energies

MDPI

Review

A Review of Reinforcement Learning Applications to Control of Heating, Ventilation and Air Conditioning Systems

Seppo Sierla [1,*], Heikki Ihasalo [1] and Valeriy Vyatkin [1,2,3]

[1] Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, FI-00076 Espoo, Finland; heikki.ihasalo@aalto.fi (H.I.); valeriy.vyatkin@aalto.fi (V.V.)
[2] Department of Computer Science, Electrical and Space Engineering, Lulea University of Technology, 97187 Lulea, Sweden
[3] International Research Laboratory of Computer Technologies, ITMO University, 197101 St. Petersburg, Russia
* Correspondence: seppo.sierla@aalto.fi

Abstract: Reinforcement learning has emerged as a potentially disruptive technology for control and optimization of HVAC systems. A reinforcement learning agent takes actions, which can be direct HVAC actuator commands or setpoints for control loops in building automation systems. The actions are taken to optimize one or more targets, such as indoor air quality, energy consumption and energy cost. The agent receives feedback from the HVAC systems to quantify how well these targets have been achieved. The feedback is captured by a reward function designed by the developer of the reinforcement learning agent. A few reviews have focused on the reward aspect of reinforcement learning applications for HVAC. However, there is a lack of reviews that assess how the actions of the reinforcement learning agent have been formulated, and how this impacts the possibilities to achieve various optimization targets in single zone or multi-zone buildings. The aim of this review is to identify the action formulations in the literature and to assess how the choice of formulation impacts the level of abstraction at which the HVAC systems are considered. Our methodology involves a search string in the Web of Science database and a list of selection criteria applied to each article in the search results. For each selected article, a three-tier categorization of the selected articles has been performed. Firstly, the applicability of the approach to buildings with one or more zones is considered. Secondly, the articles are categorized by the type of action taken by the agent, such as a binary, discrete or continuous action. Thirdly, the articles are categorized by the aspects of the indoor environment being controlled, namely temperature, humidity or air quality. The main result of the review is this three-tier categorization that reveals the community's emphasis on specific HVAC applications, as well as the readiness to interface the reinforcement learning solutions to HVAC systems. The article concludes with a discussion of trends in the field as well as challenges that require further research.

Keywords: reinforcement learning; machine learning; heating; ventilation; air conditioning; building energy simulator; indoor environment; artificial intelligence; thermal comfort

check for updates

Citation: Sierla, S.; Ihasalo, H.; Vyatkin, V. A Review of Reinforcement Learning Applications to Control of Heating, Ventilation and Air Conditioning Systems. *Energies* **2022**, *15*, 3526. https://doi.org/10.3390/en15103526

Academic Editor: Jarek Kurnitski

Received: 5 April 2022
Accepted: 10 May 2022
Published: 11 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Reinforcement learning (RL) is emerging as an advanced technique for HVAC control, due to its ability to process complex sensor information to achieve complex control objectives. Frequently, multi-objective optimization is performed to meet several targets, such as energy cost reduction [1], energy consumption reduction [2], management of thermal comfort [3] and management of indoor air quality [4]. However, various levels of abstraction are used in the problem formulation, so an assessment of this literature is needed to identify the works that are relevant for specific HVAC control problems. Furthermore, a critical look is needed to assess whether the chosen level of abstraction is justified, with respect to eventual deployment of the RL controller to control physical HVAC equipment.

Figure 1 shows a general concept of an RL agent controlling HVAC equipment. The *RL agent* is the controller. It is trained through interactions with the *RL environment*. The environment provides *state* information as input to the RL agent. The state could consist of measurements such as temperature and CO2 sensor measurements [5]. Based on the state, the agent outputs actions to the environment. For example, the actions could be setpoint values or control signals to HVAC equipment. The environment returns a *reward*, which quantifies how beneficial the outcome of the action was. The reward formulation is crafted to capture the control objectives of the HVAC application. Based on the immediate and long-term rewards, the RL agent is trained to take actions that are likely to result in better rewards in the future. Several approaches can be taken to construct the environment. One approach is to use the physical HVAC system as the environment, in which case it is practical to select a set of sensor measurements as the state and one or more actuator control signals or setpoint values as the actions. However, since the training of a RL agent can require many interactions with the environment, it is advantageous to train it in a virtual environment, such as a physics-based building simulator [6] or a data-driven model of the building [7]. The virtual environment could have the same state and action spaces as the physical environment. However, it is also possible to raise the level of abstraction, in which case the state and action spaces do not necessarily have a direct mapping to sensors and actuators [8]. As the level of abstraction is raised, important characteristics of HVAC equipment such as heat pumps or chillers may be ignored. In many of the reviewed papers, this leads to serious problems, such as failing to distinguish between kW of power consumption of a compressor and the kW of cooling or heating provided to the building. Thus, this review will critically assess the chosen level of abstraction in the reviewed works.

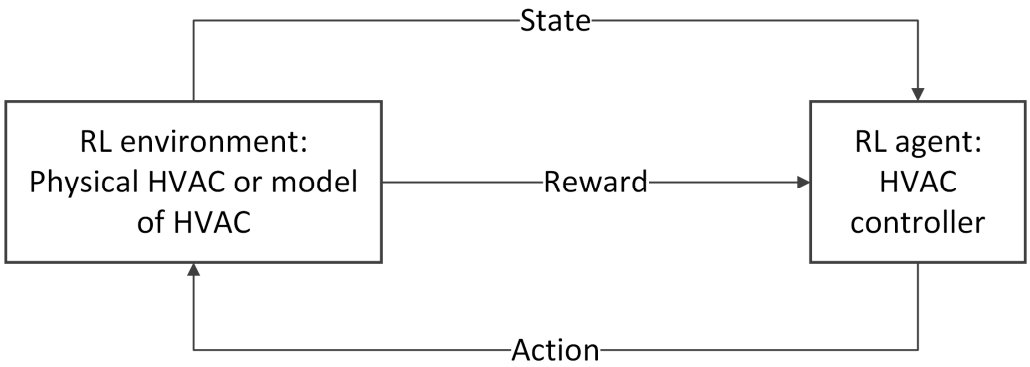


Figure 1. Concept of an RL agent controlling HVAC equipment.

Two key elements of the quality of the indoor environment are thermal comfort and indoor air quality. Thermal comfort can be a tricky concept for HVAC control, as established standards consider factors such as clothing insulation, for which there is no sensor data available [9]. Indoor air quality involves the consideration of pollutants such as carbon dioxide, particulate matter, nitrogen dioxide, ozone and volatile organic compounds [9]. Modern building automation systems often have the instrumentation in place for carbon dioxide measurement, but in general, there is limited measurement data on indoor pollutants, which would be available for HVAC control systems. For these reasons, even though RL is a technique that can handle large state spaces, RL practitioners cannot include the full complexity of the quality of the indoor environment into the RL problem formulation. This raises the question of what kind of simplifications are acceptable, to still benefit from the capabilities of RL and to achieve progress over conventional control techniques that are commonly used in building automation systems. A common approach in RL research on HVAC control is to only consider indoor temperature and to construct the RL environment with one equation that defines indoor temperature as a linear function of outdoor temperature and HVAC power (e.g., [10,11]). This approach only maintains the average indoor temperature of the building within thermal comfort limits, while optimizing energy cost or energy consumption related targets. However, this has three problems:

1. Applications for real buildings must ensure thermal comfort separately in every zone of the building, rather than maintaining the average temperature of the entire building.
2. It is unclear if a reinforcement learning agent that has been trained in a linear environment could be generalized to handle non-linear dynamics of real-world HVAC equipment, and a much larger state and action space with temperature measurements and HVAC actuators in several zones of the building.
3. Indoor air quality is ignored.

These problems could be addressed with a more sophisticated RL environment. A few reviews assume that the development of building energy models is so laborious that any control approaches that require such models will be rejected by the industry [12–14]. Thus, they exclude approaches that require the development of a data-driven or physics-based building energy model to serve as the training environment of the RL agent. Consequently, the great majority of papers applying RL to HVAC control were excluded from these reviews. The papers that were included to these reviews had unacceptably long training times, since the physical building itself was used as the training environment. For this reason, the authors concluded that RL is a problematic approach for HVAC control. Other reviews consider the use of building energy models as a legitimate approach for training a RL agent [15–17]. In each of these reviews, the focus of the analysis is on the reward. The analyses reveal that RL is well suited to multi-objective optimization for problems involving a trade-off between occupant comfort and cost saving or reduction of energy consumption.

The focus of previous reviews on RL for HVAC control can be stated with reference to Figure 1. Ma et al. [9] review the state space formulations in the literature from the perspective of capturing thermal comfort and indoor air quality. The types of algorithms used to implement the RL controller are categorized in [15–17]. Several reviews focus on the different reward formulations in the literature [15–17]. Some of the reviews only cover RL as one of many potential control techniques and do not provide an analysis that would systematically assess the reviewed works with respect to some aspects of Figure 1 [12,13].

The most relevant previous reviews are identified as [9,15–17]. In [9], authors consider RL as one of several machine learning techniques applicable to HVAC systems. The authors selected a set of articles that apply RL to simultaneously achieve energy savings while maintaining an acceptable indoor environment. The focus of the analysis is to identify the different state variables and how frequently each variable has been used in the selected set of articles. In [15], the authors review RL applications to demand a response, and identify four major categories of demand response resources. One of these categories are HVAC systems. For each paper, the authors identify the type of HVAC resource and the type of demand response mechanism. The authors observe that most works use RL to perform a trade-off between energy efficiency and occupant comfort. They further identify two general approaches: some works specify hard constraints for the quality of the indoor environment, whereas others permit the RL agent to temporarily drive the environment out of the comfortable zone in case the energy-efficiency related benefits are sufficiently high. In [16], works are categorized according to a primary and secondary optimization objective. The primary objective is either energy cost or energy consumption minimization. The secondary objective is usually thermal comfort, but in some cases, indoor air quality of lighting comfort is being optimized. In [17], authors elaborate on the various dimensions of the indoor environment that can and should be included in RL optimization. In addition to thermal comfort and indoor air quality, authors identified a number of factors considered only by the minority of the research, namely, occupancy schedules, shaping occupant behaviour, occupant feedback and lighting comfort. In summary, all of these prior reviews are structured around the optimization objectives for the RL agent. Each of them identified a common tradeoff of energy efficiency or cost minimization that must be balanced with respect to maintaining an acceptable indoor environment. The metrics for the indoor environment and the available information from the building that is available as state information for the RL agent is most comprehensively discussed in [9,17].

None of the reviews assess how the action space has been constructed and how this impacts the possibilities to apply the RL controller in a BACS (Building Automation Control System) context. The action space of a RL controller can have one or more outputs, and a single RL agent may control several zones of a building. Thus, a RL controller could have a similar input/output structure to SISO (Single Input, Single Output) controllers typically found in the basic control layer of a BACS or to MIMO (Multiple Input, Multiple Output) controllers typically found in the supervisory control layer of a BACS [18]. However, the works analyzed in this paper generally do not state whether they belong to one of these two layers or whether they partially or completely implement both layers. It is difficult for the reader to position the works in a BACS context, since most authors are vague about whether the RL actions should be mapped to actuators, setpoints of the basic control layer or something else. In the latter case, a critical analysis of the literature is required to determine if the chosen level of abstraction has retained key characteristics of the HVAC systems that need to be considered in the optimization, and if the chosen action space can support further work aiming at deployments. To assess these issues, the focus of this review is on the action space formulations.

The aim of this review is to identify the action formulations in the literature and to assess how the choice of formulation impacts the level of abstraction at which the HVAC systems are considered. The objective of the review is to organize the literature according to the action space formulation and to assess how this formulation impacts the modelling of HVAC systems and the possibilities to interface to developed RL agents to real HVAC systems.

The paper is structured as follows. Section 2 presents the methodology for searching and categorizing the articles. Section 3 provides an overview of the articles. Section 4 analyzes the articles in detail. The subsection headings of Section 4 are according to the categorization presented in Section 2. Section 5 concludes the paper with a summary and discussion of the main research gap being filled by this review.

2. Methodology

The following search string was used in the Web of Science database:

"reinforcement learning" AND (heating OR ventilation OR "air conditioning" OR cooling OR HVAC)

The search was limited to articles published since 2013.

In total, 13 review articles were found. One of these was not related to HVAC. Ntakolia et al. [19] was ignored since it focused on district heating systems without addressing HVAC. Dong et al. [20] discuss the applicability of RL for modelling occupant behavior in buildings; such techniques could be applied in the environment of Figure 1, but the paper does not identify such applications. Reviews that investigated control or machine learning research broadly, with only a very brief treatment of RL articles, were also ignored [21–23]. The remainder of the review articles have been discussed in the introduction Section 1.

The search string resulted in 278 articles, all of which has been assessed manually for inclusion in this review. The scope of this review is the applications of RL to manage HVAC systems for the purpose of controlling the indoor environment of buildings. Specifically, the following criteria were defined to exclude articles from this review:

- Articles were excluded if they discussed RL as a potential technology but did not present a solution based on RL (e.g., [24]). Applications of RL to clone other kinds of controllers such as model predictive controllers are also excluded [25].
- An article is excluded if a backup controller is used to override the RL agent, in case the agent would take an action that would violate indoor environment requirements [26–30].
- An article was excluded if it did not provide sufficient details to determine whether the action space was binary, discrete or continuous [31–34].
- Most works use a model of a building and HVAC as the environment for training the RL agent. Usually, this model is made in a building energy simulator such as

EnergyPlus, which has the capability to model heat transfer between adjacent building zones. In case a self-made building simulator was used, and it was not clear whether it had such a capability, the article was excluded [35].

- Although HVAC energy consumption forecasting is usually done with supervised learning time-series forecasting techniques (e.g., [36–38]), a few authors have used RL for this purpose [39,40]. Articles about forecasting were not selected.
- Occupant behavior is relevant to the environment in Figure 1. Approaches for using RL to model this aspect of the environment have been excluded from this review [41].
- This review is limited to HVAC applications in buildings, so other kinds of HVAC applications for systems such as batteries [42], seasonal thermal energy storage [43], vehicle cabins [44], fuel cells [45] and waste heat recovery from engines [46,47] are out of scope.
- HVAC solutions for managing the waste heat of ICT systems are in scope of this review only if they focus on the building that houses the HVAC systems (e.g., [48]). In this case, the management of the indoor environment is concerned with ensuring the lifetime of the server hardware. Solutions focusing on the internals of servers (e.g., [49]) are out of scope. Solutions that did not directly optimize the indoor temperature or other environmental variables are out of scope (e.g., Linder et al. [50] minimize total cooling airflow). It is necessary to scope this review with respect to what kind of structure housing ICT equipment is considered a building. This scoping decision was done so that data centers were considered buildings, but edge data centers [51–53] are out of scope.
- Management of district heating networks is considered out of scope [54,55], unless it involves the management of the indoor environment within the end user buildings [56]. With respect to geothermal energy, there are two approaches: to distribute the geothermal energy extracted from the wells through a district heating network [57,58] or directly to a heat exchanger in a building [59]. This article only considers systems inside buildings.
- Solutions for optimizing the operation of HVAC equipment are considered out of scope if there is no application to manage the indoor environment [60,61], or if the problem formulation is constructed so that the indoor environment is not affected in any way [62–64]. Approaches that penalized the RL agent for failing to meet heating or cooling demand were excluded if the penalty was not expressed in terms of indoor environmental variables [65]. Also, solutions for controlling physical phenomena such as thermal convection is out of scope [66] if the work is general and not applied to a building context.

Figure 2 presents the articles that were selected for inclusion in this review, organized by the publisher and publication year. A steady growth in publication activity is observed. This figure, and all the other charts in this paper, are based on manually selected articles and metadata that was collected from them. The metadata includes the publisher, year of publication, country of affiliation of the first author, as well as a 3-tier categorization that is described next.

A 3-tier categorization was applied to each article. The subsection structure of Section 4 is organized according to these tiers. Tier 1 captures how the HVAC solutions considers the possible existence of multiple zones in the building. Table 1 presents the tier 1 categories. Tier 2 captures the type of control being performed: whether the RL agent selects an action or determines a value for an action. Table 2 presents the Tier 2 categories. The Tier 2 categorization was performed by examining the action space formulation. Tier 3 captures the variables of the indoor environment that are being managed by the RL. These variables are temperature, humidity and air quality. Each combination of one or more of these variables is a Tier 3 category. Table 3 presents all such combinations that were encountered in the reviewed articles. A more detailed discussion of variables of the indoor environment is presented in Section 3. The tier 3 categorization was performed by examining the reward formulation.

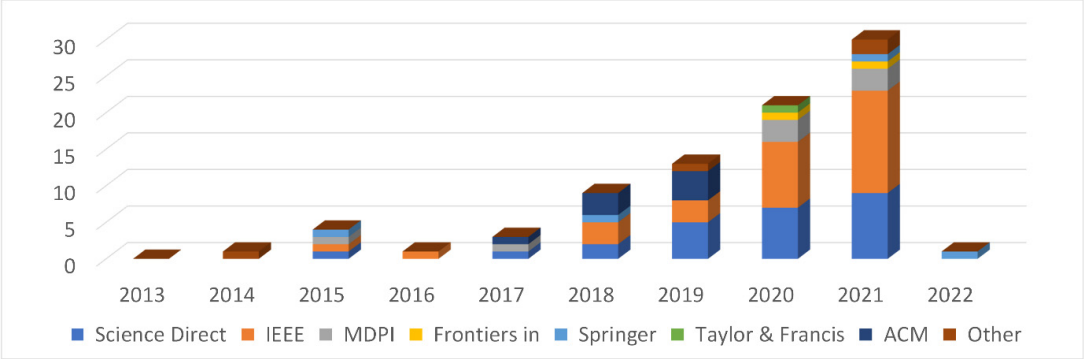


Figure 2. Articles included in the review.

Table 1. Tier 1 categories.

Tier 1 Category	Description
Single zone	The authors assume that indoor environment is uniform within the zone. In some cases, the zone is a room. In other cases, the zone can be an entire building, modelled at a level of abstraction at which conditions such as temperature are uniform throughout the building.
Independent zones	Actions taken to adjust the indoor environment in one zone do not impact the indoor environment in other zones. Heat transfer between zones is either not modelled or not possible due to the zones not being adjacent. There is no shared HVAC equipment.
Interdependent zones	Actions taken to adjust the indoor environment in one zone may impact the indoor environment in other zones, due to heat transfer between adjacent zones or due to shared HVAC equipment, for example.

Table 2. Tier 2 categories.

Tier 2 Category	Description
Binary	The RL agent selects between on/off actions or increase/decrease actions. In the latter case, the agent may also have the option of doing nothing.
Discrete	The RL agent selects one of several possible actions, such as one of several possible setpoint values.
Continuous	The RL agent determines a continuous value for one or more actions.
Hybrid	A combination of two or more of the following: binary, discrete or continuous.

Table 3. Tier 3 categories.

Tier 3 Category
Temperature
Temperature & humidity
Air quality
Temperature & air quality
Temperature, humidity & air quality

3. Overview of the Analyzed Articles

In general, the works discussed in this section use RL for multi-objective control. Common objectives involving HVAC and other distributed energy resources are renewables time shifting [67], price-based demand response [1], incentive-based demand response [68,69], electricity bill minimization under a real-time pricing scheme [70] and maximizing self-consumption of rooftop photovoltaic generation [71]. In some cases, one of the objectives

relates to the quality of the indoor environment, and such works are in the focus of this review. Unlike basic building automation systems, RL can simultaneously consider some or all of the following information: current and future weather (e.g., [72]), current and future electricity prices (e.g., [73]), occupancy (e.g., [74]), demand response programs (e.g., [26]), as well as several variables related to the indoor environment (listed in Table 3). Due to the thermal storage capacity of building structures, or phase change materials that have been installed for this purpose [75], and the possibility to allow small deviations in the indoor environment, RL opens opportunities for novel optimization approaches. As has been explained in Sections 1 and 2, some reviews have already been published in this area, and the focus of this paper is on the approaches that authors have taken to manage the indoor environment, and how these approaches are reflected in the formulation of the action space of the RL agent.

With respect to indoor environment, most of the works in this review only consider thermal comfort. A minority of authors use standard metrics for thermal comfort, such as Predicted Mean Vote (PMV), Predicted Percentage of Dissatisfied (PPD) [9], or Standard Effective Temperature (SET) [3]. These are computed based on factors such as temperature, humidity, clothing insulation and metabolic rate. In practice, the only factors that can be controlled by the RL agent are temperature and humidity, so default values are generally assumed for the other factors. Often a building energy simulator that is used to implement the environment of the RL agent is used to compute PMV or PPD. However, most researchers use a simplified and proprietary definition of thermal comfort. Frequently, an explicit definition is not provided or justified, but it can be discovered by examining the formulation of the reward of the RL agent. The most common approaches involve computing the deviation from an ideal temperature value or the deviation from a minimum or maximum temperature limit. For achieving uniform thermal comfort within a room, the variation of temperature across the room is minimized [76]. A more unusual approach is to consider thermal comfort to be satisfied if the HVAC is always on whenever indoor temperature is out of bounds [35]. In addition to thermal comfort, indoor air quality is an important aspect of the indoor air environment, with CO2 levels being the most obvious and common control target. In our review, the great majority of works ignored air quality. The control objectives impact the level of detail that should be captured in the environment. Building energy simulators are a very common approach for implementing the environment (e.g., [6]). An alternative approach is to construct a data-driven black box model of the building and its HVAC systems (e.g., [7]).

In addition to buildings occupied by humans, data centers have become a major focus of HVAC research due to their high energy consumption. Several authors have applied the concept of thermal comfort to datacenters, arguing that occasional minor deviations from ideal temperature and humidity have a tolerable impact on the lifetime of server hardware. Thus, they have applied RL to achieve a good tradeoff between HVAC energy consumption and indoor environments. None of these authors referred to any standard for datacenter indoor environment, and all of them proposed original formulations for acceptable deviations to indoor temperature, and in some cases humidity also. These works are included in this review. Two approaches emerged from this literature. In data centers with hot and cool zones, the RL agent could be used both for allocating the heat generating computational tasks to cooler zones and to control HVAC. Authors who did not model hot or cool zones were only concerned with HVAC control.

According to the majority of the reviewed works, building automation systems often employ a fairly long control time step, e.g., 15 min, so many authors use this as the control time step for the RL agent. The interaction between the RL agent and the environment in Figure 1 occurs once per control time step. When a building energy simulator is used to implement the environment, the simplest approach is to set the simulation time step equal to the control time step (e.g., [77]). However, to ensure an accurate simulation, some authors run the simulator with a shorter timestep, e.g., 5 min, in which case the simulator takes several steps adding up to 15 min before the next interaction with the RL agent is

performed. Some works use longer control timesteps such as 20 min [78] or 60 min [79]. Depending on how the environment has been implemented, and how long it takes for the system to stabilize after a control action, a shorter control time step could result in unstable feedback to the RL agent [80]. It is notable that these timesteps may be too long for some specific applications, such as adjusting the ventilation of a room in response to occupancy changes. Further research is needed to determine suitable timesteps for such applications.

Figure 3 presents the number of articles in each of the Tier 1 categories.

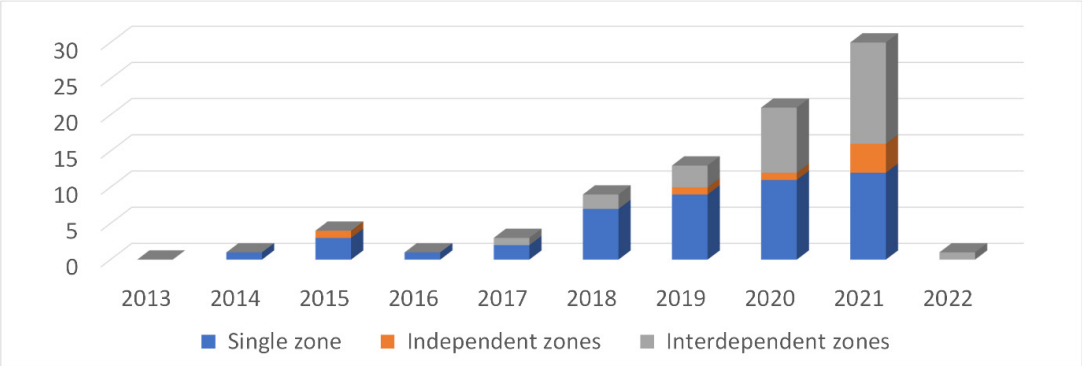


Figure 3. Number of articles in Tier 1 categories.

Figure 4 presents a sunburst chart of the 3-tier categorization of the articles.

Figure 5 presents a Sankey chart, in which the bar on the left is the country of affiliation of the first author, and the bar to the right is the Tier 1 category under which the article was categorized. The width of each flow is proportional to the number of articles that authors in the country on the left hand side of the flow contributed to the category on the right hand side of the flow.

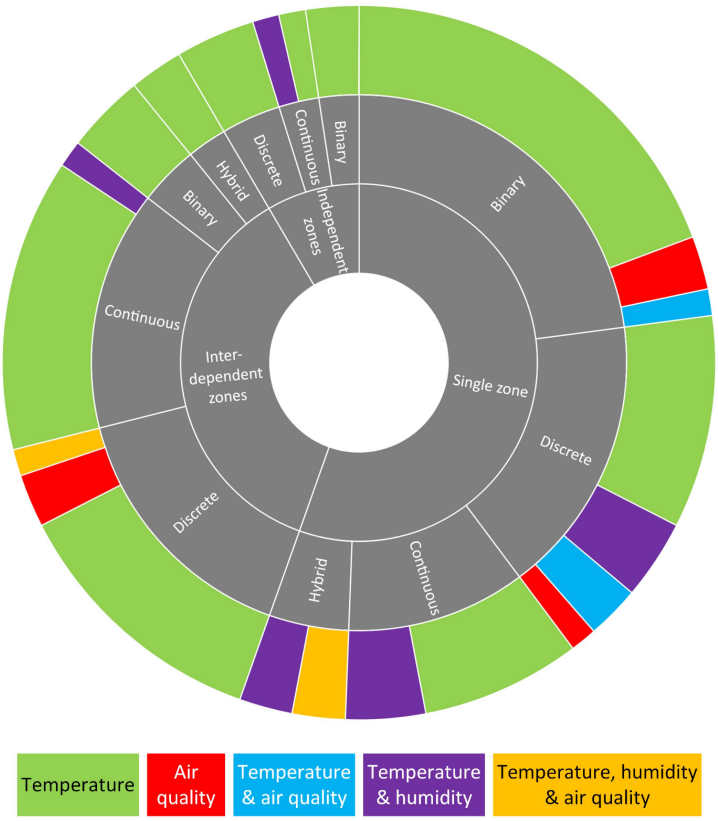


Figure 4. Sunburst chart of the 3-tier categorization of the articles.

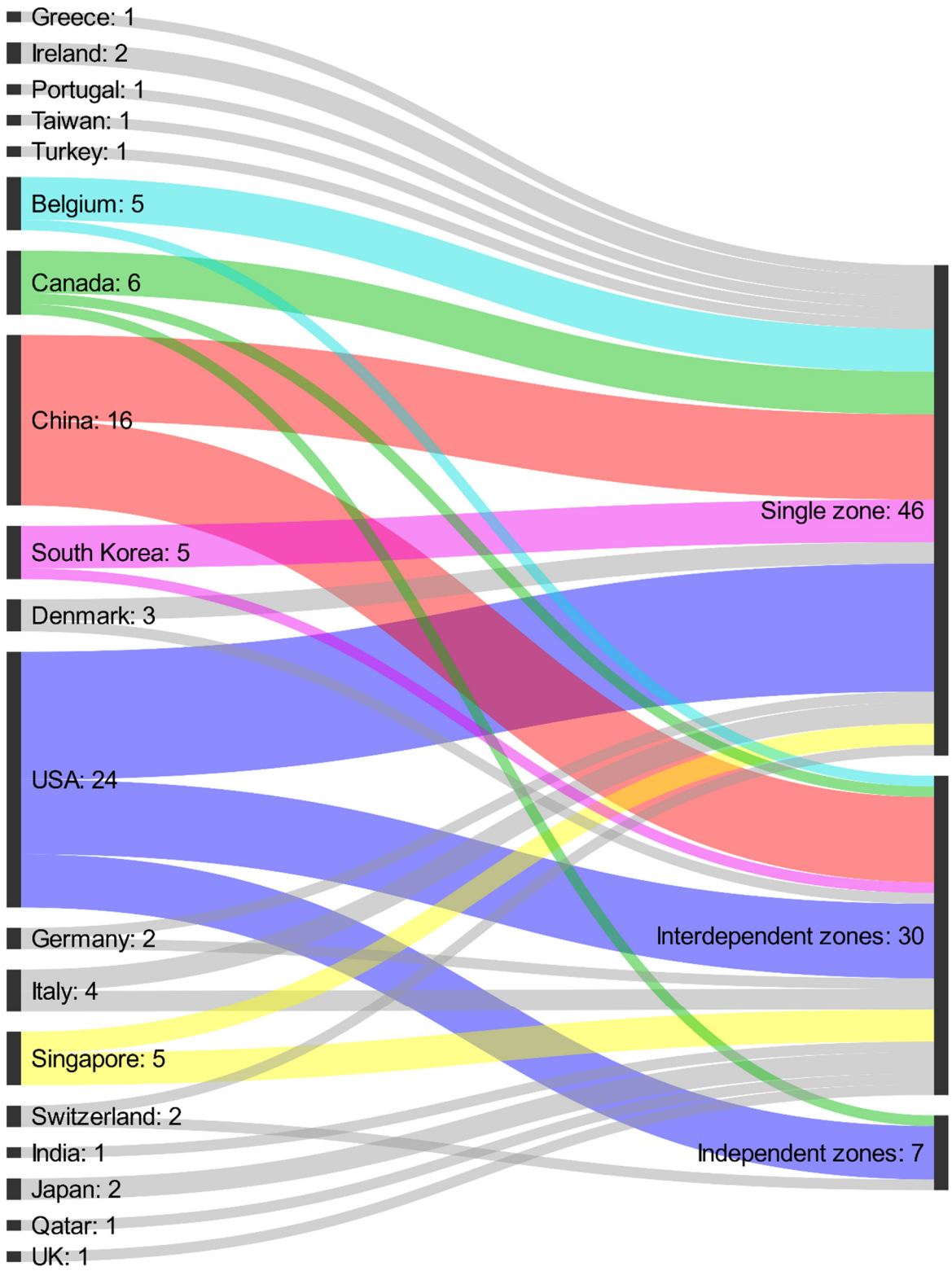


Figure 5. Sankey chart relating the country of affiliation of the first author (left) to the Tier 1, Tier 2 and Tier 3 categories.

4. Categorization of the Selected Articles

4.1. Single Zone

Figure 6 provides an overview of the works in the single zone category.

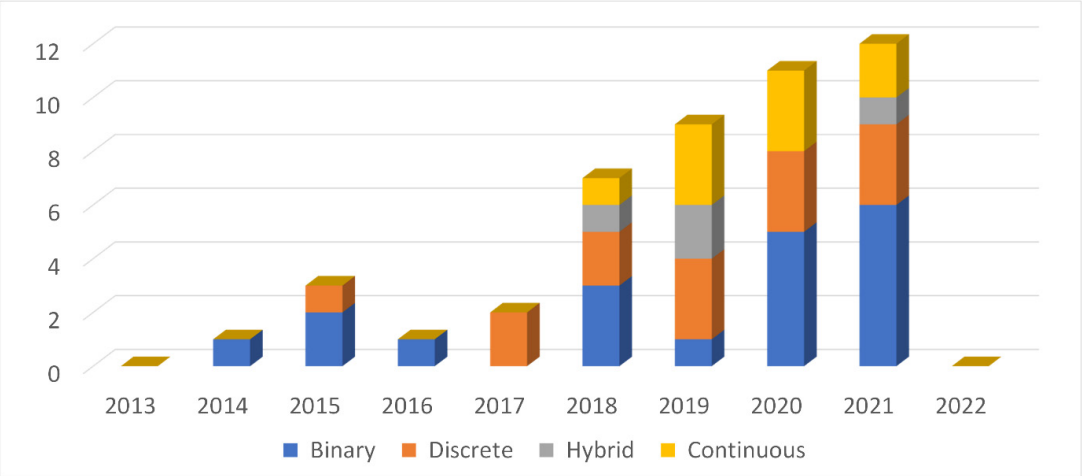


Figure 6. Distribution of single zone works by Tier 2 categories.

4.1.1. Binary

Temperature

The simplest RL HVAC controller makes on/off control decisions for an HVAC system that heats or cools a room. Some authors do not elaborate on any details of what kind of HVAC equipment is used, so the output of the RL agent is simply to turn the heating on or off [8,81]. Other authors specify the type of HVAC equipment, such as valves for underfloor heating systems [82], heat pump [83] or valves for chilled water flow [84]. Others consider the special characteristics of available HVAC equipment, such as a heat pump and an auxiliary heating element, aiming to avoid activation of the less energy efficient auxiliary heating element [85]. In some cases, a single zone may have a heating and a cooling element, both of which support binary control, resulting in four possible combinations [86]. However, there is no situation in which it makes sense to turn on the heating and cooling at the same time, so other authors have simplified the action space to three possible actions: off, heating on and cooling on [87]. In addition to penalizing temperature deviations from a setpoint, frequent switching of the on/off heating element can be penalized if it is perceived as annoying to the occupant, due to the noise involved, for example [88]. An alternative approach to binary control is to allow the RL agent to add or subtract a fixed value from the temperature setpoint [89–91]. A double binary control approach involves all the possible combinations of on/off control of two heating elements in a hot water tank [92]. Usually, binary actions from the RL agent are sent directly to on/off actuators, but another approach is to map the action to a low and high temperature set-point of a thermostat [93].

Another approach for binary temperature control involves user input for increasing or decreasing temperature. The goal of minimizing energy consumption is balanced with the goal of thermal comfort when the user is present. The RL tries to anticipate the periods of occupancy. With this problem formulation, thermal comfort is defined as a temperature at which the user's preference is satisfied. The preference is satisfied if the user does not use the user interface to request changes [2,94].

Air Quality

Indoor $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than 2.5 μm) concentrations are managed in a naturally ventilated building by an agent that takes a binary decision to open or close a window. The problem formulation is simple: the goal is to minimize indoor $PM_{2.5}$ and decisions are taken based on sensor information on indoor and outdoor $PM_{2.5}$ [95].

Temperature & Air Quality

Natural ventilation without HVAC can be used to manage both indoor air quality and indoor temperature. Especially in some large cities, concentrations of outdoor pollutants may be so high that opening the window decreases indoor air quality. In the absence of HVAC, a RL agent that decides to open or close a window is the simplest solution for achieving the best tradeoff between indoor air quality and thermal comfort [96]. Another solution combining windows and HVAC uses two possible actions for windows (open or closed) and three possible actions for HVAC (AC on, heating on, HVAC off). This results in six possible combinations, but since the system will never open the windows with the heating on, the action space consists of only five possible actions [78].

4.1.2. Discrete

Temperature

A common category of applications is a single zone temperature control, in which the RL agent selects a temperature setpoint from several predefined alternatives. The number of actions is equal to the number of possible setpoint values. An additional action may be included for turning the HVAC off [6,97]. As there are no other actions in single zone temperature control, the action space remains very manageable for RL techniques and it is practical to have many possible values for the setpoint; for example, Lork et al. use 15 actions [7]. In most cases, the setpoint is indoor temperature, but in the case of district heating, the action of the RL agent may involve setting the supply water setpoint [6]. A similar approach is applicable for water heating systems supplied by a boiler. In these cases, the indoor temperature setpoint is not affected by the action, but indoor temperature can be used in the reward function to quantify thermal comfort [98]. When RL is used to set the indoor temperature setpoint, it is assumed that a lower-level controller exists for generating the control signals to the HVAC equipment. Thus, the level of detail captured in the environment should be appropriate with respect to the control objectives. As thermal properties of buildings and the behavior of HVAC systems are complex, a black box data-driven model can capture this complexity if sufficient data from the building is available [7]. If electricity prices vary during the day, it is possible to exploit lower prices to preheat or precool the building by using the buildings thermal mass as a passive energy storage. An environment that is constructed in a building energy simulator can capture these dynamics. In these cases, thermal comfort is defined in terms of a minimum and maximum indoor temperature, so the RL agent is penalized for going out of this range. An application of this approach in hot climates is to precool the building during low energy price periods to minimize the AC energy cost [99]. An alternative for using the building as a passive heat storage is to have a dedicated water tank for this purpose [100].

The majority of works in this category involved selecting a value for the temperature setpoint. However, Overgaard et al. [101] select one of three possible values for pump speed in a mixing loop connecting district heating pipelines to the terminal unit of a building's floor heating system.

Wei et al. [102] reject established concepts of thermal comfort and argue that a more relevant temperature control target in an office environment is to reduce the sleepiness of the occupants. The facial image of an occupant is processed to extract the eyes and to detect the level of sleepiness. The RL agent is rewarded for reducing sleepiness, which it does by selecting the temperature setpoint of the incoming air flow. As the control time step is 15 min, it is unclear how changes in the facial expression during this time should be considered. The method considers only one occupant, who is expected to sit in a predefining and well-lit location.

Temperature & Humidity

Works that consider both temperature and humidity usually involve sophisticated measured of thermal comfort such as PMV [103] or PPD [104], which are calculated by

the building energy simulator used as the environment. Qiu et al. [80] use wet bulb temperature, which is a function of temperature and humidity.

Air Quality

To keep CO2 levels close to a setpoint, the RL agent chooses a percentage of the maximum ventilation rate of a HVAC device. The type of device is not specified in more detail. There are 14 possible discrete percentage values to choose from [105]. A similar approach with a more detailed description of HVAC equipment involves an underground subway station. Concentrations of particulate matter with an aerodynamic diameter less than 10 µm (PM_{10}) are managed. A variable frequency drive is used to control the ventilation fans, and the control action involves choosing one of 7 possible values for the frequency [106].

Temperature & Air Quality

An RL agent determines a power setpoint for a heating/cooling system and the ventilation air volume of a ventilation system. There are K_1 choices for the heating/cooling system and K_2 choices for the ventilation system, resulting in $K_1 \times K_2$ actions to choose from. The authors do not elaborate what kind of HVAC system is used that can provide both heating and cooling with a single power setpoint. The agent is penalized both for temperature violations and CO2 violations [4].

4.1.3. Continuous

Temperature

The simplest RL agent taking continuous actions implements a SISO temperature controller of a single zone, so that the action is one continuous control signal to control, for example, the supply water temperature of a radiant heater [72,77,107] or supply airflow to a VAV unit [108]. In some cases, authors only define a power consumption demand without elaborating on how this consumption would be split between equipment such as compressors, fans and pumps [11]. An alternative problem formulation assumes that indoor temperature in a room has a 'schedule', which defines the ideal temperature as a function of the time of day. The reward formulation minimizes deviations from this schedule [109].

Temperature & Humidity

If humidity is considered in the thermal comfort measure, a straightforward approach for managing comfort is for the RL agent to adjust HVAC setpoints for humidity and temperature [110]. Other authors have developed solutions that target specific type of HVAC equipment. The discharge temperature of an air handling unit is controlled with the objective of maintaining relative humidity at 50% [111]. Free-cooled datacenters involve management of temperature and humidity by adjusting airflow by means of opening positions of supply, mixing and exhaust dampers [112].

4.1.4. Hybrid

Temperature & Humidity

A solution is presented for a single room with a variable refrigerant flow (VRF) system and a humidifier, aiming at thermal comfort in terms of PMV. The discrete action space has 8 possible temperature setpoint values and an on/off value for the VRF, 3 setpoint values for the VRF air flow and a binary control for the humidifier. VRF systems are frequently used for building with multiple rooms, in which case there could be issues with the scalability of the address space [113]. A similar approach was used with a system consisting of an air conditioner, humidifier and ventilation system. Each of these had an on/off control and the air conditioner additionally had 3 possible setpoint values for indoor temperature and another 3 possible values for air flow rate, resulting in an action space of 40 actions [3], which also could have scalability problems for multi-zone buildings.

Temperature, Humidity & Air Quality

Two articles were found in which both binary and discrete control of various HVAC devices was used to control temperature, humidity and air quality. In the first article, a total of N discrete temperature setpoints are available to the RL agent, as well as a binary control action to turn the ventilation on/off, resulting in an action space with 2N actions [114]. In the second article, an RL agent controls the window opening, ventilation and awning. The awning supports binary control, the window supports 3 discrete setpoints and the ventilation supports 4 discrete setpoints. This results in $2 \times 3 \times 4 = 24$ possible combinations, but the authors manually excluded undesirable combinations, resulting in a total of 10 combinations, so the action space consists of these 10 options [115].

4.2. Independent Zones

Two approaches emerged in this section. One is to use a central agent to manage several zones, and another is to use one agent per zone. Figure 7 shows an overview of the works in the section.

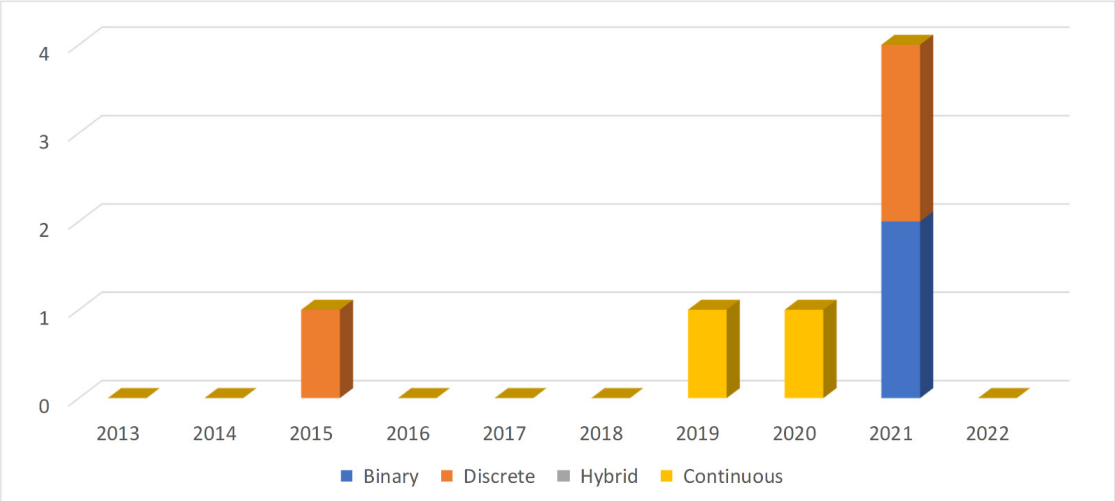


Figure 7. Distribution of works on independent zones by Tier 2 category.

4.2.1. Binary

Temperature

A straightforward extension of the single zone approach is to have one central agent controlling several zones, or several single zone buildings. The advantages of this approach are unclear if the conditions in one zone do not impact any of the other zones. Two examples of this approach are as follows. A total of six residential buildings are considered with one binary action per building: turning a heater on or off. A centralized controller decides the action for each building, resulting in an action space with 2^6 actions [116]. A very similar problem occurs in the case of air conditioning, with four zones and the RL determining a binary control signal for the air conditioner in each zone [117]. Thus, having a single agent manage several independent zones can be disadvantageous for reasons of computational complexity.

Having a central RL solution for managing several independent zones can be advantageous if advanced RL techniques are used to accelerate training of the RL agent(s). For example, domestic water heating for houses is considered, with a single RL agent in each house taking binary decisions for controlling the heating element of a hot water tank. The houses are independent and have different system states affected by occupant behavior. However, the houses and their heating systems are identical, so a multi-agent RL system is used to accelerate learning through faster state-space exploration, making use of the experiences that are obtained from all of the houses. The authors explicitly propose their

method to housing communities or other groups of houses in which these assumptions can be made [118].

4.2.2. Discrete

Temperature

A straightforward approach is to have one agent per zone, which keeps the action space manageable even if there are many zones, since the number of actions does not grow exponentially with respect to the number of zones. In case of a VAV system, each agent controls the air flow rate to the zone, which has several possible discrete values [119]. Another work managed indoor temperature and hot water tank temperature, with the RL agent choosing from a discrete set of heating power setpoints. The system is replicated to several houses, with one agent per house [120].

An uncommon problem formulation involved the use of several diverse agents in a low exergy building. Heating is provided by a ground source heat pump and solar thermal collector. RL is used to determine the mass flow rate of the water circulation loops to the solar thermal collector, the boreholes and the floor heating. 3 possible setpoint values are available for the solar thermal loop and 11 possible values are available for the other two loops. Three separate agents are trained independently to control each of these loops, and the authors justify this by explaining that each agent has independent goals [74]. This justification is questionable, since there is a shared environment that each agent affects through its actions, so a multi-agent approach could have been more appropriate.

4.2.3. Continuous

Temperature

A data center building with two independent zones with dedicated HVAC systems is considered. In each zone, a temperature setpoint and a supply air mass flow rate is adjusted by the RL agent. The authors do not discuss possible advantages of using a single agent to control both zones [121].

Temperature & Humidity

Historical building automation system data is used to train a data-driven model of the environment of the building and its HVAC. The objective is to minimize consumption and to optimize the thermal comfort in terms of PPD if the building is occupied. There are three AHUs with the following continuous actions: damper position, valve status for the pipes supplying the heating and cooling coils and fan speed. A separate RL agent is trained for each AHU. The agents are independent [122].

4.3. Interdependent Zones

Figure 8 provides an overview of the works in this category.

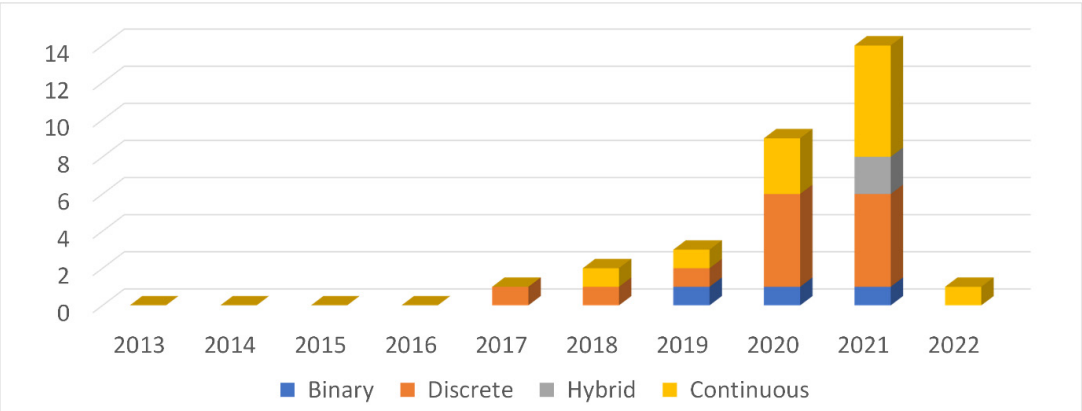


Figure 8. Distribution of works on interdependent zones according to Tier 2 categories.

4.3.1. Binary

Temperature

Binary outputs from a RL agent can be used to achieve a control similar to continuous control. This can be done by adding/subtracting a predefined small value to/from a temperature setpoint. The action of the RL agent is interpreted as add or subtract. The size of the action space is 2^N, where N is the number of thermally interdependent zones [123].

An unconventional problem formulation for interdependent zones involves separate buildings in an isolated microgrid. The microgrid has a virtual tariff, so the RL agents have a financial interdependency in their environment. This has been exploited to coordinate the control of hot water tanks in the buildings. A reward function has been defined to permit temperature fluctuations in an electric water heater tank, penalizing for deviations from a temperature preferred by the users [124].

4.3.2. Discrete

4.3.2.1. Temperature

The selection of discrete setpoint values in a building with multiple interdependent zones builds on the research in the single zone case. A separate setpoint is selected for each zone, for example to react to occupancy patterns or individual occupant preferences or to prioritize comfort in different parts of the building. In most cases, a single RL agent will be used to make the decisions for all zones, which leads to scalability issues. Consider a basic case with only two zones [125]. The acceptable user comfort is defined as a 2 °C interval, with 5 different setpoint values in this interval. For two zones, there are $5^2 = 25$ possible combinations, and each of these combinations is an action. Thus, the output layer of the neural network used to implement the RL agent has 25 nodes. This approach has scalability issues, since a building with 10 zones would require $5^{10} = 9{,}765{,}625$ nodes in the output layer, and a building with 20 zones would require $5^{20} = 95{,}367{,}431{,}640{,}625$ nodes. Another similar approach involves two zones and two possible setpoint values, resulting in $2^2 = 4$ actions [126]. Although the scalability issues are not as severe with only two possible setpoint values, there will be problems with large multi-zone buildings, since with 10 zones, the size of the output layer would be $2^{10} = 1024$, and with 20 zones it would be $2^{20} = 1{,}048{,}576$. The above-mentioned works only consider two zone buildings and do not discuss scalability. A general multi-zone approach with N zones and 4 different setpoint values is presented by Yuan et al. [127]. This results in 4^N actions, and the authors admit that with an already n = 4, a significantly longer training time was observed as a practical issue that made the research more laborious. The general case of m possible setpoint values and z zones results in an action space of m^z actions [128,129].

The interdependency of multiple building zones is due to physical dependencies between the zones. One such dependency is the heat transfer between adjacent zones. This is especially relevant in works that use the solid structures of the building as a passive thermal storage, so that the RL will precool or preheat the building during low price periods [79]. Another dependency is due to an AHU that serves several zones. Such papers model the AHU and VAV systems in some detail using a residential house model [126] or a building energy simulator [127] as the environment. To achieve uniform temperature across a room, Sakuma & Nishi [76] use computational fluid dynamics simulation software; the RL agent controls fan directions, so the action space has three actions for each fan: left, center and right. The great majority of papers do not consider how the RL trained in such virtual environments would perform when deployed to a physical building, and they do not discuss whether it would be necessary to develop a custom environment for each building. However, a few authors have investigated this with a generic house model that was used as the environment, after which the RL agent was deployed to a unique physical house, in which it was able to achieve nearly as good energy cost savings as in the training phase [126].

An alternative RL problem formulation involves choosing actions to change the behavior of occupants, or to relax the thermal comfort requirements for spaces with low

occupancy. This approach has been simulated in a multi-zone environment with dedicated HVAC for each zone. No thermal interdependency between zones has been simulated. However, since the RL agent's actions are recommendations for occupants to move between zones, and since the authors define thermal comfort requirements in terms of occupancy, the zones are interdependent [130].

In most works categorized under the multi-zone case, the zones are located in the same building. However, in the case of households connected to a district heating networked powered by a central combined heat and power production plant, the thermal interdependency occurs through the district heating network. One such work was encountered, in which each household has an energy consumption profile specifying hourly consumption. The RL agent takes actions to adjust these profiles in order to avoid consumption peaks at the central plant. The adjustments are done so that the agent is penalized for causing thermal discomfort, defined in terms of PMV [56].

4.3.2.2. Air Quality

An office building with 16 zones, two chillers and four air handling units was controlled by a single RL agent. Unlike the majority of works that were encountered in this review, the action space was not constructed with separate actions for each zone. Rather, the actions were damper positions for each air handling unit and setpoints for chilled water temperature and cooling water temperature for the chillers. Each action had 3 or 4 possible values, resulting in an action space with 972 possible combinations of values for these six actions [131].

4.3.2.3. Temperature and Air Quality

A straightforward application of single agent technology involves a school building with 21 zones of three different types: classrooms, offices, laboratories and a gym. There are 12 possible values for the temperature setpoint and 6 possible values for the CO2 setpoint [5], resulting in scalability issues as discussed in Section 4.3.2.1. A sophisticated work addressing such issues involved a multi-zone building, with AHU and VAV models in the environment. The AHU has a VFD powering a fan that supplies air to the VAV boxes and a cooling coil connected to a chiller. To keep the action space manageable, multi-agent RL is used. There is one agent with a control action for the air supply rate to each zone and one additional agent with a control action for the position of the damper at the inlet of the AHU. Each agent has a separate reward, which penalizes the agent according to its contribution to the fan and chiller power consumption as well as deviations from acceptable temperature and CO2 levels; the deviations are only penalized if occupants are present. As temperature in one zone affects neighboring zones, the agents share their observations for indoor temperature [132].

4.3.2.4. Temperature, Humidity and Air Quality

A straightforward extension of the approaches for temperature control discussed in Section 4.3.2.1 involves the joint control of HVAC and windows for control of temperature, humidity and air quality. PMV is used for thermal comfort and CO2 as a proxy for air quality [133]. The authors note severe scalability issues with the action space in a multi-zone setting, so they propose an original neural network architecture to mitigate the computational complexity.

4.3.3. Continuous

Temperature

Using continuous values for setpoints avoids much of the computational complexity encountered in Section 4.3.2. For example, the RL action can be the power percentage of the VAV unit in each zone of a building [134], air mass flow setpoint for a variable volume fan and a temperature setpoint for a cooling coil [135] or continuous values for valves that distribute cooling water to zones from a centralized chiller [136]. Further multi-zone

complexity involves considering diverse zones with different occupancy profiles. Such a solution is presented for the case of a centralized chiller and three zones [137]. Each of the zones is large, but the authors did not subdivide them further out of concerns for computational complexity. In the remainder of this subsection, novel and unconventional solutions are discussed in more detail.

A multi-agent system is used, so that one agent controls one zone with a single continuous action that is interpreted either as a cooling or heating power command, depending on whether its value is negative or positive. It is not discussed how this signal is mapped to underlying control loops or HVAC equipment [138].

A multi-zone building is considered, with thermal comfort in each zone being captured as a deviation from a reference temperature. The importance of maintaining comfort in each zone can be different, and this is expressed with a weighting factor in the reward function. The action of the RL agent involves setting the value of a tuning parameter that adjusts the relative weights of energy consumption and thermal comfort-related targets in the reward function [139].

A datacenter with two zones is modelled. One of the zones is cooled with a direct expansion system and the other one is cooled with a chiller. They are interdependent, as both systems use the same cooling tower. The intake airflow for the direct expansion system is precooled by two evaporative coolers: directive (DEC) and indirect (IEC). A single RL agent determines the values of 5 setpoints: DEC outlet air temperature, IEC outlet air temperature, chilled water-loop outlet water temperature, direct expansion cooling coil outlet air temperature (to zone 1) and chiller cooling outlet air temperature (to zone 2) [48]. An alternative approach for a two-zone datacenter involves the RL agent determining values for setpoint temperature and supply fan air mass flow rate for each zone [140]. A multi-zone datacenter approach involves assigning computational tasks to server racks and performing continuous adjustment to the airflow rate for each rack [141]. In summary, the application to data centers is similar to regular buildings from the RL perspective, unless the additional complexity of allocating computational tasks is included in the problem formulation.

Unless otherwise specified, the zones that are discussed in this paper are located within a building or a building complex. However, for power grid peak shaving, a similar RL approach can also be applied when the zones are buildings within the geographic area of the section of the grid that is being balanced [73,142]. In this case, the interdependency of the zones is due to their joint impact on the power grid.

Temperature & Humidity

A building with several zones is modelled with a building simulator that captures heat transfer through building structures. Temperature and humidity setpoints are adjusted for each zone by a dedicated agent. The agent is penalized according to energy consumption and thermal discomfort, which is quantified in terms of PMV. A multi-agent approach is used to minimize the sum of the penalties for each agent [143].

4.3.4. Hybrid

Temperature

Data center HVAC control ideally considers both the server load and indoor temperature conditions. The data center building is divided into different kinds of zones, such as cool, warm and hot zones, so the load is directed to servers in the cool zone when possible. The choice of server is a discrete action, whereas the temperature and flow setpoints are continuous actions. There are two approaches to handle this discrete-continuous action space. One approach is to use a two time-scale solution for load scheduling and HVAC control [144]. The other approach is to use a multi-agent architecture, with separate agents handling the discrete and continuous actions [145].

5. Conclusions

5.1. Summary

A minority of the works were categorized under independent zones. The entire motivation for the research in this category is questionable and poorly motivated since it would be equally possible to have a single agent for each zone. In other words, the solutions presented in the single zone Tier 1 category could simply be replicated to each zone. In this case, the computational load is proportional to the number of zones. However, in the independent zones case, when one agent is used to control several independent zones, this can result in an exponential growth of the action space, especially if the action space is discrete. Thus, the independent zones problem formulation has clear computational disadvantages. With one exception, the reviewed works did not present arguments about why the independent zones formulation would be advantageous. The exceptional work considered identical houses and heating systems, and it applied a multi-agent RL system to accelerate learning through faster state-space exploration, making use of the experiences that are obtained from all houses. As can be seen in Figure 3, only a few authors are working in this area, which is an indication of it being a less promising line of research.

According to Figure 3, the focus of the research is shifting to problem formulations involving interdependent zones. As has been discussed in Section 5, this can result in action spaces that grow exponentially with respect to the number of zones. In a laboratory context, this situation can be navigated in a brute force way by using sufficient computational resources while limiting the number of zones used in the research. However, a more elegant and scalable approach would be to use multi-agent RL. Unfortunately, only a few of the reviewed works made use of this advanced technique, so it is unclear whether multi-agent RL will become a major research trend in HVAC systems.

Only some of the works provided a clear description of the ICT architecture in which the RL agent would operate. In some cases, the actions of the RL agent could be mapped directly to HVAC actuators, and generally such works did not elaborate on what impact the innovation would have to building automation systems. Most reviewed works involved changing the setpoints of control loops managed by building automation systems, positioning the RL agent into a higher-level system. Such a higher-level system could be a HEMS (Home Energy Management System), BEMS (Building Energy Management System) or VPP (Virtual Power Plant), but there is a lack of linkage to HEMS, BEMS or VPP. Since HEMS, BEMS and VPPs are subjects of active research, establishing stronger linkage to them would be a direction for further research.

In conclusion, the applications of RL to HVAC is a body of research that has experience continued significant growth over the last several years. Most recently, the growth has been driven by applications to interdependent zones, using RL approaches supporting continuous action spaces. The action space tends to grow rapidly as the number of zones increases, resulting in a computational complexity that can be a strain for high performance computing resources. This problem is partially addressed by using continuous instead of discrete action spaces. However, a more potential approach for solving this problem would be the use of multi-agent RL, which had very limited applications in the body of research reviewed in this article. Key recommendations for further research based on this review would be the application of multi-agent RL as well as better linkage to HEMS and BEMS systems.

5.2. Contribution

In Section 1, it was stated that the aim of this review is to identify the action formulations in the literature and to assess how the choice of formulation impacts the level of abstraction at which the HVAC systems are considered. Figure 9 illustrates the main gap being filled by knowledge gained in this review. The RL system from Figure 1 is presented on the left of Figure 9. On the right of Figure 9, key HVAC related systems of smart, energy-efficient buildings are presented. The gap in the literature is how the RL agent on the left can be connected to the systems in a physical building on the right. Based on our

review, three general approaches exist, and these are illustrated with the three dashed lines in Figure 9:

- **Control signal:** the action of the RL agent may be directly sent to the actuator through the building automation system's PLC (programmable logic controller). This applies to the 'binary' and 'continuous' action categories in Table 2. In case of a binary action, mapping the action to the binary I/O (input/output) of the PLC is straightforward. In case of a continuous action, an analog actuator must be used, and some scaling is required in the PLC according to the specification of the actuator. Some of the reviewed articles clearly specified an actuator such as a heating element or valve, so this approach is directly applicable to deploying the RL agent to a real building. However, some articles were more vague and just discussed turning heating or ventilation on or off, without specifying the actuator or type of HVAC equipment, even though several actuators may be involved in a large building.
- **Setpoint adjustment:** in some cases, the action of the RL agent can be directly connected to the setpoint input of a building automation control loop, such as a temperature control loop. This applies to the 'discrete' and 'continuous' categories in Table 2. However, for some articles using a continuous action space, it is not clearly stated whether the action should be mapped directly to the analog I/O or to the setpoint input of a control loop.
- **Planning:** in some cases, the RL agent takes actions to anticipate future situations. These are generally financial incentives to shift energy consumption to certain times of the day to benefit from tariffs, real-time electricity prices or demand response programs. Intelligent RL agents can learn the thermal dynamics of buildings and develop strategies such as precooling a dedicated cold storage or, in the absence of such a storage, using the building's thermal mass as a storage. The strategies respect the requirements for the indoor environment.

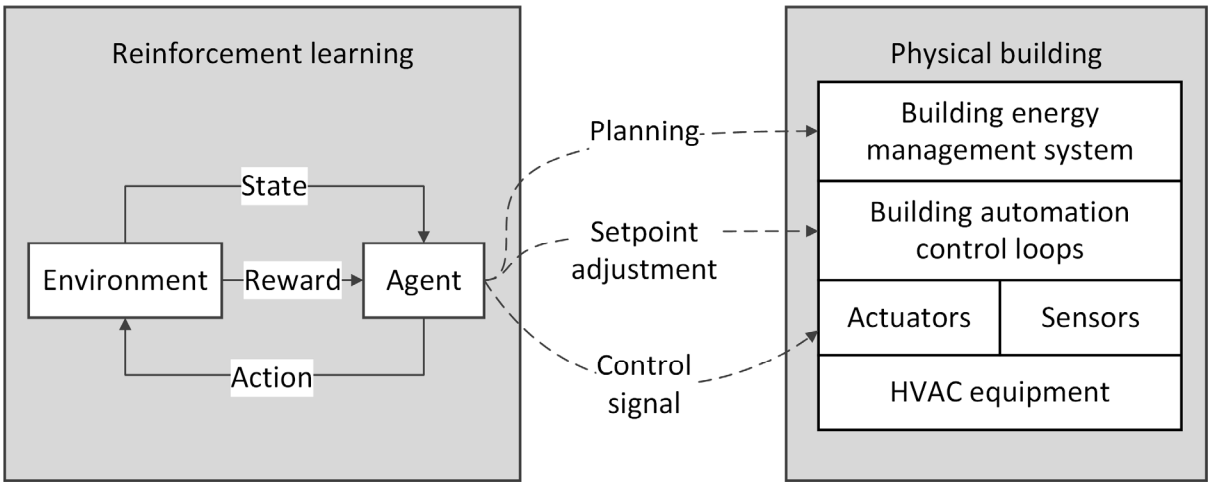


Figure 9. Main gap being filled by this review.

The main gap being filled by this review is illustrated by the dashed lines in Figure 9, which map the action output of the RL agent to different systems in the physical building. In general, this mapping has not been explicitly specified in the reviewed articles, although in some cases it is straightforward to infer. For the purpose of stimulating applied research, it is desirable that the authors specify this mapping explicitly. For the coherence of the body of research in RL applications to HVAC, it is desirable that the research community establish an understanding of whether the three mappings presented in Figure 9 are the only ones or if additional kinds of mappings exist.

Author Contributions: Conceptualization, S.S.; methodology, S.S.; software, S.S.; validation, S.S., H.I. and V.V.; formal analysis, S.S.; investigation, S.S., H.I. and V.V.; resources, S.S., H.I. and V.V.; data

curation, S.S.; writing—original draft preparation, S.S.; writing—review and editing, S.S., H.I. and V.V.; visualization, S.S.; supervision, V.V.; project administration, S.S.; funding acquisition, S.S. and V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Business Finland grant 7439/31/2018.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Azuatalam, D.; Lee, W.-L.; de Nijs, F.; Liebman, A. Reinforcement learning for whole-building HVAC control and demand response. *Energy AI* **2020**, *2*, 100020. [CrossRef]
2. Fazenda, P.; Veeramachaneni, K.; Lima, P.; O'Reilly, U.-M. Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems. *J. Ambient Intell. Smart Environ.* **2014**, *6*, 675–690. [CrossRef]
3. Kim, S.-H.; Yoon, Y.-R.; Kim, J.-W.; Moon, H.-J. Novel Integrated and Optimal Control of Indoor Environmental Devices for Thermal Comfort Using Double Deep Q-Network. *Atmosphere* **2021**, *12*, 629. [CrossRef]
4. Yang, T.; Zhao, L.; Li, W.; Wu, J.; Zomaya, A.Y. Towards healthy and cost-effective indoor environment management in smart homes: A deep reinforcement learning approach. *Appl. Energy* **2021**, *300*, 117335. [CrossRef]
5. Chemingui, Y.; Gastli, A.; Ellabban, O. Reinforcement Learning-Based School Energy Management System. *Energies* **2020**, *13*, 6354. [CrossRef]
6. Zhang, Z.; Chong, A.; Pan, Y.; Zhang, C.; Lam, K.P. Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning. *Energy Build.* **2019**, *199*, 472–490. [CrossRef]
7. Lork, C.; Li, W.-T.; Qin, Y.; Zhou, Y.; Yuen, C.; Tushar, W.; Saha, T.K. An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management. *Appl. Energy* **2020**, *276*, 115426. [CrossRef]
8. Faddel, S.; Tian, G.; Zhou, Q.; Aburub, H. On the Performance of Data-Driven Reinforcement Learning for Commercial HVAC Control. In Proceedings of the 2020 IEEE Industry Applications Society Annual Meeting, Detroit, MI, USA, 10–16 October 2020; pp. 1–7. [CrossRef]
9. Ma, N.; Aviv, D.; Guo, H.; Braham, W.W. Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110436. [CrossRef]
10. Li, H.; Wan, Z.; He, H. Real-Time Residential Demand Response. *IEEE Trans. Smart Grid* **2020**, *11*, 4144–4154. [CrossRef]
11. Yu, L.; Xie, W.; Xie, D.; Zou, Y.; Zhang, D.; Sun, Z.; Zhang, L.; Zhang, Y.; Jiang, T. Deep Reinforcement Learning for Smart Home Energy Management. *IEEE Internet Things J.* **2019**, *7*, 2751–2762. [CrossRef]
12. Afram, A.; Janabi-Sharifi, F. Theory and Applications of HVAC Control systems–A Review of Model Predictive Control (MPC). *Build. Environ.* **2014**, *72*, 343–355. [CrossRef]
13. Maddalena, E.T.; Lian, Y.; Jones, C.N. Data-driven methods for building control—A review and promising future directions. *Control Eng. Pract.* **2019**, *95*, 104211. [CrossRef]
14. Royapoor, M.; Antony, A.; Roskilly, T. A review of building climate and plant controls, and a survey of industry perspectives. *Energy Build.* **2018**, *158*, 453–465. [CrossRef]
15. Vázquez-Canteli, J.R.; Nagy, Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Appl. Energy* **2018**, *235*, 1072–1089. [CrossRef]
16. Yu, L.; Qin, S.; Zhang, M.; Shen, C.; Jiang, T.; Guan, X. A Review of Deep Reinforcement Learning for Smart Building Energy Management. *IEEE Internet Things J.* **2021**, *8*, 12046–12063. [CrossRef]
17. Han, M.; May, R.; Zhang, X.; Wang, X.; Pan, S.; Yan, D.; Jin, Y.; Xu, L. A review of reinforcement learning methodologies for controlling occupant comfort in buildings. *Sustain. Cities Soc.* **2019**, *51*, 101748. [CrossRef]
18. Aste, N.; Manfren, M.; Marenzi, G. Building Automation and Control Systems and performance optimization: A framework for analysis. *Renew. Sustain. Energy Rev.* **2017**, *75*, 313–330. [CrossRef]
19. Ntakolia, C.; Anagnostis, A.; Moustakidis, S.; Karcanias, N. Machine learning applied on the district heating and cooling sector: A review. *Energy Syst.* **2021**, *13*, 1–30. [CrossRef]
20. Dong, B.; Liu, Y.; Fontenot, H.; Ouf, M.; Osman, M.; Chong, A.; Qin, S.; Salim, F.; Xue, H.; Yan, D.; et al. Occupant behavior modeling methods for resilient building design, operation and policy at urban scale: A review. *Appl. Energy* **2021**, *293*, 116856. [CrossRef]
21. Yu, Z.; Huang, G.; Haghighat, F.; Li, H.; Zhang, G. Control strategies for integration of thermal energy storage into buildings: State-of-the-art review. *Energy Build.* **2015**, *106*, 203–215. [CrossRef]
22. Hasan, Z.; Roy, N. Trending machine learning models in cyber-physical building environment: A survey. *WIREs Data Min. Knowl. Discov.* **2021**, *11*, e1422. [CrossRef]
23. Thieblemont, H.; Haghighat, F.; Ooka, R.; Moreau, A. Predictive control strategies based on weather forecast in buildings with energy storage system: A review of the state-of-the art. *Energy Build.* **2017**, *153*, 485–500. [CrossRef]
24. Chen, J.; Sun, Y. A new multiplexed optimization with enhanced performance for complex air conditioning systems. *Energy Build.* **2017**, *156*, 85–95. [CrossRef]
25. Lee, Z.E.; Zhang, K.M. Generalized reinforcement learning for building control using Behavioral Cloning. *Appl. Energy* **2021**, *304*, 117602. [CrossRef]

26. Ruelens, F.; Claessens, B.J.; Vandael, S.; De Schutter, B.; Babuska, R.; Belmans, R. Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning. *IEEE Trans. Smart Grid* **2016**, *8*, 2149–2159. [CrossRef]
27. Ruelens, F.; Claessens, B.J.; Vrancx, P.; Spiessens, F.; Deconinck, G. Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning. *CSEE J. Power Energy Syst.* **2019**, *5*, 423–432. [CrossRef]
28. Leurs, T.; Claessens, B.J.; Ruelens, F.; Weckx, S.; Deconinck, G. Beyond theory: Experimental results of a self-learning air conditioning unit. In Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016; pp. 1–6. [CrossRef]
29. Patyn, C.; Ruelens, F.; Deconinck, G. Comparing neural architectures for demand response through model-free reinforcement learning for heat pump control. In Proceedings of the 2018 IEEE International Energy Conference (ENERGYCON), Limassol, Cyprus, 3–7 June 2018; pp. 1–6. [CrossRef]
30. De Somer, O.; Soares, A.; Vanthournout, K.; Spiessens, F.; Kuijpers, T.; Vossen, K. Using reinforcement learning for demand response of domestic hot water buffers: A real-life demonstration. In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Turin, Italy, 26–29 September 2017; pp. 1–7. [CrossRef]
31. Yu, K.-H.; Chen, Y.-A.; Jaimes, E.; Wu, W.-C.; Liao, K.-K.; Liao, J.-C.; Lu, K.-C.; Sheu, W.-J.; Wang, C.-C. Optimization of thermal comfort, indoor quality, and energy-saving in campus classroom through deep Q learning. *Case Stud. Therm. Eng.* **2021**, *24*, 100842. [CrossRef]
32. Yu, Z.; Yang, X.; Gao, F.; Huang, J.; Tu, R.; Cui, J. A Knowledge-based reinforcement learning control approach using deep Q network for cooling tower in HVAC systems. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 1721–1726. [CrossRef]
33. Mastropietro, A.; Castiglione, F.; Ballesio, S.; Fabrizio, E. Reinforcement Learning Control Algorithm for HVAC Retrofitting: Application to a Supermarket Building Model by Dynamic Simulation. In Proceedings of the Building Simulation 2019: 16th Conference of IBPSA, Rome, Italy, 2–4 September 2019; pp. 1412–1419. [CrossRef]
34. Yu, K.-H.; Jaimes, E.; Wang, C.-C. AI Based Energy Optimization in Association With Class Environment. In Proceedings of the ASME 2020 14th International Conference on Energy Sustainability, American Society of Mechanical Engineers, Virtual, Online, 17–18 June 2020; pp. 1–6. [CrossRef]
35. McKee, E.; Du, Y.; Li, F.; Munk, J.; Johnston, T.; Kurte, K.; Kotevska, O.; Amasyali, K.; Zandi, H. Deep Reinforcement Learning for Residential HVAC Control with Consideration of Human Occupancy. In Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2–6 August 2020. [CrossRef]
36. Deng, H.; Fannon, D.; Eckelman, M.J. Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy Build.* **2018**, *163*, 34–43. [CrossRef]
37. Ding, Z.; Chen, W.; Hu, T.; Xu, X. Evolutionary double attention-based long short-term memory model for building energy prediction: Case study of a green building. *Appl. Energy* **2021**, *288*, 116660. [CrossRef]
38. Fan, S. Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network. *IEEE Access* **2019**, *7*, 165143–165154. [CrossRef]
39. Liu, T.; Xu, C.; Guo, Y.; Chen, H. A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction. *Int. J. Refrig.* **2019**, *107*, 39–51. [CrossRef]
40. Liu, T.; Tan, Z.; Xu, C.; Chen, H.; Li, Z. Study on deep reinforcement learning techniques for building energy consumption forecasting. *Energy Build.* **2019**, *208*, 109675. [CrossRef]
41. Deng, Z.; Chen, Q. Reinforcement learning of occupant behavior model for cross-building transfer learning to various HVAC control systems. *Energy Build.* **2021**, *238*, 110860. [CrossRef]
42. Xie, Q.; Yue, S.; Pedram, M.; Shin, D.; Chang, N.; Qing, X. Adaptive Thermal Management for Portable System Batteries by Forced Convection Cooling. In Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013, Grenoble, France, 18–22 March 2013; IEEE Conference Publications: Piscataway, NJ, USA; pp. 1225–1228. [CrossRef]
43. Lago, J.; Suryanarayana, G.; Sogancioglu, E.; De Schutter, B. Optimal Control Strategies for Seasonal Thermal Energy Storage Systems With Market Interaction. *IEEE Trans. Control Syst. Technol.* **2020**, *29*, 1891–1906. [CrossRef]
44. Brusey, J.; Hintea, D.; Gaura, E.; Beloe, N. Reinforcement learning-based thermal comfort control for vehicle cabins. *Mechatronics* **2018**, *50*, 413–421. [CrossRef]
45. Li, J.; Li, Y.; Yu, T. Distributed deep reinforcement learning-based multi-objective integrated heat management method for water-cooling proton exchange membrane fuel cell. *Case Stud. Therm. Eng.* **2021**, *27*, 101284. [CrossRef]
46. Wang, X.; Wang, R.; Shu, G.; Tian, H.; Zhang, X. Energy management strategy for hybrid electric vehicle integrated with waste heat recovery system based on deep reinforcement learning. *Sci. China Technol. Sci.* **2021**, *65*, 713–725. [CrossRef]
47. Wang, X.; Wang, R.; Jin, M.; Shu, G.; Tian, H.; Pan, J. Control of superheat of organic Rankine cycle under transient heat source based on deep reinforcement learning. *Appl. Energy* **2020**, *278*, 115637. [CrossRef]
48. Li, Y.; Wen, Y.; Tao, D.; Guan, K. Transforming Cooling Optimization for Green Data Center via Deep Reinforcement Learning. *IEEE Trans. Cybern.* **2019**, *50*, 2002–2013. [CrossRef]
49. Chu, W.-X.; Lien, Y.-H.; Huang, K.-R.; Wang, C.-C. Energy saving of fans in air-cooled server via deep reinforcement learning algorithm. *Energy Rep.* **2021**, *7*, 3437–3448. [CrossRef]

50. Linder, S.P.; Van Gilder, J.; Zhang, Y.; Barrett, E. Dynamic Control of Airflow Balance in Data Centers. In Proceedings of the ASME 2019 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems, American Society of Mechanical Engineers, Hilton Anaheim, CA, USA, 7–9 October 2019; pp. 1–6. [CrossRef]
51. Pérez, S.; Arroba, P.; Moya, J.M. Energy-conscious optimization of Edge Computing through Deep Reinforcement Learning and two-phase immersion cooling. *Futur. Gener. Comput. Syst.* **2021**, *125*, 891–907. [CrossRef]
52. Shao, Z.; Islam, M.A.; Ren, S. DeepPM: Efficient Power Management in Edge Data Centers using Energy Storage. In Proceedings of the 2020 IEEE 13th International Conference on Cloud Computing (CLOUD), Beijing, China, 19–23 October 2020; pp. 370–379. [CrossRef]
53. Shao, Z.; Islam, M.A.; Ren, S. Heat Behind the Meter: A Hidden Threat of Thermal Attacks in Edge Colocation Data Centers. In Proceedings of the 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Seoul, Korea, 27 February–3 March 2021; pp. 318–331. [CrossRef]
54. Zhou, S.; Hu, Z.; Gu, W.; Jiang, M.; Chen, M.; Hong, Q.; Booth, C. Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach. *Int. J. Electr. Power Energy Syst.* **2020**, *120*, 106016. [CrossRef]
55. Idowu, S.; Ahlund, C.; Schelen, O. Machine learning in district heating system energy optimization. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS), Budapest, Hungary, 24–28 March 2014; pp. 224–227. [CrossRef]
56. Solinas, F.M.; Bottaccioli, L.; Guelpa, E.; Verda, V.; Patti, E. Peak shaving in district heating exploiting reinforcement learning and agent-based modelling. *Eng. Appl. Artif. Intell.* **2021**, *102*, 104235. [CrossRef]
57. Weinand, J.M.; Kleinebrahm, M.; McKenna, R.; Mainzer, K.; Fichtner, W. Developing a combinatorial optimisation approach to design district heating networks based on deep geothermal energy. *Appl. Energy* **2019**, *251*, 113367. [CrossRef]
58. Ceglia, F.; Macaluso, A.; Marrasso, E.; Roselli, C.; Vanoli, L. Energy, Environmental, and Economic Analyses of Geothermal Polygeneration System Using Dynamic Simulations. *Energies* **2020**, *13*, 4603. [CrossRef]
59. Carotenuto, A.; Ceglia, F.; Marrasso, E.; Sasso, M.; Vanoli, L. Exergoeconomic Optimization of Polymeric Heat Exchangers for Geothermal Direct Applications. *Energies* **2021**, *14*, 6994. [CrossRef]
60. Zhang, D.; Gao, Z. Improvement of Refrigeration Efficiency by Combining Reinforcement Learning with a Coarse Model. *Processes* **2019**, *7*, 967. [CrossRef]
61. Gellrich, T.; Min, Y.; Schwab, S.; Hohmann, S. Model-Free Control Design for Loop Heat Pipes Using Deep Deterministic Policy Gradient. *IFAC-PapersOnLine* **2020**, *53*, 1575–1580. [CrossRef]
62. Amasyali, K.; Munk, J.; Kurte, K.; Kuruganti, T.; Zandi, H. Deep Reinforcement Learning for Autonomous Water Heater Control. *Buildings* **2021**, *11*, 548. [CrossRef]
63. Kazmi, H.; Mehmood, F.; Lodeweyckx, S.; Driesen, J. Gigawatt-hour scale savings on a budget of zero: Deep reinforcement learning based optimal control of hot water systems. *Energy* **2018**, *144*, 159–168. [CrossRef]
64. Vázquez-Canteli, J.R.; Ulyanin, S.; Kämpf, J.; Nagy, Z. Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities. *Sustain. Cities Soc.* **2018**, *45*, 243–257. [CrossRef]
65. Zsembinszki, G.; Fernández, C.; Vérez, D.; Cabeza, L. Deep Learning Optimal Control for a Complex Hybrid Energy Storage System. *Buildings* **2021**, *11*, 194. [CrossRef]
66. Beintema, G.; Corbetta, A.; Biferale, L.; Toschi, F. Controlling Rayleigh–Bénard convection via reinforcement learning. *J. Turbul.* **2020**, *21*, 585–605. [CrossRef]
67. Abedi, S.; Yoon, S.W.; Kwon, S. Battery energy storage control using a reinforcement learning approach with cyclic time-dependent Markov process. *Int. J. Electr. Power Energy Syst.* **2021**, *134*, 107368. [CrossRef]
68. Wen, L.; Zhou, K.; Li, J.; Wang, S. Modified deep learning and reinforcement learning for an incentive-based demand response model. *Energy* **2020**, *205*, 118019. [CrossRef]
69. Lu, R.; Hong, S.H. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. *Appl. Energy* **2019**, *236*, 937–949. [CrossRef]
70. Zhao, H.; Wang, B.; Liu, H.; Sun, H.; Pan, Z.; Guo, Q. Exploiting the Flexibility Inside Park-Level Commercial Buildings Considering Heat Transfer Time Delay: A Memory-Augmented Deep Reinforcement Learning Approach. *IEEE Trans. Sustain. Energy* **2021**, *13*, 207–219. [CrossRef]
71. Lissa, P.; Deane, C.; Schukat, M.; Seri, F.; Keane, M.; Barrett, E. Deep reinforcement learning for home energy management system control. *Energy AI* **2020**, *3*, 100043. [CrossRef]
72. Coraci, D.; Brandi, S.; Piscitelli, M.S.; Capozzoli, A. Online Implementation of a Soft Actor-Critic Agent to Enhance Indoor Temperature Control and Energy Efficiency in Buildings. *Energies* **2021**, *14*, 997. [CrossRef]
73. Pinto, G.; Deltetto, D.; Capozzoli, A. Data-driven district energy management with surrogate models and deep reinforcement learning. *Appl. Energy* **2021**, *304*, 117642. [CrossRef]
74. Yang, L.; Nagy, Z.; Goffin, P.; Schlueter, A. Reinforcement learning for optimal control of low exergy buildings. *Appl. Energy* **2015**, *156*, 577–586. [CrossRef]
75. de Gracia, A.; Fernandez, C.; Castell, A.; Mateu, C.; Cabeza, L.F. Control of a PCM ventilated facade using reinforcement learning techniques. *Energy Build.* **2015**, *106*, 234–242. [CrossRef]
76. Sakuma, Y.; Nishi, H. Airflow Direction Control of Air Conditioners Using Deep Reinforcement Learning. In Proceedings of the 2020 SICE International Symposium on Control Systems (SICE ISCS), Tokushima, Japan, 3–5 March 2020; pp. 61–68. [CrossRef]

77. Chen, B.; Cai, Z.; Bergés, M. Gnu-RL: A Practical and Scalable Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy. *Front. Built Environ.* **2020**, *6*, 562239. [CrossRef]
78. Chen, Y.; Norford, L.K.; Samuelson, H.W.; Malkawi, A. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy Build.* **2018**, *169*, 195–205. [CrossRef]
79. Fu, C.; Zhang, Y. Research and Application of Predictive Control Method Based on Deep Reinforcement Learning for HVAC Systems. *IEEE Access* **2021**, *9*, 130845–130852. [CrossRef]
80. Qiu, S.; Li, Z.; Li, Z.; Li, J.; Long, S.; Li, X. Model-free control method based on reinforcement learning for building cooling water systems: Validation by measured data-based simulation. *Energy Build.* **2020**, *218*, 110055. [CrossRef]
81. Mason, K.; Grijalva, S. Building HVAC Control via Neural Networks and Natural Evolution Strategies. In Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, Poland, 28 June–1 July 2021; pp. 2483–2490. [CrossRef]
82. Blad, C.; Kallesoe, C.S.; Bogh, S. Control of HVAC-Systems Using Reinforcement Learning With Hysteresis and Tolerance Control. In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 938–942. [CrossRef]
83. Heidari, A.; Marechal, F.; Khovalyg, D. An adaptive control framework based on Reinforcement learning to balance energy, comfort and hygiene in heat pump water heating systems. *J. Phys. Conf. Ser.* **2021**, *2042*, 012006. [CrossRef]
84. Faddel, S.; Tian, G.; Zhou, Q.; Aburub, H. Data Driven Q-Learning for Commercial HVAC Control. In Proceedings of the 2020 SoutheastCon, Raleigh, NC, USA, 28–29 March 2020; pp. 1–6. [CrossRef]
85. Ruelens, F.; Iacovella, S.; Claessens, B.J.; Belmans, R. Learning Agent for a Heat-Pump Thermostat with a Set-Back Strategy Using Model-Free Reinforcement Learning. *Energies* **2015**, *8*, 8300–8318. [CrossRef]
86. Barrett, E.; Linder, S. Autonomous HVAC Control, A Reinforcement Learning Approach. In *Machine Learning and Knowledge Discovery in Databases*; Bifet, A., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9286, pp. 3–19. [CrossRef]
87. Huchuk, B.; Sanner, S.; O'Brien, W. Development and evaluation of data-driven controls for residential smart thermostats. *Energy Build.* **2021**, *249*, 111201. [CrossRef]
88. Hosseinloo, A.H.; Ryzhov, A.; Bischi, A.; Ouerdane, H.; Turitsyn, K.; Dahleh, M.A. Data-driven control of micro-climate in buildings: An event-triggered reinforcement learning approach. *Appl. Energy* **2020**, *277*, 115451. [CrossRef]
89. Schreiber, T.; Schwartz, A.; Muller, D. Towards an intelligent HVAC system automation using Reinforcement Learning. *J. Phys. Conf. Ser.* **2021**, *2042*, 012028. [CrossRef]
90. Marantos, C.; Lamprakos, C.P.; Tsoutsouras, V.; Siozios, K.; Soudris, D. Towards plug&play smart thermostats inspired by reinforcement learning. In Proceedings of the Workshop on INTelligent Embedded Systems Architectures and Applications, New York, NY, USA, 4 October 2018; pp. 39–44. [CrossRef]
91. Dermardiros, V.; Bucking, S.; Athienitis, A.K. A Simplified Building Controls Environment with a Reinforcement Learning Application. In Proceedings of the 16th Conference of the International-Building-Performance-Simulation-Association (IBPSA), Rome, Italy, 2–4 September 2019; pp. 956–964. [CrossRef]
92. Amasyali, K.; Kurte, K.; Zandi, H.; Munk, J.; Kotevska, O.; Smith, R. Double Deep Q-Networks for Optimizing Electricity Cost of a Water Heater. In Proceedings of the 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 16–18 February 2021; pp. 1–5. [CrossRef]
93. Peirelinck, T.; Ruelens, F.; Decnoninck, G. Using reinforcement learning for optimizing heat pump control in a building model in Modelica. In Proceedings of the 2018 IEEE International Energy Conference (ENERGYCON), Limassol, Cyprus, 3–7 June 2018; pp. 1–6. [CrossRef]
94. Park, S.; Park, S.; Choi, M.-I.; Lee, S.; Lee, T.; Kim, S.; Cho, K.; Park, S. Reinforcement Learning-Based BEMS Architecture for Energy Usage Optimization. *Sensors* **2020**, *20*, 4918. [CrossRef] [PubMed]
95. An, Y.; Xia, T.; You, R.; Lai, D.; Liu, J.; Chen, C. A reinforcement learning approach for control of window behavior to reduce indoor PM2.5 concentrations in naturally ventilated buildings. *Build. Environ.* **2021**, *200*, 107978. [CrossRef]
96. Han, M.; May, R.; Zhang, X.; Wang, X.; Pan, S.; Da, Y.; Jin, Y. A novel reinforcement learning method for improving occupant comfort via window opening and closing. *Sustain. Cities Soc.* **2020**, *61*, 102247. [CrossRef]
97. Zhang, Z.; Lam, K.P. Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. In Proceedings of the 5th Conference on Systems for Built Environments, New York, NY, USA, 7–8 November 2018; pp. 148–157. [CrossRef]
98. Brandi, S.; Piscitelli, M.S.; Martellacci, M.; Capozzoli, A. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. *Energy Build.* **2020**, *224*, 110225. [CrossRef]
99. Jiang, Z.; Risbeck, M.J.; Ramamurti, V.; Murugesan, S.; Amores, J.; Zhang, C.; Lee, Y.M.; Drees, K.H. Building HVAC control with reinforcement learning for reduction of energy cost and demand charge. *Energy Build.* **2021**, *239*, 110833. [CrossRef]
100. Vázquez-Canteli, J.R.; Kämpf, J.; Nagy, Z. Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration. *Energy Procedia* **2017**, *122*, 415–420. [CrossRef]
101. Overgaard, A.; Nielsen, B.K.; Kallesoe, C.S.; Bendtsen, J.D. Reinforcement Learning for Mixing Loop Control with Flow Variable Eligibility Trace. In Proceedings of the 2019 IEEE Conference on Control Technology and Applications (CCTA), Hong Kong, China, 19–21 August 2019; pp. 1043–1048. [CrossRef]

102. Wei, Q.; Li, T.; Liu, D. Learning Control for Air Conditioning Systems via Human Expressions. *IEEE Trans. Ind. Electron.* **2020**, *68*, 7662–7671. [CrossRef]
103. Wang, Y.; Velswamy, K.; Huang, B. A Long-Short Term Memory Recurrent Neural Network Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems. *Processes* **2017**, *5*, 46. [CrossRef]
104. Li, B.; Xia, L. A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings. In Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden, 24–28 August 2015; pp. 444–449. [CrossRef]
105. Baghaee, S.; Ulusoy, I. User comfort and energy efficiency in HVAC systems by Q-learning. In Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4. [CrossRef]
106. Heo, S.; Nam, K.; Loy-Benitez, J.; Li, Q.; Lee, S.; Yoo, C. A deep reinforcement learning-based autonomous ventilation control system for smart indoor air quality management in a subway station. *Energy Build.* **2019**, *202*, 109440. [CrossRef]
107. Wang, Y.; Velswamy, K.; Huang, B. A Novel Approach to Feedback Control with Deep Reinforcement Learning. *IFAC-PapersOnLine* **2018**, *51*, 31–36. [CrossRef]
108. Chen, B.; Cai, Z.; Bergés, M. Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13 November 2019; pp. 316–325. [CrossRef]
109. Liu, B.; Akcakaya, M.; Mcdermott, T.E. Automated Control of Transactive HVACs in Energy Distribution Systems. *IEEE Trans. Smart Grid* **2020**, *12*, 2462–2471. [CrossRef]
110. Gao, G.; Li, J.; Wen, Y. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning. *IEEE Internet Things J.* **2020**, *7*, 8472–8484. [CrossRef]
111. Naug, A.; Ahmed, I.; Biswas, G. Online Energy Management in Commercial Buildings using Deep Reinforcement Learning. In Proceedings of the 2019 IEEE International Conference on Smart Computing (SMARTCOMP), Washington, DC, USA, 12–15 June 2019; pp. 249–257. [CrossRef]
112. Van Le, D.; Liu, Y.; Wang, R.; Tan, R.; Wong, Y.-W.; Wen, Y. Control of Air Free-Cooled Data Centers in Tropics via Deep Reinforcement Learning. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13 November 2019; pp. 306–315. [CrossRef]
113. Yoon, Y.R.; Moon, H.J. Performance based thermal comfort control (PTCC) using deep reinforcement learning for space cooling. *Energy Build.* **2019**, *203*, 109420. [CrossRef]
114. Valladares, W.; Galindo, M.; Gutiérrez, J.; Wu, W.-C.; Liao, K.-K.; Liao, J.-C.; Lu, K.-C.; Wang, C.-C. Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. *Build. Environ.* **2019**, *155*, 105–117. [CrossRef]
115. Avendano, D.N.; Ruyssinck, J.; Vandekerckhove, S.; Van Hoecke, S.; Deschrijver, D. Data-driven Optimization of Energy Efficiency and Comfort in an Apartment. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 174–182. [CrossRef]
116. Gupta, A.; Badr, Y.; Negahban, A.; Qiu, R.G. Energy-efficient heating control for smart buildings with deep reinforcement learning. *J. Build. Eng.* **2020**, *34*, 101739. [CrossRef]
117. Zhang, X.; Biagioni, D.; Cai, M.; Graf, P.; Rahman, S. An Edge-Cloud Integrated Solution for Buildings Demand Response Using Reinforcement Learning. *IEEE Trans. Smart Grid* **2020**, *12*, 420–431. [CrossRef]
118. Kazmi, H.; Suykens, J.; Balint, A.; Driesen, J. Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads. *Appl. Energy* **2019**, *238*, 1022–1035. [CrossRef]
119. Wei, T.; Ren, S.; Zhu, Q. Deep Reinforcement Learning for Joint Datacenter and HVAC Load Control in Distributed Mixed-Use Buildings. *IEEE Trans. Sustain. Comput.* **2019**, *6*, 370–384. [CrossRef]
120. Ojand, K.; Dagdougui, H. Q-Learning-Based Model Predictive Control for Energy Management in Residential Aggregator. *IEEE Trans. Autom. Sci. Eng.* **2021**, *19*, 70–81. [CrossRef]
121. Zhang, C.; Kuppannagari, S.R.; Kannan, R.; Prasanna, V.K. Building HVAC Scheduling Using Reinforcement Learning via Neural Network Based Model Approximation. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; pp. 287–296. [CrossRef]
122. Zou, Z.; Yu, X.; Ergan, S. Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network. *Build. Environ.* **2019**, *168*, 106535. [CrossRef]
123. Kotevska, O.; Munk, J.; Kurte, K.; Du, Y.; Amasyali, K.; Smith, R.W.; Zandi, H. Methodology for Interpretable Reinforcement Learning Model for HVAC Energy Control. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 1555–1564. [CrossRef]
124. Xu, J.; Mahmood, H.; Xiao, H.; Anderlini, E.; Abusara, M. Electric Water Heaters Management via Reinforcement Learning With Time-Delay in Isolated Microgrids. *IEEE Access* **2021**, *9*, 132569–132579. [CrossRef]
125. Du, Y.; Zandi, H.; Kotevska, O.; Kurte, K.; Munk, J.; Amasyali, K.; Mckee, E.; Li, F. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Appl. Energy* **2020**, *281*, 116117. [CrossRef]
126. Kurte, K.; Munk, J.; Kotevska, O.; Amasyali, K.; Smith, R.; McKee, E.; Du, Y.; Cui, B.; Kuruganti, T.; Zandi, H. Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses. *Sustainability* **2020**, *12*, 7727. [CrossRef]

127. Yuan, X.; Pan, Y.; Yang, J.; Wang, W.; Huang, Z. Study on the application of reinforcement learning in the operation optimization of HVAC system. *Build. Simul.* **2020**, *14*, 75–87. [CrossRef]
128. Wei, T.; Wang, Y.; Zhu, Q. Deep Reinforcement Learning for Building HVAC Control. In Proceedings of the 54th Annual Design Automation Conference 2017, Austin, TX, USA, 18–22 June 2017; pp. 1–6. [CrossRef]
129. Wei, T.; Chen, X.; Li, X.; Zhu, Q. Model-based and data-driven approaches for building automation and control. In Proceedings of the International Conference on Computer-Aided Design, San Diego, CA, USA, 5–8 November 2018; pp. 1–8. [CrossRef]
130. Wei, P.; Xia, S.; Chen, R.; Qian, J.; Li, C.; Jiang, X. A Deep-Reinforcement-Learning-Based Recommender System for Occupant-Driven Energy Optimization in Commercial Buildings. *IEEE Internet Things J.* **2020**, *7*, 6402–6413. [CrossRef]
131. Ahn, K.U.; Park, C.S. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. *Sci. Technol. Built Environ.* **2019**, *26*, 61–74. [CrossRef]
132. Yu, L.; Sun, Y.; Xu, Z.; Shen, C.; Yue, D.; Jiang, T.; Guan, X. Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings. *IEEE Trans. Smart Grid* **2020**, *12*, 407–419. [CrossRef]
133. Ding, X.; Du, W.; Cerpa, A. Octopus: Deep reinforcement learning for holistic smart building control. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; pp. 326–335. [CrossRef]
134. Zhao, H.; Zhao, J.; Shu, T.; Pan, Z. Hybrid-Model-Based Deep Reinforcement Learning for Heating, Ventilation, and Air-Conditioning Control. *Front. Energy Res.* **2021**, *8*, 610518. [CrossRef]
135. Biemann, M.; Scheller, F.; Liu, X.; Huang, L. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. *Appl. Energy* **2021**, *298*, 117164. [CrossRef]
136. Schreiber, T.; Eschweiler, S.; Baranski, M.; Müller, D. Application of two promising Reinforcement Learning algorithms for load shifting in a cooling supply system. *Energy Build.* **2020**, *229*, 110490. [CrossRef]
137. Zhang, X.; Li, Z.; Li, Z.; Qiu, S.; Wang, H. Differential pressure reset strategy based on reinforcement learning for chilled water systems. *Build. Simul.* **2021**, *15*, 233–248. [CrossRef]
138. Taboga, V.; Bellahsen, A.; Dagdougui, H. An Enhanced Adaptivity of Reinforcement Learning-Based Temperature Control in Buildings Using Generalized Training. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *6*, 255–266. [CrossRef]
139. Masburah, R.; Sinha, S.; Jana, R.L.; Dey, S.; Zhu, Q. Co-designing Intelligent Control of Building HVACs and Microgrids. In Proceedings of the 24th Euromicro Conference on Digital System Design (DSD), Palermo, Spain, 1–3 September 2021; pp. 457–464. [CrossRef]
140. Moriyama, T.; De Magistris, G.; Tatsubori, M.; Pham, T.-H.; Munawar, A.; Tachibana, R. Reinforcement Learning Testbed for Power-Consumption Optimization. In *Methods and Applications for Modeling and Simulation of Complex Systems*; Li, L., Ed.; Communications in Computer and Information Science; Springer: Berlin/Heidelberg, Germany, 2018; Volume 946, pp. 45–59. [CrossRef]
141. Ran, Y.; Hu, H.; Zhou, X.; Wen, Y. DeepEE: Joint Optimization of Job Scheduling and Cooling Control for Data Center Energy Efficiency Using Deep Reinforcement Learning. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–10 July 2019; pp. 645–655. [CrossRef]
142. Jin, H.; Teng, Y.; Zhang, T.; Wang, Z.; Chen, Z. A deep neural network coordination model for electric heating and cooling loads based on IoT data. *CSEE J. Power Energy Syst.* **2020**, *6*, 22–30. [CrossRef]
143. Li, J.; Zhang, W.; Gao, G.; Wen, Y.; Jin, G.; Christopoulos, G. Toward Intelligent Multizone Thermal Control With Multiagent Deep Reinforcement Learning. *IEEE Internet Things J.* **2021**, *8*, 11150–11162. [CrossRef]
144. Zhou, X.; Wang, R.; Wen, Y.; Tan, R. Joint IT-Facility Optimization for Green Data Centers via Deep Reinforcement Learning. *IEEE Netw.* **2021**, *35*, 255–262. [CrossRef]
145. Chi, C.; Ji, K.; Song, P.; Marahatta, A.; Zhang, S.; Zhang, F.; Qiu, D.; Liu, Z. Cooperatively Improving Data Center Energy Efficiency Based on Multi-Agent Deep Reinforcement Learning. *Energies* **2021**, *14*, 2071. [CrossRef]