}^{(k)}$	electrical efficiency of unit k considering the lower heating value [%]
c_{EPEX}	EPEX spot price [EUR/MWh]
p_{Steam}^{HRSG}	steam pressure of HRSG [bar]
T_{EG}^{GT}	exhaust gas temperature of GT [K]
T_{Steam}^{HRSG}	steam temperature of HRSG [K]

Identifiers and indices

$\overline{X}$	nominal capacity
$X_{\_}$	minimum capacity
k	Com: compressor; GT: gas turbine; Ind: industrial facility; EZ: electrolyzer; St: storage; WF: wind farm

Abbreviations

ANN	Artificial Neural Network
CHP	combined heat and power
CO_2	carbon dioxide
DMO	discrete mathematical optimization
DP	Dynamic Programming
DRL	Deep Reinforcement Learning
EMS	energy management system
ESS	energy storage system
GHG	green house gas
HRSG	heat recovery steam generator
MDP	Markov Decision Process
MES	multi energy system
ML	Machine Learning
SoC	State of Charge
SGD	stochastic gradient descent
RL	Reinforcement Learning
RPG	renewable power generation
RB	rule-based
UCP	Unit Commitment Problem
PPO	Proximal Policy Optimization

most recently explored subfields shows a high potential for solving different tasks in the operational energy management context, with multiple benefits. One of these benefits is the ability of self-learning, which can prove useful in changing and uncertain environments. This is especially relevant for RPG portfolios, which are usually affected by uncertainties regarding power generation. DMO approaches often have: i) high requirements concerning computing capacity, ii) high computational costs and iii) require perfect input data information for the optimization horizon. In contrast, trained ML models can be deployed where computing and storage preconditions are limited, enabling sequential online optimization. Further, updated data can be entered in real-time operation, thus reducing the impact of forecasting errors. RL approaches allow higher economic efficiency than rule-based models, since past experience is included which improves operating decisions. Thus, RL-based EMSs are able to comprehend complex information such as time-variable prices. In comparison, RB approaches often require a human expert to define the decision-making sequences in complex systems. Within this paper, a Deep Reinforcement Learning (DRL) approach[1] for the energy management of a hydrogen-based energy storage system is developed and compared to the performance of both an RB and a DP approach. The investigated approach differs from previous studies not only in (i) the application of RL-based EMS to a hydrogen production and storage context but also in (ii) the regard that variable spot prices for electricity are taken into account, significantly increasing the complexity of the RL-agent's decision making. Furthermore (iii), forecasts of prices, RPG, and energy demand are included in the agent's observation space, which is shown to significantly improve the results. At the time of submission, there is no known paper in the energy management context that considers full-day forecasts for single or multiple

[1] For simplicity, we will refer to the approach mostly as RL in the following, whereas DRL would be more specific, addressing the depth of the agent's neural network structure.

time series in RL[2]. The remaining paper is structured as follows: Section 2 provides background information regarding hydrogen-based ESS and their operational management. Section 3 introduces the object of investigation and presents the methodology of the three considered operational management approaches. Section 4 describes the design of the conducted experiments. The obtained results are presented in Section 5 and discussed and concluded in Section 6.

2. Background

2.1. Hydrogen-based energy storage for greenhouse gas-neutral energy sectors

In hydrogen-based ESS, electrolysis is employed to convert electricity and water into gaseous hydrogen, which is subsequently stored[3]. Using gas turbines, internal combustion engines or fuel cells, the stored hydrogen can be reconverted into dispatchable carbon dioxide (CO_2)-neutral power when needed [5]. Gas turbines are especially suited for this task as they feature short startup times, high load gradients and a wide load range [18]. In addition, gas turbines provide not only dispatchable power but also high-temperature heat, which can be utilized in various combined heat and power (CHP) applications, such as process steam generation. Technical challenges associated with the combustion of hydrogen, e.g., a high flame speed, require the design of gas turbines adapted to the use of hydrogen [5]. Research and development activities to overcome these challenges are currently being carried out by all major gas turbine manufacturers [19].[4]

The positioning of production, storage and reconversion facilities is an important differentiator of hydrogen-based ESS. When considering electrolyzers and hydrogen-fired gas turbines, two main alternatives are conceivable [20]. First, electrolyzers and gas turbines can be installed at independent sites. This enables the selection of electrolyzer sites with a favorable supply of renewable power and allows for the utilization of large-scale hydrogen storage facilities, e.g., salt caverns, resulting in low specific storage costs [5]. On the downside, this approach presupposes the availability of dedicated infrastructures that enable the transport of large quantities of hydrogen. While various concepts for hydrogen transport infrastructures have been proposed in the past, e.g., the European Hydrogen Backbone [21], none of these concepts has been realized to date. The second option is the direct installation of electrolyzers at gas turbine sites. The resulting spatial proximity eliminates the need to transport large quantities of hydrogen. This approach presupposes the installation of on-site hydrogen storage facilities, which generally limits the choice of available storage technologies. Considering the current lack of large-scale hydrogen transport infrastructures, it is likely that ESS employing hydrogen-fired gas turbines will initially incorporate on-site electrolysis and on-site hydrogen storage. Consequently, the present study focuses on this type of system configuration. This choice is supported by various demonstration projects investigating similar set-ups, e.g., the “Hyflexpower” project in France [22].

2.2. Operational management of hydrogen-based energy storage systems

ESS comprising of hydrogen-fired gas turbines, on-site electrolysis and on-site hydrogen storage require sophisticated operational management. This primarily results from the volatility of RPG, the limited on-site hydrogen storage capacity and the time-dependent demand to be met by the gas turbine [23]. In general, the interaction between gas turbines and volatile RPG has already been studied extensively. Xia et al. investigated the ability of a hybrid wind and gas turbine power system to guarantee constant power production during a 24-hour period [24]. A similar system configuration was analyzed by Branchini et al., focusing on the unit commitment of a hybrid system incorporating two types of gas turbines [25]. However, both studies only considered a natural gas-based operation of gas turbines. In contrast, Ebaid et al. conducted a model-based analysis of the operation of a $100MW_{el}$ photovoltaic and hydrogen-fired gas turbine power plant [26]. Similarly, Colbertaldo et al. analyzed the potential of employing hydrogen-fired combined cycle power plants to balance electricity grids with a high share of RPG [27]. In other studies, prior investigations of CHP systems featuring hydrogen-fired industrial gas turbines, on-site electrolyzers and on-site hydrogen storage facilities were conducted [20,23,28].[5] A common denominator of the studies mentioned above is the consideration of RB approaches or DMO methods for the operational management of the investigated system configurations. As outlined in the following section, the application of RL-based approaches has been limited so far.

2.3. Reinforcement Learning based operational energy management

In the late 1980s, the modern field of RL originated based on research from both psychology and optimal control problems [29]. Within the RL approach, an agent interacts with an environment by receiving environmental observations, taking actions that affect the state of the environment, and receiving a reward-based on the effectiveness of the performed action. This approach was further enhanced through the combination of RL and Deep Neural Networks creating the field of DRL, which was first introduced by Mnih et al. [30,31].[6]

In terms of energy management, RL has the capacity to solve a variety of optimization problems at a fraction of the computational cost of conventional optimization approaches. Zhang et al. surveyed implementations of RL and DRL in the energy management context [34]. They find that the investigated problems range from the management of microgrids and ESS to operational control to demand response and fault detection. Ye et al. propose a model-free DRL approach to solve the optimization problem for a residential multi energy system (MES), achieving near-optimal results [35]. In their analysis, the authors are able to manage multiple electrical assets to meet the energy demand of households, including PV arrays, ESS, electric heat pumps and gas boilers. They reformulate the optimization problem into a Markov Decision Process (MDP) and optimize it through a deep deterministic policy gradient with an experience replay strategy. In similar research by Mbuwir et al., microgrid in different scenarios was optimized based on DRL [36]. They concentrate on battery storage management and use a fitted Q-iteration to solve the problem. Their DRL proved beneficial for the management of the microgrid. Zhang et al. further find that researchers often combine DRL algorithms with other Deep Learning methods, e.g., to predict required time series [34]. As example Lu et al. optimize a home energy management system and include an Artificial Neural Network (ANN) to predict time series [37]. They are able to predict future electricity prices and feed them into the DRL algorithm in order to provide an hour-ahead demand response.

[2] In this study, forecasts with a time horizon of 1 to 24 h are considered, whereas known approaches only consider very limited time horizons of one or two time steps [37,60].

[3] Buttler et al. and Preuster et al. provide comprehensive overviews of the state of the art of water electrolysis technologies and hydrogen storage technologies [61,62].

[4] Preliminary results, e.g., by Kawasaki Heavy Industries [63], already indicate that fully hydrogen-capable gas turbines will become commercially available in the near future.

[5] These prior studies mainly address the technical feasibility and economic viability of CHP systems. The authors additionally investigated the impact of optimized operational strategies in Ref. [28]. However, they did not consider the use of DRL which is the main novelty of the present study.

[6] Since then, DRL has demonstrated human-like or even superhuman performance in board games such as chess (with AlphaZero [32]) and Go (AlphaGo [33]), indicating the high potential of self-learning algorithms.

Aside from residential MES, there have been first propositions to apply DRL for industrial purposes. According to Huang et al., the required frameworks differ because there are interactions between the individual machines and a solely price-based optimization may not suffice [38]. Accordingly, the authors developed a DRL algorithm for an industrial demand response optimization, based on the Actor Critic approach. They create a simulation of a real-world manufacturing process in which the agent is able to lower energy costs, meet production goals, and balance the factory's energy demand [38]. Another approach is proposed by Lin et al., in which they utilize the DRL approach to combine multiple remote distributed energy units to a virtual power plant, to manage industrial demand [39]. In their virtual power plant, different energy sources such as solar and wind generation and a gas powered turbine are managed. Huang et al. follw a similar approach and also base their RL algorithm on the Actor Critic method with the A3C algorithm [40]. Lastly, as presented by Zhou et al., DRL has also been used to manage the heat and power economical dispatch [41]. In summary, DRL has proven very adaptable in its applications.

As a consequence, DRL may provide additional benefits to hydrogen-related research, especially when considering volatile RPG and weather data. In recent years only a few studies in this field have been published, e.g, by Francois-Lavet et al. [42]. For the study, the authors built a microgrid with photovoltaic production and various energy storage options, then tested it on an off-grid microgrid in Belgium [42]. Next to a short-term battery storage unit, the microgrid includes hydrogen electrolysis as a long-term storage unit. However, in order to include the costly hydrogen storage option in their analysis, they assume a fixed external electricity price of 2 EUR/kWh, strongly simplifying the degree of diffculty for the optimization problem. Even though they have to combine stochastic weather and consumer data with the deterministic microgrid, Francois-Lavet et al. are able to achieve acceptable results with the RL-agent [42]. Following up on these promising results, Tomin et al. published on a similar microgrid [43]. However, they expanded the research by considering a diesel generator as additional microgrid option. With the diesel generator included, it is possible to compare hydrogen storage with a less expensive alternative and to optimize using a possible hydrogen-diesel hybrid. Tomin et al. included an off-grid microgrid in Russia in addition to the Belgian use-case, using similar simplifying assumptions as Francois-Lavet et al. [42]. Nyong-Bassey et al. published a study on hydrogen-related optimization that considers a hybrid RL approach to optimize their hybrid ESS framework in the face of uncertainty [44]. The authors propose three different Power Pinch Analysis approaches, one of which is a combination with a Q-Learning RL approach. Nyong-Bassey et al. are able to demonstrate that the RL hybrid was the best grid management strategy using data from an isolated microgrid in Greece. However, the authors deduce that RL increases the required complexity of the grid [44].

Overall, DRL has been present in various energy topics and is showing reliable results. Yet, with regard to the optimized operational management of hydrogen-based ESS, there are still many questions that have to be analyzed in order to correctly assert the advantages of DRL. This paper intends to answer some of these questions and place them into context through the direct comparison with both an RB and a DP approach. Further, through the inclusion of real external data, the applicability of DRL is analyzed more accurately.

3. Methodology

3.1. Object of investigation

To investigate the potential benefits of applying DRL to the operational management of ESS that incorporate hydrogen-fired gas turbines with on-site electrolysis and on-site hydrogen storage, the present study considers the system configuration visualized in Fig. 1. This system configuration has been the subject of prior studies (see Ref. [20,23,28]) and is similar to the industrial CHP application currently being investigated within the scope of the "Hyflexpower" demonstration project [22].

The energy unit portfolio under consideration is comprised of three subsystems: an industry subsystem, a wind subsystem and a hydrogen subsystem. The industry subsystem represents a manufacturing plant with an electrical load and aF high-temperature thermal load, with the latter beeing met exclusively by the hydrogen-fired gas turbine. Consequently, its operating point is defined by the industrial heat demand. The resulting power generation is used to partially meet the electrical load of the industry subsystem. The wind farm is a stand-alone system that generates volatile renewable electricity, which is primarily used to meet the remaining power demand of the industry subsystem. When the combined power generation of the gas turbine and the wind farm is insufficient to meet the electrical load of the manufacturing plant, additional grid power is consumed. On the other hand, when the non-dispatchable power generation of the wind farm exceeds the remaining industrial power demand,

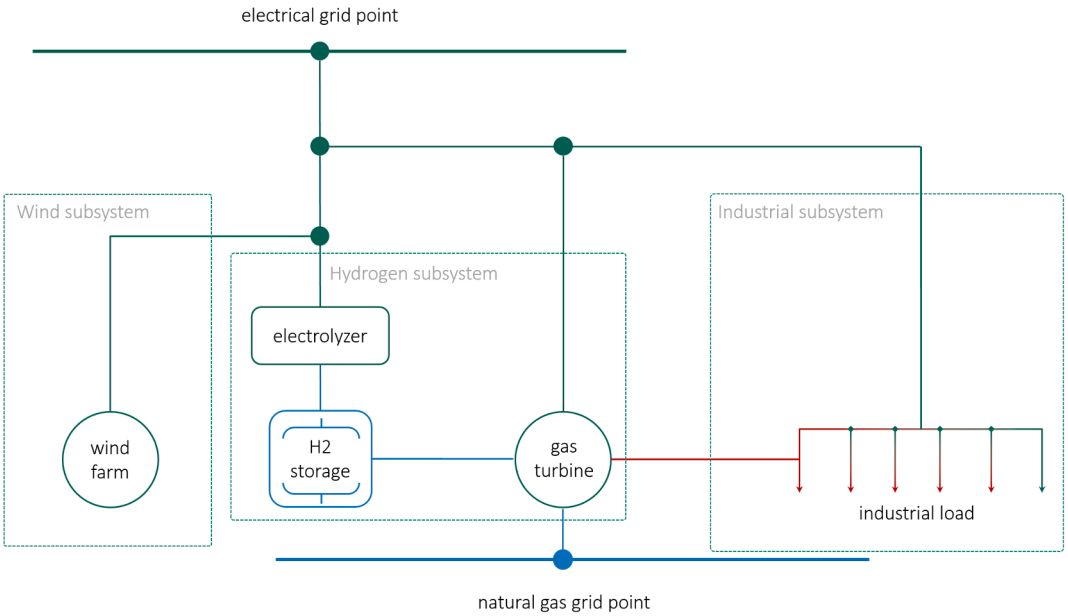

Fig. 1. Investigated system configuration with three subsystems.

local renewable power surpluses are generated. These power surpluses can either be utilized by the hydrogen subsystem or fed into the grid, resulting in a time-dependent compensation at market price. In addition to the hydrogen-fired gas turbine, the hydrogen subsystem consists of the on-site electrolyzer and the on-site hydrogen storage facility. The storage facility comprises pressure vessels and a compressor. Depending on the implemented operational strategy, the electrolyzer is powered either by the locally generated renewable power surpluses or by a combination of local power surpluses and additional grid power. The exclusive use of renewable power surpluses results in the production of fully CO_2-neutral hydrogen. When additional grid power is utilized for the operation of the electrolyzer, the CO_2 emissions associated with the hydrogen production are defined by the CO_2 emission factor of the grid.[7] Regardless of the operational strategy considered for the electrolyzer, the generated hydrogen is temporarily stored in the on-site pressure vessels and subsequently used by the gas turbine to provide dispatchable power and high-temperature heat without the emission of CO_2. When needed, the gas turbine can also use natural gas as backup fuel. However, during such periods, the operation of the gas turbine is no longer CO_2-free.

3.2. Simulation models

The present study employs a detailed MATLAB simulation model of the system configuration described in Sec. 3.1. The steady-state model is able to simulate the operation of each system component during a one-year period with a 15-minute temporal resolution.[8] This allows for the consideration of both short-term and long-term fluctuations in wind power generation.

The RL model is developed in Python building the environment upon the OpenAI Gym and the algorithm framework RLlib. The MATLAB submodels of the individual system components provide the functional correlations for the design of the RL agent's learning environment (compare Sec. 3.5). Furthermore, the MATLAB model is used to simulate a RB unit commitment as well as an optimized operational management based on DP (compare Sec. 3.3 and 3.4). The results of these simulations serve as comparative indicators for the evaluation of the RL agent's operational management.

3.3. Rule-based operational management

The definition of individual rules for the operation of each component constitutes the simplest approach to the operational management of any system configuration. Consequently, the results of the RB unit commitment serve as baseline reference for the evaluation of the operational strategies optimized by means of DP and RL. Within the present study, the RB operational management aims to produce completely CO_2-neutral hydrogen. Thus, no grid power is used for the operation of the on-site electrolyzer and the associated hydrogen compressor. The electrolyzer and compressor are only operated when local renewable power surpluses are available. Further, the considered RB operational management foresees that the gas turbine uses the stored hydrogen as primary fuel. When the storage vessels are depleted, natural gas is utilized as backup fuel. Note that when the maximum allowable pressure of the hydrogen storage vessels is reached, the operation of the electrolyzer has to be suspended. In this case, local power surpluses are fed into the grid entirely.

3.4. Optimized operational management with Reinforcement Learning and Dynamic Programming

While the described RB operational management results in the maximum production and usage of CO_2-neutral hydrogen, it does not consider the time-dependent character of grid feed-in tariffs and wholesale electricity prices. The available on-site hydrogen storage capacity limits the operating time of the electrolyzer. This limitation creates a potential for economic optimization as the revenues from the grid feed-in of local power surpluses can potentially be increased while both the amount of hydrogen produced and consumed remain unaltered.[9] The exploitation of this potential requires a forecast-based optimization, which corresponds to a non-linear, time-variant optimal control problem with a constraint final state. The general form of this control problem is provided in Eqn. (1) to (7). The overall goal is to minimize the cost function J by applying an optimal set of control signals u within a given time period while simultaneously respecting the control signal constraints and the constraints of the state variable x.

$$\min_{u(t)} J(u(t)), \qquad \text{s.t.} \tag{1}$$

$$\dot{x} = F(x(t), u(t), t) \tag{2}$$

$$x(0) = x_0 \tag{3}$$

$$x(t_{\text{final}}) \in [x_{\text{final,min}}\ x_{\text{final,max}}] \tag{4}$$

$$x(t) \in [x_{\min} x_{\max}] \tag{5}$$

$$u(t) \in [u_{\min} u_{\max}] \tag{6}$$

$$J(u(t)) = G(x(t_{\text{final}})) + \int_0^{t_{\text{final}}} H(x(t), u(t), t)\,dt \tag{7}$$

The present study primarily aims to investigate whether the described optimal control problem can be solved by a trained RL agent. To generate a reference for the performance of the RL agent, the control problem was additionally solved by applying the deterministic DP algorithm first introduced by Bellman [45].[10]

Within the DP framework, the cost function accounts for the time-dependent revenues resulting from the grid feed-in of surplus power. In line with the RB unit commitment, no grid power shall be used for the operation of the electrolyzer and the hydrogen compressor. Correspondingly, the cost function penalizes any potential grid power consumption of these components. The pressure level of the hydrogen storage vessels represents the state variable of the optimal control problem. For any time period considered for optimization, both the amounts of hydrogen produced and consumed are predetermined by the RB unit commitment. As a result, the final state constraint of the optimal control problem corresponds to the storage pressure level at the final time step calculated by the RB unit commitment. The state variable is impacted by the hydrogen production of the electrolyzer and the hydrogen consumption of the gas turbine. The operating points of both components are potential control inputs for the optimization within the

[7] For example, in 2019, the generation mix for grid power in Germany yielded a CO_2 emission factor of 401 g/kWh [56]. Under these circumstances, the sole use of grid power for the operation of an electrolyzer and the subsequent substitution of natural gas as fuel would increase the overall CO_2 emissions by a factor of up to three [20].

[8] Information regarding the implemented modelling approach is provided in Ref. [23].

[9] It is important to note that this optimization does not consider the economic viability of using hydrogen as gas turbine fuel, as both the amounts of hydrogen produced and consumed are predefined by the RB unit commitment. The general economic viability of the fuel switch from natural gas to hydrogen is briey discussed in Sec. 4.1.3. In summary, the described optimization can be regarded as an extension of the RB unit commitment and not as a standalone operational strategy.

[10] The implemented DP optimization algorithm is based on the work of Sundström et al. [57,58]. Additional information regarding the application of DP for the optimized operational management of a hydrogen-fired gas turbine with on-site electrolysis and on-site hydrogen storage is provided in Ref. [28].

DP framework. The ability to switch between a hydrogen-based and a natural gas-based operation of the gas turbine constitutes a third potential control input. Within the investigated system configuration, the gas turbine's operating point is determined by the heat demand of the manufacturing plant. Consequently, the operating point cannot be used as a control input for the optimization. Regarding the ability to switch between a hydrogen-based and a natural gas-based operation of the gas turbine, the RB operational management already considers the primary use of hydrogen. In comparison to other potential strategies, this prioritization results in the lowest average pressure levels of the hydrogen storage vessels. Lower pressure levels correlate with a reduced power consumption of the hydrogen compressor. Considering this effect and the predetermined amount of hydrogen to be consumed by the gas turbine, the RB operational management already provides the optimal control input trajectory. This leaves the operating point of the electrolyzer, represented by its level of power consumption, as the sole control input for the optimization within the considered DP framework.

3.5. Design of the Reinforcement Learning agent's learning environment

In order to map the outlined optimal control problem as an RL problem, the technical and physical constraints, as well as the economic incentives, are described in a MDP. The MDP is a time-discrete stochastic control process that defines the agent's interaction with its environment and consists of the tuple $\langle \mathscr{S}, \mathscr{A}, \mathscr{P}, \mathscr{R}, \theta \rangle$, described in the following.

$\mathscr{S}$ denotes the state space, which includes all of the possible environmental states. Accordingly, s_t denotes the state of each time step t, including the environment information on the operating point of each component, the current State of Charge (SoC) of the storage vessels, and the current tariff for the grid feed-in of power. $\mathscr{A}$ is the agent's action space, which dentoes the actions available in the environment through unit control. While the environment consists of rule-based, passive and smart units, only the latter are operated by an agent through a control signal. In the investigated scenario, the electrolyzer is the only smart unit, with the agent's actions controlling the power consumption at each time step. The rule-based units are those that are influenced indirectly by the agent through the smart unit operation, e.g., the hydrogen storage vessels. Passive units are assumed to be uncontrollable, which are the wind and the industy subsystem. $\mathscr{R}$ is the reward function that provides a reward r_t to the agent after the operation at has been completed and the transition to the new states has occurred. Equivalent to the cost function in Eq. (7), the implemented incentive mechanism includes economic signals from the grid interaction. Further, the reward function considers incentives for the production and consumption of hydrogen. This approach ensures the comparability between the approaches. Lastly, the transition probabilities $\mathscr{P}$ are directly derived from the MDP. In the case of the DRL, $\mathscr{P}$ are referred to as policies $\pi(a_t|s_t)$ that evaluate the future reward for the given action a_t. Thus, the policies describe the probability of choosing action at given the current state s_t and is dependent on the parameter θ. In DRL, the policies are described through a neural network, where the weights of the network correspond to the parameter θ.

The following sections describe the environment and how the agent's chosen actions are taken into account in the UCP. The environment can be described as a "parachute environment", in which the agent is free to choose technically inadmissible actions which are corrected by the environment. This approach avoids the learned policy being distorted by strong negative reward signals.

3.6. Hydrogen subsystem

As described in Sec. 3.1, the hydrogen subsystem consists of the smart electrolyzer, the rule-based hydrogen storage facility and the hydrogen-fired gas turbine.

3.6.1. Smart electrolyzer

The hydrogen generated in period t is given as a function of the electrolyzer's power consumption, such that[11]:

$$\dot{m}_{\mathrm{H_2,t}}^{\mathrm{EZ}} = f_{\mathrm{EZ}}\left(P_{\mathrm{el,t}}^{\mathrm{EZ}}\right) \tag{8}$$

The electrolyzer is controlled by the action of the agent using the action interval $a_t \in [0, 1]$. Eqn. (9) describes how the agents action is translated into the requested electrolyzer power consumption:

$$P_{\mathrm{el,t}}^{\mathrm{EZ,req}} = \min\left(a_t \cdot \overline{P}_{\mathrm{el}}^{\mathrm{EZ}}, \left(\mathrm{P}_{\mathrm{el,t}}^{\mathrm{WF}} + P_{\mathrm{el,t}}^{\mathrm{GT}} - \mathrm{P}_{\mathrm{el,t}}^{\mathrm{Ind}} - P_{\mathrm{el,t}}^{\mathrm{Com}}\right), f_{\mathrm{EZ}}^{-1}\left(\overline{\dot{m}}_{\mathrm{H_2,t}}^{\mathrm{St}}\right)\right) \tag{9}$$

where in the first term the operating point chosen by the agent is defined through the nominal power consumption capacity of the electrolyzer $\overline{P}_{\mathrm{el}}^{\mathrm{EZ}}$. Similar to the other approaches, the electrolyzer and the associated hydrogen compressor are only intended to consume local renewable power surpluses. The availability of surplus power is influenced by the power generation of the wind farm $\mathrm{P}_{\mathrm{el,t}}^{\mathrm{WF}}$ and the gas turbine $P_{\mathrm{el,t}}^{\mathrm{GT}}$, as well as the electrical load of the industrial subsystem $\mathrm{P}_{\mathrm{el,t}}^{\mathrm{Ind}}$. Additionally, the potential power consumption of the hydrogen compressor $P_{\mathrm{el,t}}^{\mathrm{Com}}$ is of relevance. Lastly, the electrolyzer's operating point is limited by the maximum permissible charging rate of the hydrogen storage vessels $\overline{\dot{m}}_{\mathrm{H_2,t}}^{\mathrm{St}}$. This restriction is considered by an inversion of the functional correlation provided in Eq. (8). Due to technical limitations, electrolyzers typically cannot use the full load range during nominal operation [5]. To account for the minimum load point of the electrolyzer $P_{\underline{\mathrm{el}}}^{\mathrm{EZ}}$, Eq. (10) is defined as:

$$P_{\mathrm{el,t}}^{\mathrm{EZ,set}} = \begin{cases} P_{\mathrm{el,t}}^{\mathrm{EZ,req}} & \text{if } P_{\mathrm{el,t}}^{\mathrm{EZ,req}} \geq \underline{P}_{\mathrm{el}}^{\mathrm{EZ}} \\ 0 & otherwise \end{cases} \tag{10}$$

Where $P_{\mathrm{el,t}}^{\mathrm{EZ,req}}$ is the requested electrolyzer operation and $P_{\mathrm{el,t}}^{\mathrm{EZ,set}}$ is the applicable operational set point.

3.6.2. Rule-based hydrogen storage facility

The SoC of the hydrogen storage vessels is defined by the stored hydrogen mass of the previous period plus the charged or discharged mass during period t:

$$m_{\mathrm{H_2,t}}^{\mathrm{St}} = m_{\mathrm{H_2,t-1}}^{\mathrm{St}} + \dot{m}_{\mathrm{H_2,t}}^{\mathrm{St}} \cdot \Delta_t \tag{11}$$

To store hydrogen, it must be compressed to a pressure level correlating with the SoC. The storage pressure can be obtained from Eqn. (12), which is an extended version of the ideal gas law that accounts for real gas behavior.

$$p_t^{\mathrm{St}} = \frac{R_{\mathrm{H_2}} \cdot T^{\mathrm{St}} \cdot Z_{\mathrm{H_2}}\left(T^{\mathrm{St}}, p^{\mathrm{St}}\right) \cdot m_{\mathrm{H_2,t}}^{\mathrm{St}}}{V_{\mathrm{geo}}^{\mathrm{St}}} \tag{12}$$

Eqn. (13) describes the maximum permissible charging rate of the hydrogen storage vessels at time step t, which is determined by the SoC at the previous time step and the maximum SoC $\overline{m}_{\mathrm{H_2}}^{\mathrm{St}}$. As indicated by Eq. (9), the maximum charging rate limits the potential power consumption of the electrolyzer. Similarly, the maximum permissible discharging rate determines whether the gas turbine can operate on hydrogen or whether natural gas has to be used as backup fuel (compare Eq. (19), (20)). As described by Eq. (14), the maximum discharging rate is defined considering the minimum SoC $m_{\underline{\mathrm{H_2}}}^{\mathrm{St}}$ as well as the SoC of the previous time step. Both the maximum and the minimum SoC are design parameters of the hydrogen storage facility.

[11] As mentioned in Sec. 3.2, all functional correlations used for the design of the agent's learning environment are derived from the MATLAB model introduced in Ref. [23]. The parameterization of this model is described in Sec. 4.1.

$$\overline{\dot{m}}^{St}_{H_2,t} = \frac{\overline{m}^{St}_{H_2} - m^{St}_{H_2,t-1}}{\Delta_t} \tag{13}$$

$$\underline{\dot{m}}^{St}_{H_2,t} = \frac{m^{St}_{H_2,t-1} - \underline{m}^{St}_{H_2}}{\Delta_t} \tag{14}$$

The hydrogen mass flow during one period is determined by the hydrogen produced through electrolysis and the hydrogen consumed by the gas turbine:

$$\dot{m}^{St}_{H_2,t} = \dot{m}^{EZ}_{H_2,t} - \dot{m}^{GT}_{H_2,t} \tag{15}$$

Finally, the power demand of the hydrogen compressor in time step t is a function of the current hydrogen production of the electrolyzer, the current pressure level of the hydrogen storage vessels and the storage pressure at the previous time step:

$$P^{Com}_{el,t} = f\left(\dot{m}^{EZ}_{H_2,t}, p^{St}_{t}, p^{St}_{t-1}\right) \tag{16}$$

3.6.3. Hydrogen-fired gas turbine

The hydrogen-fired gas turbine is primarily employed to supply high-temperature heat for the manufacturing plant, as described in Eqn. (17). To determine the corresponding electrical power output, the known correlation between power and heat generation of the gas turbine is used. As indicated by Eqn. (18), the developed environment distinguishes between hydrogen-based and natural gas-based operation as the thermal efficiency of the gas turbine process depends on the fuel used.[12]

$$\dot{Q}^{GT}_{th,t} = \dot{Q}^{Ind}_{th,t} \tag{17}$$

$$P^{GT}_{el,t} = \begin{cases} f^{-1}_{GT,H_2}\left(\dot{Q}^{GT}_{th,t}\right) & \text{if } \dot{m}^{GT}_{H_2,t} > 0 \\ f^{-1}_{GT,NG}\left(\dot{Q}^{GT}_{th,t}\right) & \text{if } \dot{m}^{GT}_{NG,t} > 0 \\ 0 & otherwise \end{cases} \tag{18}$$

With the aim to utilize hydrogen as primary fuel, the individual decision whether hydrogen can be used depends on the sufficiency of the stored hydrogen to operate the gas turbine during the considered period at nominal capacity (Eqn. (19)).[13] If the condition indicated in Eqn. (19) cannot be met, natural gas is used as backup fuel (Eqn. (20)).

$$\dot{m}^{GT}_{H_2,t} = \begin{cases} f_{GT,H_2}\left(P^{GT}_{el,t}\right) & \text{if } f_{GT,H_2}\left(\overline{P^{GT}_{el}}\right) \leq \underline{\dot{m}}^{St}_{H_2,t} \\ 0 & otherwise, \end{cases} \tag{19}$$

$$\dot{m}^{GT}_{NG,t} = \begin{cases} f_{GT,NG}\left(P^{GT}_{el,t}\right) & \text{if } \dot{m}^{GT}_{H_2,t} = 0 \\ 0 & otherwise. \end{cases} \tag{20}$$

3.7. Wind subsystem, industrial subsystem and grid point consideration

The wind farm is classiffied as a passive unit, which means that its power generation $P^{WF}_{el,t}$ is not regulated by any action or rule. The same holds for the industry subsystem, which includes the electrical load $P^{Ind}_{el,t}$ and the high-temperature thermal load $\dot{Q}^{Ind}_{th,t}$. Finally, the electrical grid point closes the electrical balance of the system, in that all remaining power surpluses are fed into the grid and all surplus demand is covered by grid consumption. The latter can be reduced to the remaining electricity demand of the manufacturing plant, since the electrolyzer and hydrogen compressor shall only consume the locally generated renewable power surpluses (Eqn. (9)).

3.8. Reward function definition

Defining the reward function based on purely economic parameters, e.g., revenues, is unsuitable for the considered RL environment for two reasons. Firstly, the revenues from the grid feed-in of surplus power significantly outweigh the reward signals received by the agent. The second reason relates to the structure of the investigated optimal control problem. To obtain results that are comparable with the RB operational management and the optimized operational management based on DP, a target SoC for the end of each optimization period has to be specified in addition to the economic parameters. For the same reason, the gas turbine has to be provided with a target for using hydrogen as fuel within the considered period.[14] Both the target SoC and the targeted hydrogen consumption are defined by the RB operational management (compare Sec. 3.3).

To account for these objectives, a three-part reward function has been defined. The first part is an instantaneous reward, returned at each time step after transition of the state following the agents action. Secondly, rewards for reaching the episode targets, relative to the degree of fulfilment, are returned after the termination of the trajectory. Thirdly, a reward is returned at the end of the trajectory if and only if the target is reached in full.

$$r^{(1)}_t = \underbrace{\left(P^{EZ,set}_{el,t} + P^{Com}_{el,t}\right) \times \Phi_t \times W^{(1)}}_{\mathbb{R}}, \tag{21}$$

$$r^{(2)} = \underbrace{\min\left(\frac{m^{St}_{H_2,\tau} - m^{St}_{H_2-target,\tau}}{\overline{m}^{St}_{H_2} - \underline{m}^{St}_{H_2}} \cdot W^{(2)}, 0\right)}_{\mathbb{R}^+_0} + \underbrace{\max\left(\frac{m^{GT}_{H_2-\ target,\tau} - m^{GT}_{H_2,\ \tau}}{\overline{\dot{m}}^{GT}_{H_2,t} \cdot \tau \cdot \Delta_t} \cdot W^{(3)}, 0\right)}_{\mathbb{R}^+_0}. \tag{22}$$

Eqn. (21) defines the instantaneous reward, where Φ_t respresents the hourly varying electricity spot pices. Because the consumption of the electrolyzer ($P^{EZ,set}_{el,t}$) and the consumption of the compressor ($P^{Com}_{el,t}$) are both negative, this term has a negative impact on the agent's reward. As a result, the agents objective is to reduce costs by shifting consumption to low-cost periods. The parameters $W^{(1)}$, $W^{(2)}$ and $W^{(3)}$ represent the weights for the reward function terms which are determined by experiment. Eqn. (21) incentivizes the operation of the electrolyzer and the associated hydrogen compressor during times of low grid feed-in tariffs to minimize opportunity costs. The first term in Eqn. (22) addresses the objective of reaching the final SoC of the hydrogen storage vessels, while the second term incentivizes the use of hydrogen for the gas turbine operation. The resulting reward function is defined as follows:

$$r_t = \begin{cases} r^{(1)}_t & \text{if } t < \tau \\ r^{(1)}_t + r^{(2)} + R^{(3)} \times b^{St} + R^{(4)} \times b^{GT} & \text{if } t = \tau, with \end{cases} \tag{23}$$

$$b^{St} = \begin{cases} 1 & \text{if } m^{St}_{H_2,\tau} \geq m^{St}_{H_2-\ target,\tau} \\ 0 & otherwise \end{cases} \tag{24}$$

[12] Additional information regarding the impact of hydrogen on the thermodynamics of the gas turbine process is provided in Ref. [59].

[13] This constraint limits the frequency of fuel switches. A similar constraint is considered for the RB operational management as well as the optimized operational management based on DP.

[14] As explained in Sec. 3.4, the RB operational management already provides the optimal control input trajectory regarding the operation of the gas turbine. However, in order to increase the flexibility of the RL agent, the implemented reward function only considers the overall hydrogen consumption of the gas turbine.

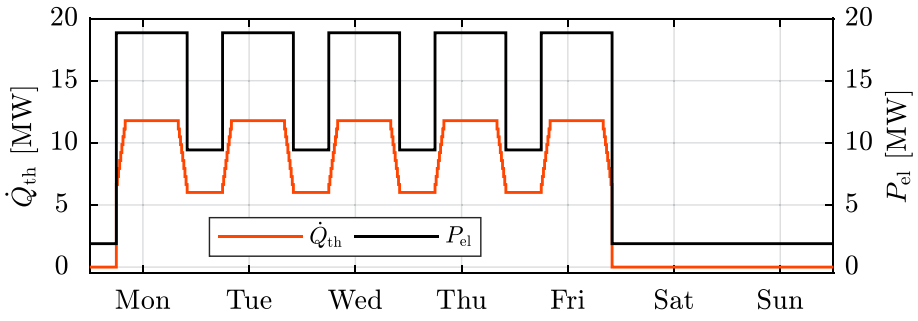

Fig. 2. Power and high-temperature heat demand of the industy subsystem.

$$b^{\mathrm{GT}} = \begin{cases} 1 & \text{if } m^{GT}_{H_2-\ \mathrm{measured},\tau} \geq m^{\mathrm{GT}}_{\mathrm{H_2}-\ \mathrm{target},\tau} \\ 0 & otherwise \end{cases} \tag{25}$$

where b^{St} and b^{GT} are binary variables yielding a reward if and only if the target is reached in full. The corresponding rewards $\mathrm{R}^{(3)}$ and $\mathrm{R}^{(4)}$ are determined by experiment.

3.9. Reinforcement Learning algorithm

The optimization algorithm is of major importance for the success of the RL agent. Considering the experimental setting, the Proximal Policy Optimization (PPO) algorithm was chosen for this task, which was first introduced by Schulman et al. [46]. The PPO is an advanced policy gradient method that addresses the problem of performance degradation caused by excessively large policy updates in policy gradient methods. Targets of the algorithm are the parameters θ of a neural network that selects the actions in the environment through the policies $\pi(a_t|s_t)$ (compare Sec. 3.5). The PPO optimizes the parameters θ through an iterative stochastic gradient descent (SGD), e.g., with the Adam algorithm [47], but ensures stable performance through further restrictions in the optimization. This restriction consists primarily of a clipping mechanism, that evaluates the updates between θ_{old} and θ through a ratio $\gamma_t(\theta)$:

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}. \tag{26}$$

The ratio describes the proportional update of the SGD between each iteration step. Considering that the PPO penalizes large updates, a smaller ratio is preferred. This also translates in the loss function of the algorithm, where ϵ is a clipping parameter and $\widehat{A}_t(\theta)$ describes the advantage function:

$$L^{\mathrm{CLIP}}_t(\theta) = \widehat{E}_t[\min(r_t(\theta)\widehat{\mathrm{A}}_t(\theta), clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\widehat{\mathrm{A}}_t(\theta))]. \tag{27}$$

Alongside the major restrictions, the PPO also has further algorithmic improvements, e.g., efficient sampling of mini batches with parallel workers and thus proves suitable for the analyses of this paper. In the practical application, the PPO was implemented with the open-source library RLlib[15].

4. Experiment design

4.1. System design parameters

A direct comparison between the different approaches requires a consistent parameterization of the investigated system configuration, which is briefly introduced in this section.[16]

4.1.1. Wind subsystem

To calculate the performance of the wind farm, the present investigation uses wind speed data from the year 2019 provided by the German Weather Service [48]. The data correspond to an on-shore site in Western Germany. Employing the approach presented in [49], the temporal resolution of the wind data was increased to match the 15-minute temporal resolution of the overall simulation model. The modified wind speed data serve as input for a wind turbine performance curve. The present study uses the performance curve of a state-of-the-art 3.05 $\mathrm{MW_{el}}$ wind turbine [50]. The employed wind farm model is able to simulate the performance of multiple turbines at a common site accounting for wind farm effects. Information regarding the considered wind farm effects is provided in [49].

Within the investigated system configuration, the number of wind turbines affects the economic viability of the wind farm as well as the availability of surplus renewable power that can be used for the production of hydrogen. Accounting for capital expenditures, operational expenditures and potential revenues, Ref. [20] provides a comprehensive techno-economic assessment of this trade-off and reveals that the installation of up to 10 wind turbines is economically viable within the considered economic and regulatory framework. Based on these findings, the investigated wind farm comprises 10 wind turbines, resulting in an installed power generation capacity of 30.5 $\mathrm{MW_{el}}$.[17]

4.1.2. Industry subsystem

The considered power and high-temperature heat demand profiles are shown in Fig. 2. It is assumed that the illustrated profiles are identical for each week of the investigated one-year period. The displayed power demand profile is a simplified representation of the demand characteristics of an automotive manufacturing plant provided by Ruangpattana et al. [51]. Regarding the thermal load, Giamperi et al. indicate that automotive manufacturing plants typically require high-temperature heat in the form of process steam [52]. The authors also reveal that the total power demand of these plants usually exceeds their total heat demand. Taking this into account, it is assumed that the peak high-temperature heat demand of the considered automotive manufacturing plant is significantly lower than its peak power demand and that both the power and the heat demand characteristic follow the same general pattern. Within the investigated system configuration, the demand for process steam is met by a heat recovery steam generator (HRSG), which utilizes the exhaust gas heat of the gas turbine. Thus, the peak heat demand displayed in Fig. 2 matches the nominal steam

[15] See for more information: https://docs.ray.io/en/master/rllib-algorithms.html.

[16] This section only introduces the parameterization of the individual system components at their respective design point. Information regarding the off-design performance of each component is provided in Ref. [23].

[17] The installation of 10 wind turbines at a single on-shore site is assumed to be the upper limit regarding technical and regulatory feasibility.

Table 1
Design parameters of hydrogen subsystem.

	Parameter	Value	Unit
Gas turbine (GT)	P_{el}	7900	[kW]
	$\eta_{el,LHV}$	30.6	[%]
	T_{EG}	815.15	[K]
	$\dot{m}_{EG}$	30.2	[kg/s]
HRSG	p_{Steam}	21	[bar]
	T_{Steam}	673.15	[K]
	$\dot{m}_{Steam}$	4.17	[kg/s]
Electrolyzer (EZ)	P_{el}	12,000	[kW]
	$\dot{m}_{H_2}$	0.06	[kg/s]
	$\eta_{el,LHV}$	60	[%]
Hydrogen storage (St)	p_{max}	300	[bar]
	p_{min}	25	[bar]
	V_{geo}	200	[m^3]
	T	288.15	[K]
Hydrogen compressor (Com)	P_{el}	980	[kW]

generation capacity of the considered HRSG (compare Sec. 4.1.3).

4.1.3. Hydrogen subsystem

The main design parameters of the considered hydrogen subsystem are provided in Table 1. The listed gas turbine performance data correspond to a state-of-the-art industrial gas turbine commonly employed for CHP applications [53]. It is assumed that the gas turbine is able to operate with 100% hydrogen as well as 100% natural gas. As discussed in Sec. 2.1, this assumption goes beyond the state of the art and presupposes the successful conclusion of current research and development activities.

The specified process steam parameters are based on the exhaust gas parameters of the industrial gas turbine, the performance characteristic of a state-of-the-art HRSG and additional assumptions regarding the requirements of the manufacturing plant [53].

The indicated electrolyzer efficiency is derived from the design data of an alkaline electrolyzer that is currently operated within a power-to-gas demonstration plant in Germany [5]. The listed nominal power consumption capacity of the electrolyzer and the corresponding hydrogen production capacity are based on the results of a prior techno-economic analysis provided in Ref. [20]. Focusing on a system configuration similar to the one described in Sec. 3.1, this prior study considers electrolyzer power consumption capacities between 600 kW_{el} and 12 MW_{el}. The study reveals that, within the investigated economic and regulatory framework, none of the investigated power consumption capacities allows the production of hydrogen to be economically viable. Thus, no distinctive techno-economic optimum could be determined. On the other hand, the prior study underscores that an increased power consumption capacity of the electrolyzer correlates with and increased production of CO_2-neutral hydrogen. For the present investigation, an increased level of hydrogen production offers more opportunities regarding optimized operational management. Consequently, the nominal power consumption capacity selected for the present study corresponds to the maximum power consumption capacity investigated in Ref. [20].

The geometric storage volume of the on-site hydrogen storage vessels was selected for similar reasons. Considering the results provided in Ref. [20], the selected geometric storage volume is regarded as a viable compromise between the potential for operational optimization and the general ability to operate the on-site electroyzer. The listed maximum allowable pressure of the storage vessels corresponds to the state of the art of steel vessels employed for gas storage applications [5]. The specified minimum storage pressure correlates with the stable injection of hydrogen into the gas turbine combustor.

Finally, the indicated nominal power demand of the hydrogen compressor is derived from the hydrogen production capacity of the electrolyzer, the maximum storage pressure and additional assumptions regarding the compressor efficiency and design. These assumptions are based on a data set of an inter-cooled multi-stage diaphragm compressor provided in Ref. [54].

4.1.4. Economic framework

The present study considers the current regulatory and economic framework of the German energy sector.[18] As mentioned in Sec. 4.1.3, the on-site hydrogen production for a hydrogen-fired gas turbine is generally not economically viable within this framework. Nevertheless, investigations regarding an optimized operational management of these system configurations are of high relevance. Firstly, lessons learned can

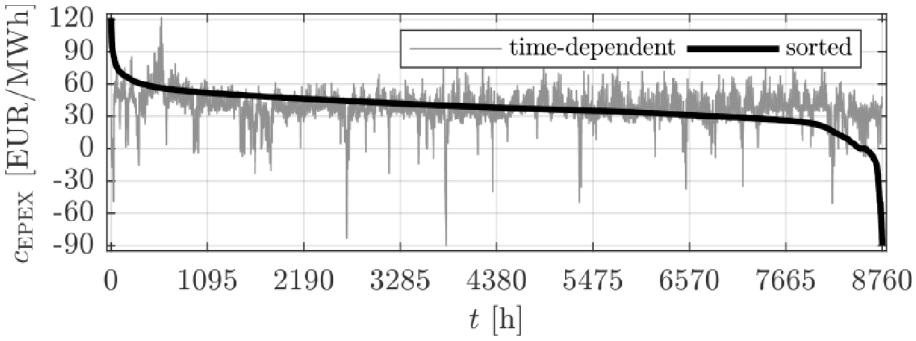

Fig. 3. EPEX spot price data.

[18] Detailed information regarding this framework is provided in Ref. [20].

be directly applied to ongoing demonstration projects. Secondly, changing economic conditions may result in an economically viable on-site hydrogen production for gas turbines in the foreseeable future [20]. As described in Sec. 3.4 and 3.5, both the optimized operational management using DP and RL require information regarding the time-dependent revenues resulting from the grid feed-in of surplus power. In line with the selected economic framework, the present investigation uses day-ahead spot price data from the European Power Exchange (EPEX) for this purpose. The employed data from the year 2019 are illustrated in Fig. 3.

4.2. Design of comparative analysis

To conduct the comparative analysis between the proposed three methods, the MATLAB simulation model introduced in Sec. 3.2 was parameterized using the data outlined in Sec. 4.1. In a following step, the required functional correlations were transferred into the RL agent's learning environment described in Sec. 3.5.

Since it does not consider the time-dependent character of revenues resulting from feeding surplus power into the grid, the RB unit commitment constitutes the simplest approach to the operational management of the investigated system configuration. Consequently, the results of the RB operational management serve as baseline reference for the comparative analysis. To obtain this reference, the parameterized MATLAB model is employed to simulate the operation of the system configuration for a one-year period with a 15-minute temporal resolution.

In a next step, the MATLAB model is utilized to simulate the optimized operational management based on DP for the same one-year period. In accordance with the day-ahead spot price data introduced in Sec. 4.1.4, the DP algorithm provides a day-ahead optimization for a 24-hour period. Consequently, a forecast for the day-ahead spot price data is required. The optimized operational management aims to achieve the same level of hydrogen production and consumption as the RB operational management. Therefore, the optimization algorithm requires information regarding the RB unit commitment. Furthermore, the DP algorithm requires forecasts regarding the non-dispatchable power generation of the wind farm as well as the power and heat demand of the manufacturing plant.[19] In general, all forecasts are prone to errors, e.g., see Ref. [55]. However, the consideration of forecast errors would prevent any direct comparison between the results of the RB unit commitment and the optimized operational management. Therefore, assuming perfect foresight for the considered 24-hour periods, forecast errors are neglected within the present investigation. As a result, the simulation outcome obtained for the optimized operational management based on DP can be regarded as a theoretical optimum for the comparative analysis.

Considering that the DP optimization algorithm provides the theoretical optimum and the RB unit commitment serves as the baseline reference, it is expected that the RL agent's performance ranges between these boundaries. As with the RL approach, the RB approach takes decisions concerning the electrolyzer's operational state in each time step sequentially based only on the observable environment state. An hourly resolution is chosen for all time series, which is different to the DP approach but necessary in order to reduce the data complexity and obtain viable results. The annual data set is divided into 365 one-day minibatches for the training, with each minibatch representing a training or test episode. Seventy-five percent of the data is used for training and twenty-five percent for testing. The RL agent's observation space is expanded by forecasts of the day-ahead spot prices, the manufacturing plant's energy demands, and the wind farm's power generation. As a result, the observations contain perfect foresight data until the end of the currently optimized period.[20] This enhances the quality of the results and improves the comparbility to the DP approach. A brief analysis of the application of forecasts for DRL is included in Sec. 5.3.

5. Results

In this section, the results of the comparative analysis outlined in Sec. 4.2 are discussed. To provide a reference for the performance of the RL agent, the results of the RB unit commitment and the optimized operational management based on DP are initially examined in Sec. 5.1 and 5.2. Subsequently, the RL results are presented in Sec. 5.3. Finally, Sec. 5.4 focuses on the comparison between the RL agent's performance and the reference approaches.

5.1. Rule-based unit commitment

The analysis of the RB operational management focuses on two exemplary 24-hour periods within the simulated one year period. The graphical evaluation is based on these exemplary days because the challenges of reaching a targeted final SoC and a targeted hydrogen consumption are adequately represented Fig. 4 illustrates the RB unit commitment during the first exemplary period ("period A").[21]

The low power demand of the manufacturing plant and the lacking gas turbine operation indicate that the considered period corresponds to a weekend day (compare Sec. 4.1.2). The investigated day features a high level of power generation by the wind farm. Consequently, the non-dispatchable RPG significantly exceeds the power demand of the manufacturing plant. Until approximately 7 a.m., the resulting power surplus is almost fully consumed by the on-site electrolyzer and the associated hydrogen compressor.[22] At approximately 7 a.m., the level of wind power generation increases significantly. As a result, the electrolyzer starts to operate at its nominal power consumption capacity and the unused share of surplus renewable power is fed into the grid. At approximately 4p.m., the electrolyzer operation is suspended and the surplus power is fed into the grid entirely. The reason for the electrolyzer shutdown can be derived from Fig. 5, which illustrates the impact of the hydrogen production and consumption on the pressure level of the hydrogen storage vessels.[23] At approximately 3p.m., the maximum allowable pressure level of the storage vessels is reached, which results in a suspension of the electrolyzer operation.

Fig. 6 illustrates the RB unit commitment during the second 24-hour period under consideration ("period B"). In contrast to Fig. 4, Fig. 6 indicates the significant power demand of the manufacturing plant as well as the heat-driven operation of the gas turbine. Accordingly, the considered exemplary period corresponds to a business day (compare Sec. 4.1.2). During the entire 24-hour period, the combined power generation of the gas turbine and wind farm exceeds the power demand of the manufacturing plant. The resulting power surplus is utilized by

[19] In a real-world application, the operator of the investigated system configuration would also require these forecasts for the RB operational management. This is due to the fact that the grid feed-in of power typically has to be reported to the responsible Transmission System Operator beforehand [49].

[20] Equivalent to the DP optimization algorithm, the RL approach also requires information regarding the RB unit commitment and corresponding forecasts to define the targeted hydrogen production and consumption.

[21] Positive values indicate the consumption of power or the supply of additional power by the grid. Correspondingly, negative values indicate the generation of power or the feed-in of surplus power into the grid.

[22] Fig. 4 indicates a low level of grid feed-in (illustrated as negative residual grid load) during this period. This grid feed-in is a result of the implemented RB unit commitment, which employs a conservative estimation of the power demand of the hydrogen compressor to determine the operating point of the electrolyzer. This conservative approach eliminates the need for an iterative calculation.

[23] Positive values indicate the production of hydrogen, while negative values correlate with the consumption of hydrogen.

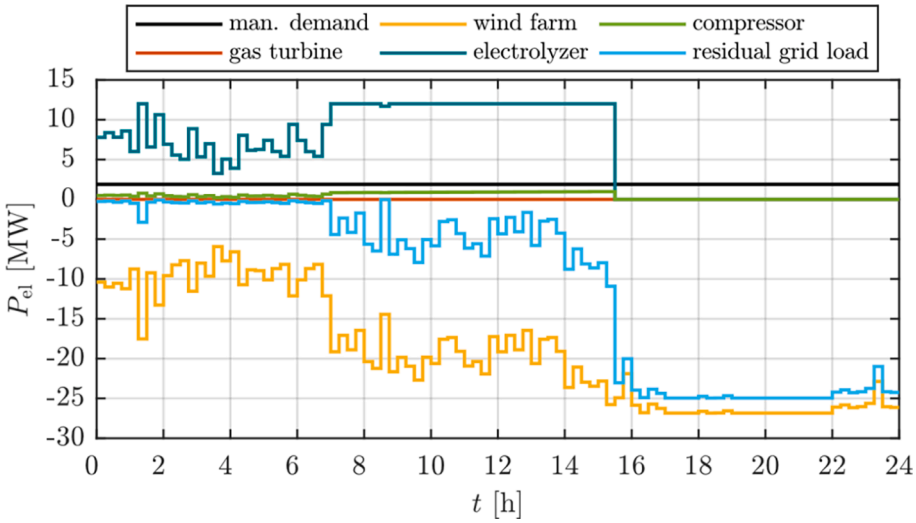

Fig. 4. RB unit commitment during exemplary period A.

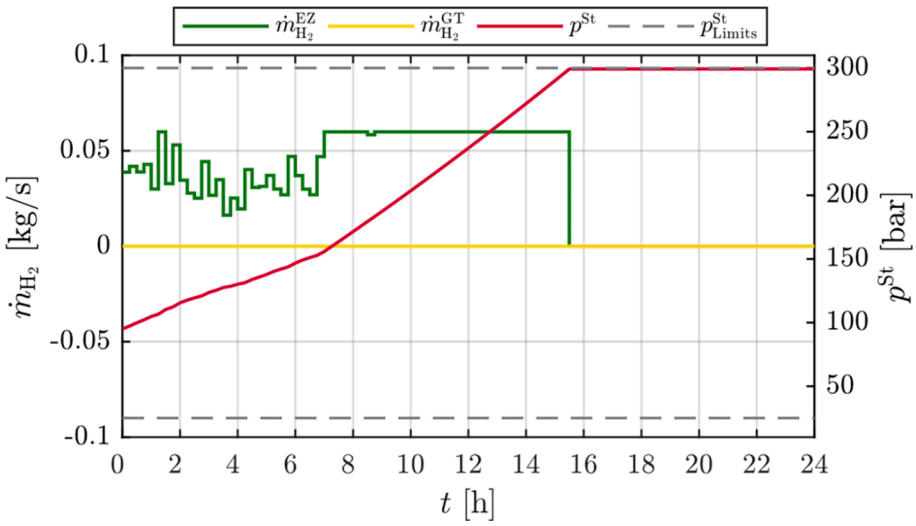

Fig. 5. Hydrogen storage pressure level during exemplary period A.

the hydrogen compressor and the electrolyzer, which operates at its nominal power consumption capacity during prolonged periods. Regarding the pressure level of the hydrogen storage vessels Fig. 7 indicates a comparatively low initial SoC. In addition, the nominal hydrogen consumption capacity of the gas turbine significantly exceeds the nominal hydrogen production capacity of the on-site electrolyzer. The amount of hydrogen available is insufficient to enable the gas turbine to be fully operated using hydrogen. Consequently, natural gas is utilized as the backup fuel when the minimum storage pressure is approached.[24]

5.2. Optimized operational management based on Dynamic Programming

The analysis of the operational management based on DP considers the same exemplary 24-hour periods as before. To highlight the differences between the two approaches, Figs. 8 and 9 illustrate the respective power consumption trajectories of the electrolyzer and the corresponding pressure level trajectories of the hydrogen storage vessels. Both figures depict the boundary lines of the optimal control problem calculated by the DP algorithm. As explained in Ref. [57], the upper and lower boundary lines separate the state space into feasible and unfeasible regions. For the optimal control problem under consideration, the boundary line trajectories are determined by the maximum and minimum allowable pressure of the storage vessels, the nominal hydrogen production capacity of the electrolyzer and the final state constraint. Moreover, the boundary line trajectories incorporate the time-dependent hydrogen consumption of the gas turbine as defined by the RB unit commitment.[25]

Since optimized operational management is related to the time-dependency of grid feed-in tariffs, Figs. 8 and 9 additionally display the trajectories of the considered EPEX spot prices during the examined 24-hour periods. Depicting the results for the exemplary period A, Fig. 8

[24] As mentioned in Sec. 3.6.3, the RB operational management limits the frequency of fuel switches. The use of hydrogen is only initiated when the stored hydrogen mass allows for at least one hour of gas turbine operation at nominal load. After the use of hydrogen is initiated, the gas turbine is continuously operated on hydrogen until the minimum storage pressure level is approached.

[25] As outlined in Sec. 3.4, the RB operational management already provides the optimal control trajectory regarding the operation of the gas turbine.

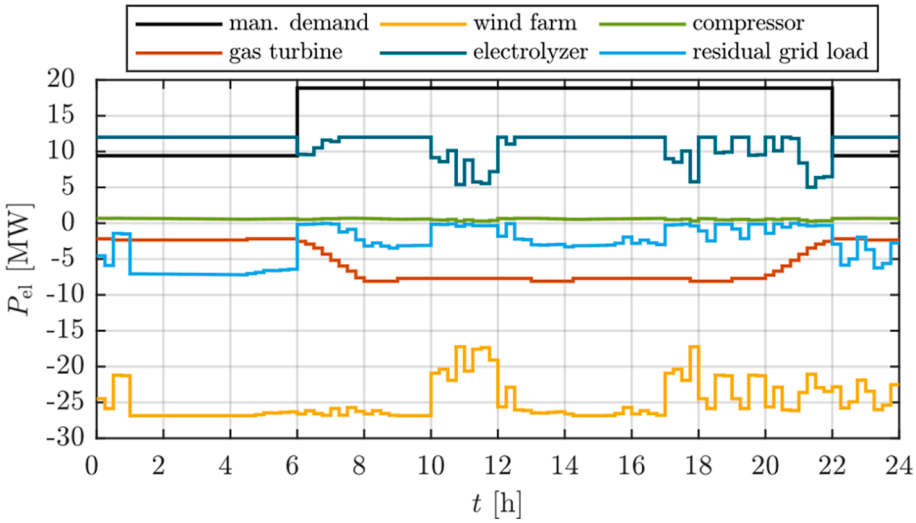

Fig. 6. RB unit commitment during exemplary period B.

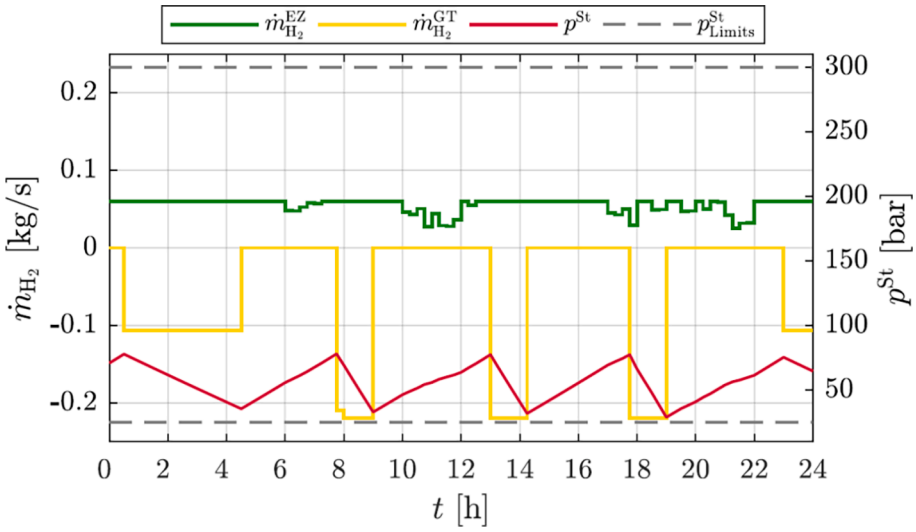

Fig. 7. Hydrogen storage pressure level during exemplary period B.

reveals significant differences between the RB operation of the electrolyzer and the optimized operation based on DP. Until approximately 9 a.m., the optimized trajectory of the electrolyzer's power consumption runs slightly above the trajectory corresponding to the RB unit commitment. Accounting for the comparatively low spot prices, the optimized operational management aims to fully utilize the available power surpluses during this period. In contrast, the RB operational strategy is not able to fully utilize the available surplus power. This is a result of the previously mentioned conservative estimation of the hydrogen compressor's power demand (compare Sec. 5.1). At approximately 8 a.m., Fig. 8 indicates a significant increase in the spot prices. Consequently, the optimized operational management foresees a reduction in the power consumption of the electrolyzer to reduce the corresponding opportunity costs. In contrast, the power consumption trajectory determined by the RB unit commitment shows a continued operation at nominal capacity, which leads to the maximum allowable storage pressure level being reached at approximately 3 p.m. The optimized operational management based on DP aims to produce the same amount of CO_2-neutral hydrogen as the RB operational management.[26] Consequently, the final state constraint of the considered optimal control problem corresponds to the maximum allowable pressure of the storage vessels. To comply with this final state constraint, the optimized operational management foresees the displayed operation of the electrolyzer between 2 p.m. and 5 p.m. and between 8 p.m. and 12 a.m. As illustrated in Figs. 4 and 8, both periods are characterized by the suffcient availability of renewable power surpluses as well as comparatively low spot prices. In summary, Fig. 8 reveals that the optimized operational management based on DP is able to comply with the final state constraint while shifting the operating time of the electrolyzer towards periods of low grid feed-in tariffs, resulting in a reduction of opportunity costs.

However, Fig. 8 also highlights that the utilized potential for optimization mainly derives from the fact that the operating time of the electrolyzer is limited by the available hydrogen storage capacity. As

[26] As mentioned in Sec. 5.1, the exemplary period A corresponds to a weekend day. Thus, the hydrogen-fired gas turbine is not operated and the consumption of hydrogen does not have to be considered.

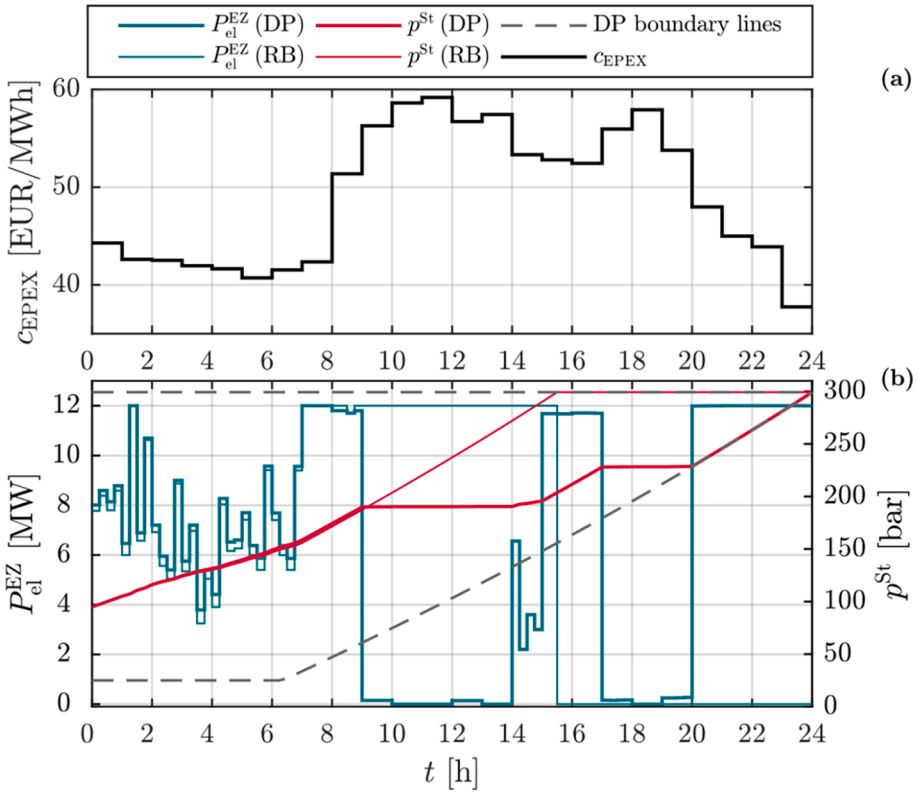

Fig. 8. Optimized operational management based on DP during exemplary period A.

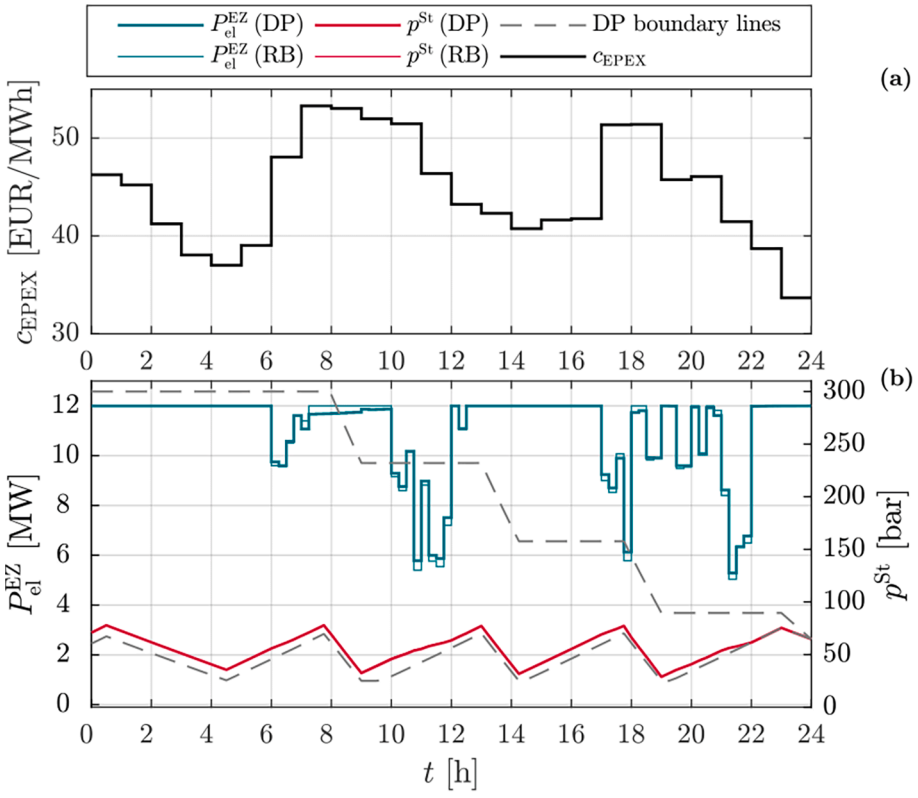

Fig. 9. Optimized operational management based on DP during exemplary period B.

shown in Fig. 7, the available on-site hydrogen storage capacity does not have a limiting impact on the operation of the electrolyzer during the investigated exemplary period B. Consequently, Fig. 9 shows no significant differences between the RB operation of the electrolyzer and the optimized operation based on DP. In order to comply with the boundary lines as well as the resulting final state constraint, the optimized operational management based on DP foresees a power consumption trajectory for the electrolyzer that is very similar to the RB unit commitment. This trajectory is mainly defined by the time-dependent availability of renewable power surpluses (see Fig. 6). However, the previously described conservative estimation of the hydrogen compressor's power demand for the RB unit commitment leaves a limited potential for optimization. As highlighted in Fig. 9, the implemented DP algorithm is able to exploit this potential by fully utilizing the available power surpluses during periods of low spot prices. As a result, the power consumption of the electrolyzer can be slightly reduced during periods of high spot prices. In summary, the accruing opportunity costs are reduced while the amount of hydrogen produced and consumed remain unaltered in comparison to the RB unit commitment.

5.3. Operational management based on Deep Reinforcement Learning and forecasts

Fig. 10 depicts the operation of the elecrolyzer based on the RL algorithm's selected actions during the previously introduced exemplary period A. The resulting impact on the SoC, represented by the pressure level of the storage vessels, and the corresponding spot prices are also illustrated. The figure reveals significant differences between the electrolyzer set points chosen by the agent with (with FC) and without forecast information (w/o FC). Despite the lack of forecast information, the RL-based operational management is able to reach the targeted final SoC. However, the displayed electrolyzer power consumption indicates no consideration of the time-dependent spot prices in this case. As a result, the RL-based operational management without forecasts is very similar to the RB-based unit commitment displayed in Fig. 8. In contrast, the results corresponding to the inclusion of forecasts in the agent's observations are promising. It is evident that the objective of a final SoC of 300 bar is met. Furthermore, the electrolyzer's operation is noticeably oriented towards the spot prices, i.e. when prices are high, the surplus power is not used to operate the electrolyzer but instead fed into the grid, and vice versa. The agent's decision-making appears to be improved by the forecast information concerning the wind power generation, spot prices, and the manufacturing plant's energy demands. Accounting for foreseeable renewable power surpluses, the agent is able to consider the number of future time steps the electrolyzer will be able to produce hydrogen. Simultaneously, the agent utilizes information regarding the time-dependent development of the spot prices. Due to these auspicious results, the following comparative analysis only considers the inclusion of forecasts in the agent's observations space.

5.4. Comparative analysis of Reinforcement Learning, Dynamic Programming, and Rule-based operational management

Considering exemplary period A, Fig. 11 depicts the comparison between the RL-based optimized operational management, the RB unit commitment and the optimized operational management based on DP. The area between the respective storage pressure trajectories displayed in Fig. 11 can be roughly regarded as the optimality gap. Overall, both optimization approaches are able to meet the targeted final storage pressure defined by the RB unit commitment. While this condition is binding for the DP approach, the RL agent is rewarded for meeting this target. When considering the RL agent's results, it's noticeable that a delay in filling the storage vessels causes the RL-based SoC trajectory to shift towards the DP trajectory. Consequently, the RL and DP trajectories indicate coinciding plateaus. The plateaus derive from the reduction in the electrolyzer's power consumption during periods of high spot prices. In summary, the RL agent is able to comply with the targeted final SoC while shifting the operating time of the elctrolyzer towards periods of

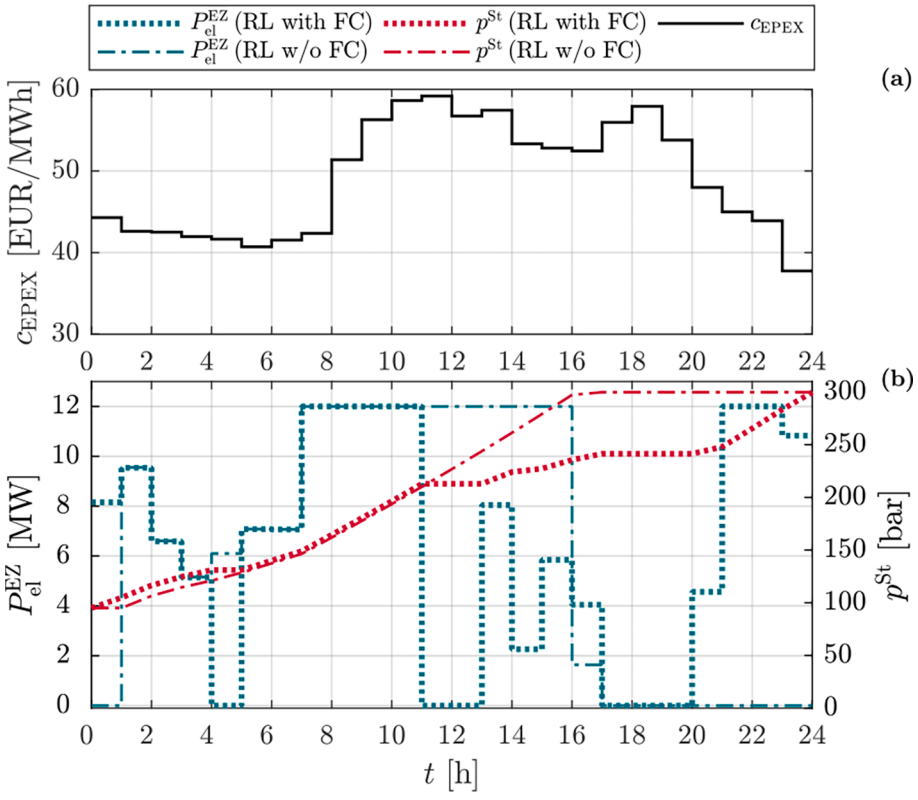

Fig. 10. RL agent's performance with and without forecasts during exemplary period A.

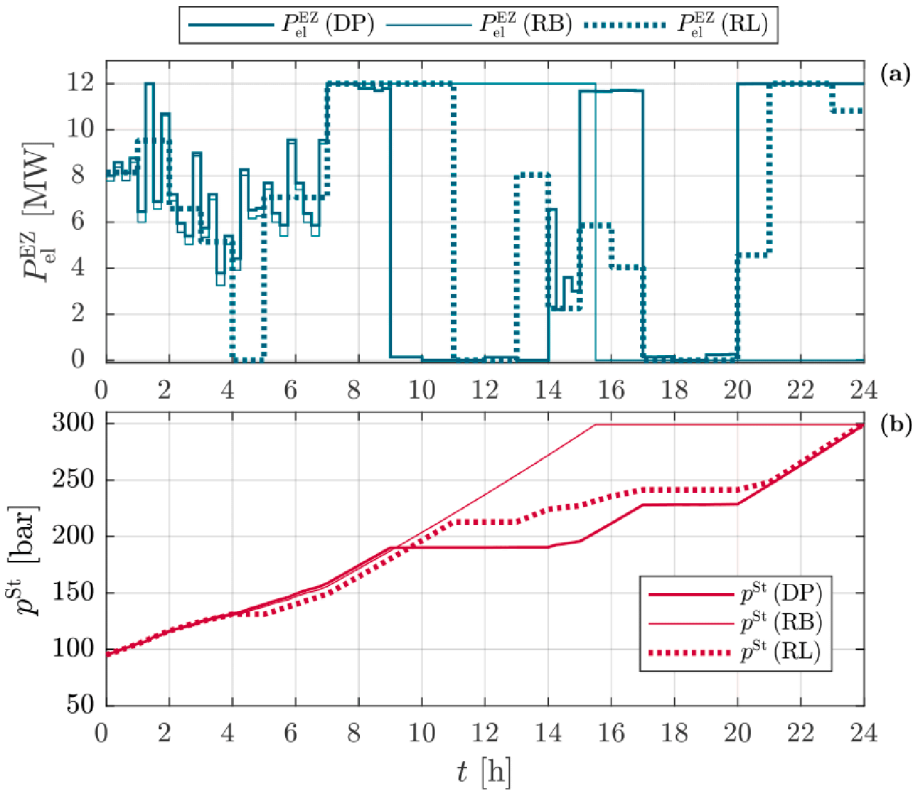

Fig. 11. Comparison between the RL-based approach, the RB unit commitment and the optimized operational management based on DP during exemplary period A.

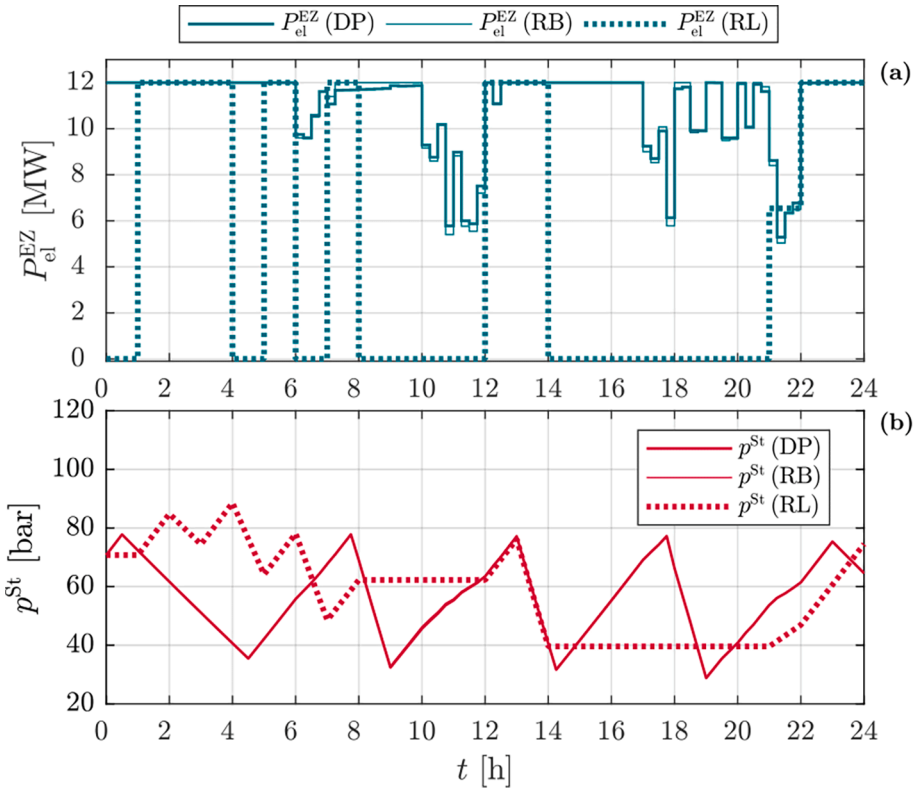

Fig. 12. Comparison between the RL-based approach, the RB unit commitment and the optimized operational management based on DP during exemplary period B.

Table 2
Numerical results of the RL-based approach, the RB unit commitment and the optimized operational management based on DP during exemplary periods A and B.

Parameter	Description	Per. A	Per. B	Unit
$m^{St}_{H_2,\tau,RB}$	Final SoC (RB)	4221	1046	[kg]
$m^{St}_{H_2,\tau,DP}$	Final SoC (DP)	4221	1046	[kg]
$m^{St}_{H_2,\tau,RL}$	Final SoC (RL)	4228	1200	[kg]
$m^{GT}_{H_2,\tau,RB}$	GT H_2 consumption (RB)	0	4862	[kg]
$m^{GT}_{H_2,\tau,DP}$	GT H_2 consumption (DP)	0	4862	[kg]
$m^{GT}_{H_2,\tau,RL}$	GT H_2 consumption (RL)	0	2001	[kg]
$R_{feed-in,RB}$	Grid feed-in revenues (RB)	12,733	2965	[EUR]
$R_{feed-in,DP}$	Grid feed-in revenues (DP)	13,556	2972	[EUR]
$R_{feed-in,RL}$	Grid feed-in revenues (RL)	13,225	10,302	[EUR]

low spot prices, resulting in an increase in the electrolyzer's economic performance.

Fig. 12 displays the comparison between the investigated operational management approaches during the exemplary period B. In contrast to the exemplary period A, this exemplary period also includes the consumption of hydrogen by the gas turbine (compare Sec. 5.1). When the stored amount of hydrogen suffices for one hour of gas turbine operation at nominal load, the gas turbine switches to the use of hydrogen. As outlined in Sec. 3.5, the RL agent receives additional rewards to match the level of hydrogen consumption determined by the RB unit commitment. Fig. 12 indicates no significant differences between the optimized operational management based on the DP and the RB approach (compare Sec. 5.2).

Consequently, the RL-based optimized operational management should foresee a similar operational strategy for the electrolyzer to reach both the predefined level of hydrogen consumption and the targeted final SoC. In general, Fig. 12 shows promising results regarding the RL agent's ability to consider the utilization of hydrogen as fuel for the gas turbine. Until approximately 8 p.m., the RL-based operational management is comparable to the DP-based management strategy.[27] Furthermore, the RL agent is able to reach the targeted final SoC with only a slight overshoot. However, Fig. 12 also reveals that RL agent suspends the operation of the electrolyzer during prolonged periods (e.g. between 2 p.m. and 9 p.m.). As a result, a hydrogen-based operation of the gas turbine cannot be initiated and the targeted hydrogen consumption defined by the RB unit commitment is not achieved.

Fig. 12 highlights the challenge of contradicting objectives associated with the RL-based operational management. While the agent is incentivized to operate the electrolyzer in a way that results in the consumption of hydrogen by the gas turbine, the risk of not reaching the targeted SoC is considered as well. This trade-off is superimposed by the incentive to reduce the time-dependent opportunity costs resulting from the power consumption of the electrolyzer.

The visualization of results provided in Figs. 11 and 12, is underscored by the numerical results listed in Table 2. In line with Fig. 11, the results corresponding to the exemplary period A indicate that both the optimized operational management based on DP and RL are able to meet the targeted final SoC.[28] Further, Table 2 confirms that both optimization approaches lead to a reduction in opportunity costs, which is equivalent to an increase in revenues resulting from feeding surplus power into the grid. Using the grid feed-in of the RB unit commitment as reference, the RL-based operational management increases the revenues by 3.9 % while the optimized operational management based on DP increases the revenues by 6.5 %. This confirms the initial hypothesis that the RL agent's performance ranges between the RB unit commitment and the operational management based on DP.

Considering the numerical results corresponding to the exemplary period B, Table 2 highlights the previously described overshoot of the SoC target, which amounts to 14.7 %. More significantly, Table 2 quantifies the mentioned lack of hydrogen consumption by the gas turbine resulting from the RL-based operational management. In comparison to the RB unit commitment, the hydrogen consumption of the gas turbine drops by 58.9 %. On the other hand, Table 2 reveals that the revenues from the grid feed-in increase by 246.7 %. It is important to note that this increase in revenues cannot be attributed to the optimized operational management of the RL agent as the predefined target for the hydrogen consumption of the gas turbine is not met.[29] Consequently, the results of the RL-based operational management and the optimized operational management based on DP cannot be directly compared for the exemplary period B.

An additional analysis of all 109 RL testing episodes reveals that 70 offer the general potential for an optimized operational management. In the other cases, there is either no surplus power to operate the electrolyzer or insufficient hydrogen storage capacity. Of the episodes with optimization potential, the RL agent achieves a reduction in opportunity costs in 43 cases. Within these episodes, the targeted final SoC is met in 18 cases. Both the targeted final SoC and the desired hydrogen consumption by the gas turbine are met in three cases. Note that these three cases correspond to weekend days featuring a target hydrogen consumption of zero (see exemplary period A). The other 15 episodes show similar performance characteristics as exemplary period B. Within these episodes, the RL agent achieves 0% of the targeted hydrogen utilization in the worst case and 78% in the best case.

In summary, the comparative analysis shows that an equilibrium in the agent's parameterization, which achieves both the desired hydrogen utilization and the targeted final SoC in each episode, could not be found for the investigated system configuration and data set. Main hindrances are assumably the low degrees of freedom available, as well as the very different initial and target states of the episodes, which result in strongly differing training experiences for the ANN.

6. Discussion and conclusion

This paper evaluates a DRL-based approach for energy management in the context of CO_2-neutral hydrogen production and storage for an industrial CHP application. To assess the technical feasibility, an RB energy management approach was established as reference. For the evaluation of the economic performance, a DP approach, featuring an optimized day-ahead unit commitment, was developed as an upper benchmark. A comparative analysis outlines both potential advantages and weaknesses of the implemented RL approach.

The RL-based simulation results are promising in that they show notable actions of the RL agent to control the electrolyser. The challenging task of achieving a targeted end-of-day SoC while accounting for variable wind power generation, spot prices, and gas turbine operation is accomplished commendably. The introduction of forecasts for energy generation, consumption and variable feed-in-tariffs to the observation space significantly improves the quality of results and is novel in scientific literature. The most arduous task, however, is balancing multiple conicting objectives in a setting with sparse degrees of freedom. As a result, no parameterization of the reward function could be found that satisfies the targeted SoC, the consideration of variable prices, and the desired hydrogen consumption by the gas turbine. The high variability of the daily targets set by the RB approach is undoubtedly a significant impediment, since it hinders the availability of sufficient available

[27] Due to the different temporal resolutions of one hour and 15 min, the corresponding trajectories of the storage pressure level cannot be directly compared.

[28] For the exemplary period A, Tab. 2 indicates a slight (0.2 %) overproduction of hydrogen by the RL-agent that is not visible in Fig. 11.

[29] In contrast, the optimized operational management based on DP increases the revenues only by 0.2 %. This increase derives from the potential for optimization outlined in Sec. 5.2.

training experience as well as a convergence of the ANN parameter vector towards a singular optimum.

As a result, the challenges of applying RL algorithms, particularly in highly complex energy management applications, were identified in this work. Other research in the context of microgrids [36], residential applications [35], or systems with simplifying assumptions like constant prices [42] or short planning time horizons [37] have produced very favorable findings (compare also [34]). Note that the present study did not apply any simplifying assumptions. Thus, it highlights that self-learning algorithms face a significant obstacle in meeting multiple divergent objectives such as hydrogen generation and consumption while simultaneously considering the volatiltiy of market prices and renewable power generation. Conclusively, the real-world application of RL-based operational management systems for hydrogen-based energy storage applications requires additional research. The present study demonstrates potential ways to achieve divergent objectives in a complex system, particularly through a good definition of the environment, the definition of a suitable reward function and the consideration of energy and price forecasts in the algorithm's observation space. The strengths of this approach can come into focus once the indentified entry hurdles have been overcome. Strengths include advantages in dealing with uncertainty, the ability of the algorithms to improve over time and experience, and, last but not least, low computing capacity requirements once the agents have been trained.

Further research should, on the one hand, concentrate on refining the objectives with the goal of increasing the degrees of freedom and thus the agents' ability to optimize their actions. On the other hand, the promising introduction of forecasts to the observation space should be further addressed.

CRediT authorship contribution statement

Alexander Dreher: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Thomas Bexten:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Tobias Sieker:** Data curation, Validation. **Malte Lehna:** Conceptualization, Validation, Writing – original draft, Writing – review & editing. **Jonathan Schütt:** Validation, Writing – review & editing. **Christoph Scholz:** Funding acquisition, Resources, Supervision, Validation. **Manfred Wirsum:** Funding acquisition, Resources, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Hessian Ministry of Higher Education, Research, Science and the Arts through the Competence Center for Cognitive Energy Systems (K-ES) project under reference number: 511/17.001. The methods utilized in this study were partially developed within the research project "Future Municipal Energy Supply Systems" supported by the RWTH Aachen University Strategy Fund.

Appendix

Figures

The following Fig. 13 depicts an exemplary DRL training progress. The increasing mean reward in the number of iterations demonstrates a clear improvement of the policy. Furthermore, the curve displays an unambiguous convergence. Nonetheless, the algorithm achieves the highest training reward at iteration 1488. The optimal algorithm for testing is determined in an post-training evaluation.

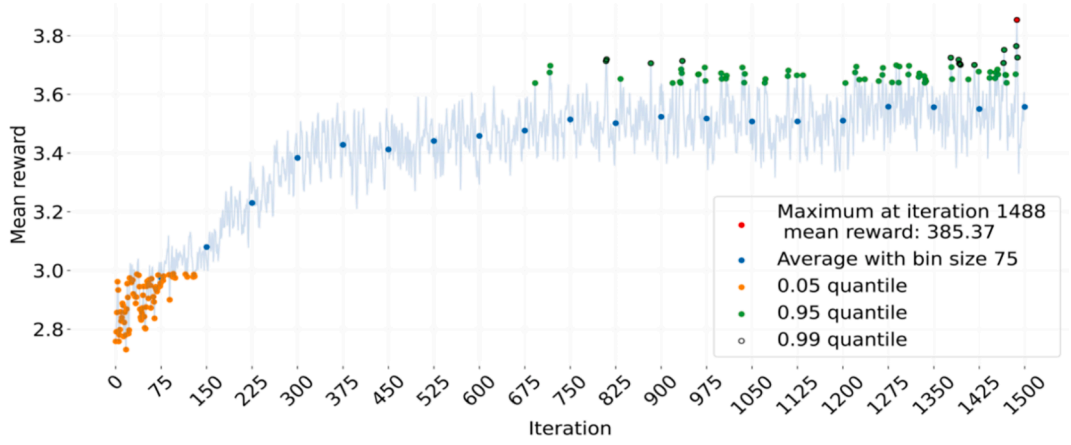

Fig. 13. Exemplary results of the training progress.

Hyperparameters RLlib (Table 3)

Table 3

Hyperparameter	Description	Value
γ	gamma	0.9998
ϵ	clipping parameter	0.3
lr	SGD learning rate	5e-05
num_sgd_iter	SGD iterations in each outer loop	10
train_batch_size	time steps collected for each SGD round	4000
entropy_coeff	coefficient of entropy regularizer	0.001

References

[1] International Energy Agency. Data and statistics. (online) https://www.iea.org/data-and-statistics, 2021. (accessed 2021/06/21).
[2] Schellnhuber HJ, Rahmstorf S, Winkelmann R. Why the right climate target was agreed in Paris. Nat Clim Change 2016;6(7):649–53.
[3] International Energy Agency. Net Zero by 2020. International Energy Agency: A Roadmap for the Global Energy Sector Report; 2021.
[4] Holttinen H. Impact of Hourly Wind Power Variations on the System Operation in the Nordic Countries. Wind Energy 2005;8(2):197–218.
[5] Stolten D, Emonts B. Hydrogen Science and Engineering, volume 1. Weinheim, Germany: Wiley-VCH Verlag; 2016.
[6] Robinius M, Otto A, Heuser P, Welder L, Syranidis K, Ryberg D, et al. Linking the Power and Transport Sectors - Part 1: The Principle of Sector Coupling. Energies 2017;10(7):956. https://doi.org/10.3390/en10070956.
[7] European Commission, 2020, A european green deal: Striving to be the first climate-neutral continent.
[8] European Commission, 2020, A hydrogen strategy for a climate-neutral europe.
[9] German Federal Ministry for Economic Affairs and Energy, 2021, Entwurf eines Gesetzes zur Umsetzung unionsrechtlicher Vorgaben und zur Regelung reiner Wasserstoffnetze im Energiewirtschaftsrecht: Energiewirtschaftsrechtsänderungsgesetz (energy industry law amendment act).
[10] Meng L, Sanseverino ER, Luna A, Dragicevic T, Vasquez JC, Guerrero JM. Microgrid supervisory controllers and energy management systems: A literature review. Renew Sustain Energy Rev 2016;60:1263–73.
[11] Zia MF, Elbouchikhi E, Benbouzid M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Appl Energy 2018;222:1033–55.
[12] Lee D, Cheng C-C. Energy savings by energy management systems: A review. Renew Sustain Energy Rev 2016;56:760–77.
[13] Shareef H, Ahmed MS, Mohamed A, Al Hassan E. Review on home energy management system considering demand responses, smart technologies, and intelligent controllers. IEEE Access 2018;6:24498–509.
[14] Rouzbahani HM, Karimipour H, Lei L. A review on virtual power plant for energy management. Sustainable Energy Technol Assess 2021;47:101370. https://doi.org/10.1016/j.seta.2021.101370.
[15] Karavas C-S, Kyriakarakos G, Arvanitis KG, Papadakis G. A multi-agent decentralized energy management system based on distributed intelligence for the design and control of autonomous polygeneration microgrids. Energy Convers Manage 2015;103:166–79.
[16] García Vera YE, Dufo-López R, Bernal-Agustín JL. Energy management in microgrids with renewable energy sources: A literature review. Appl Sci 2019;9(18):3854.
[17] Marchand, S., Richter, L., Scholz, C., Dreher, A., Lehna, M., Lenk, S., 2021, Artificial intelligence for energy supply chain automation. *in review*.
[18] Lechner C, Seume J. Stationäre Gasturbinen, volume 2. Berlin Heidelberg, Germany: Springer-Verlag; 2010.
[19] Global ETN. Hydrogen Gas Turbines. ETN Global, Brussels, Belgium: Technical report; 2020.
[20] Bexten T, Sieker T, Wirsum M. Techno-economic Analysis of a Hydrogen Production and Storage System for the On-site Fuel Supply of Hydrogen-fired Gas Turbines. J Eng Gas Turbine Power 2021;143(12):121020.
[21] Wang A, Van der Leun K, Peters D, Buseman M. European Hydrogen Backbone. Guidehouse, Utrecht, The Netherlands: Technical report; 2020.
[22] European Commission. HYdrogen as a FLEXible energy storage for a fully renewable European POWER system. (online) https://cordis.europa.eu/project/id/884229, 2021. (accessed 2021/06/22.
[23] Bexten T, Wirsum M, Roscher B, Schelenz R, Jacobs G. Model-based analysis of a combined heat and power system featuring a hydrogen-fired gas turbine with on-site hydrogen production and storage. J Eng Gas Turbines Power 2021;143(8).
[24] Xia J, Zhao P, Dai Y. Operation and simulation of hybrid wind and gas turbine power system employing wind power forecasting. American Society of Mechanical Engineers; 2012.
[25] Branchini L, Bianchi M, Cavina N, Cerofolini A, De Pascale A, Melino F. Wind-hydro-gas turbine unit commitment to guarantee firm dispatchable power. Proceedings of the ASME Turbo Expo 2014: Turbomachinery Technical Conference and Exposition, volume 3B: Oil and Gas Applications Organic Rankine Cycle Power Systems Supercritical CO2 Power Cycles Wind Energy. American Society of Mechanical Engineers; 2014.
[26] Ebaid MSY, Hammad M, Alghamdi T. Thermo-economic analysis of PV and hydrogen gas turbine hybrid power plant of 100 MW power output. Int J Hydrogen Energy 2015;40(36):12120–43.
[27] Colbertaldo, P., Guandalini, G., Crespi, E., Campanari, S.,2020, Balancing a High-Renewables Electric Grid With Hydrogen-Fuelled Combined Cycles: A Country Scale Analysis. In: Proceedings of the ASME Turbo Expo 2020: Turbomachinery Technical Conference and Exposition, volume 6: Education; Electric Power. Virtual, Online. September 21-25, 2020, page V006T09A006. ASME.
[28] Bexten T, Wirsum M, Roscher B, Schelenz R, Jacobs G, Weintraub D, et al. Optimal operation of a gas turbine cogeneration unit with energy storage for wind power system integration. J Eng Gas Turbines Power 2018;141(1).
[29] Sutton R. S., Barto, A. G., 2018, *Reinforcement learning: An introduction*. Adaptive computation and machine learning series. The MIT Press, Cambridge Massachusetts, second edition.
[30] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., 2013, Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
[31] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015;518(7540):529–33.
[32] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science (New York, N.Y.) 2018;362(6419):1140–4.
[33] Silverm D, Huang A, Maddison CJ, Guez A, Sifre L, Van den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature 2016;529(7587):484–9.
[34] Zhang D, Han X, Deng C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE J Power Energy Syst 2018;4(3):362–70.
[35] Ye Y, Qiu D, Wu X, Strbac G, Ward J. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. IEEE Trans Smart Grid 2020;11(4):3068–82.
[36] Mbuwir B, Ruelens F, Spiessens F, Deconinck G. 2017, Battery energy management in a microgrid using batch reinforcement learning. Energies 1846;10(11).
[37] Lu R, Hong SH, Yu M. Demand response for home energy management using reinforcement learning and artificial neural network. IEEE Trans Smart Grid 2019;10(6):6629–39.
[38] Huang X, Hong SH, Yu M, Ding Y, Jiang J. Demand response management for industrial facilities: A deep reinforcement learning approach. IEEE Access 2019;7:82194–205.
[39] Lin L, Guan X, Peng Y, Wang N, Maharjan S, Ohtsuki T. Deep reinforcement learning for economic dispatch of virtual power plant in internet of energy. IEEE Internet Things J 2020;7(7):6288–301.
[40] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K., 2016, Asynchronous methods for deep reinforcement learning. In: International conference on machine learning, pages 1928–1937. PMLR.
[41] Zhou S, Hu Z, Gu W, Jiang M, Chen M, Hong Q, et al. Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach. Int J Electr Power Energy Syst 2020;120:106016.
[42] François-Lavet V, Taralla D, Ernst D, Fonteneau R. 2016. European Workshop on Reinforcement Learning (EWRL: Deep reinforcement learning solutions for energy microgrids management; 2016.
[43] Tomin, N., Zhukov, A., Domyshev, A., 2019, Deep reinforcement learning for energy microgrids management considering flexible energy sources. In EPJ Web of Conferences, volume 217, page 01016. EDP Sciences.
[44] Nyong-Bassey BE, Giaouris D, Patsios C, Papadopoulou S, Papadopoulos AI, Walker S, et al. Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty. Energy 2020;193:116622.
[45] Bellman R. Dynamic programming. Science 1966;153(3731):34–7.
[46] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017, Proximal policy optimization algorithms.
[47] Kingma, D.P., Adam, J. B., 2014, A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[48] Deutscher Wetterdienst, 2020, CDC-Climate Data Center. (online) https://cdc.dwd.de. (accessed 2020/10/12).
[49] Bexten, T., Wirsum, M., Roscher, B., Schelenz, R., Jacobs, G., Weintraub, D., Jeschke, P., 2017 Techno-Economic Study of Wind Farm Forecast Error Compensation by Flexible Heat-Driven CHP Units. In Proceedings of the ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition, volume 9: Oil and Gas Applications; Supercritical CO2 Power Cycles; Wind Energy. Charlotte, North Carolina, USA. June 26-30, 2017, page V009T49A004. ASME.
[50] ENERCON GmbH, 2016, ENERCON Produktübersicht. Technical report, ENERCON GmbH, Aurich, Germany.
[51] Ruangpattana S, Klabjan D, Arinez J, Biller S. Optimization of on-site renewable energy generation for industrial sites. 2011 IEEE/PES Power Systems Conference and Exposition. IEEE. 2011.
[52] Giampieri A, Ling-Chin J, Ma Z, Smallbone A, Roskilly AP. A review of the current automotive manufacturing practice from an energy perspective. Appl Energy 2020;261:114074.
[53] SIEMENS AG. SGT-300 Industrial Gas Turbine. SIEMENS AG, Munich, Germany: Technical report; 2015.

[54] Vetter G. Leckfreie Pumpen. Verdichter und Vakuumpumpen: Vulkan-Verlag, Essen, Germany; 1998.
[55] Bludszuweit H, Dominguez-Navarro J-A, Llombart A. Statistical analysis of wind power forecast error. IEEE Trans Power Syst 2008;23(3):983–91.
[56] Icha P, Kuhs G. Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990–2019. Umweltbundesamt, Dessau-Roßlau, Germany: Technical report; 2020.
[57] Sundstrom O, Ambühl D, Guzzella L. On implementation of dynamic programming for optimal control problems with final state constraints. Oil & Gas Science and Technology Revue de l'Institut Français du Pétrole 2009;65(1):91–102.
[58] Sundstrom O, Guzzella L. A generic dynamic programming matlab function. 2009 IEEE International Conference on Control Applications. IEEE. 2009.
[59] Bexten T, Jörg S, Petersen N, Wirsum M, Liu P, Li Z. Model-based thermodynamic analysis of a hydrogen-fired gas turbine with external exhaust gas recirculation. J Eng Gas Turbines Power 2021;143(8).
[60] Kuznetsova E, Li Y-F, Ruiz C, Zio E, Ault G, Bell K. Reinforcement learning for microgrid energy management. Energy 2013;59:133–46.
[61] Buttler A, Spliethoff H. Current status of water electrolysis for energy storage, grid balancing and sector coupling via power-to-gas and power-to liquids: A review. Renewable Sustain Energy Rev 2018;82.
[62] Preuster P, Alekseev A, Wasserscheid P. Hydrogen Storage Technologies for Future Energy Systems. Ann Rev Chem Biomol Eng 2017;8(1).
[63] Funke HH-W, Beckmann N, Keinz J, Horikawa A. 30 years of dry-low-NOx micromix combustor research for hydrogen-rich fuels -an overview of past and present activities. J Eng Gas Turbines Power 2021;143(7).