



Article An Applied Framework for Smarter Buildings Exploiting a Self-Adapted Advantage Weighted Actor-Critic

Ioannis Papaioannou ^{1,2,*}, Asimina Dimara ^{1,3}, Christos Korkas ^{1,2,*,†}, Iakovos Michailidis ^{1,2,†}, Alexios Papaioannou ^{1,4,†}, Christos-Nikolaos Anagnostopoulos ^{3,†}, Elias Kosmatopoulos ^{1,2,†} and Stelios Krinidis ^{1,4,†} and Dimitrios Tzovaras ^{1,†}

- ¹ Centre for Research and Technology Hellas, Information Technologies Institute, 57001 Thessaloniki, Greece; adimara@iti.gr (A.D.); michaild@iti.gr (I.M.); alexiopa@iti.gr (A.P.);
- kosmatop@iti.gr (E.K.); krinidis@mst.ihu.gr (S.K.); dimitrios.tzovaras@iti.gr (D.T.)
- ² Electrical and Computer Engineering Department, Democritus University of Thrace, 67100 Xanthi, Greece
- ³ Intelligent Systems Lab, Department of Cultural Technology and Communication, University of the Aegean, 81100 Mytilene, Greece; canag@aegean.gr
- ⁴ Management Science and Technology Department, International Hellenic University (IHU), 65404 Kavala, Greece
- * Correspondence: johnpap@iti.gr (I.P.); chriskorkas@iti.gr (C.K.); Tel.: +30-2310-464160 (C.K.)
- [†] These authors contributed equally to this work.

Abstract: Smart buildings are rapidly becoming more prevalent, aiming to create energy-efficient and comfortable living spaces. Nevertheless, the design of a smart building is a multifaceted approach that faces numerous challenges, with the primary one being the algorithm needed for energy management. In this paper, the design of a smart building, with a particular emphasis on the algorithm for controlling the indoor environment, is addressed. The implementation and evaluation of the Advantage-Weighted Actor-Critic algorithm is examined in a four-unit residential simulated building. Moreover, a novel self-adapted Advantage-Weighted Actor-Critic algorithm is proposed, tested, and evaluated in both the simulated and real building. The results underscore the effectiveness of the proposed control strategy compared to Rule-Based Controllers, Deep Deterministic Policy Gradient, and Advantage-Weighted Actor-Critic Experimental results demonstrate a 34.91% improvement compared to the Deep Deterministic Policy Gradient and a 2.50% increase compared to the best Advantage-Weighted Actor-Critic method in the first epoch during a real-life scenario. These findings solidify the Self-Adapted Advantage-Weighted Actor-Critic algorithm's efficacy, positioning it as a promising and advanced solution in the realm of smart building optimization.

Keywords: reinforcement learning; building controls; smart buildings; occupants well-being; AWAC

1. Introduction

Green energy and renewable power are gaining increasing significance in the efforts to reduce dependence on fossil fuels and combat climate change. According to a recent report, global energy consumption grew by 2.3% in 2018, at nearly twice the average growth rate since 2010 [1]. As the world transitions towards a more sustainable future with a focus on renewable energy sources, the need for secure and reliable energy systems has become more important than ever [2]. The incorporation of artificial intelligence (AI) and security applications in green energy and renewable power has the potential to revolutionize the way energy is generated, distributed, and consumed. To address these challenges, researchers have turned to AI to improve the security of green energy and renewable power systems [3]. AI can provide advanced threat detection and response capabilities, analyze large amounts of data to identify abnormalities and patterns, and automate security tasks [4].

Complementary achievements in smart building technology are applying these technical advancements to enhance building performance and improve the overall occupant



Citation: Papaioannou, I.; Dimara, A.; Korkas, C.; Michailidis, I.; Papaioannou A.; Anagnostopoulos C.-N.; Kosmatopoulos, E.; Krinidis S.; Tzovaras D. An Applied Framework for Smarter Buildings Exploiting a Self-Adapted Advantage Weighted Actor-Critic. *Energies* 2024, *17*, 616. https://doi.org/10.3390/en17030616

Academic Editors: Pedro Faria and Ramiro Barbosa

Received: 12 December 2023 Revised: 12 January 2024 Accepted: 25 January 2024 Published: 27 January 2024



Copyright: © 2024 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/). experience, while AI's capabilities continue to change digital information [5]. The emergence of smart buildings has led to the development of numerous tools and technologies that can be used to optimize building performance while enhancing energy efficiency [6–8]. Among these tools are many methods, like reinforcement learning (RL) algorithms that accelerate offline learning [9] and optimization techniques like evolutionary algorithms [10].

Numerous challenges emerge when integrating AI-driven techniques for real-time and realistic energy management in smart homes [11]. The requirement for immediate decisionmaking is one of the major problems since standard algorithms frequently find it difficult to swiftly adapt to patterns of dynamic energy usage. Algorithms for smart homes must be able to react rapidly in real-world events, such as when energy consumption suddenly changes or renewable energy sources become intermittent. In addition, a major difficulty is the traditional algorithms' dependence on historical data for training. Extensive datasets are frequently required by traditional methods to accurately estimate and predict energy consumption trends [12]. This presents a significant gap in their application, particularly in newly developed smart home situations where it may be difficult to obtain a large volume of real-time data for training. There is an urgent need for more sophisticated AI techniques in smart home energy management systems because of the constraints associated with historical data requirements and the desire for real-time adaptation. These difficulties drive academics to investigate novel approaches that make use of AI capabilities, such as learning algorithms that train offline, to improve the responsiveness and efficiency of smart home energy management in practical contexts.

1.1. Related Work

In recent years, there has been a surge of diverse and extensive research endeavors exploring green energy, renewable power, and their integration into smart home applications. In the pursuit of sustainable land space utilization, solar farms, rooftop solar installations, and floating photovoltaic (PV) power systems are identified as optimal solutions [13]. Notably, an example is found in the investigation of a 47.5 MW floating PV plant connected to the national utility grid (110 kV) at the Da Mi hydroelectric reservoir in Binh Thuan, Vietnam. The viability of this hydropower park is analyzed through comprehensive economic, financial, and environmental studies, revealing payback periods of 9.3 and 14.4 years, respectively.

A new architecture for energy-efficient homes is proposed in the paper [14]. In this paper, energy and communication units are used in outlets and lights and a home server to collect both the energy consumption data through ZigBee and energy generation data through a Programmable Logic Controller (PLC)-based renewable energy gateway. The home server optimizes household energy utilization by considering both consumption and generation. Based on a weather forecast, the home server can determine the energy production. To reduce energy costs, the home server can regulate the home's energy use schedule. Finally, users can use smart devices to access information about household energy. Likewise, in the work of [15], the information is transmitted to the home server using the generated power. It also makes use of EMCU and ZigBee, which are used to transmit data for relay function on/off and power line communication devices. The information about how much power is being utilized and how much is being generated is updated on the home server using the PLC. Then, the home server uploads the most recent information to the internet. Moreover, the internet server, which also manages their operations, updates the home consumption.

In [16], a look into the most recent advancements in building technology that can be applied to the construction of smart homes and buildings is addressed. An Internet of Things (IoT) architecture for smart homes to support older people's independent living in their living environments is presented in [17], highlighting various sectors in which IoT applications are beneficial. Tao et al. [18] also developed a home automation anthology powered by the IoT and a multilayered architecture built on the cloud. Elkhorchani and Grayaa [19] proposed an architecture for a smart home energy management system and

a shedding algorithm for household energy use. Domestic renewable energy sources, wireless domotic device communication, a control system, a home management system, and a grid management system were the foundations for this effort.

Control of Heating, Ventilation, and Air Conditioning (HVAC) systems, according to Xi Fang [20], depends on finding the right balance between energy consumption and indoor thermal comfort. It used a Deep Q-Learning (DQN) based multi-objective optimum control technique for temperature setpoint real-time reset. The simulation results show that the DQN control strategy effectively balances energy consumption and indoor air temperature. The study concludes that the DQN control strategy is a powerful deep reinforcement learning algorithm for controlling temperature setpoint in real-time in multi-zone building HVAC systems.

One commonly employed technique in smart home automation systems is Rule-Based Control (RBC) [21]. RBC uses upper and lower set points to control operations within predetermined bounds, based on an established set of rules. HVAC subsystems commonly use this technique, especially for temperature management. RBC can only modulate in simple actions such as setting the temperature to a specific value. Thus, actions are limited to simple tasks like turning on or off a switch or thermostat. Modern smart home automation systems, on the other hand, tend to use more sophisticated strategies, such as the Deep Deterministic Policy Gradient (DDPG) algorithm [22]. When compared to RBC, DDPG is superior because it can create efficient daily schedules without relying on the predictions of stochastic factors. It is essential to take into account that the DDPG algorithm is usually trained online when utilizing it for smart home control systems [23]. But in this particular study, the AWAC method, which has the unique benefit of being offline trainable, is adapted [24]. This feature distinguishes AWAC since it allows for the algorithm to grow and change without requiring constant online interactions. As shown in the cited study, this offline training feature gives the smart home automation system an extra degree of adaptability and effectiveness.

1.2. Main Contributions

New approaches are necessary to address the numerous challenges related to realtime energy management in smart homes. RBC and DDPG algorithms provide distinct approaches to smart home automation, building on previous research. RBC is limited to basic actions within preset limitations because it depends on predefined rules and fixed points. Conversely, DDPG presents a more advanced approach that generates effective daily schedules independent of stochastic forecasts. However, in some situations, the online training component of DDPG presents issues. It appears that offline algorithm training is the most effective way to tackle this issue and improve adaptability. Without the limitations of ongoing online interactions, offline training enables continual growth, offering a promising avenue to improve the effectiveness of smart home energy management systems.

Within this context, this paper provides a thorough and broadly applicable smart building design methodology. Everything from the bottom up, which includes the IoT infrastructure, to the highest levels of management, which include advanced energy management algorithms, is addressed. The IoT layer is the fundamental component that supports the smart building design. The complicated system of linked sensors, actuators, and devices that comprise the IoT is explored. The foundational elements of the intelligent building system enable instantaneous data gathering and exchange, serving as its central nervous system. The emphasis then shifts to the algorithms responsible for the smart building's energy management at the upper management layer. These algorithms optimize energy usage, improve operational efficiency, and support sustainability goals by utilizing the insights obtained from IoT data.

As a result, the main goal of this work is summarized in the following key points:

- Address the main gap by exploiting offline training through reinforcement learning.
- Apply offline training utilizing the AWAC as the main algorithm in simulated use case scenarios as close as possible to a real building replica.

- Propose an adaptive AWAC (SA-AWAC) to enhance the accuracy and possibilities of AWAC.
- Apply SA-AWAC in a real life use case scenario to test both its accuracy with the AWAC and other methods (RBC, DDPG).

The remainder of this paper structure is organized as follows: In Section 2, multiple layers are covered presenting the design of a smart building, beginning with the IoT layer and going on to the protocol, middleware, data, and management layers, where the AWAC framework is essential. Section 3 discusses the experimental findings in two primary contexts: a real-world use case with its setup configurations and outcomes, and a simulated building use case with predetermined configurations and outcomes. Lastly, in Section 4, the article highlights the importance of the suggested SA-AWAC configuration in improving the efficiency and adaptability of HVAC systems in smart buildings and summarizes the main conclusions and insights from the experiments carried out.

2. Methodology

A comprehensive approach to designing an intelligent solution for smart buildings is covered in this section. The design is based on a hierarchical manner comprised of distinct layers. The protocol layer, which standardizes communication, comes after the IoT layer, which enables smooth device connectivity. The data layer manages information processing and storage, whereas the middleware layer enables effective data interchange. The management layer is in charge of managing the smart building as a whole. A new development in the area, the incorporation of self-adaptation into the AWAC algorithm, offers a sophisticated and effective method for optimizing smart building systems. It is tailored to prioritize benefits, improving decision-making in the setting of smart buildings. This research analyzes the complicated architectural elements of this smart solution and provides a thorough framework for the real-world use of smarter building technology.

2.1. Design Approach for a Smart Building

2.1.1. IoT (Things) Layer

Within a smart building IoT network, a "thing" can refer to a sensor, actuator, or energy analyzer. However, the scope of IoT domestication has expanded to include a broader range of smart devices and appliances such as smart mirrors, smart windows, and robot vacuum cleaners [25]. In a broader context, a smart home can be defined as an environment that is sensed and monitored, with the option for automation. Monitoring involves the use of diverse sensors and energy analyzers, while automation involves the utilization of actuators. Commonly utilized sensors in smart homes encompass a wide range of functionalities, including temperature, humidity, motion, illuminance, CO₂, smoke, occupancy, weather, water leak, freeze, and window/door sensors [26]. The specific sensors incorporated within a smart home depend on the intended application, cost considerations, and desired usage scenario.

Energy/power analyzers, or smart meters, are crucial in monitoring electrical qualities such as power, energy, current, voltage, harmonics, and more. Smart meters are essential components in Building Energy Management Systems (BEMSs) as they enable the monitoring of energy consumption within a building. Actuators play a vital role in processing sensor information and executing controls based on recommendations from the home automation system or BEMS [27]. These actuators are capable of performing specific actions on various home elements. Examples of such actuators include motorized blinds, automated pergolas, smart thermostats, intelligent valves, and other similar devices.

2.1.2. Protocol Layer

In smart homes, communication between sensors, energy analyzers, and actuators relies on a variety of communication protocols, which can be categorized into wired and wireless options [28]. Wired protocols include Modbus, which facilitates reliable transmission of information over serial lines. Wireless protocols offer greater flexibility

and convenience. Infrared devices enable one-way communication and are commonly used for remote controls, while WiFi, based on the IEEE 802.11 standard [29], provides wireless internet connectivity within a range of approximately 25 m. Bluetooth, with its range of around 10 m, is popular in mobile phones and portable devices. Thread enables devices to communicate even when the WiFi network is unavailable. Zigbee and Z-Wave both operate in mesh networks, allowing devices to form self-organizing networks for communication. KNX, with its decentralized topology, is commonly used for home automation and building control systems. NFC employs radio waves for short-range data transmission and is commonly used for contactless transactions and identification. These protocols are instrumental in facilitating seamless and efficient communication between smart home devices, enhancing the overall functionality and automation capabilities of the IoT ecosystem.

2.1.3. Middleware

Middleware is a software layer that extends the capabilities and common services provided by the operating system to applications. It primarily handles application services, authentication, API management, and data management. Additionally, it serves as a framework that connects users, data, and applications, enabling the creation of applications by developers. Among the messaging protocols used as middleware for bidirectional communication between IoT components, MQTT is widely popular [30].

In the context of IoT development, MQTT is often preferred over other protocols like HTTP due to its ease of use, quick response time, high throughput, and lower battery and bandwidth usage. MQTT employs a publish/subscribe messaging transport protocol, allowing a client to publish data to a specific topic on a server, and all interested parties can subscribe to receive that data. The MQTT protocol specification can be found on the MQTT webpage [31].

To implement MQTT, a server hosts the MQTT broker, and clients connect to the broker for publish/subscribe functionality. In the scope of this work, MQTT is used to establish local and offline communication between the system's components, leveraging its ability to facilitate communication across different programming languages. The broker is installed as a Linux system service on the edge controller, binding to the local host IP. All system components, including the Z-Wave network manager, energy analyzers, sensor APIs, weather stations (if applicable), and other occasional entities, are programmed to communicate their data to the broker. This approach ensures robustness and allows for easier integration of additional components in the future.

2.1.4. Data Layer

The data layer encompasses all the processes and applications used for managing, processing, and collecting data and information from sensors, devices, and actuators. Every device that generates data, such as sensors, energy meters, and edge devices, is considered an integral part of the system. To process this data efficiently, a data handler service utilizes a configuration file that contains various variables, including a map of the entities present in the specific instance of the system. This map is utilized by the data handler to automatically generate the appropriate data format. To facilitate the seamless integration of new system entities into the firmware, each entity is accompanied by a data format template. This template, along with the map, is utilized by the data handler. Initially, when a new entity is encountered, the template must be manually created. It contains predefined information about the entity, which enables the data handler to process the data accurately.

The primary component employed for data management is a local MQTT broker, which is installed as a Linux service. MQTT offers a significant advantage in data management as it serves as an abstraction layer, allowing various services implemented in different programming languages to publish their data to the broker. This ensures language independence for all data-related services as long as they interact with the broker to publish and retrieve data. To facilitate the concurrent development of applications that communicate bidirectionally, a predefined topic structure is employed. Each piece of information is published under a specific topic. By adhering to this topic structure, applications can easily subscribe to the relevant topics and retrieve the required data for their intended functionality. Once the data is collected, the data handling service proceeds to format it into a predefined JSON structure. This formatting ensures consistency and standardization of the data for further processing [32]. The structured data is then posted to a dedicated database, specifically designed to store and organize the system's data. By following this approach, the data handling service effectively gathers data from various system entities, transforms it into a consistent format, and stores it in a dedicated database for easy retrieval and utilization by other components or applications within the system.

2.2. Management Layer

In the context of a smart home system, the management layer acts as a crucial component responsible for making optimal decisions inspired by the AWAC algorithm. This sub-section describes the main features of the AWAC algorithm and the modifications introduced in the proposed paper introducing the Self-Adapted Advantage Weighted Actor-Critic (SA-AWAC) to enhance its efficiency in real-life scenarios.

2.2.1. Advantage Weighted Actor-Critic

AWAC utilizes off-policy temporal-difference learning for Q-function estimation, showcasing adaptability by leveraging prior datasets. To revisit and learn from previous interactions, the algorithm makes use of a replay buffer, which is a data storage mechanism containing past experiences. During policy improvement, it maximizes the critic's value through temporal difference bootstrapping. To navigate potential challenges, AWAC harmonizes its policy distribution with observed data during actor updates. AWAC is characterized as a reinforcement learning method due to its dual optimization approach, which involves maximizing the Q-function while adhering to observed actions [33]. In essence, AWAC adeptly harmonizes policy optimization with data adherence, offering a nuanced solution suitable for diverse environments.

In the context of reinforcement learning, the AWAC algorithm is applied to address problems modeled as Markov Decision Processes (MDPs). The MDP is a model of predicting outcomes and attempts to predict an outcome given only information provided by the current state. According to [34], the framework of MDPs represents the environment as a tuple $\langle S, A, P : S \times A \times S \rightarrow [0,1], R : S \times A \rightarrow IR, \gamma \rangle$ where *S* denotes the set of states ($S = (s_1, s_2, s_3, \dots, s_t, \dots)$), *A* is the action space ($A = (a_1, a_2, a_3, \dots, a_t, \dots)$), *P* is the transition model, with $\mathcal{P}^a_{ss'}$ denotes the reward function as the reward that the agent receives ($R = (r_1, r_2, r_3, \dots, r_t, \dots)$), and γ is the discount factor. The variable γ takes a price between 0 and 1 and using this is rationalized by the idea that future rewards are uncertain.

The actor-critic method is a well-known technique for training agents in reinforcement learning that aims to combine the benefits of actor-only and critic-only approaches [35]. The actor is the policy and the critic is the value function. The critic's responsibility is to process the rewards and the state it receives and feeds into an actor. Given the current state, the actor generates a control input, and after several policy evaluation steps by a critic, it is updated using the information from the critic.

The standard reinforcement learning notation was used for the proposed method, with states *s*, actions *a*, policy $\pi(a|s)$, rewards r(s, a), and dynamics p(s'|s, a). Various algorithms aim to reduce variance by making use of the value function $V^{\pi}(s) = \mathbb{E}_{p^{\pi}(\tau)}[R_t|s]$, action-value function $Q^{\pi}(s, a) = \mathbb{E}_{p^{\pi}(\tau)}[R_t|s, a]$, or advantage $A^{\pi}(s, a) = Q^{\pi}(s, a) - V^{\pi}(s)$. At iteration *k*, the AWAC algorithm optimizes the policy to maximize the estimated Q-function $Q^{\pi_k}(s, a)$ at every state, while constraining it to stay close to the actions observed in the data. This is similar to prior offline RL methods, although this constraint will be enforced

differently. Note that optimizing $Q^{\pi_k}(s, a)$ is equivalent to $A^{\pi_k}(s, a)$. Mathematically, it can be described as follows:

$$\operatorname{argmax}_{\pi} E_{a \sim \pi(*|s)}[A^{\pi_{\kappa}}(s,a)] \quad s.t. \quad D_{KL}(\pi(*|s)||\pi_{\beta}(*|s)) \leq \epsilon, \tag{1}$$

where $D_{KL}(\pi(*|s)||\pi_{\beta}(*|s))$ shows the *KL* divergence of π_{θ} from the optimal non-parametric solution π^* . The Lagrangian was used to introduce a penalty term into the optimization objective that penalizes deviations from the observed actions. It can be described as follows:

$$L(\pi,\lambda) = \mathbb{E}_{a \sim \pi(\cdot|s)}[A^{\pi_k}(s,a)] + \lambda(\epsilon - D_{KL}(\pi(\cdot|s)||\pi_\beta(\cdot|s)))$$
(2)

As with deep neural networks, function approximators require the projection of the non-parametric solution into our policy space. Minimizing the *KL* divergence of π_{θ} from the optimal non-parametric solution π^* under the data distribution $\rho_{\pi_{\beta}}(s)$ can achieve this for a policy π_{θ} with parameters θ :

$$\arg\min_{\theta} \mathbb{E}_{\rho_{\pi_{\beta}}(s)}[D_{KL}(\pi^{*}(\cdot|s)||\pi_{\theta}(\cdot|s))]$$
(3)

$$\arg\min_{\theta} \mathbb{E}_{\rho_{\pi_{\beta}}(s)} \left[\mathbb{E}_{\pi^{*}(\cdot|s)} [-\log \pi_{\theta}(\cdot|s)] \right]$$
(4)

An illustration of the AWAC's fundamental functioning can be found in Figure 1.





The main steps of the AWAC can be described in Algorithm 1, where T is the indoor set temperature that is optional.

Algorithm 1 AWAC(T)

- 1: **Initialize** actor and critic networks: π_{θ} , Q_{ϕ}
- 2: **Initialize** replay buffer: $D = \{(s, a, s', r, T = None)\}$
- 3: **Initialize** iteration counter: i = 1
- 4: while convergence not achieved and *i* < max_iterations do
- 5: **Sample** a batch $(s, a, s', r) \sim D$
- 6: **Compute** target value $y = r + \gamma \cdot \mathbb{E}[Q_{\phi}(s', a')]$ using critic network
- 7: **Update** critic network: $\phi \leftarrow \arg \min_{\phi} \mathbb{E}[(Q_{\phi}(s, a) y)^2]$
- 8: **Update** actor network using importance sampling: $\theta \leftarrow \arg \max_{\theta} \mathbb{E}[\log \pi_{\theta}(a|s) \cdot \exp(\frac{1}{\lambda}A^{\pi_k}(s, a))]$
- 9: **Compute** target actor policy: $\pi_{\text{target}} \leftarrow \text{ModifiedActorPolicy}(\pi_{\theta}, Q_{\phi})$
- 10: **Increment** iteration counter: $i \leftarrow i + 1$
- 11: end while
- 12: **Return** π_{θ} , Q_{ϕ}

2.2.2. Self-Adapted Advantage Weighted Actor-Critic

The Self-Adapted Advantage Weighted Actor-Critic (SA-AWAC) algorithm is presented in this section. SA-AWAC improves efficiency and effectiveness over the original AWAC algorithm by introducing a selection mechanism among different strategies within the AWAC framework. Another difference between SA-AWAC and standard AWAC is the adaptive initialization of the replay buffer, which is based on the desired thermostat set temperature value. An in-depth explanation of the proposed SA-AWAC is presented in the following subsections.

Even under ideal indoor conditions, 80% of people are anticipated to be dissatisfied with interior environment settings, according to thermal comfort principles, specifically the Predicted Mean Vote (PMV) scale [36]. This emphasizes the challenge to offer allencompassing indoor comfort and how customized approaches must be used in indoor environmental design. Nevertheless, there is proof supporting the prospect of optimal indoor conditions [37]. This demonstrates the dynamic nature of indoor comfort, as different conditions may cause a greater percentage of people to be more satisfied overall. Indepth research into these optimal conditions and the variables that affect them may be very helpful in the creation of customized and successful interior environmental design plans.

Within this text, a possible approach to enhance efficiency is to develop use case scenarios for the SA-AWAC that take into consideration ideal indoor conditions for the thermostat set temperature value. A basis for adapting AWAC functionality to match occupant preferences and comfort levels is provided by using particular scenarios derived from the data shown in Table 1. The different scenarios are produced based on optimal conditions as suggested in the literature [38]. A satisfying user experience could be achieved by coordinating AWAC operations with ideal interior conditions. This integration helps to improve AWAC's overall effectiveness and efficiency in a variety of contexts, in addition to addressing issues related to indoor comfort.

Table 1. AWAC's scenarios based on optimal conditions.

Outdoor Temperature	Season	Optimal Conditions	Scenario 1	Scenario 2	Scenario 3
<15 °C	Winter	21–23 °C	21 °C	22 °C	23 °C
15–28 °C	Spring, Autumn	22–24 °C	22 °C	23 °C	24 °C
28 °C	Summer	23–25 °C	23 °C	24 °C	25 °C

The proposed SA-AWAC is presented in Algorithm 2. The algorithm dynamically selects different configurations of AWAC based on the current outdoor temperature. In

colder conditions (below 15 °C, it runs AWAC with parameters 21, 22, and 23, selecting the model with the highest cumulative reward. For temperatures between 15 and 28 °C, it employs AWAC with parameters 22, 23, and 24. In warmer conditions (above 28°C), the algorithm utilizes AWAC with parameters 23, 24, and 25. The model with the highest cumulative reward is then chosen from each scenario. This adaptability allows SA-AWAC to optimize its performance across diverse environmental conditions.

Algorithm 2 SA-AWAC

Choose scenario based on current outdoor temperature:

	1
1:	if outdoor temperature < 15 °C then
2:	Run AWAC(21), AWAC(22), AWAC(23).
3:	Select the model with the highest cumulative reward.
4:	end if
5:	if 15 °C \leq outdoor temperature \leq 28 °C then
6:	Run AWAC(22), AWAC(23), AWAC(24).
7:	Select the model with the highest cumulative reward.
8:	end if
9:	if outdoor temperature $>$ 28 °C then
10:	Run AWAC(23), AWAC(24), AWAC(25).
11:	Select the model with the highest cumulative reward.
12:	end if
13:	Return Optimal π_{θ} , Q_{ϕ}

Figure 2 describes the process of the SA-AWAC. The scenario is selected based on the Out_T and three environments are running with the optimal conditions with set temperatures T1, T2, T3 which are used to select the optimal AWAC configuration. The system module is in charge of assessing each of the three environments' performances and choosing the ones that have the highest cumulative reward. The environment with the highest cumulative reward is selected as optimal and is executed by the system giving as output the action.





2.3. Overall Architecture of a Smart Building Intelligent Solution

The bottom-up layered conceptual architecture, as depicted in Figure 3, is structured as a series of interconnected layers, each serving a specific purpose in the IoT system. At the foundation of the architecture are the devices, sensors, and actuators layer. This layer comprises the physical "things" that make up the IoT network. Devices, sensors, and actuators are the nodes of the network, responsible for generating data and performing actions. They form the fundamental building blocks of the system. Moving up the architecture

stack, the protocol layer is encountered. This layer encompasses the various communication protocols utilized within the IoT network. It provides the necessary standards and protocols for devices to exchange data and communicate with each other. The protocol layer ensures smooth and standardized communication between devices in the network, enabling interoperability and seamless data transfer.



Figure 3. Overall IoT conceptual architecture.

Ascending further, the data layer is reached. The focus shifts towards processing, analyzing, and deriving insights from the data generated by the devices. This layer employs data analysis techniques and algorithms to transform raw data into meaningful information. Sitting on the site of the architecture is the middleware layer, which serves as an intermediary for interlinking and communication within the IoT system. This layer facilitates communication between devices, applications, and external systems or services. It enables seamless data exchange, coordination, and integration, both at the edge-to-edge level within the IoT network and in connecting with the "outer" world. The management layer functions as the command center, overseeing the SA-AWAC algorithm, the replay buffer, and the decision actions. SA-AWAC independently adapts to optimize smart building performance to meet changing conditions, thereby enhancing overall efficiency.

The IoT system achieves a solid basis for managing and exploiting device-generated data, applying standardized communication protocols, performing data analysis, and facilitating application development by utilizing this layered architecture. Every layer in this architectural design provides an important contribution to the IoT system's overall functionality and capacities. At the highest level of this hierarchical structure, the management layer is responsible for managing and supervising the operation of the system as a whole. By including elements such as the replay buffer, decision actions, and SA-AWAC algorithm, it offers an extra degree of intelligence and autonomy.

3. Experimental Results

The purpose of this section is to conduct experiments and evaluate the performance of the SA-AWAC algorithm in comparison to alternative control strategies, such as RBCs and DDPG. A comparison is made between RBCs, DDPG control strategy, AWAC, and the proposed SA-AWAC. The RBCs use a straightforward control strategy that entails keeping each thermal zone's HVAC set points constant at (19–23 °C), respectively, during occupancy hours.

3.1. SA-AWAC Configuration for Modeling Heating Ventilation Air Conditioning Systems

This subsection presents the design of the system state, control actions, and reward function. The AWAC-based control model of building the HVAC system diagram is shown in Figure 4. Better results were observed when the value of the cumulative reward was bigger. The AWAC control method is the agent and the building HVAC system is the environment.



Figure 4. SA-AWAC basic configuration for an HVAC system.

- State: The state represents the information about the current state of the building energy system that is used as input for the control strategy. The state variables capture various aspects of the building and its environment, providing a comprehensive representation of the system's dynamics. To describe all of the information, state space needs to be established with the appropriate variables. The variables under examination in this study, as presented in Table 2, encompass the indoor environment conditions. This includes the air temperature and humidity levels for each zone. Additionally, the study considers the state of the external environment, which involves outdoor temperature and humidity. Furthermore, the analysis considers elements related to the battery, including the battery's state of charge and the energy used for charging the battery.
- Action: The four air-to-water heat pumps (HPs) make up the building's thermal system. Every unit has an integrated water tank that supplies hot water for the fan, heating coils, and domestic hot water (DHW) use. Since the DHW consumption profiles were investigated using a fixed temperature (50 °C), neither the HP supply temperature nor the temperature of the storage tanks may be controlled due to the system setup (they change over time). As seen in Table 3, a single set point is assumed for each thermal zone to regulate its temperature. To minimize the impact on the building, the suggested strategy controls energy consumption and maintains the control set points.
- Reward: In the context of this work, the AWAC control strategy is designed to optimize the energy consumption of the HVAC system, representing the cost and thermal discomfort factors in our environment. A reward function is formulated, integrating both energy and comfort to guide the optimization process.

$$Reward = -\alpha \times Consumption - \beta \times Comfort,$$
(5)

where:

- α and β are weighting factors that balance the importance of energy and comfort in the optimization process.
- Consumption represents the energy component, which takes into account elements such as energy usage and operating costs, reflecting the economic aspect of the control approach.
- Comfort represents the comfort component, measuring the variance between predicted mean vote (PMV) and predicted percentage of dissatisfied (PPD) [36]. Minimizing thermal discomfort prioritizes occupant comfort, striving to maintain temperatures within a comfortable range.

... The values of α and β can be adjusted based on the specific emphasis intended for minimizing costs and thermal discomfort in the chosen control strategy. The cost component considers elements such as energy usage and operating costs, reflecting the economic aspect of the control approach. The primary objective is to minimize costs, aiming for energy-efficient operation and cost savings. The thermal discomfort component measures the variance between indoor temperatures and the ideal setpoints. Minimizing thermal discomfort prioritizes occupant comfort, striving to maintain temperatures within a comfortable range. The integration of cost and thermal discomfort allows us to strike a balance between energy economy and occupant comfort, enabling a comprehensive assessment and comparison of the effectiveness of various control systems.

3.2. Simulated Building Use Case

A widely used open-source tool for building simulation was used (i.e., Energym [39]) in an attempt to replicate a real building as closely as possible.

3.2.1. Set-Up Configurations

The case's general characteristics are shown in Figure 5. The building used has four apartments ("Apartments2 Grid"), located in an eight-zone complex in Tarragona, Spain. Its entire volume is 1042.84 m³ and its total surface area is 417.12 m². In terms of electricity, the "Apartments2 Grid" system consists of an electric vehicle, a community battery, and a photovoltaic array. The PV panels have a 40° slant and a 58 m² active surface area and they are facing south. The PV generator has a rated electrical power output of 10,750 W and the community battery has a 10 kWh capacity. The day and night thermal zone visualization of the simulated building is displayed in Figure 5.



Figure 5. Examined building case.

The four-unit simulation building incorporates a wide variety of sensors, actuators, and devices designed for commercial research and development purposes. In Table 2,

the simulation outputs utilized in the building are presented. These outputs include information about the building's internal and external conditions, battery charge levels, and thermostat settings for each floor. Additionally, Table 3 illustrates the simulation inputs used in this study, highlighting the four setpoints corresponding to the thermostats on each floor.

Lower Upper Variable Name Type Description Bound Bound Outdoor Ext_T -1040 scalar temperature (°C) Outdoor relative 0 Ext_RH scalar 100 humidity (%RH) Direct normal 0 1000 Ext_Irr scalar radiation (W/m²) Outdoor air pressure 80,000.0 130,000.0 Ext_P scalar (Pa) Floor 1 thermostat P1_T_Thermostat_sp_out scalar 16 26 temperature (°C) Floor 2 thermostat P2_T_Thermostat_sp_out scalar 16 26 temperature (°C) Floor 3 thermostat P3_T_Thermostat_sp_out scalar 16 26 temperature (°C) Floor 4 thermostat P4_T_Thermostat_sp_out scalar 16 26 temperature (°C) Zone 1 temperature 10 40 Z01_T scalar $(^{\circ}C)$ Zone 1 relative 0 100 Z01_RH scalar humidity (%RH) Zone 2 temperature Z02_T 10 40 scalar (°C) Zone 2 relative Z02_RH scalar 0 100 humidity (%RH) Zone 3 temperature 40 Z03_T scalar 10 $(^{\circ}C)$ Zone 3 relative 0 100 Z03_RH scalar humidity (%RH) Zone 4 temperature 10 40 Z04_T scalar $(^{\circ}C)$ Zone 4 relative Z04_RH 0 100 scalar humidity (%RH) Zone 5 temperature Z05_T scalar 10 40 $(^{\circ}C)$ Zone 5 relative 0 100 Z05_RH scalar humidity (%RH) Zone 6 temperature Z06_T scalar 10 40 $(^{\circ}C)$

Table 2. Simulation outputs of the building with four apartments.

Variable Name	Туре	Lower Bound	Upper Bound	Description
Z06_RH	scalar	0	100	Zone 6 relative humidity (%RH)
Z07_T	scalar	10	40	Zone 7 temperature (°C)
Z07_RH	scalar	0	100	Zone 7 relative humidity (%RH)
Z08_T	scalar	10	40	Zone 8 temperature (°C)
Z08_RH	scalar	0	100	Zone 8 relative humidity (%RH)
FA_ECh_Bat	scalar	0	4000.0	Battery charging energy (Wh)
Bd_FracCh_Bat	scalar	0	1	Battery state of charge

Table 2. Cont.

Table 3. Simulation inputs of the building with four apartments.

Variabel Name	Туре	Lower Bound	Upper Bound	Description
P1_T_Thermostat_sp	scalar	16	26	Floor 1 thermostat setpoint (°C)
P2_T_Thermostat_sp	scalar	16	26	Floor 2 thermostat setpoint (°C)
P3_T_Thermostat_sp	scalar	16	26	Floor 3 thermostat setpoint (°C)
P4_T_Thermostat_sp	scalar	16	26	Floor 4 thermostat setpoint (°C)

3.2.2. Results

In this section, the results of the simulation data are presented. The experiments were conducted in October 2023, indicating that the season for these experiments was autumn. Three different versions of AWAC were implemented to compare with SA-AWAC and DDPG, each with a different thermostat value set to fill the replay buffers, as described below. Also three different AWAC versions were implemented for SA-AWAC, each with a different thermostat value set to fill the replay buffers, as described in (Section 2.2.2). All of the algorithms underwent 60 epochs of training, with 4000 steps in each epoch. The thermostat setpoint for the DDPG was 21 °C. For the RBC scenarios, denoted as RBC₁₉, RBC₂₀, RBC₂₁, RBC₂₂, and RBC₂₃, the thermostat set points are configured to 19 °C, 20 °C, 21 °C, 22 °C, and 23 °C respectively. The three versions of AWAC are presented below:

- 1. AWAC₁, all thermostat setpoints were randomly selected between [21 °C, 22 °C], and a 2000 experience AWAC replay buffer was created.
- 2. AWAC₂, all thermostat setpoints were randomly selected between [22 °C, 23 °C], and a 2000 experience AWAC replay buffer was created.
- 3. AWAC₃, all thermostat setpoints were randomly selected between [23 °C, 24 °C], and a 2000 experience AWAC replay buffer was created.

The cumulative reward for the first epoch of the proposed SA-AWAC is displayed in Table 4, providing a comparison with DDPG, AWAC₁, AWAC₂, and AWAC₃. As a dual metric, cumulative reward is estimated utilizing Equation (5), where $\alpha = \beta = 0.5$. DDPG achieved a cumulative reward of -48,882.70 in the first epoch, while AWAC₁, AWAC₂, and

AWAC₃ obtained corresponding values of -34,205.5, -31,523.4, and -32,424.65. The three versions of AWAC exhibited an increase of [30–38%] compared to DDPG. Additionally, SA-AWAC showed an increase of 44.16% compared to DDPG and 20.19%, 13.40%, and 15.81% compared to the three versions of AWAC. In comparison to DDPG and the three versions of AWAC, the results consistently demonstrate that SA-AWAC outperforms other approaches, indicating its efficacy in achieving faster convergence and superior initial performance.

Table 4. Cumulative reward of epoch one for the simulation building (DDPG, AWAC₁, AWAC₂, AWAC₃, SA-AWAC).

DI Algorithm	Cumulative	Percentage Diff (%)
KL Algorithm	Reward	SA-AWAC
DDPG	-48,883	44.16%
AWAC ₁	-34,205	20.19%
AWAC ₂	-31,523	13.40%
AWAC ₃	-32,424	15.81%
SA-AWAC	-27,296	-

In Figure 6, a comparison of the DDPG with AWAC₂ and DDPG with SA-AWAC is presented. The figure showcases the evolution of the average episode reward. The SA-AWAC algorithm converges significantly faster than the centralized approach of DDPG, achieving similar reward levels by the end of the training process. More specifically, SA-AWAC achieves a similar or greater level of performance compared to DDPG in less than five epochs during the training.



Figure 6. Cumulative reward over epochs for the simulation building.

Additionally, a comparison was made between the RBC controller, which is based on predefined rules, and the model-free algorithms (DDPG, AWAC₁, AWAC₂, AWAC₃, and SA-AWAC). For each pricing profile, ten days were simulated with various configurations for the RBC rules to test the effectiveness of the RL algorithms. The evaluation comparison between the RBC and the model-free algorithms involves distinct configurations between episodes (days), but the same-day configuration with identical operating conditions for all algorithms in each episode. The mean of the episodic reward for each algorithm (the average daily reward) was collected and is presented in Table 5. It can be observed that all model-free algorithms outperform the RBC controller. Furthermore, SA-AWAC achieves better performance levels than the regular version of DDPG and the three versions of AWAC. More specifically, in the case of SA-AWAC, there is a significant improvement of

approximately 37.34% in the algorithm's performance compared to RBC₂₂, 5.71% compared to DDPG, and 16.9% compared to AWAC₁.

Table 5. Evaluation reward on 10 days for RBCs, DDPG, AWAC₁, AWAC₂, AWAC₃, SA-AWAC (simulation building).

DI Alexanthan	Cumulative	Percentage Diff (%)
KL Algorithm	Reward	SA-AWAC
RBC ₁₉	-114,138	>100%
RBC ₂₀	-112,201	>100%
RBC ₂₁	-42,165	>100%
RBC ₂₂	-1633	37.34%
RBC ₂₃	-2417	>100%
DDPG	-1257	5.71%
AWAC ₁	-1390	16.9%
AWAC ₂	-1239	4.2%
AWAC ₃	-1241	4.37%
SA-AWAC	-1189	-

3.3. Real-Life Use Case

The CERTH/ITI nZEB Smart Home is a realistic residential building that enables occupants to immerse themselves in authentic living situations while discovering various innovative smart technologies based on the IoT. It offers services related to Energy, Health, Big Data, Robotics, and AI, demonstrating rapid prototyping and showcasing novel technologies. Being Greece's inaugural Smart Near-Zero Energy Building, it incorporates advanced construction materials and intelligent ICT solutions, resulting in a sustainable and forward-thinking environment for testing, validating, and evaluating various solutions. This infrastructure ensures a future-proof setting that supports the development of cutting-edge and eco-friendly practices. In early 2017, the Digital Innovation Hub (DIH) became officially registered in the JRC Catalogue, making it the first DIH in the Central Macedonia region. Since then, it has been functioning effectively, providing a platform for digital innovation activities in the field. Figure 7 shows the thermal zone of the smart home building.



Figure 7. CERTH/ITI nZEB Smart Home thermal zone.

3.3.1. Set-Up Configuration

The CERTH/ITI house incorporates an extensive array of sensors, actuators, and devices utilized for research and development purposes and in commercial contracts. Automated and artificial intelligence algorithms are crucial in supporting automation processes and implementing multi-factor efficiency scenarios. These scenarios consider factors such as occupancy and comfort levels, ensuring optimal energy usage and resource management within the house. Integrating these technologies and intelligent algorithms enhances energy efficiency, making the CERTH/ITI house a cutting-edge platform for exploring innovative energy solutions.

To apply the above technologies and framework, the CERTH/ITI house is equipped with a range of energy-related IoT equipment such as smart meters, dimming and on/off actuators, environmental sensors, occupancy sensors, smart plugs, smart appliances, photovoltaics, batteries/storage systems, and more. These devices enable the monitoring of energy consumption, production, and the overall condition of the building. Indicatively, some of those devices are presented in Table 6.

The outputs used in the smart home are shown in Table 7. Furthermore, Table 8 presents the inputs utilized in the research, emphasizing the single setpoint that corresponds to the building's room thermostat.

Туре	Description	Model	Protocol
Actuator	Wireless actuator Light Dimmer switch	EnOcean Wireless Dimmer 1–10 V model FSG71/	EnOcean
Sensor	Mutisensor	Fibaro Fgms-001-zw5 Motion Detector	Zwave
Sensor	Temperature- humidity sensor	Plugwise	Zigbee
Sensor	luminance sensor	FIH65B EnOcean Wireless Indoor Luminance	EnOcean
Device/Actuator	Heating ventilation air conditioning system (HVAC)	LG Ceiling mountain Cassette and HVAC controller	WiFi
Energy analyzer	Energy meter	Carlo Gavazzi 3phase	Modbus

Table 6. Overview of smart home equipment.

Table 7. Outputs of the real building.

Variable Name	Туре	Lower Bound	Upper Bound	Description
Ext_T	scalar	-10	40	Outdoor temperature(°C)
Ext_RH	scalar	0	100	Outdoor relative humidity (%RH)
P1_T_Thermostat_sp_out	scalar	16	26	Room 1 thermostat temperature (°C)
Z01_T	scalar	10	40	Zone 1 temperature (°C)
Z01_RH	scalar	0	100	Zone 1 relative humidity (%RH)

Variabel Name	Туре	Lower Bound	Upper Bound	Description
P1_T_Thermostat_sp	scalar	16	26	Floor 1 thermostat setpoint (°C)

3.3.2. Results

The cumulative reward for the first epoch of the proposed SA-AWAC is presented in Table 9, with a comparative analysis against DDPG, AWAC₁, AWAC₂, and AWAC₃. DDPG achieved a cumulative reward of -65,498.43 during the first epoch, while AWAC₁, AWAC₂, and AWAC₃ recorded corresponding values of -51,434.5, -49,766.34, and -50,325.4. The three AWAC variants exhibited an increase ranging from 27% to 31% compared to DDPG. Moreover, SA-AWAC displayed a rise of 34.91% compared to DDPG and 5.94%, 2.50%, and 3.65% relative to the three versions of AWAC. These results suggest that SA-AWAC outperforms other approaches, demonstrating its effectiveness in achieving swifter convergence and superior initial performance.

Table 9. Cumulative reward of epoch one for the real building (DDPG, AWAC₁, AWAC₂, AWAC₃, SA-AWAC).

DI Alcorithm	Cumulative	Percentage Diff (%)
KL Algorithm	Reward	SA-AWAC
DDPG	-65,498	34.91%
AWAC ₁	-51,434	5.94%
AWAC ₂	-49,766	2.50%
AWAC ₃	-50,325	3.65%
SA-AWAC	-48,549	-

A comparison of the DDPG with SA-AWAC and AWAC₂ and the average episode reward's evolution are depicted in Figure 8. By the end of the training process, the SA-AWAC algorithm achieves similar reward levels and converges much faster than the centralized approach of DDPG. More precisely, in less than five training epochs, SA-AWAC outperforms DDPG to an equivalent or higher degree of performance.



Figure 8. Cumulative reward over epochs for the real building.

Furthermore, a comparative analysis was conducted between model-free algorithms (DDPG, AWAC₁, AWAC₂, AWAC₃, and SA-AWAC) and the RBC controller, relying on predefined rules. The same setup configuration is used as in the simulated data (Section 3.2.2). For every algorithm, the mean daily reward was gathered and is presented in Table 10. Interestingly, the RBC controller was outperformed by all model-free algorithms. Additionally, SA-AWAC outperformed three different versions of AWAC and the standard DDPG. In particular, SA-AWAC demonstrated a noteworthy enhancement of roughly 31.2% in comparison to RBC₂₂, 8.9% in comparison to DDPG, and 12.81% in comparison to AWAC₁.

DI Alexandthan	Cumulative	Percentage Diff (%)
KL Algorithm	Reward	SA-AWAC
RBC ₁₉	-136,328	>100%
RBC ₂₀	-135,267	>100%
RBC ₂₁	-67,249	>100%
RBC ₂₂	-1997	31.2%
RBC ₂₃	-3060	>100%
DDPG	-1658	8.9%
AWAC ₁	-1717	12.81%
AWAC ₂	-1577	3.61%
AWAC ₃	-1673	9.92%
SA-AWAC	-1522	-

Table 10. Evaluation reward for DDPG and AWAC and SA-AWAC (real building).

4. Conclusions

This paper introduces a novel adaptive algorithm called the SA-AWAC based on offline training while applying it to a real building, thus providing the basis of an applied framework. The design is based on a hierarchical manner comprised of distinct layers. The design follows a hierarchical structure with distinct layers: the protocol layer, the IoT layer, the middleware, the data layer, and the management layer, which incorporates the proposed SA-AWAC. The effectiveness of the proposed SA-AWAC was evaluated using simulated data from a building replica, as well as real data from a CERTH/ITI smart home. The results demonstrated consistently high cumulative rewards in both datasets. A comparison with non-RL and RL methods was conducted, revealing that the SA-AWAC consistently outperformed other methods in both simulated and real datasets. Specifically, on simulated data, the SA-AWAC exhibited a 44.16% improvement compared to DDPG and a 13.40% increase over the best-performing AWAC method in the first epoch. Similarly, on real data, the proposed SA-AWAC showed a 34.91% improvement compared to DDPG and a 2.50% increase compared to the best-performing AWAC method in the first epoch. These results highlight the superior performance of the SA-AWAC model in addressing the challenges posed by both simulated and real smart home environments.

In terms of future work, an exploration of various intelligent behaviors is planned, encompassing pre-cooling strategies for discomfort alleviation, HVAC modulation to reduce energy consumption peaks, and the enhanced integration of renewable sources.

Author Contributions: Conceptualization, I.P., A.D., A.P.; Methodology, I.P., A.D., C.K., A.P., C.-N.A. and E.K.; Software, I.P., A.D. and A.P.; Validation, A.D. and C.K.; Formal analysis, C.K., I.M., A.P., C.-N.A., E.K., S.K. and D.T.; Investigation, A.D. and C.K.; Resources, S.K. and D.T.; Data curation, C.K.; Writing—original draft, I.P. and A.D.; Writing—review & editing, I.P., A.D., C.K., I.M., A.P., C.-N.A., E.K., S.K. and D.T.; Visualization, I.M., A.P. and C.-N.A.; Supervision, A.D.; Project administration, C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the PRECEPT project which is funded by the EU H2020 under grant agreement No. 958284.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

- 1. US EIA. Global Energy & CO₂ Status Report: The Latest Trends in Energy and Emissions in 2018; US EIA: Washington, DC, USA, 2018.
- Korkas, C.; Dimara, A.; Michailidis, I.; Krinidis, S.; Marin-Perez, R.; Martínez García, A.I.; Skarmeta, A.; Kitsikoudis, K.; Kosmatopoulos, E.; Anagnostopoulos, C.N.; et al. Integration and Verification of PLUG-N-HARVEST ICT Platform for Intelligent Management of Buildings. *Energies* 2022, 15, 2610. [CrossRef]
- Papaioannou, A.; Dimara, A.; Michailidis, I.; Stefanopoulou, A.; Karatzinis, G.; Krinidis, S.; Anagnostopoulos, C.N.; Kosmatopoulos, E.; Ioannidis, D.; Tzovaras, D. Self-protection of IoT Gateways Against Breakdowns and Failures Enabling Automated Sensing and Control. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, León, Spain, 14–17 June 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 231–241.
- Şerban, A.C.; Lytras, M.D. Artificial intelligence for smart renewable energy sector in europe—Smart energy infrastructures for next generation smart cities. *IEEE Access* 2020, *8*, 77364–77377. [CrossRef]
- 5. Alanne, K.; Sierla, S. An overview of machine learning applications for smart buildings. *Sustain. Cities Soc.* 2022, 76, 103445. [CrossRef]
- 6. Lamnatou, C.; Chemisana, D.; Cristofari, C. Smart grids and smart technologies in relation to photovoltaics, storage systems, buildings and the environment. *Renew. Energy* **2022**, *185*, 1376–1391. [CrossRef]
- Labonnote, N.; Høyland, K. Smart home technologies that support independent living: Challenges and opportunities for the building industry—A systematic mapping study. *Intell. Build. Int.* 2017, 9, 40–63. [CrossRef]
- 8. Huang, Q. Energy-Efficient Smart Building Driven by Emerging Sensing, Communication, and Machine Learning Technologies. *Eng. Lett.* **2018**, 26, 3.
- 9. Nair, A.; Gupta, A.; Dalal, M.; Levine, S. Awac: Accelerating online reinforcement learning with offline datasets. *arXiv* 2020, arXiv:2006.09359.
- Rager, M.; Gahm, C.; Denz, F. Energy-oriented scheduling based on evolutionary algorithms. *Comput. Oper. Res.* 2015, 54, 218–231. [CrossRef]
- Touqeer, H.; Zaman, S.; Amin, R.; Hussain, M.; Al-Turjman, F.; Bilal, M. Smart home security: Challenges, issues and solutions at different IoT layers. J. Supercomput. 2021, 77, 14053–14089. [CrossRef]
- 12. Zaidan, A.; Zaidan, B. A review on intelligent process for smart home applications based on IoT: Coherent taxonomy, motivation, open challenges, and recommendations. *Artif. Intell. Rev.* 2020, *53*, 141–165. [CrossRef]
- 13. Nguyen, N.H.; Le, B.C.; Bui, T.T. Benefit Analysis of Grid-Connected Floating Photovoltaic System on the Hydropower Reservoir. *Appl. Sci.* 2023, 13, 2948. [CrossRef]
- 14. Han, J.; Choi, C.S.; Park, W.K.; Lee, I.; Kim, S.H. Smart home energy management system including renewable energy based on ZigBee and PLC. *IEEE Trans. Consum. Electron.* **2014**, *60*, 198–202. [CrossRef]
- Sathisshkumar, A.; Jayamani, S. Renewable energy management system in home appliance. In Proceedings of the 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, India, 19–20 March 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–4.
- 16. Kim, D.; Yoon, Y.; Lee, J.; Mago, P.J.; Lee, K.; Cho, H. Design and implementation of smart buildings: A review of current research trend. *Energies* **2022**, *15*, 4278. [CrossRef]
- Chernbumroong, S.; Atkins, A.; Yu, H. Perception of smart home technologies to assist elderly people. In Proceedings of the 4th International Conference on Software, Knowledge, Information Management and Applications, Suwon, Republic of Korea, 14–15 January 2010; pp. 90–97.
- 18. Tao, M.; Zuo, J.; Liu, Z.; Castiglione, A.; Palmieri, F. Multi-layer cloud architectural model and ontology-based security service framework for IoT-based smart homes. *Future Gener. Comput. Syst.* **2018**, *78*, 1040–1051. [CrossRef]
- 19. Elkhorchani, H.; Grayaa, K. Novel home energy management system using wireless communication technologies for carbon emission reduction within a smart grid. *J. Clean. Prod.* **2016**, *135*, 950–962. [CrossRef]
- 20. Fang, X.; Gong, G.; Li, G.; Chun, L.; Peng, P.; Li, W.; Shi, X.; Chen, X. Deep reinforcement learning optimal control strategy for temperature setpoint real-time reset in multi-zone building HVAC system. *Appl. Therm. Eng.* **2022**, 212, 118552. [CrossRef]
- Badar, A.Q.; Anvari-Moghaddam, A. Smart home energy management system—A review. Adv. Build. Energy Res. 2022, 16, 118–143. [CrossRef]
- 22. 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]

- 23. Ye, Y.; Qiu, D.; Wang, H.; Tang, Y.; Strbac, G. Real-time autonomous residential demand response management based on twin delayed deep deterministic policy gradient learning. *Energies* **2021**, *14*, 531. [CrossRef]
- 24. Zanette, A.; Wainwright, M.J.; Brunskill, E. Provable benefits of actor-critic methods for offline reinforcement learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 13626–13640.
- Koohang, A.; Sargent, C.S.; Nord, J.H.; Paliszkiewicz, J. Internet of Things (IoT): From awareness to continued use. Int. J. Inf. Manag. 2022, 62, 102442. [CrossRef]
- Friess, P.; Ibanez, F. Putting the Internet of Things forward to the next nevel. In Internet of Things Applications—From Research and Innovation to Market Deployment; River Publishers: Aalborg, Denmark , 2022; pp. 3–6.
- 27. Mahapatra, B.; Nayyar, A. Home energy management system (HEMS): Concept, architecture, infrastructure, challenges and energy management schemes. *Energy Syst.* 2022, 13, 643–669. [CrossRef]
- Ramalingam, S.P.; Shanmugam, P.K. A Comprehensive Review on Wired and Wireless Communication Technologies and Challenges in Smart Residential Buildings. *Recent Adv. Comput. Sci. Commun. (Former. Recent Patents Comput. Sci.)* 2022, 15, 1140–1167. [CrossRef]
- 29. Banerji, S.; Chowdhury, R.S. On IEEE 802.11: Wireless LAN technology. arXiv 2013, arXiv:1307.2661.
- D'Ortona, C.; Tarchi, D.; Raffaelli, C. Open-Source MQTT-Based End-to-End IoT System for Smart City Scenarios. *Future Internet* 2022, 14, 57. [CrossRef]
- 31. MQTT: The Standard for IoT Messaging. Available online: https://mqtt.org/ (accessed on 15 May 2023).
- Dimara, A.; Vasilopoulos, V.G.; Papaioannou, A.; Angelis, S.; Kotis, K.; Anagnostopoulos, C.N.; Krinidis, S.; Ioannidis, D.; Tzovaras, D. Self-healing of semantically interoperable smart and prescriptive edge devices in IoT. *Appl. Sci.* 2022, 12, 11650. [CrossRef]
- Villaflor, A.; Dolan, J.; Schneider, J. Fine-Tuning Offline Reinforcement Learning with Model-Based Policy Optimization, 2020. Available online: https://openreview.net/forum?id=wiSgdeJ29ee (accessed on 19 December 2023)
- Comanici, G.; Precup, D. Optimal policy switching algorithms for reinforcement learning. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, ON, Canada, 10–14 May 2010; Volume 1, pp. 709–714.
- Grondman, I.; Busoniu, L.; Lopes, G.A.D.; Babuska, R. A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* 2012, 42, 1291–1307. [CrossRef]
- 36. Kaushal, A.A.; Anand, P.; Aithal, B.H. Assessment of the Impact of Building Orientation on PMV and PPD in Naturally Ventilated Rooms during Summers in Warm and Humid Climate of Kharagpur, India. In Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Istanbul, Turkey, 15–16 November 2023; pp. 528–533.
- 37. Wu, J.; Hou, Z.; Shen, J.; Lian, Z. A method for the determination of optimal indoor environmental parameters range considering work performance. *J. Build. Eng.* **2021**, *35*, 101976. [CrossRef]
- Dimara, A.; Timplalexis, C.; Krinidis, S.; Schneider, C.; Bertocchi, M.; Tzovaras, D. Optimal comfort conditions in residential houses. In Proceedings of the 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 23–26 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
- Scharnhorst, P.; Schubnel, B.; Fernández Bandera, C.; Salom, J.; Taddeo, P.; Boegli, M.; Gorecki, T.; Stauffer, Y.; Peppas, A.; Politi, C. Energym: A building model library for controller benchmarking. *Appl. Sci.* 2021, *11*, 3518. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.