POLITECNICO DI TORINO
Repository ISTITUZIONALE

A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings

Original
A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings / Brandi, S.; Gallo, A.; Capozzoli, A.. - In: ENERGY REPORTS. - ISSN 2352-4847. - 8:(2022), pp. 1550-1567. [10.1016/j.egyr.2021.12.058]

Availability:
This version is available at: 11583/2955660 since: 2022-02-17T18:42:06Z

Publisher:
Elsevier Ltd

Published
DOI:10.1016/j.egyr.2021.12.058

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

30 December 2025

Energy Reports 8 (2022) 1550–1567

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings

Silvio Brandi, Antonio Gallo, Alfonso Capozzoli *

Dipartimento Energia "Galileo Ferraris", Politecnico di Torino, TEBE Research Group, BAEDA lab, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Article history:
Received 24 September 2021
Received in revised form 8 December 2021
Accepted 18 December 2021
Available online xxxx

Keywords:
Deep reinforcement learning
Integrated energy systems in buildings
Battery storage
Thermal storage
Energy management

ABSTRACT

The management of integrated energy systems in buildings is a challenging task that classical control approaches usually fail to address. The present paper analyzes the effect of the implementation of a reinforcement learning-based control strategy in an office building characterized by integrated energy systems with on-site electricity generation and storage technologies. The objective of the proposed controller is to minimize the operational cost to meet the cooling demand exploiting thermal energy storage and battery system considering a time-of-use electricity price schedule and local PV production. Two control solutions, a Soft-Actor-Critic agent coupled with a rule-based controller, and a fully rule-based control strategy, used as a baseline, are tested and compared considering various configurations of battery energy storage system capacities, and thermal energy storage sizes. Results show that the proposed control strategy leads to a reduction of operational energy costs respect to the fully rule-based control ranging from 39.5% and 84.3% among different configurations. Moreover the advanced control strategy improves the on-site PV utilization leading to an average increasing of self-sufficiency and self-consumption of 40% among different scenarios. The baseline control strategy results more sensitive to the size of storage whereas the proposed control achieves high savings also when smaller capacities of battery energy storage systems and sizes of thermal energy storage are implemented. The outcomes of the work prove the impact of implementation of advanced control as a way to optimize energy costs with a comprehensive view of the whole integrated energy system considering both thermal and electrical energy storage operation.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Building sector accounts for 40% of global energy consumption playing a pivotal role in the energy transition and global warming mitigation processes (IEA, 2019). Renewable Energy Sources (RES), especially photovoltaic (PV) systems, have been widespread adopted and promoted to sustain the growing energy demand in buildings (Martinopoulos et al., 2018). Due to weather-dependent nature of RES, electrical storage solutions have been introduced to increase Self-Consumption (SC) of on-site renewable energy production providing benefits to both end-users and grid operators (Baniasadi et al., 2020).

However, Battery Energy Storage System (BESS) (e.g., Lead-acid and Li-ion batteries) are characterized by high investment cost making their adoption unfeasible for many applications (Shabani and Mahmoudimehr, 2018) if incentives provided by policymakers are not foreseen (Koskela et al., 2019). Nevertheless, BESS improves PV energy utilization, by addressing the problem of the solar energy flexibility. On the other hand, Thermal Energy Storage (TES) proved to be a sustainable solution making the Heating, Ventilation and Air Conditioning (HVAC) systems more flexible to time-varying electricity prices improving capabilities of the system to shift its demand patterns (Das et al., 2018; Terlouw et al., 2019).

Considering the dependency of BESS capacity and operation strategy from building electrical demand patterns, coupling BESS with TES can lead to several economical and environmental advantages compared to the implementation of only BESS (Baniasadi et al., 2020).

In this context, the identification of optimal management strategies capable to increase the profitability of storage systems is a key aspect to address. The optimal operation of storage systems in buildings with Integrated Energy Systems (IES) is affected by exogenous factors such as weather, energy demand patterns and electricity prices which all vary over time. Classical control strategies are usually not able to consider trade-offs between multiple and contrasting objectives, such as thermal comfort, energy consumption, energy flexibility and Self-Sufficiency (SS), and are not capable to adapt to an evolving system characterized

* Corresponding author.
E-mail addresses: silvio.brandi@polito.it (S. Brandi), antonio.gallo@polito.it (A. Gallo), alfonso.capozzoli@polito.it (A. Capozzoli).

https://doi.org/10.1016/j.egyr.2021.12.058
2352-4847/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Nomenclature

Δt	Control sampling time (h)
η	Efficiency of photo-voltaic module ()
η_{rte}	Round-Trip efficiency of battery ()
$\hat{U}_{op}$	Global heat exchange coefficient of opaque envelope (W/m^2 K)
$\hat{U}_{tr}$	Global heat exchange coefficient of transparent envelope (W/m^2 K)
C_{buy}	Electricity buying price (€/kWh)
C_B	Nominal capacity of battery (kWh)
C_{sell}	Electricity selling price (€/kWh)
E_{dem}	Building electrical energy demand (kWh)
$E_{grid,buy}$	Electricity bought from the grid (kWh)
$E_{grid,sell}$	Electricity sold to the grid (kWh)
$Grid_{frac}$	Fraction of electricity supplied by grid ()
$P_{B,ch,max}$	Maximum battery charging power (kW)
$P_{B,ch}$	Battery charging power (kW)
$P_{B,dis,max}$	Maximum battery discharging power (kW)
$P_{B,dis}$	Battery discharging power (kW)
P_{dem}	Building electrical load (kW)
P_{PV}	Photo-voltaic power generation (kW)
Q_{cap}	Capacity of chiller (kW)
Q_d	Building heat demand (kW)
SOC_B	State-Of-Charge of the battery ()
SOC_T	State-Of-Charge of the water storage ()
$SOC_{B,max}$	Maximum State-Of-Charge of the battery ()
$SOC_{B,min}$	Minimum State-Of-Charge of the battery ()
T_o	Outdoor air temperature (°C)
T_s	Storage temperature (°C)
T_{ch}	Chiller supply temperature (°C)
$T_{s,max}$	Storage temperature upper boundary (°C)
$T_{s,min}$	Storage temperature lower boundary (°C)

Acronyms

HVAC	Heating, Ventilation and Air Conditioning
RES	Renewable Energy Sources
RBC	Rule-Based Control
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
SAC	Soft-Actor-Critic
DQN	Deep Q-Network
MDP	Markov Decision Process
POMDP	Partially Observable Markov Decision Process
DNN	Deep Neural Networks
DC	Direct Current
AC	Alternate Current
TES	Thermal Energy Storage
BESS	Battery Energy Storage System
SOC	State-Of-Charge
TOU	Time-Of-Use
SS	Self-Sufficiency
SC	Self-Consumption

by dynamic boundary conditions, including grid requirements, and constraints (May, 2019; Finck et al., 2018).

To overcome these limitations, researchers worldwide have recently focused their efforts in the development and implementation of advanced control strategies to improve the management of IES in buildings based on predictive architectures or optimization processes. Comodi et al. (2016) assessed the viability of introducing a Cold Thermal Energy Storages (CTES) for demand side management strategies into an existing cooling system of an institutional building under a Time of Use (ToU) pricing scheme. The storage was charged during night time to exploit higher chiller Coefficient of Performance (COP) and lower electricity price. It was demonstrated that a CTES could increase the overall energy efficiency and decrease the energy cost by being charged during off-peak hours with a payback period between 8.9 and 16 years. Arteconi et al. (2015) analyzed a factory building equipped with Heat Pump (HP) and TES. The TES was charged during low price periods to cover the whole cooling demand during occupancy periods. This strategy was able to save about 54% of the electricity cost related to the cooling process. Ioli et al. (2015) proposed a novel convex constrained optimization to optimize the operational cost of cooling system coupled with a TES into a single zone office building by controlling the storage operation and zone temperature. The proposed approach achieved 14.8% cost saving and 6.5% energy saving with respect to the strategy where zone temperature is fixed. Other strategies have been developed recently, as in Ren et al. (2021) that analyzed an IES with an HVAC assisted by a photovoltaic thermal hybrid collector and a TES. The results showed that using the PV panels to power the heat pump to charge the TES provided additional energy flexibility respect to the use of only Demand Side Management (DSM) strategies. Comodi et al. (2015) managed the integration of electrical and thermal storage into a nearly Zero Energy Building (nZEB). Thermal flows were optimized by a Mixed-Integer Linear Programming (MILP) algorithm to reduce the grid exchange electricity, while BESS were managed by a Rule-Based Control (RBC). In that way, the building achieved an SS level of 100% even though the cost of electrical storage did not justify the investment. A fuzzy rule control logic was developed by Dimitroulis and Alamaniotis (2022) for the charging scheduling of a BESS within an IES with renewable generation and Electric Vehicle (EV). The results showed a reduction of the monthly bill as compared to the linear optimization approach, and to an RBC. Biyik and Kahraman (2019) proposed a Model Predictive Control (MPC) for an IES with HVAC system, renewable generation and BESS to reduce the peak load. The controller provided an average reduction of 23% in peak electrical demand compared to a baseline where indoor temperature is kept fixed. Predictive management for energy supply networks using PV, HP and battery units was developed by Wakui et al. (2019) by combining two-stage stochastic schedule programming and RBC to reduce operating cost. The proposed approach performed better than the management based on the deterministic schedule planning and the rule-based management without schedule planning. Recently, Reinforcement Learning (RL) has gained popularity with the promise to revolutionize building control applications (Wang and Hong, 2020). In opposition to well-established model based approaches (i.e., model predictive control), in which a model of the system is embedded within the controller (Serale et al., 2018;

Tarragona et al., 2020), RL follows a model-free approach where an agent directly learns the optimal control policy by interacting with the system through a trial-and-error approach (Sutton and Barto, 2018). In the field of building energy management, RL was successfully applied to control TES systems (Liu and Henze, 2007; Henze and Schoenmann, 2003), thermostat set-points (Barrett and Linder, 2015), lighting devices (Park et al., 2019) and BESS (Abedi et al., 2022).

Deep Reinforcement Learning (DRL) couples RL framework with the feature extraction capabilities of Deep Neural Networks (DNN), enhancing the capability to solve complex control problems (Mnih et al., 2015). With regard to building control, DRL was applied for the management of supply water temperature set-point control (Brandi et al., 2020; Coraci et al., 2021), fan regulation (Chu et al., 2021; Valladares et al., 2019), indoor temperature set-point control (Du et al., 2021; Gao et al., 2019), TES systems both at single (Vázquez-Canteli et al., 2019; Wang et al., 2016) and multiple building level (Pinto et al., 2021).

Comprehensive review works recently published can support the interested reader in discovering details on the potentialities of DRL applied to building energy management challenges. In Vázquez-Canteli and Nagy (2019) was analyzed the application of DRL to support Demand Response (DR) policies in buildings highlighting the lack of real-world case studies in the current scientific literature. Han et al. (2019) focused on the application of DRL for controlling occupant comfort identifying the necessity to introduce human-feedback within the control loop. A DRL agent was used by Sanaye and Sarrafi (2021) to control the operation of the Combined Heat and Power (CHP) generation unit and the gas-fired boiler in an hybrid system with PV panels, solar collectors, wind turbines, a hot water storage tank and batteries. This control strategy reduced the operational cost of a residential complex with respect to two different RBC strategies. Anvari-Moghaddam et al. (2017) proposed an energy management based on a multi-agent system for IES in a microgrid to reduce operational cost and to ensure user's needs. Particularly, a Bayesian Reinforcement Learning (BRL) control was used for the battery operation, which was coordinated with other agents in charge of collecting and sharing information, making predictions of renewable generation and providing computation services. Another work focused on the use of a Q-Learning algorithm to reduce the bill of a smart-building with PV generation, electrical storage and EV station (Kim and Lim, 2018). Wang and Hong (2020) discussed significant findings related to utilization of prediction of external disturbances, the definition of the control actions and the lack of real implementation of DRL strategy in buildings. Moreover, from the same work (Wang and Hong, 2020) emerged the lack of studies in which both thermal and electrical storage are coordinated simultaneously with an advanced control strategy. In their analysis, Kathirgamanathan et al. (2021) emphasized the lack of studies on data-driven controllers considering multiple source of energy flexibility in buildings including BESS, TES and local renewable generation. Considering these aspects, the present work analyzes the application of a control strategy based on a DRL agent coupled with a RBC to optimally manage the IES of an office building characterized by on-site PV generation and the simultaneous presence of BESS and TES equipment. The works intends to highlight the impact of advanced control strategies on the storage operation in integrated energy systems also considering different configurations of BESS capacities and TES sizes.

1.1. Contribution and structure of the work

The management of storage systems is a key factor to consider in buildings with IES to enhance energy flexibility and reduce operational costs. Traditional controls may behave sub-optimally due to their lack of adaptability and their reactive approach. The design and implementation of storage solutions, including BESS and TES, is usually performed by different actors which are also responsible of the definition of their control logic. Failure to consider proper control strategies in the design stage may result in oversized storage systems and consequently higher investment costs (Liu et al., 2020; Sharma et al., 2019).

The introduction of advanced control strategies based on a predictive and adaptive approach can enable a better management of multiple storage technologies in buildings. These controllers, thanks to their predictive and adaptive nature, can increase the effectiveness of storage equipment during building operation making competitive also solutions characterized by relatively low sizes and capacities. Accounting for the effect of advanced control strategies in the design stage can limit investment costs by adopting storage solutions that otherwise could be considered not suitable. For instance, in Medved et al. (2021) the authors highlighted that most approaches to storage system sizing do not take into account storage daily performance which could contribute to determine appropriate sizes and capacities of storage equipment.

With this in mind, the present paper aims to analyze the performance of a DRL strategy coupled with a RBC against a fully RBC to manage the operation of a chiller system coupled with a cold-water storage tank for an office building with on-site electricity generation and battery system. The analysis was carried out for multiple configurations of the energy systems including different sizes of TES and different capacities of BESS. A state-of-the-art DRL algorithm, namely Soft-Actor-Critic (SAC), was implemented. SAC proved to be a promising solution in building energy management applications thanks to its innovative formulation and rapid convergence time (Biemann et al., 2021). While most of papers make use of Deep Q-Network (DQN) framework to handle discrete action spaces, this paper exploits a novel formulation of the SAC algorithm recently introduced in Christodoulou (2019) capable to deal with discrete action settings. The analysis was carried out in a simulation environment which employs EnergyPlus (Crawley et al., 2000) as dynamic simulation software coupled with Python.

On the basis of the reasoning presented in this section, the main contributions of the present paper can be summarized as follows:

- Demonstrate the energy and cost benefits of adopting advanced DRL-based control strategies in IES characterized by BESS and TES equipment over classical RBC approaches.
- Evaluate the flexibility potential and the storage management which can be achieved with the adoption of advanced control strategies integrating a comprehensive management of the whole IES.
- Analyze the effectiveness of advanced control strategies with the variation of sizes of TES and capacities of BESS equipment highlighting the impact of the control also on the selection of storage in buildings.
- Adopting a novel formulation of the SAC algorithm specifically designed for discrete control actions differently from the commonly implemented DQN framework.

The paper is organized as follows: Section 2 introduces the case study and the control problem, Section 3 describes the methodological framework and provides information about DRL control, Section 4 reports implementation details of the different control strategies and configuration of the energy system. Section 5 reports the results obtained while the last two sections include discussion and conclusions.

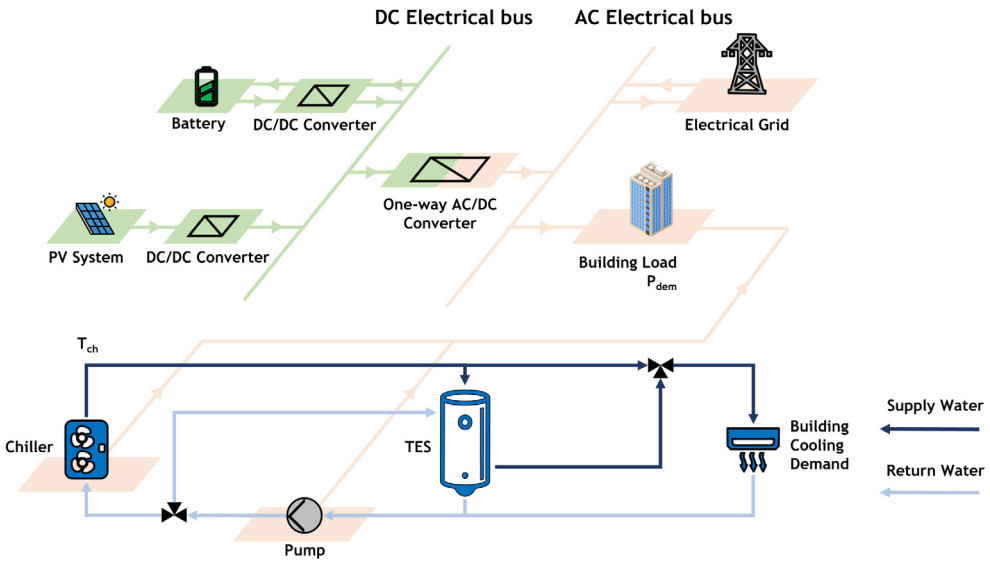

Fig. 1. Schematics of the electrical and cooling systems of the analyzed case study.

2. Formulation of the control problem

In the present work, the effect of the adoption of advanced control strategies on the operation of IES in buildings considering different configurations of storage was evaluated for an office building located in Turin, Italy. The building is equipped with a TES system (i.e., a cold water storage tank) that is operated as a buffer between the building and an air-to-water chiller. The IES also includes a mono-crystalline silicon PV module and a lithium-ion electrical battery (i.e. BESS). Further, technical specifications of the components is provided in Section 4.

Fig. 1 shows a simplified schema of the electrical and cooling systems of the analyzed case study. The building electrical load (P_{dem}) is determined by the electrical demand of the chiller and circulation pump. The electrical system is formed by a Direct Current (DC) bus and Alternate Current (AC) bus interfaced by a mono directional AC/DC inverter. On the DC bus a PV system and a BESS are installed. The PV and the battery are connected to the DC bus by a DC/DC converter. Grid is not allowed to charge the BESS according to the normative of many European Countries, but it is used to assist in matching electricity demand of the building and renewable power generation at each time-step (Cui et al., 2017). At each step if local RES production is not zero the PV injects energy into the system according to the following priority: (i) building, (ii) BESS, (iii) grid.

The electric chiller supplies cold water at constant set-point value (T_{ch}). The thermal storage can be operated in a temperature range between $T_{s,min}$ and $T_{s,max}$. The thermostatic control of the building was not considered in this application as the building cooling demand is considered as an external disturbance of the system along with weather conditions and electricity prices. To this purpose building cooling demand is evaluated in advance to maintain fixed conditions of indoor air temperature and relative humidity given the influence of weather and occupancy schedules.

The aim of the controller is to minimize the electricity cost of the chiller and circulation pump by managing three different cooling operation modes and BESS operation at each time step.

The three different cooling operation modes showed in Fig. 2 are (i) charging mode, where cooling energy is provided to both storage tank and building (if requested) simultaneously, (ii) discharging mode, where cooling energy is provided to the building to meet the demand only through the storage and (iii) chiller cooling mode, where cooling energy is provided to the building exclusively through the electric chiller.

Discharging mode and chiller cooling mode were introduced considering that the system configuration was not conceived to provide cooling to the building via two separate sources at the same time. However the two modes were introduced to allow the control agent to select during building operation at each control step the one that is optimal according to boundary conditions (i.e. employ chiller also during high price periods due to high PV production).

The proposed control strategy couples DRL to manage the cooling system operation with RBC which is employed to manage the BESS. Conversely, the baseline employs a fully RBC strategy to manage both BESS and cooling system operation. The case study was designed to assess the effect of adopting advanced control strategies also considering the performance for different sizes and capacities of TES and BESS, respectively.

3. Methodology

This section describes the methodological steps and the main methods adopted in the present work. The case study introduced in Section 2 was used as a testbed to assess the effectiveness of an advanced control strategy consisting of a DRL coupled with a RBC for an office building with IES.

A DRL control agent was developed and trained in order to identify the optimal control policy for the management of the cooling modes. The performance of the proposed control strategy was evaluated against a baseline consisting of a fully RBC for different configurations of storage systems. A different DRL control agent was trained for each configuration resulting from the combination of BESS capacity and TES size.

3.1. Design of baseline and proposed control strategies

As introduced in the previous section, the baseline controller employs a fully RBC approach. This strategy was conceived to simulate the performance of classical control approaches applied to manage TES and BESS as two distinct systems. Two RBC strategies were separately designed to control the cooling operation modes and BESS without sharing mutual information between the two. This hypothesis was deemed legitimate since

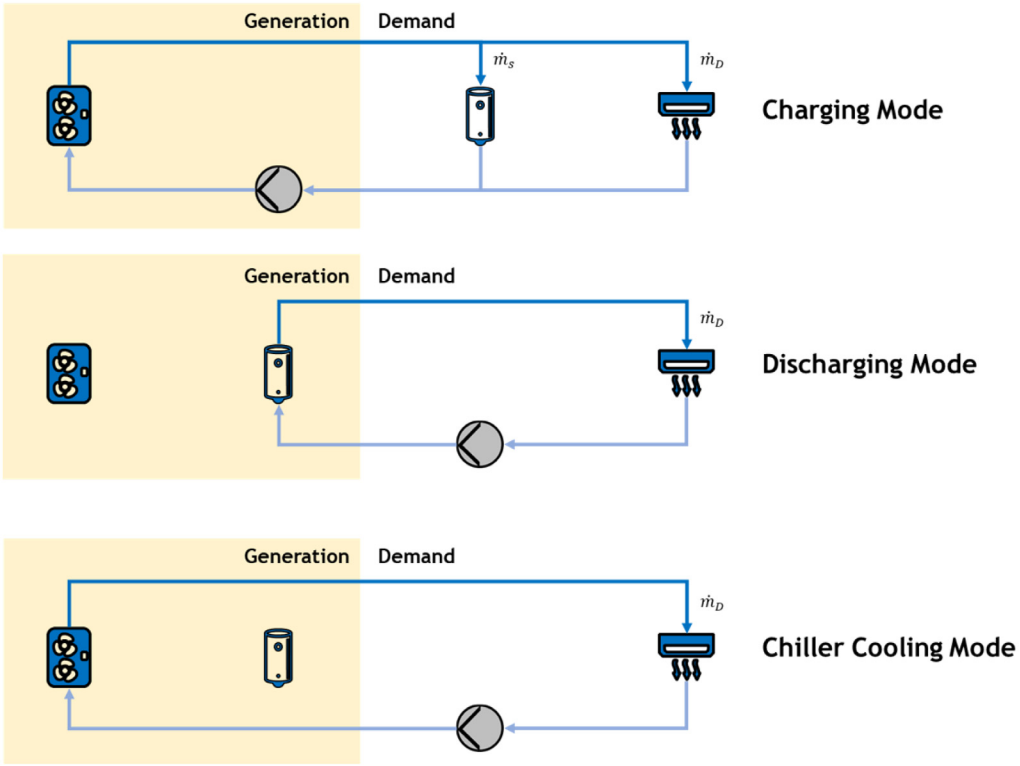

Fig. 2. Schematics of the three different modes of the cooling system analyzed.

BESS and TES equipment are usually implemented by different stakeholders.

On the other hand, the proposed controller employs an approach where DRL control agent was coupled with an RBC strategy. The BESS system was managed by the same RBC strategy employed by the baseline controller. Conversely, the management of cooling operation modes which involves TES was implemented through an advanced DRL controller which exploits also information on PV production and BESS status. The reason behind the choice to couple SAC with an RBC controller is that this latter strategy is very effective in managing BESS considering building demand, electricity price, and PV production (Amato et al., 2021; Ruusu et al., 2019). However, the management of cooling modes requires an advanced controller capable of considering also the boundary conditions determined by the PV system and the BESS in selecting the optimal action. Thanks to this approach, the proposed controller operates with a comprehensive perspective of the whole energy system.

3.2. Deep reinforcement learning control

In the RL framework, a control agent learns the optimal control policy by interacting with the controlled environment through a trial-and-error process. RL can be mathematically formulated as a Markov Decision Process (MDP), which is normally characterized by a 4-values tuple, including: *state*, *action*, *transition probabilities* and *reward function*. The state is a mathematical representation of the controlled environment, including the set of features that a RL agent receives in order to determine a control action, which is defined as *observation*. If the observation is a subset of the state, this results in a Partially Observable Markov Decision Process (POMDP). The action corresponds to the control signal that the agent has identified as the best to be applied to the system. The transition probabilities defines the probability that the environment has to move from a certain state (s) to another (s'), if a defined action a is applied to the system. The reward function measures the control performance of the agent in achieving the desired objectives.

The final goal of an RL control agent is to efficiently learn the optimal control policy (π). The control policy maps the relationships between state of the environment and the control action to maximize the cumulative sum of future rewards (Sutton and Barto, 2018).

There are two methods in the RL framework to identify the optimal control policy: *value-based* and *policy-based*. Value-based methods aim at learning the value function, which estimates the effect and benefit of taking a specific action a starting from state s. Policy-based methods do not employ the value function as a proxy, but attempt to learn directly the optimal control policy π. In general, value-based methods are more simple and efficient, while policy-based methods have better convergence properties and are capable to handle highly stochastic continuous problems.

Another characterizing aspect of RL algorithms is the policy method, which can be divided between *on-policy* and *off-policy* methods. On-policy RL algorithms attempt to directly improve the policy that is used by the agent to generate decisions. Off-policy methods evaluate a policy that is different from the one used to select actions (Sutton and Barto, 2018).

In this work a modified version of SAC was implemented (Christodoulou, 2019). SAC is a state-of-the-art off-policy DRL algorithm which showed excellent performance in solving several control tasks (Haarnoja et al., 2019). SAC was originally developed for continuous action spaces, thus, a modified version, derived for discrete action settings, was implemented to solve the proposed discrete control problem. The discrete SAC algorithm was chosen over the DQN algorithm, which is normally used for discrete control problem. The reason lies in the lower dependence on the hyperparameters tuning of discrete SAC algorithm while achieving state-of-art performance in terms of sample efficiency. Fig. 3 describes the discrete SAC algorithm implemented in this work.

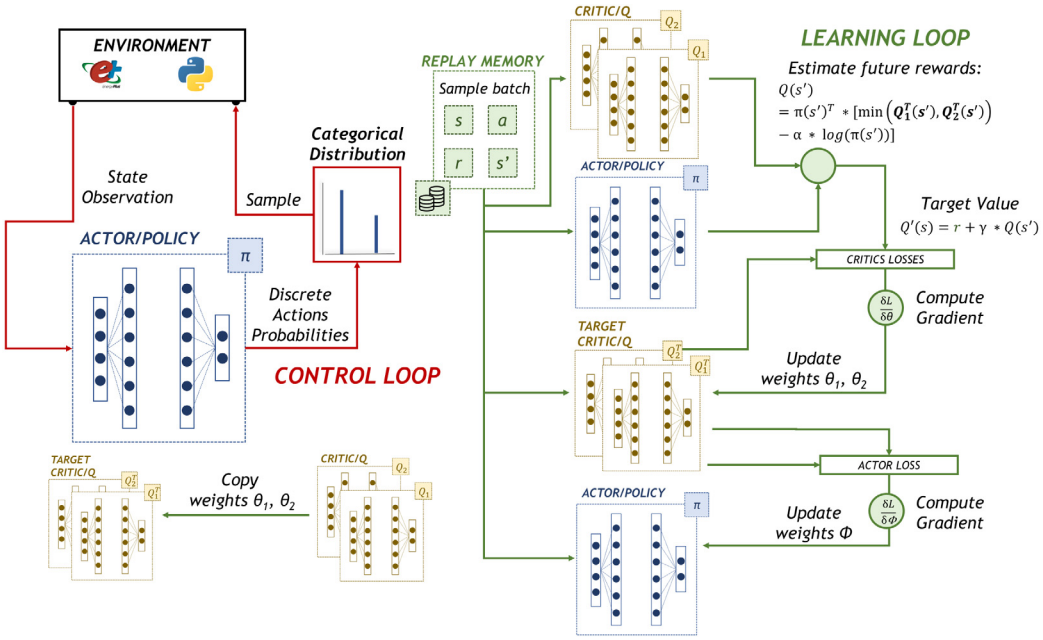

Fig. 3. Discrete SAC schema.

The figure shows that the Actor-Critic architecture employs two function approximators. The Actor has the aim to determine the optimal action for a given specific state of the controlled environment (policy-based), while the Critic evaluates the decisions made by the actor (value-based). The actor and the critic are parametrized as DNN. The actor is employed in both the control loop and learning loop while the critic is employed only during learning. This framework is generally coupled with an off-policy implementation, enabling the re-utilization of the previous experience collected by the agent in order to improve the control policy (i.e. replay memory). Moreover, the SAC policy is trained to maximize the expected sum of future rewards and the expected entropy of the policy at the same time, as defined in Eq. (1):

$$\pi^* = argmax_{\pi_\phi} E[\sum_{t=0}^{\infty} \gamma^t (r_t + \alpha H_t^\pi)] \tag{1}$$

where H_t^π is the Shannon entropy term, which is a constant term which associates to each state a probability distribution over the possible actions. Through this approach, the agent has the possibility to explore during the training phase, while, during the deployment phase, the mean value of the distribution is used to select deterministic actions, ensuring a robust control policy. The term α is the entropy regularization coefficient which indicates the relative importance of the entropy term with respect to reward term. γ represents the discount factor for future rewards and r_t is the reward obtained by the agent at the timestep t.

In the modified version of the SAC algorithm the critic network, also called soft-Q network, outputs directly the Q-value of each possible action. The parameters of the critic network are updated in order to minimize the error J_Q expressed as follows:

$$J_Q(\theta) = E_{(s_t,a_t)\sim D}[\frac{1}{2}(Q_\theta(s_t,a_t) - (r(s_t,a_t) + \gamma E_{s_{t+1}\sim p(s_t,a_t)}[V_{\bar{\theta}}(s_{t+1})]))^2] \tag{2}$$

where D is the replay buffer and $V_{\bar{\theta}} s_{t+1}$ is estimated by means of a target network. In practice two different critic networks are employed and the minimum of their two outputs is employed to compute the above objective. The actor network, also called policy network, directly outputs the action probabilities. The losses employed to update the policy network are calculated according to the following formula:

$$J_\pi(\phi) = E_{s_t\sim D}[\pi_t(s_t)^T [\alpha log(\pi_\phi(s_t)) - Q_\theta(s_t)]] \tag{3}$$

In the following sub-sections the design of the action-space, of the reward function and of the state space are introduced along with the training strategy employed.

3.2.1. Action-space design

The control action determines the operation mode of the cooling system at each time step. Since three operation modes were defined for the proposed case study, the action-space was designed as a discrete space as follows:

$$A_{(t)} = [0, 1, 2] \tag{4}$$

where 0 correspond to discharging mode, 1 to chiller cooling mode and 2 to charging mode as described in Section 2.

3.2.2. Reward function design

The reward measures the performance of the controller after selecting an action at each time step. The controller operates with the aim to minimize the energy cost related to the energy exchanged between the electrical grid and the system (E_{grid}). Electrical energy can be imported from the grid when there is no PV power generation and the BESS system is out of charge. Electrical energy is injected to the grid when PV power generation excesses the building electrical demand and the BESS system is fully charged. The electrical energy exchanged with the grid was defined as negative when it is imported from the grid and positive when injected. The reward function was defined as follows:

$$\begin{cases} r(t) = \beta E_{grid}(t) \cdot C_{buy}(t) & if\ E_{grid}(t) < 0 \\ r(t) = \beta E_{grid}(t) \cdot C_{sell}(t) & if\ E_{grid}(t) > 0 \end{cases} \tag{5}$$

where $C_{buy}(t)$ and $C_{sell}(t)$ are defined according to the schedule price for buying and selling electricity and β is a factor introduced to weight the magnitude of the reward, namely reward scale, and it is considered an hyperparameter of the algorithm.

3.2.3. State-space design

The state-space includes all the variables employed by the SAC control agent to determine at each time step the optimal control action capable to maximize the stream of future rewards. Moreover, the state-space may include information relative to historical values of the variables describing the behavior of the system and future values of external disturbances. In this work, information about historical values were introduced to account for slow-responsive thermal dynamics of the components of the controlled system (Wang and Hong, 2020). At the same time, future values of external disturbances were introduced since they can provide crucial information that the agent can leverage to optimally solve the control problem. In the present paper perfect predictions of external disturbance were employed.

More detailed information on the variables included within the state-space are provided in Section 4.

3.2.4. DRL training

The control policy of the SAC agent was trained on a model of the proposed case study described in Section 2. During the training process a specific period called episode was presented multiple times to the control agent in order to gradually improve its control policy by enabling the exploration of different trajectories. At the end of this process the trained agent was statically deployed on the same episode in order to evaluate its control performance. The static deployment of a SAC agent was achieved by stopping the update of the parameters determining the control policy and employing the actor network to select the optimal control actions given the state of the environment.

3.3. Design of BESS and TES configurations

Different configurations consisting in the combination of various volumes of the cold-water storage tank and nominal capacities of the BESS were investigated. The aim is to find out how the proposed advanced control strategy can improve the performance with respect to a classical control strategy while implementing storage equipment with various sizes and capacities. The objective is to evaluate if the introduction of advanced control strategies could support the introduction of equipment characterized by lower sizes and capacities. Thus, reducing the initial investment cost which is decisive to guarantee the spread of the storage technologies.

More detailed information on the configurations of BESS and TES are provided in Section 4.

4. Implementation

The test facility analyzed in this work consists of two study rooms, one control room and a technical room. The technical room is not served by the air-conditioning system and the storage tank is placed within it.

The facility is a prefabricated building with a rectangular layout. The floor area is 196.3 m^2 (11.25 × 17.45 m). The interior gross floor conditioned area is around 96.8 m^2. The ceiling height is 2.8 m at the minimum and 3.7 m at the maximum above the floor level, due to the different tilt angles of the roof, which are 13.4° on SE side and 15° on NW side. The features of the building envelope are reported in Table 1.

The chiller has a reference capacity Q_{cap} of 12 kW and reference COP of 2.67. The reference COP is provided by the employed chiller model provided by EnergyPlus and it is calculated considering a reference leaving chilled water temperature of 6.67 °C and a reference entering condenser fluid temperature of 35 °C. The design water mass flow rate during charging phase ($\dot{m}_S$) is 0.2 kg/s while during discharging phase ($\dot{m}_S$) is 0.35 kg/s. This latter value corresponds to the sum of the design mass flow rates of the three air-conditioned zones. The supply water temperature at the outlet of the chiller was set equal to 7 °C. The TES operates in the range between 10 °C and 18 °C which correspond to a state-of-charge (SOC_T) of 1 and 0, respectively.

Table 1
Features of the building envelope.

Feature	Value
Conditioned floor area	96.8 m^2
Conditioned volume	501 m^3
Envelope surface/conditioned volume ratio	0.85 m^{-1}
Transparent/opaque envelope surface ratio	6.6%
Opaque envelope surface	400 m^2
$\hat{U}_{op}$	0.16 W/m^2 K
$\hat{U}_{tr}$	0.55 W/m^2 K

The HVAC system serving the building can meet the cooling demand through the electric chiller or the TES. The building cooling demand was considered as an external disturbance of the system and was calculated through EnergyPlus considering an indoor air temperature of 26 °C and a relative humidity of 55% during occupancy periods which occur between 09:00 and 18:00 from Monday to Friday. During these periods, the zones were supposed to be occupied at their maximum capacity (i.e. 3 people for the control room and 10 people for the two study rooms). No regular occupancy was expected for the technical room. The air infiltration rate was set to 0.15 h^{-1}, a typical value for office buildings. The air ventilation rate for the control room and the study rooms was set to 10 L/s per person resulting in 30 L/s and 100 L/s, respectively.

The price of the electrical energy drawn from the grid to operate the chiller unit and auxiliary equipment is based on a Time-Of-Use (TOU) tariff structure commonly implemented in Italy. The weekly period is divided into low price, medium price and high price periods, corresponding to 0.03 €/kWh, 0.165 €/kWh and 0.3 €/kWh respectively. The tariff rates of the electricity were designed in order to discriminate the values for the optimization application starting from a real value of the high price period. This approach has been found to be effective in ensuring better discrimination of time periods of the day based on the price of electricity providing the agent with faster convergence to the optimal control policy. Specifically the low and medium price values were chosen to be respectively 1/10 and 1/2 of the highest one. Table 2 reports a summary of electricity prices used in this work.

The price of the electrical energy sold to the grid from the PV overproduction was assumed equal to 0.01 €/kWh according to data extracted from the Italian regulator.

The weather file used is the reference weather file (ITA TORINO-CASELLE IGDG.epw) available in EnergyPlus for Torino, Italy. Considering that the system under investigation involves the optimization of a cooling system the simulation period was limited from June to August. Both the control and simulation time steps were set equal to 1 h.

The efficiency of mono-directional DC/AC was assumed to be equal to 90% and the efficiency of DC/DC converters to 95%.

The experiments were carried out in a co-simulation environment combining Python and EnergyPlus (Brandi et al., 2020). Building dynamics and the cooling system were implemented in EnergyPlus while the electrical system including PV and BESS was developed in Python along with the different control strategies. The interaction between Python and EnergyPlus was handled through Building Control Virtual Test Bed (BCVTB) (Wetter, 2011). Further details on the co-simulation environment are provided in Section 4.6.

Table 2
Details of electricity prices used in this work in €/kWh.

Day	Hour of the day				
	00:00–07:00	07:00–08:00	08:00–19:00	19:00–23:00	23:00–24:00
Mon–Fri	0.03	0.0165	0.3	0.0165	0.03
Sat	0.03		0.0165		0.03
Sun			0.03		

Table 3
PV parameters.

Parameter	Value
Nominal power	3 kW
Surface	22 m^2
η_{STC}	0.15
Tilt angle	33°
Azimuth angle	116°

Table 4
BESS characteristics.

Parameter	Value
Round-Trip Efficiency	0.96
Maximum discharging power	1C
Maximum charging power	0.5C
$SOC_{B,min}$	10%
$SOC_{B,max}$	90%

Table 5
TES configurations.

Volume [m^3]	UA-value [W/K]
10.0	12.0
8.0	10.3
6.0	8.5
3.0	6.0

4.1. Modeling of the PV system

The model of the PV system was implemented through a Python class. Solar position was imported from the pvlib package (Holmgren et al., 2018). A commercial mono-crystalline silicon photo-voltaic module was modeled in the proposed environment. The selected module has a specific power of about 80 W/m^2 and an efficiency (η) of 15% under standard conditions (solar irradiance $G_{STC} = 1000$ W/m^2, cell temperature $T_{STC} = 25$ °C, Air Mass $AM_{STC} = 1.5$), as described by Durisch et al. (2007) and reported in Eq. (6).

$$\eta = f(G, AM, T_{out}) \tag{6}$$

The PV panels tilt angle has been chosen from the world dataset provided by Jacobson and Jadhav (2018). Thus, the tilt angle was set to 33°, whereas the azimuth is constrained by the orientation of the test facility. These inputs along with solar radiation and incidence angle allow to compute the PV power generation (P_{PV}) at each time-step which was calculated as the product of the efficiency and incident solar radiation. Table 3 recaps the parameters of the PV module.

The nominal power of the PV system of 3 kW was chosen in order to match up the peak power of the building total electrical demand.

4.2. Modeling of the BESS system

The battery system was simulated through a Python class. A simple and widely adopted model was implemented according to Amato et al. (2021). The model involves the estimation of the State-Of-Charge (SOC), which it was considered sufficiently accurate for carrying out a preliminary evaluation of the impact of BESS installation, even though the degradation of the battery is not taken into account. The calculation of the SOC at each time step t was performed according to the set of equations reported in Eq. (7):

$$\begin{cases} SOC_B(t) = SOC_B(t-1) + \eta_{rte}\frac{P_{B,ch}(t)*\Delta t}{C_B} & (charge) \\ SOC_B(t) = SOC_B(t-1) - \frac{P_{B,dis}(t)*\Delta t}{C_B} & (discharge) \end{cases} \tag{7}$$

where $SOC_B(t-1)$ is the SOC at the previous time step and η_{rte} is the round-trip efficiency. $P_{B,ch}$ and $P_{B,dis}$ are the average power exchanged in the period between two consecutive the time steps (Δt) between the BESS and the system during charging and discharging process respectively. C_B is the battery nominal capacity. Safety constraints were introduced in order to preserve battery lifetime. Charging and discharging processes have to respect two limits defined by $P_{B,ch,max}$ and $P_{B,dis,max}$. These values are introduced in the technical specifications to avoid too rapid charging/discharging operations. Typically, maximum charging and discharging power are different and when the power exceeds these thresholds, the controller limits it to the maximum recommended values. In order to preserve the health of the battery, the levels of the SOC were constrained by the minimum and maximum values provided by the manufacturer (i.e. $SOC_{B,min}$, $SOC_{B,max}$).

The characteristics of the BESS considered in this work were gathered from the data sheet of a modular Li-ion battery available on the market and reported in Table 4.

In compliance with the typical values for the lithium-ion technology the minimum SOC value ($SOC_{B,min}$) was set equal to 10% and the maximum SOC value ($SOC_{B,max}$) was set equal to 90% for a total Depth of Charge of 80% (Amato et al., 2021). An initial SOC of 50% was imposed. The maximum charging power ($P_{B,ch,max}$) and maximum discharging power ($P_{B,dis,max}$) were set equal to 0.5 times and 1 time the nominal capacity of the battery (C_B) respectively.

4.3. Setup of BESS and TES configurations

As introduced in Section 3 the baseline and the proposed control strategies were implemented considering different capacities of BESS and different sizes of TES.

Table 5 reports for each size of TES the total volume and the corresponding UA-value considered to estimate heat losses. The largest size of 10 m^3 was chosen considering 3-times the maximum daily cooling demand of the building. The smallest size of 3 m^3 was chosen considering 2-times the maximum hourly cooling demand of the building. The intermediate values were picked up according to commercial sizes between minimum and maximum values.

Table 6 reports the features of the various configurations of the BESS. A commercial capacity for the battery unit of 2.4 kWh has been chosen as a reference. This value was selected according to the maximum value of the building electrical demand on an hourly basis. The other two capacities of BESS are supposed as obtained by connecting in series two and three units respectively.

Eventually, Table 7 summarizes all the configurations resulting from the combination of the different capacities of BESS and sizes

Table 6
BESS configurations.

Capacity [kWh]	Max charging power [kW]	Max discharging power [kW]	Units in series
2.4	1.2	2.4	1
4.8	2.4	4.8	2
7.2	3.6	7.2	3

Table 7
Configurations simulated for the experiment.

Configuration	BESS capacity [kWh]	TES volume [m^3]
1	2.4	10.0
2	4.8	10.0
3	7.2	10.0
4	2.4	8.0
5	4.8	8.0
6	7.2	8.0
7	2.4	6.0
8	4.8	6.0
9	7.2	6.0
10	2.4	3.0
11	4.8	3.0
12	7.2	3.0

of TES that have been tested with both baseline and proposed control strategy.

4.4. Implementation of the baseline fully rule-based control

As introduced in Section 3, the baseline strategy manages both the operational modes of the cooling system (and consequently the TES) and BESS through two different RBC strategies. The baseline RBC strategy operates the cooling system in charging mode whenever the price of electricity is low (i.e. between 11 p.m and 7 a.m during Mondays and Saturdays and between 0 a.m and 24 p.m during Sundays) and the temperature of the TES is greater than 12 °C. During these periods the storage is charged until its temperature reaches 10 °C or the price of electricity rises. The cooling system is operated in discharging mode whenever the building cooling demand is not zero until this value returns to zero or the temperature of the TES is greater than 18 °C. If the temperature of the TES is greater than 18 °C and building cooling demand is not zero the cooling system is operated in chiller cooling mode.

A simple still effective controller inspired from previous scientific literature (Ruusu et al., 2019; Amato et al., 2021) was implemented for BESS management. The BESS is charged when PV generation is greater than the building electrical demand, otherwise it is discharged. More specifically, during charging process the PV surplus is diverted to the BESS if it is allowed by the constraints on charging power ($P_{B,ch,max}$) and maximum SOC ($SOC_{B,max}$). If PV generation is greater than the sum of building electrical demand and BESS capacity the remaining overproduction is diverted to the grid. During discharging, the BESS works in parallel with the PV to meet the electrical demand. If the contribution from both PV and BESS is not sufficient to meet the building electrical load the grid is employed to meet the demand.

4.5. Implementation of the proposed control strategy based on DRL coupled with RBC

As introduced in Section 3 the SAC agent manages the three cooling operation modes (i.e. charging mode, discharging mode and chiller cooling mode) while the BESS is managed by the same RBC strategy described in the section above. The SAC agent is defined through the reward function, the action space and the state space. Table 8 reports the variables included in the state-space.

Table 8
Variables included in the state space.

Variable	Min value	Max value	Unit	Timestep
Outdoor Air Temperature (T_o)	7.0	40.0	°C	t
TES SOC (SOC_T)	0.0	1.0	–	t, t − 1, t − 2
BESS SOC (SOC_B)	0.0	1.0	–	t
Building Cooling Demand (Q_d)	0.0	10.0	kW	t, t + 1, …,t + 24
PV power generation (P_{PV})	0.0	3.0	kW	t, t + 1, …,t + 24
Electricity price (C_{buy})	0.03	0.3	€/kWh	t, t + 1, …,t + 24

Table 9
Hyperparameters of the SAC control agent.

Hyperparameter	Value
Discount factor (γ)	0.99
Learning rate	0.001
Boltzmann temperature coefficient (α)	0.2
Number of hidden layers	2
Number of neurons per hidden layer	256
Activation Function	ReLu
Optimizer	Adam
Batch size	32
Number of training episodes	30
Reward magnitude weight-factor (β)	100

The state-space was conceived to provide to the agent comprehensive information about the whole IES including PV production and BESS status. Observations of the storage tank including the SOC (SOC_T) evaluated at the current timestep t and up to two timestep ($t - 2$) in the past were provided to the agent. These values carry information about the amount of cooling energy actually stored and its evolution over time.

The SOC of the BESS is also a key-information provided to the agent to correctly manage the operation of the cooling system. BESS is operated to provide electricity to the chiller and the pumping system during high price periods. This value was provided only at the current timestep t due to lower inertia of BESS compared to TES.

The electricity price is the main driver of the agent choices since it strongly influences the reward. Current value was provided to the agent along with the exact values for 24 h ahead. The electricity price schedules were supposed to be always known.

The building cooling demand together with the PV power generation is a fundamental information to optimally manage the controlled system. Also, the values related to time step t to time step $t+24$ were provided to the agent. The predictions of building cooling demand and PV power generation were assumed to be perfectly known.

Eventually, information about outdoor air temperature were included in order to provide knowledge about its influence on the COP of the chiller unit. Despite being a key information, the solar irradiation was not included in the state-space since the PV power generation is directly related to this variable.

Table 8 reports the maximum and the minimum values that were employed to re-scale the state space through a min–max normalization before providing the variables to neural network models.

Besides the definition of state-space, action-space and reward function, the SAC algorithm is characterized by a series of hyperparameters. The settings of these hyperparameters adopted in this application are reported in Table 9.

Each episode (i.e. one cooling season lasting from June to August) is presented to the SAC control agent 30 times in order

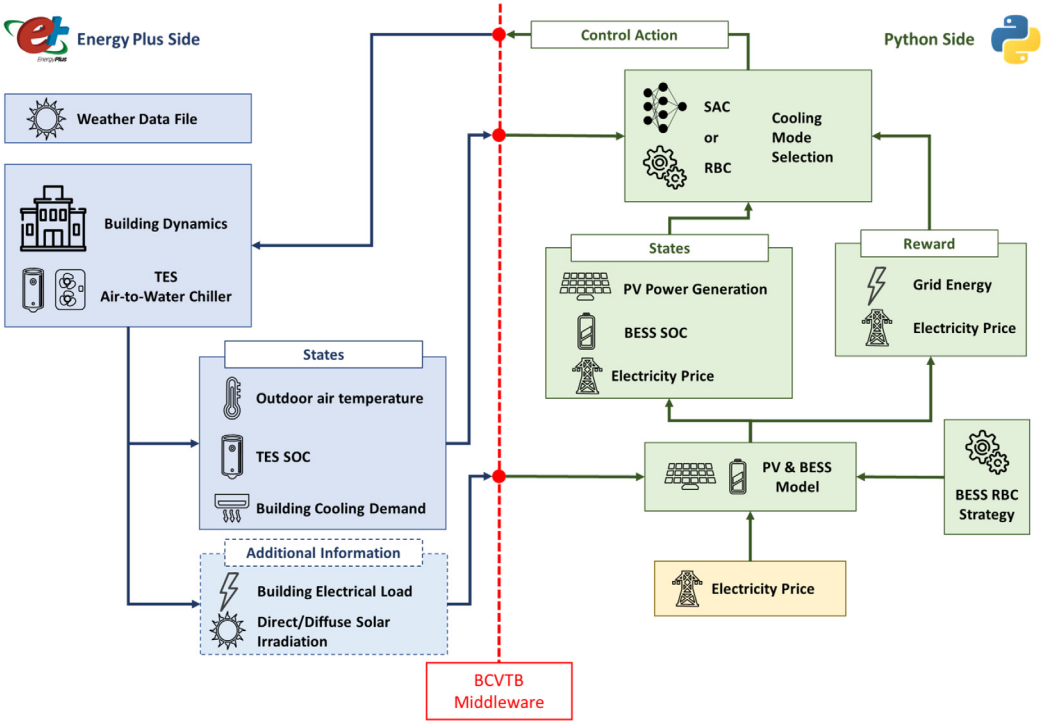

Fig. 4. Architecture of the co-simulation environment.

to train the control policy for each configuration. At the end of the training process the SAC agent was statically deployed for one single deployment episode corresponding to the same cooling season as the training episode. In the static deployment approach, the agent is implemented as a static entity, meaning that the control policy is no longer updated, and any learning goes on. When the SAC agent is statically deployed the control policy is determined by the weights resulting from the last update (i.e., the last control step of the last training episode) of the training phase. For this reason, as common practice, the static deployment of the SAC agent was performed on the same period (i.e., from June to August) of the training. In fact, during the deployment process the performance of the agent during the training period could provide a good indication of the stability of the learned control policy.

4.6. Co-simulation environment

The experiments were carried out in a co-simulation environment combining Python and EnergyPlus through the BCVTB. Fig. 4 describes the architecture of the co-simulation environment. The architecture is organized in two sides.

The EnergyPlus side is formed by a model of the building dynamics and of the cooling system (comprising of the air-to-water chiller and TES) receiving at each time-step information from the weather data file and a controller which selects the cooling mode. This model provides in output the state variables (i.e. outdoor air temperature, TES SOC and Building Cooling Demand) employed by SAC agent evaluated at each time-step. Moreover, the Energy Plus model produces additional information such as building electrical load and direct and diffuse solar irradiation employed by the PV and BESS models.

The Python side of the co-simulation environment is formed by the PV and BESS models and by the control strategies employed to manage the integrated energy system. The PV model employs solar irradiation to calculate the PV power generation which is one of the state variables provided to the SAC agent. The BESS SOC is evaluated through the BESS model which receives information to whether charge or discharge the battery from the BESS RBC strategy described in Section 4.4. This strategy manages the BESS according to building electrical load (provided by EnergyPlus and determined by chiller and pump operation), PV power generation and electricity price (provided to Python through a csv file). The electricity price is furtherly forwarded as a state variable employed by both the SAC agent and the RBC strategy to select the cooling mode at each time-step. Once the BESS operation is evaluated, the environment evaluates the energy exchanged with the electrical grid which determines the reward obtained by the SAC agent along with the electricity price. The final component of the Python side is represented by either the SAC agent or the RBC strategy employed to select the cooling mode. The SAC strategy makes use of all the information included within the state space and the reward function to learn the optimal control policy. The RBC strategy, as described in Section 4.4, employs only TES SOC and electricity price to determine the control action.

Eventually, the control action is forwarded to EnergyPlus in order to advance the simulation to next time-step. The interface between python and EnergyPlus is managed through BCVTB and the *ExternalInterface* function of EnergyPlus. BCVTB is a software environment that allows the users to combine different softwares for co-simulation. The interaction orchestrated by BCVTB between Python and EnergyPlus is dynamic, and take place at each time-step of the simulation process.

5. Results

This section reports the results of the implementation of the methodology introduced in Section 3.

A SAC control agent coupled with RBC was simulated together with a baseline fully RBC strategy during the cooling season in the period ranging from June to August for different sizes and capacities of TES and BESS, respectively. For the sake of simplicity, in the following sections the proposed controller which couples SAC with RBC is indicated as *SAC*, while the baseline fully RBC strategy is simply indicated as *RBC*.

Table 10
Energy imported from grid ($E_{grid,buy}$), energy sold to grid ($E_{grid,sell}$), Cost of electricity and cost savings obtained from the implementation of SAC agent and RBC strategy.

Config	$E_{grid,buy}$ [kWh]		$E_{grid,sell}$ [kWh]		Cost [€]		Cost savings [%]
	SAC	RBC	SAC	RBC	SAC	RBC	
1	314.70	871.40	380.90	919.10	6.0	16.9	64.7
2	223.70	749.10	274.60	776.80	6.5	14.7	55.8
3	172.40	628.60	222.30	636.50	3.9	12.5	68.8
4	292.20	872.70	357.50	928.10	6.9	16.9	59.2
5	310.60	750.90	355.90	786.40	8.9	14.7	39.5
6	147.90	632.00	193.70	648.00	3.4	12.5	72.8
7	355.40	861.10	420.80	928.90	8.2	18.1	54.7
8	231.10	747.20	281.90	796.20	5.2	14.9	65.1
9	188.20	636.60	230.40	667.30	5.3	12.5	57.3
10	281.20	797.00	358.50	862.00	7.7	49.2	84.3
11	209.10	693.00	271.70	740.70	4.9	24.5	80.0
12	178.00	591.50	233.20	622.40	6.1	12.5	51.2

Table 11
Contribution of the different sources (PV_{frac}, $BESS_{frac}$ and $Grid_{frac}$) to the building electrical demand (E_{dem}) obtained by SAC and RBC strategy for the different configurations.

Config	E_{dem} [kWh]		PV_{frac}		$BESS_{frac}$		$Grid_{frac}$	
	SAC	RBC	SAC	RBC	SAC	RBC	SAC	RBC
1	1070.5		0.56		0.14	0.12	0.30	0.80
2	1075.4	1090.7	0.58	0.08	0.21	0.23	0.21	0.69
3	1064.2		0.55		0.29	0.34	0.16	0.58
4	1073.10		0.60		0.13	0.12	0.27	0.80
5	1078.70	1083.0	0.50	0.08	0.21	0.23	0.29	0.69
6	1069.10		0.58		0.28	0.34	0.14	0.58
7	1070.20		0.52		0.15	0.11	0.33	0.80
8	1064.30	1072.2	0.53	0.09	0.25	0.22	0.22	0.69
9	1063.60		0.51		0.31	0.32	0.18	0.59
10	1055.70		0.57		0.16	0.11	0.27	0.74
11	1053.30	1075.1	0.54	0.15	0.26	0.20	0.20	0.65
12	1058.00		0.54		0.29	0.30	0.17	0.55

Table 10 reports both electrical energy imported from and sold to the grid together with the electricity costs achieved by implementing SAC and RBC strategies for each configuration during the whole simulation period. The last column of the table reports the monetary savings achieved through the implementation of SAC strategy.

The results in Table 10 show that SAC control policy learnt to minimize the interactions with the electrical grid with respect to RBC strategy. Across all configurations the energy imported from grid and energy sold to grid were on average 67% and 61% lower for SAC strategy compared to RBC strategy. RBC performance in terms of operational cost improved with the increasing of BESS size. A cost reduction between 26.1% and 74.3% was achieved by the baseline strategy when nominal capacity was increased from 2.4 kWh to 7.2 kWh.

Independently from TES size, RBC achieved the best performance with a BESS capacity of 7.2 kWh (i.e. configurations 3, 6, 9 and 12). The increase of TES size beyond 6 m^3 did not lead to significant improvements in terms of operational costs of RBC strategy for the configurations implementing the same BESS capacity (i.e. configurations from 1 to 6).

Similarly to RBC, the operational cost with the SAC control agents decreased with the increase of BESS capacity. However, due to their intrinsic stochastic nature in the training process and initialization of the neural network policy their performance did not show a linear pattern.

SAC strategy led to the best performance with the configurations implementing 8 m^3 and 10 m^3 leading to a monetary expense of 3.4 € and 3.9€, respectively.

The SAC control agents led to a better performance than the RBC with an economic savings ranging from 39.5% to 84.3%. The highest difference between the two control strategies were achieved for configuration 10 implementing both TES and BESS with the lowest sizes.

Table 11 reports the building electrical consumption over the simulation period (E_{dem}) along with the percentages indicating the contribution of each source by implementing SAC and RBC strategies. PV_{frac}, $BESS_{frac}$ and $Grid_{frac}$ indicate the percentage of electrical demand satisfied by PV generation directly provided to the building, by BESS and through the grid, respectively.

In the case of RBC strategy, independently from TES size, the implementation of different BESS capacities had no influence on the percentage contribution of PV generation directly feeding the building and the electrical energy demand as can be seen for the configurations 1–3, 4–6, 7–9 and 10–12, respectively. Generally, SAC led to lower energy consumption compared to RBC, as shown by second and third column (i.e. E_{dem}) suggesting that SAC learnt a better management strategy. Moreover, as shown by column PV_{frac}, the SAC strategy was capable to better exploit PV generation to feed the building with respect to RBC. In the case of baseline controller the percentage contribution of PV generation directly feeding the building ranges between 8% and 15% increasing with the reduction of TES size. SAC outperformed RBC exploiting the PV production in a range between 50% and 60% across all configurations.

With the increasing of the BESS capacity RBC was capable to shift the contribution from the grid to the BESS. SAC and RBC showed similar utilization of the BESS system among all configurations.

Considering the configurations implementing the smallest BESS capacity of 2.4 kWh (i.e. configuration 1, 4, 7, 10) the configuration 10 is the one which led to the highest operational cost despite the lowest percentage of electricity drawn from the grid with respect to configurations 1, 4 and 7. This pattern suggests that in that case the RBC controller was forced to rely on electrical grid to operate the chiller during high-price periods due to not enough thermal or electrical energy stored.

Key indicators to assess the performance of PV-BESS systems are the SS and the SC, the former describing the amount of the demand which is satisfied by the local generation, the latter the amount of the local generation which is consumed in place. SC also indicates the economic viability of the PV systems which is usually increased through the introduction of BESS. Since the BESS is charged only through PV, the value of PV generation employed to calculate SS and SC comprises the PV generation directly feeding the building and the electricity provided to the building by the BESS.

Fig. 5 shows the SS and SC resulted from the implementation of SAC and RBC strategies for all the configurations of storage analyzed.

The results show that TES volume did not significantly affect SS and SC values. SAC performed significantly better than RBC, increasing SS and SC with an average value of 40% considering all the configurations. Moreover, RBC performance was affected by BESS capacity both in terms of SS and SC, whereas SAC managed to maintain their values almost constant among the configurations.

Table 12 reports the TES operation in terms of thermal energy charged (*Charge*) and discharged (*Discharge*) along with the percentage of the building cooling demand (*Demand*) satisfied through storage discharging by implementing SAC and RBC strategies.

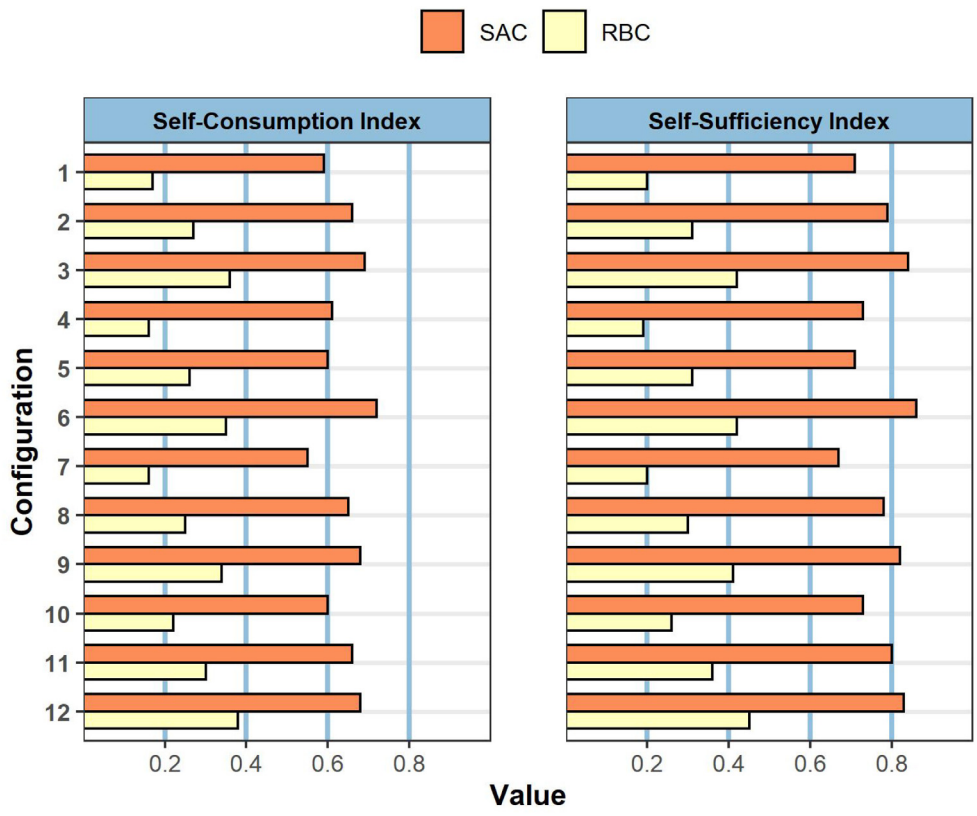

Fig. 5. SS and SC indices obtained by implementing SAC and RBC for all configurations of TES and BESS analyzed.

Table 12
Thermal energy exchanged by the TES during charging (Charge) and discharging (Discharge) phases and percentage of building cooling demand satisfied (Demand) by implementing the different control strategies.

Config	Charge [kWh$_{th}$]		Discharge [kWh$_{th}$]		Demand [%]	
	SAC	RBC	SAC	RBC	SAC	RBC
1	1825.0		1703.0		54.30	
2	1630.1	3307.4	1494.3	3132.8	47.65	99.96
3	1594.4		1488.2		47.43	
4	1643.3		1518.6		48.40	
5	1975.9	3281.0	1829.3	3129.0	58.34	99.77
6	1622.1		1508.6		48.06	
7	1895.2		1775.4		56.57	
8	1649.4	3160.2	1538.1	3046.3	48.99	97.06
9	1548.0		1435.6		45.74	
10	1429.2		1351.4		43.00	
11	1444.8	2234.8	1372.6	2131.9	43.68	67.88
12	1420.9		1339.9		42.65	

The results show that the operation of the thermal storage was not influenced by the capacity of BESS when the RBC is employed. On the other hand, SAC managed the system by charging less the TES when the capacity of the BESS is higher. Moreover, while the RBC almost fully met the building cooling demand through TES discharging for the configurations implementing a TES size greater than 6 m^3, SAC met only the 48.7% on average among all configurations.

These patterns along with the results presented in Tables 10 and 11 suggest that SAC learnt to optimally manage the cooling system and the thermal storage in coordination with local PV production and BESS.

Figs. 6 and 7 report the SOC profiles for both BESS and TES resulted from RBC and SAC implementation for configuration 3 and 10 respectively during the month of August. The black dotted lines indicates the beginning of a different week (i.e. from Monday to Sunday).

Configuration 3 implements the highest sizes for both TES and BESS (i.e., 10 m^3 and 7.2 kWh). It can be observed that SAC learnt to manage the thermal storage to maintain in average a lower SOC of the system compared to RBC. In particular, the SAC agent charged the TES at the beginning of the week and gradually released this energy during the first days of the week. Despite the controllers directly act only on the operational state of the cooling system, the control strategies affected also the operation of the BESS. The BESS was charged and discharged more frequently when the SAC strategy is adopted compared to the case implementing RBC strategy.

The variation of the SOC of BESS and TES for configuration 10 which implements the lowest sizes for both TES and BESS (i.e. 3 m^3 and 2.4 kWh) is reported in Fig. 7. Also in this case SAC managed the cooling system in order to maintain the SOC of the thermal storage as low as possible. This pattern is particularly evident during weekends in which RBC maintained a SOC close to 1 while SAC maintained it close to zero until the beginning of the successive week. Also for this configuration SAC showed a more variable use of the BESS system than RBC strategy.

Figs. 8 and 9 better depict how the different management strategies of the cooling modes affected the behavior of the whole energy system. The figures show in three subplots the trend of several variables on hourly basis for five days of the simulation period (i.e. between Friday 14-08 and Tuesday 18-08). For the sake of simplicity, only the results obtained for configuration 10 implementing a TES size of 3 m^3 and a BESS capacity of 2.4 kWh are presented. This configuration was chosen since it resulted as particularly representative of the difference between SAC and RBC strategies. The top subplot reports the building total electrical load and the sources through which it is met. Moreover, the subplot reports the PV power production and its dispatchment.

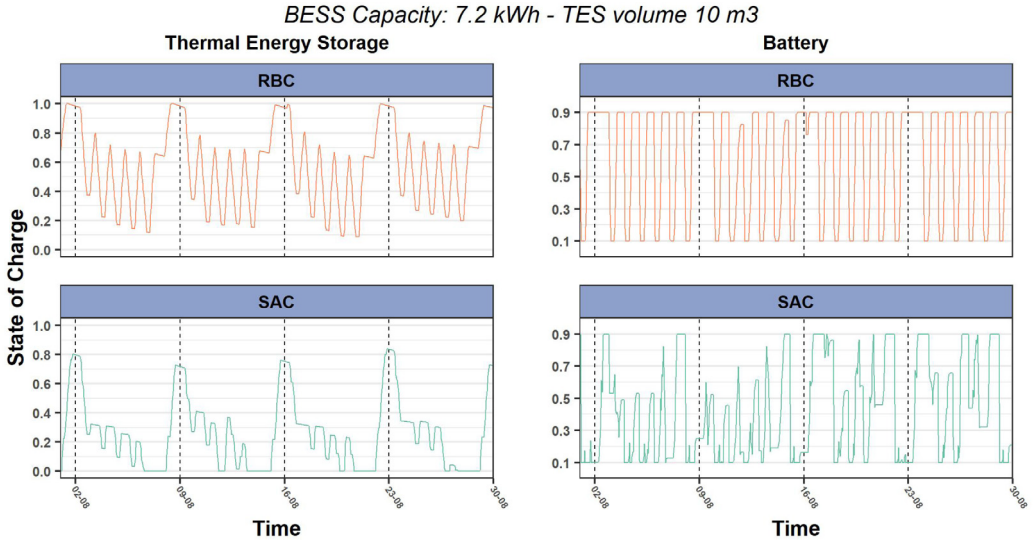

Fig. 6. TES and BESS SOC resulted by SAC and RBC implementation during the whole simulation period for system configuration 3.

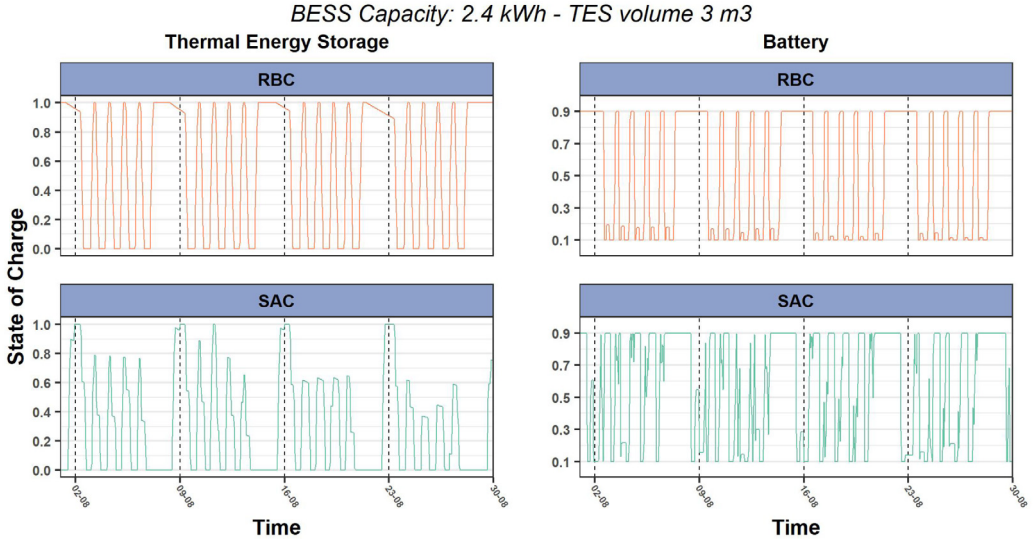

Fig. 7. TES and BESS SOC obtained by SAC and RBC during the whole simulation period for system configuration 10.

The central subplot shows the building cooling demand and the sources employed to meet it along with the cooling energy provided by the chiller to charge the TES. The bottom subplot depicts the trend of SOC for TES and BESS along with electricity price values scaled with a min–max normalization.

Fig. 8 presents the results with reference to RBC strategy. According to this strategy the TES is charged whenever the price of electricity drop to its minimum value. This behavior generated an electricity demand due to chiller operations mainly during night hours when the PV production is null. As a consequence, the system was forced to import energy from the grid during low-price periods. Until 09:00 AM there is no electrical demand from the building and the PV fed the BESS. When the building is occupied, the TES was discharged to meet the cooling demand while the building electrical load is determined only by circulation pumps which were powered by PV production. Through this approach, the import of electricity from the grid during high-price periods was avoided. When the BESS was fully charged the PV overproduction was sold to the grid. Since for configuration 10 the BESS capacity is relatively small, the amount of energy sold to the grid during this period is considerable. During the last hours of the day the thermal energy stored within the TES is exhausted and the systems was forced to use the chiller to meet the cooling load. The PV generation was not sufficient to meet the electrical load, and as a consequence, BESS and grid were employed during high-price periods as shown in the bottom subplot. Moreover, it can be noticed that at the beginning of the weekend the RBC strategy immediately charged the TES due to the occurrence of a low price period. The TES was fully charged after few hours and was not discharged until the beginning of the next week. In these periods TES lost part of its thermal energy to the ambient, resulting in a sub-optimal management of the system.

Fig. 9 shows the results obtained by SAC control strategy. The agent tried to charge the thermal storage during low-price periods close to arrival time of occupants in order to minimize heat losses to the ambient due to storage inactivity. Through this approach, the SAC strategy was capable to reduce electrical energy consumption due to TES charging and, consequently, to consume less electrical energy than RBC as reported in Table 11. During the first hours of occupancy in working days the SAC agent followed a similar policy to RBC powering circulation auxiliaries through PV production and charging the BESS at the same time.

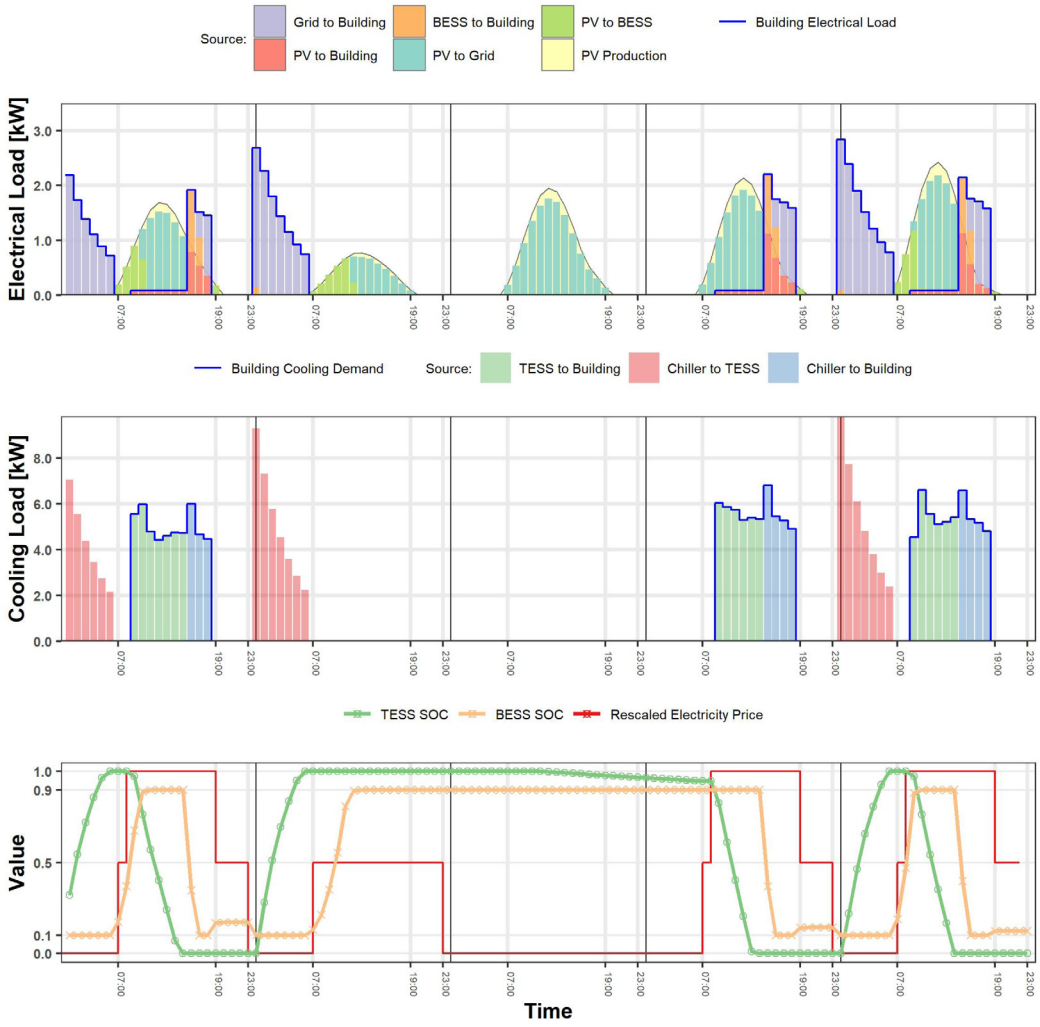

Fig. 8. Trends of the electrical load, cooling load and SOC obtained by RBC strategy between Friday 14-08 and Tuesday 18-08 for configuration 10.

However, during the central hours of working days, the control policy learnt by the SAC agent is completely different from RBC. The agent switched the system in chiller cooling mode in order to leverage PV production to feed the chiller avoiding to sell renewable energy to the grid and maximizing SC. When the PV production was not sufficient, the BESS previously charged was employed. During the last hours of the occupancy period which are still characterized by high electricity prices, the cooling system was switched again to discharging mode since the PV and BESS could not meet the electrical load of the chiller. In this period the PV production was employed to operate the circulation pumps and charge the BESS while the excess of energy was sold to the grid. Moreover, SAC control strategy during weekend awaits Sundays to charge the TES in order to minimize electricity cost even during low-price periods and maximizing SC. Through this approach the SAC agent was also capable to limit TES heat losses compared to RBC strategy.

6. Discussion

The results obtained by applying RBC and SAC strategies for an IES of an office building provided interesting information about the impact of an advanced control strategy on the sizing and operation of energy storage solutions.

SAC was capable to outperform RBC in terms of operating cost for all the configurations of TES and BESS tested. RBC proved to be very sensitive to storage capacities resulting in a huge impact on the operational cost. This aspect is particularly relevant for the BESS capacity.

On the other hand, SAC strategy was able to achieve considerable economic savings also with small capacities, but as the storage capacities increase, the improvement achieved was lower than those achieved by RBC. SAC did not show a clear dependency of the operating cost from the capacities of the storage systems, rather it learnt effective control policies for each configuration. Larger BESS helped SAC in reducing the TES utilization while a similar pattern was not observed for RBC.

BESS is largely considered as the best way to increase SC. However, the operating costs decrease as long as the PV production is sold to the grid leaving no room of improvement of SC levels. Advanced control strategies such as SAC proved to be a viable solution to increase SS and SC levels also with relatively low capacity of the BESS. This is an important aspect to consider given that BESS has a great impact on the total investment cost of energy systems. Reducing the energy exchanged with the electrical grid results in higher profitability of storage technologies and higher flexibility of the building IES. When PV production is sold to the grid the performance of the system in terms of SC degrades. For this reason, SAC aimed at matching PV production and chiller operation as much as possible. In this way, SAC not only avoided unnecessary BESS operations, which would have involved electrical losses due to the round-trip efficiency and

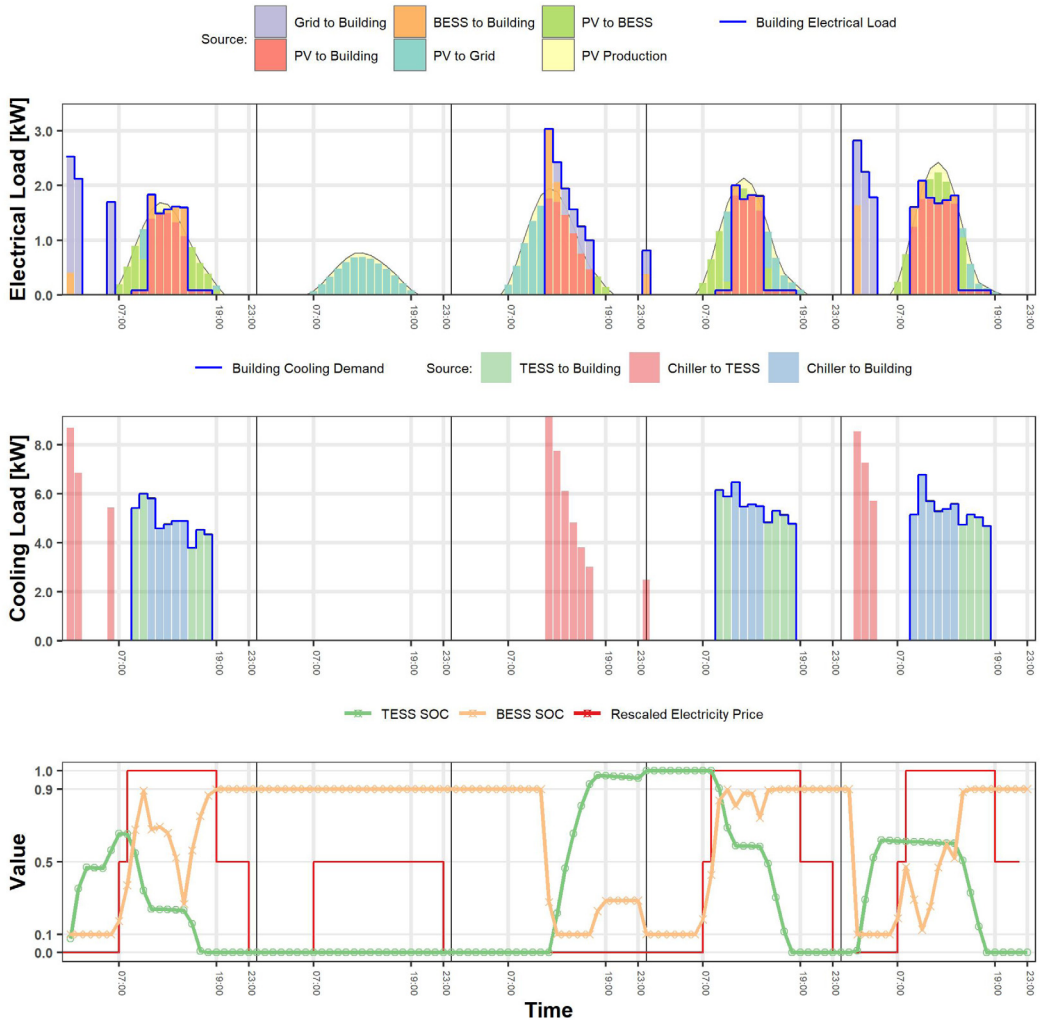

Fig. 9. Trends of the electrical load, cooling load and SOC obtained by SAC control strategy between Friday 14-08 and Tuesday 18-08 for configuration 10.

converter efficiency, but it also managed effectively the system with smaller capacity of BESS.

Eventually the energy consumption and PV contribution to building electrical demand was not affected by the capacity of BESS when RBC was employed. The reason might be that RBC strategy employed two distinctive strategies for both BESS and TES. These controllers were responsible only for their relative system and did not share information between each other. As a consequence, the RBC controller managing the cooling system operation was not aware of PV production and BESS SOC and vice versa. This aspect strongly limited the capability of the RBC to optimally control the proposed system despite the reasonable control rules implemented. This fact clearly shows the limitations of traditional control approaches.

In fact, the RBC strategy implemented in this work was developed making the hypothesis that in classical control approaches the different storage solutions are managed by control laws unaware of other systems. This assumption was deemed reasonable considering that BESS and TES are usually implemented in existing buildings by different stakeholders in different periods of time. Moreover, installers and maintainers usually lack of competences to design an integrated control system capable to coordinate multiple storage equipment.

Conversely, SAC based its decision process on a set of information including those relative to PV production and BESS status. This approach provided to the agent with a comprehensive view of the operation of the whole IES, enabling the identification of a better control policy compared to RBC. Moreover, SAC leveraged predictions of external disturbances to furtherly optimize the decision process.

SAC outperformed the RBC in managing the PV-BESS system, even though the control action does not directly act on the battery. High levels of SS and SC make the use of electricity storage technologies much more desirable from the point of view of the building flexibility. Moreover, SAC was capable to achieve appreciable levels of SS and SC with configurations implementing small sizes of the storage systems.

These results suggest that advanced control strategies are necessary elements to be integrated in a system where the number of connections and prosumers is drastically increasing.

7. Conclusion and future work

The paper explored the effect of the implementation of advanced control strategies in a office building with on-site electricity generation and storage technologies. The objective of the proposed controller was to minimize the operational cost of the system during the cooling season by exploiting TES and BESS as energy flexibility sources to shift cooling and electrical demand according to price schedules and local PV production.

Two control solutions, a DRL agent coupled with a RBC and fully RBC strategy, were tested and analyzed for different BESS

capacities and TES sizes. The controllers were designed to adjust the operation mode of the cooling system deciding whether to charge/discharge the thermal storage to satisfy building cooling demand or to directly employ the chiller. The BESS was in both cases managed through a rule-based strategy. A state-of-the-art SAC algorithm modified for discrete action settings was implemented as DRL control strategy.

The implemented fully RBC strategy based its decisions only on TES SOC and electricity price. Thus, it was not aware of local PV production or BESS status. This resulted in sub-optimal control policy especially when the capacities of TES and BESS were small.

SAC proved to be capable to learn better control policy compared to RBC given the same storage capacities reducing the operating cost between 39.5% and 84.3%. RBC resulted more sensitive to the storage size, giving greater importance to the initial design, whereas SAC achieved high savings also when smaller capacities were implemented. The advantage with respect to RBC narrows down as the capacities were increased. For the same BESS capacity installed, SAC control strategy was capable to notably increase the levels of SS and SC, reducing the energy exchanged with the grid and increasing building energy flexibility.

In conclusion, the results obtained highlighted the importance of implementing advanced control strategies in the design framework of IES in buildings. However, the proposed SAC control strategy despite its model-free definition is not completely independent by a modeling effort since it was trained for several episodes before converging to the final solution. Therefore, future works will be focused on the following directions:

- Extending the present work in order to analyze the performance of the proposed strategy during the whole year including the heating season. Moreover, the present work will be broaden to include a more detailed modeling of the PV-BESS system also considering Maximum Power Point Tracking (MPPT) techniques to maximize PV power output to the BESS system.
- Analyzing the proposed strategy applied to different buildings and case studies in order to better characterize the effectiveness of advanced control strategies applied to integrated energy systems.
- Benchmarking the proposed method against other advanced control strategies (i.e. MPC) in order to provide a more comprehensive view of the benefits provided by its implementation.
- Evaluating the opportunities to share the control policy learnt from one building to target buildings characterized by similar features (i.e. transfer learning) enhancing the scalability and generalizability of the proposed solution.

Eventually, the analysis of DRL control policies provided useful information to improve the operation of storage technologies defining guidelines for the design of more efficient RBC strategies through the definition of an optimized set of rules.

CRediT authorship contribution statement

Silvio Brandi: Conceptualization, Methodology, Investigation, Visualization, Software, Writing – original draft. **Antonio Gallo:** Investigation, Visualization, Software, Writing – original draft. **Alfonso Capozzoli:** Conceptualization, Methodology, Supervision, Formal analysis, Writing – review & editing, Project administration.

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.

Acknowledgment

The work of Silvio Brandi was made in the context of a Ph.D. scholarship at Politecnico di Torino, Italy funded by Enerbrain s.r.l.

References

Abedi, S., Yoon, S.W., Kwon, S., 2022. Battery energy storage control using a reinforcement learning approach with cyclic time-dependent Markov process. Int. J. Electr. Power Energy Syst. 134, 107368. http://dx.doi.org/10.1016/j.ijepes.2021.107368, URL: https://www.sciencedirect.com/science/article/pii/S0142061521006074.

Amato, A., Bilardo, M., Fabrizio, E., Serra, V., Spertino, F., 2021. Energy evaluation of a PV-based test facility for assessing future self-sufficient buildings. Energies 14, 329. http://dx.doi.org/10.3390/en14020329.

Anvari-Moghaddam, A., Rahimi-Kian, A., Mirian, M.S., Guerrero, J.M., 2017. A multi-agent based energy management solution for integrated buildings and microgrid system. Appl. Energy 203, 41–56. http://dx.doi.org/10.1016/j.apenergy.2017.06.007, URL: https://www.sciencedirect.com/science/article/pii/S0306261917307572.

Arteconi, A., Xu, J., Ciarrocchi, E., Paciello, L., Comodi, G., Polonara, F., Wang, R., 2015. Demand side management of a building summer cooling load by means of a thermal energy storage. Energy Procedia 75, 3277–3283. http://dx.doi.org/10.1016/j.egypro.2015.07.705, URL: https://www.sciencedirect.com/science/article/pii/S1876610215014733, Clean, Efficient and Affordable Energy for a Sustainable Future: The 7th International Conference on Applied Energy (ICAE2015).

Baniasadi, A., Habibi, D., Al-Saedi, W., Masoum, M.A., Das, C.K., Mousavi, N., 2020. Optimal sizing design and operation of electrical and thermal energy storage systems in smart buildings. J. Energy Storage 28, 101186. http://dx.doi.org/10.1016/j.est.2019.101186, URL: https://www.sciencedirect.com/science/article/pii/S2352152X19311545.

Barrett, E., Linder, S., 2015. Autonomous HVAC control, a reinforcement learning approach. In: Bifet, A., May, M., Zadrozny, B., Gavalda, R., Pedreschi, D., Bonchi, F., Cardoso, J., Spiliopoulou, M. (Eds.), Machine Learning and Knowledge Discovery in Databases. Springer International Publishing, Cham, pp. 3–19.

Biemann, M., Scheller, F., Liu, X., Huang, L., 2021. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. Appl. Energy 298, 117164. http://dx.doi.org/10.1016/j.apenergy.2021.117164, URL: https://www.sciencedirect.com/science/article/pii/S0306261921005961.

Biyik, E., Kahraman, A., 2019. A predictive control strategy for optimal management of peak load, thermal comfort, energy storage and renewables in multi-zone buildings. J. Build. Eng. 25, 100826. http://dx.doi.org/10.1016/j.jobe.2019.100826, URL: https://www.sciencedirect.com/science/article/pii/S2352710219302165.

Brandi, S., Piscitelli, M.S., Martellacci, M., Capozzoli, A., 2020. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. Energy Build. 224, 110225. http://dx.doi.org/10.1016/j.enbuild.2020.110225, URL: http://www.sciencedirect.com/science/article/pii/S0378778820308963.

Christodoulou, P., 2019. Soft actor-critic for discrete action settings. CoRR, URL: http://arxiv.org/abs/1910.07207. arXiv:1910.07207.

Chu, W.-X., Lien, Y.-H., Huang, K.-R., Wang, C.-C., 2021. Energy saving of fans in air-cooled server via deep reinforcement learning algorithm. Energy Rep. 7, 3437–3448. http://dx.doi.org/10.1016/j.egyr.2021.06.003, URL: https://www.sciencedirect.com/science/article/pii/S2352484721003607.

Comodi, G., Carducci, F., Nagarajan, B., Romagnoli, A., 2016. Application of cold thermal energy storage (CTES) for building demand management in hot climates. Appl. Therm. Eng. 103, 1186–1195. http://dx.doi.org/10.1016/j.applthermaleng.2016.02.035, URL: https://www.sciencedirect.com/science/article/pii/S1359431116301788.

Comodi, G., Giantomassi, A., Severini, M., Squartini, S., Ferracuti, F., Fonti, A., Nardi Cesarini, D., Morodo, M., Polonara, F., 2015. Multi-apartment residential microgrid with electrical and thermal storage devices: Experimental analysis and simulation of energy management strategies. Appl. Energy 137, 854–866. http://dx.doi.org/10.1016/j.apenergy.2014.07.068, URL: https://www.sciencedirect.com/science/article/pii/S030626191400751X.

Coraci, D., Brandi, S., Piscitelli, M.S., Capozzoli, A., 2021. Online implementation of a soft actor-critic agent to enhance indoor temperature control and energy efficiency in buildings. Energies 14 (4), http://dx.doi.org/10.3390/en14040997, URL: https://www.mdpi.com/1996-1073/14/4/997.

Crawley, D., Pedersen, C., Lawrie, L., Winkelmann, F., 2000. EnergyPlus: Energy simulation program. Ashrae J. 42, 49–56.

Cui, T., Chen, S., Wang, Y., Zhu, Q., Nazarian, S., Pedram, M., 2017. An optimal energy co-scheduling framework for smart buildings. Integration 58, 528–537. http://dx.doi.org/10.1016/j.vlsi.2016.10.009.

Das, C.K., Bass, O., Kothapalli, G., Mahmoud, T.S., Habibi, D., 2018. Overview of energy storage systems in distribution networks: Placement, sizing, operation, and power quality. Renew. Sustain. Energy Rev. 91, 1205–1230. http://dx.doi.org/10.1016/j.rser.2018.03.068, URL: https://www.sciencedirect.com/science/article/pii/S1364032118301606.

Dimitroulis, P., Alamaniotis, M., 2022. A fuzzy logic energy management system of on-grid electrical system for residential prosumers. Electr. Power Syst. Res. 202, 107621. http://dx.doi.org/10.1016/j.epsr.2021.107621, URL: https://www.sciencedirect.com/science/article/pii/S0378779621006027.

Du, Y., Zandi, H., Kotevska, O., Kurte, K., Munk, J., Amasyali, K., Mckee, E., Li, F., 2021. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. Appl. Energy 281, 116117. http://dx.doi.org/10.1016/j.apenergy.2020.116117, URL: http://www.sciencedirect.com/science/article/pii/S030626192031535X.

Durisch, W., Bitnar, B., Mayor, J.-C., Kiess, H., hang Lam, K., Close, J., 2007. Efficiency model for photovoltaic modules and demonstration of its application to energy yield estimation. Sol. Energy Mater. Sol. Cells 91 (1), 79–84. http://dx.doi.org/10.1016/j.solmat.2006.05.011, URL: https://www.sciencedirect.com/science/article/pii/S0927024806003345.

Finck, C., Beagon, P., Clauß, J., Péan, T., Vogler-Finck, P., Zhang, K., Kazmi, H., 2018. Review of applied and tested control possibilities for energy flexibility in buildings - a technical report from IEA ebc annex 67 energy flexible buildings. http://dx.doi.org/10.13140/RG.2.2.28740.73609.

Gao, G., Li, J., Wen, Y., 2019. Energy-efficient thermal comfort control in smart buildings via deep reinforcement learning. arXiv:1901.04693.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., Levine, S., 2019. Soft actor-critic algorithms and applications. arXiv:1812.05905.

Han, M., May, R., Zhang, X., Wang, X., Pan, S., Yan, D., Jin, Y., Xu, L., 2019. A review of reinforcement learning methodologies for controlling occupant comfort in buildings. Sustainable Cities Soc. 51, 101748. http://dx.doi.org/10.1016/j.scs.2019.101748, URL: http://www.sciencedirect.com/science/article/pii/S2210670719307589.

Henze, G.P., Schoenmann, J., 2003. Evaluation of reinforcement learning control for thermal energy storage systems. HVAC&R Res. 9 (3), 259–275. http://dx.doi.org/10.1080/10789669.2003.10391069, URL: https://www.tandfonline.com/doi/abs/10.1080/10789669.2003.10391069.

Holmgren, W.F., Hansen, C.W., Mikofski, M.A., 2018. Pvlib python: a python package for modeling solar energy systems. J. Open Source Softw. 3 (29), 884. http://dx.doi.org/10.21105/joss.00884.

IEA, 2019. World energy outlook 2019, IEA, Paris. IEA, URL: https://www.iea.org/reports/world-energy-outlook-2019.

Ioli, D., Falsone, A., Prandini, M., 2015. Optimal energy management of a building cooling system with thermal storage: A convex formulation. IFAC-PapersOnLine 48 (8), 1150–1155. http://dx.doi.org/10.1016/j.ifacol.2015.09.123, URL: https://www.sciencedirect.com/science/article/pii/S2405896315012045, 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2015.

Jacobson, M.Z., Jadhav, V., 2018. World estimates of PV optimal tilt angles and ratios of sunlight incident upon tilted and tracked PV panels relative to horizontal panels. Sol. Energy 169, 55–66. http://dx.doi.org/10.1016/j.solener.2018.04.030, URL: https://www.sciencedirect.com/science/article/pii/S0038092X1830375X.

Kathirgamanathan, A., De Rosa, M., Mangina, E., Finn, D.P., 2021. Data-driven predictive control for unlocking building energy flexibility: A review. Renew. Sustain. Energy Rev. 135, 110120. http://dx.doi.org/10.1016/j.rser.2020.110120, URL: https://www.sciencedirect.com/science/article/pii/S1364032120304111.

Kim, S., Lim, H., 2018. Reinforcement learning based energy management algorithm for smart energy buildings. Energies 11 (8), http://dx.doi.org/10.3390/en11082010, URL: https://www.mdpi.com/1996-1073/11/8/2010.

Koskela, J., Rautiainen, A., Järventausta, P., 2019. Using electrical energy storage in residential buildings – sizing of battery and photovoltaic panels based on electricity cost optimization. Appl. Energy 239, 1175–1189. http://dx.doi.org/10.1016/j.apenergy.2019.02.021, URL: https://www.sciencedirect.com/science/article/pii/S0306261919303113.

Liu, J., Chen, X., Yang, H., Li, Y., 2020. Energy storage and management system design optimization for a photovoltaic integrated low-energy building. Energy 190, 116424. http://dx.doi.org/10.1016/j.energy.2019.116424, URL: https://www.sciencedirect.com/science/article/pii/S036054421932119X.

Liu, S., Henze, G., 2007. Evaluation of reinforcement learning for optimal control of building active and passive thermal storage inventory. J. Solar Energy Eng. Trans. Asme - J. Sol. Energy Eng. 129, http://dx.doi.org/10.1115/1.2710491.

Martinopoulos, G., Papakostas, K.T., Papadopoulos, A.M., 2018. A comparative review of heating systems in EU countries, based on efficiency and fuel cost. Renew. Sustain. Energy Rev. 90, 687–699. http://dx.doi.org/10.1016/j.rser.2018.03.060, URL: http://www.sciencedirect.com/science/article/pii/S1364032118301333.

May, R., 2019. The reinforcement learning method : A feasible and sustainable control strategy for efficient occupant-centred building operation in smart cities.

Medved, S., Domjan, S., Arkar, C., 2021. Contribution of energy storage to the transition from net zero to zero energy buildings. Energy Build. 236, 110751. http://dx.doi.org/10.1016/j.enbuild.2021.110751, URL: https://www.sciencedirect.com/science/article/pii/S0378778821000359.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533, URL: http://dx.doi.org/10.1038/nature14236.

Park, J.Y., Dougherty, T., Fritz, H., Nagy, Z., 2019. Lightlearn: An adaptive and occupant centered controller for lighting based on reinforcement learning. Build. Environ. 147, 397–414. http://dx.doi.org/10.1016/j.buildenv.2018.10.028, URL: https://www.sciencedirect.com/science/article/pii/S0360132318306462.

Pinto, G., Piscitelli, M.S., Vázquez-Canteli, J.R., Nagy, Z., Capozzoli, A., 2021. Coordinated energy management for a cluster of buildings through deep reinforcement learning. Energy 229, 120725. http://dx.doi.org/10.1016/j.energy.2021.120725, URL: https://www.sciencedirect.com/science/article/pii/S0360544221009737.

Ren, H., Sun, Y., Albdoor, A.K., Tyagi, V., Pandey, A., Ma, Z., 2021. Improving energy flexibility of a net-zero energy house using a solar-assisted air conditioning system with thermal energy storage and demand-side management. Appl. Energy 285, 116433. http://dx.doi.org/10.1016/j.apenergy.2021.116433, URL: https://www.sciencedirect.com/science/article/pii/S0306261921000027.

Ruusu, R., Cao, S., Manrique Delgado, B., Hasan, A., 2019. Direct quantification of multiple-source energy flexibility in a residential building using a new model predictive high-level controller. Energy Convers. Manage. 180, 1109–1128. http://dx.doi.org/10.1016/j.enconman.2018.11.026.

Sanaye, S., Sarrafi, A., 2021. A novel energy management method based on deep q network algorithm for low operating cost of an integrated hybrid system. Energy Rep. 7, 2647–2663. http://dx.doi.org/10.1016/j.egyr.2021.04.055, URL: https://www.sciencedirect.com/science/article/pii/S2352484721002730.

Serale, G., Fiorentini, M., Capozzoli, A., Bernardini, D., Bemporad, A., 2018. Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. Energies 11 (3), http://dx.doi.org/10.3390/en11030631, URL: https://www.mdpi.com/1996-1073/11/3/631.

Shabani, M., Mahmoudimehr, J., 2018. Techno-economic role of PV tracking technology in a hybrid PV-hydroelectric standalone power system. Appl. Energy 212, 84–108. http://dx.doi.org/10.1016/j.apenergy.2017.12.030, URL: https://www.sciencedirect.com/science/article/pii/S0306261917317452.

Sharma, V., Haque, M.H., Aziz, S.M., 2019. Energy cost minimization for net zero energy homes through optimal sizing of battery storage system. Renew. Energy 141, 278–286. http://dx.doi.org/10.1016/j.renene.2019.03.144, URL: https://www.sciencedirect.com/science/article/pii/S0960148119304653.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction, second ed. The MIT Press, URL: http://incompleteideas.net/book/the-book-2nd.html.

Tarragona, J., de Gracia, A., Cabeza, L.F., 2020. Bibliometric analysis of smart control applications in thermal energy storage systems. a model predictive control approach. J. Energy Storage 32, 101704. http://dx.doi.org/10.1016/j.est.2020.101704, URL: https://www.sciencedirect.com/science/article/pii/S2352152X20315413.

Terlouw, T., AlSkaif, T., Bauer, C., van Sark, W., 2019. Optimal energy management in all-electric residential energy systems with heat and electricity storage. Appl. Energy 254, 113580. http://dx.doi.org/10.1016/j.apenergy.2019.113580, URL: https://www.sciencedirect.com/science/article/pii/S0306261919312541.

Valladares, W., Galindo, M., Gutiérrez, J., Wu, W.-C., Liao, K.-K., Liao, J.-C., Lu, K.-C., Wang, C.-C., 2019. Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. Build. Environ. 155, 105–117. http://dx.doi.org/10.1016/j.buildenv.2019.03.038, URL: https://www.sciencedirect.com/science/article/pii/S0360132319302008.

Vázquez-Canteli, J.R., Nagy, Z., 2019. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Appl. Energy 235, 1072–1089. http://dx.doi.org/10.1016/j.apenergy.2018.11.002, URL: http://www.sciencedirect.com/science/article/pii/S0306261918317082.

Vázquez-Canteli, J.R., Ulyanin, S., Kämpf, J., Nagy, Z., 2019. Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities. Sustainable Cities Soc. 45, 243–257. http://dx.doi.org/10.1016/j.scs.2018.11.021, URL: http://www.sciencedirect.com/science/article/pii/S2210670718314380.

Wakui, T., Sawada, K., Yokoyama, R., Aki, H., 2019. Predictive management for energy supply networks using photovoltaics, heat pumps, and

battery by two-stage stochastic programming and rule-based control. Energy 179, 1302–1319. http://dx.doi.org/10.1016/j.energy.2019.04.148, URL: https://www.sciencedirect.com/science/article/pii/S0360544219307819.

Wang, Z., Hong, T., 2020. Reinforcement learning for building controls: The opportunities and challenges. Appl. Energy 269, 115036. http://dx.doi.org/10.1016/j.apenergy.2020.115036, URL: http://www.sciencedirect.com/science/article/pii/S0306261920305481.

Wang, Y., Lin, X., Pedram, M., 2016. A near-optimal model-based control algorithm for households equipped with residential photovoltaic power generation and energy storage systems. IEEE Trans. Sustain. Energy 7 (1), 77–86. http://dx.doi.org/10.1109/TSTE.2015.2467190.

Wetter, M., 2011. Co-simulation of building energy and control systems with the building controls virtual test bed. J. Build. Perform. Simul. 4 (3), 185–203. http://dx.doi.org/10.1080/19401493.2010.518631.