Reinforcement Learning for Building Energy System Control in Multi-Family Buildings

Michael Bachseitz, Muhammad Sheryar, David Schmitt, Thorsten Summ, Christoph Trinkl and Wilfried Zörner

Institute of new Energy Systems, Technische Hochschule Ingolstadt, Ingolstadt (Germany)

Abstract

Replacing the widespread use of fossil fuel heating systems in the existing multi-family building stock with systems powered by renewable energy is one of the most important contributions to the energy transition in Germany. Heat pumps can also be a suitable solution for older buildings, especially in combination with photovoltaic systems. Advanced control strategies are needed to optimize the self-consumption rate and reduce the load on the electricity grid. Due to the increase in available computing power, artificial intelligence methods such as reinforcement learning, which can be used to control (energy) systems, have become very popular in the recent years. A literature review was conducted to provide an up-to-date overview of the use of reinforcement learning for building energy system control, with a focus on multi-family buildings. Such a building with an energy system including a heat pump, thermal and electrical storage, and a photovoltaic system was modeled in MATLAB/Simulink. The model is used to investigate typical control strategies as a reference and to develop a heat pump control strategy using the MATLAB Reinforcement Learning Toolbox. Two reinforcement learning agents with different numbers of observations were trained to control the heat pump and studied in annual simulations. Both RL agents provide the required supply water temperature. Adding one observation improves the annual reward by 160 %. Nevertheless, a potential to improve learning the RL agent switching of the heat pump at times of high bottom storage temperatures was identified.

Keywords: multi-family building, energy system, optimization, self-consumption, reinforcement learning

1. Introduction

The decarbonization of the residential building sector is one of the most important contributions to the energy transition in Germany. In addition to reducing the energy consumption, especially for heating, through the energy renovation of existing buildings, the widespread use of fossil fuel heating systems must be replaced by more efficient systems powered by renewable energy. The scenario of a climate-neutral Germany by 2045 introduced in Prognos et a. (2021) shows a need of in total 14 million heat pumps in the building sector.

Heat pumps are a highly efficient solution for supplying heat to buildings based on (partially) on-site renewable energy, such as rooftop photovoltaic (PV) systems. Bongs et al. (2023) investigated low-exergy concepts for multi-family buildings, which account for about 41 % of the living space in Germany (Loga et al., 2015). They developed, analyzed, and demonstrated solutions for the efficient use of heat pumps, heat transfer systems, and ventilation systems for energy renovations of multi-family buildings. Their results show that by replacing only a few radiators in critical rooms (e.g., 2 - 7 % of the existing radiators in their use case), the heat supply temperatures can be reduced to 60 °C or less, and therefore, heat pumps can be a suitable solution for heat supply, especially in combination with PV systems.

In order to use most of the electricity generated on site in the building itself and to reduce the load on the electricity grid, battery storage is often used to better match demand and generation. As battery storage is still expensive, the conversion of electricity into thermal energy via heat pumps and the use of thermal storage can provide an even cheaper alternative or additional flexibility to battery storage (Zator and Skomudek, 2020). A heat pump in combination with a thermal storage is typically operated in an on/off mode, depending on the temperatures in the thermal storage, without considering the PV electricity generation. Sometimes simple rule-

10.18086/swc.2023.08.01 Available at http://proceedings.ises.org

based approaches are used to increase the PV self-consumption rate. They switch on the heat pump and/or increase the set point for the supply water temperature when the PV power exceeds the electricity demand of the households. More advanced control strategies using model predictive control (MPC) have also been developed, e.g., Kuboth et al. (2019). However, they cannot easily be implemented due to the high effort of designing and parameterizing each energy system model required for the variety of existing multi-family buildings (Dounis and Caraiscos, 2009). In recent years, the increase in available computing power has led to an intense development and exploration of artificial intelligence (AI) approaches for a variety of applications. Reinforcement learning (RL) is one of the three types of machine learning that is used for system control (Sutton and Barto, 2018). RL agents learn control strategies by interacting with the environment (building energy system) and can learn preferences, such as comfort requirements, by interacting with users. This method is potentially model-free, meaning that models of the environment are not required, as is the case with mixed integer linear programming (MILP) approaches. Forecasts - for weather and heat/electricity demand - are not required but can be implemented and will improve the results. (Yu et al., 2020, 2021, Vázquez-Canteli and Nagy, 2019) Due to these advantages, RL became attractive for the energy management in residential buildings and has been investigated for the control of a variety of building energy (sub)systems. Wu et al. (2018) write that reinforcement learning algorithms "…could be promising candidates for home energy management…" and Yu et al. (2021) want to "…raise the attention of SBEM [smart building energy management; author's note] research community to explore and exploit DRL, as another alternative or even a better solution for SBEM.". Therefore, this paper investigates an RL approach for controlling the energy system of a conventionally energy renovated multi-family building that includes an air source heat pump, thermal storage, PV system, and optional battery storage. The focus is on controlling the heat pump according to the temperatures in a stratified thermal storage to ensure the required supply water temperatures of the heating system, as a conventional controller does. The development of this RL agent will be the basis for further developments for the implementation of PV optimization.

The novelty of this work is the investigation of a multi-family building using a detailed building energy system model in MATLAB Simulink in combination with RL algorithms provided by the Reinforcement Learning Toolbox of MATLAB (The MathWorks, Inc., 2023) to control heat pump operation.

In Section 2, an overview of the use and development of reinforcement learning approaches for the control of building energy (sub)systems with a focus on residential buildings is given. Section 3 introduces the building energy system model developed in MATLAB/Simulink using the CARNOT toolbox (Solar Institut Jülich, 2022), describes typical control strategies, and shows the development of RL-based heat pump controller using the Reinforcement Learning Toolbox (The MathWorks, Inc., 2023). The results of annual simulations using i) the typical control strategies and ii) the developed RL-based control strategies are presented and discussed in Section 4. A conclusion and outlook are given in Section 5.

2. Literature Review

A literature review was conducted to obtain an up-to-date overview of related scientific work and to investigate the use of RL for energy systems in multi-family buildings. The most popular databases of the scientific journals *Science Direct* and *EIII* were analyzed. Keywords used for the research included machine learning, deep reinforcement learning, building/home energy management, building energy system, system control, building control, smart buildings, smart home, multi-family buildings. The most relevant papers are presented here.

Yu et al. (2021) provide a comprehensive review of publications on deep reinforcement learning (DRL) for smart building energy management (SBEM). They also identify existing open issues and suggest possible future research directions. A distinction is made between studies on single building energy subsystems, multi energy subsystems in buildings, and microgrids. Only six of the 15 publications mentioned for single building energy subsystems deal with residential buildings. Three of them examine electric water heaters and another three examine heating, ventilation, and air conditioning (HVAC) systems. Their primary objective is to reduce energy costs or consumption. Six of the eight publications on multi energy subsystems in buildings deal with residential buildings and have energy costs as an objective. Only one of them (Ye et al., 2020) investigates a system similar to the one described in this paper (Section 3), but with an additional back-up system of a gas boiler.

Vázquez-Canteli and Nagy (2019) reviewed reinforcement learning algorithms and modeling techniques for demand response, ultimately selecting and summarizing 105 articles. Most of the studies were conducted for only one or two (combined) systems, especially for HVAC systems, electric vehicle charging strategies, or distributed generators combined with electric storage. They identified HVAC and domestic hot water (DHW) systems as one of four major groups of energy systems with significant demand response (DR) potential. However, many publications on HVAC systems examine only the reduction of thermal energy demand while maintaining thermal comfort for building occupants. None of the publications cited examine a system like the one described in Section 3.

Langer and Volling (2022) study a system configuration identical to the one used in this paper. They transform a mixed integer linear program (MILP) into a DRL implementation. Through numerical analyses of self-sufficiency, they compare the results of a deep deterministic policy gradient (DDPG) algorithm with a model predictive controller (MPC) under full information (theoretical optimum) and a rule-based approach. Their best DDPG algorithm achieves an overall self-sufficiency of 75 % which outperforms the rule-based approach with 66 % and reaches almost the theoretical optimum of 79 %.

However, the buildings considered in Ye et al. (2020) and Langer and Volling (2022) are very energy efficient single-family homes due to their small system dimensions. The investigations there are based on simplified linear models developed in previous MILP investigations with time step sizes of one hour, where mass flows and temperatures in hydraulic circuits and thermal storages are not considered in detail.

To the best of the author's knowledge, there are no studies on the energy system of an energetically renovated multi-family building that includes an air source heat pump, a thermal storage, a PV system, and optional battery storage that are modeled in detail. In the next two sections, such a model is introduced, the development of RL agents for heat pump control is described and results of annual simulations are shown.

3. Building Energy System Model

For the simulation of typical control strategies and the development of RL-based control strategies in MATLAB/Simulink, a multi-family building was chosen according to the building age category E of the IWU building typology of Germany (Loga et al., 2015). There, the large variety of architecture and construction methods of the German residential building stock is described, analyzed, and classified. The rough classification of the energy quality of the buildings is based on specific parameters. The definition of these parameters, the assignment of the buildings to age categories and to the state of renovation, a summary of the number of buildings as well as the presentation of typical specifications and energy saving potentials of each category are part of this typology. A building in age category E was chosen because it represents 19 % (225 million m²) of the total multi-family buildings living space in Germany (1,168 million m²) and is therefore the largest category for multi-family buildings. Such buildings were built between 1958 and 1968 and typically have 3 to 5 floors with a heated floor area of about 2,850 m² divided into 32 apartments. For each category, three types of energy quality are described: the original unrenovated state, a conventional renovated state, and a future-oriented renovated state with a calculated space heating energy demand of 145.9 $\text{kWh/m}^2_{\text{net floor area}}$, 67.3 $\text{kWh/m}^2_{\text{net floor area}}$ and 42.0 $\text{kWh/m}^2_{\text{net floor area}}$ respectively. According to the results presented in Bongs et al. (2023), the conventionally renovated version of this building was chosen for the investigations because the heat supply temperatures can be reduced to 60 °C or less, which allows the use of heat pumps for the thermal energy supply.

3.1. Building energy system model

The investigated building with its heat and electricity demand and the energy system consisting of an air source heat pump (hp), a thermal storage, a PV system and, in an extended version, a stationary battery storage was modeled in MATLAB/Simulink using the CARNOT toolbox (Solar Institut Jülich, 2022). CARNOT is a toolbox extension that provides an open-source library of models for the calculation and simulation of typical components of building energy systems in MATLAB/Simulink. The models can be individually parametrized or even redesigned due to the open-source code that can be modified.

The system configuration modeled for the investigations is shown in Figure 1.

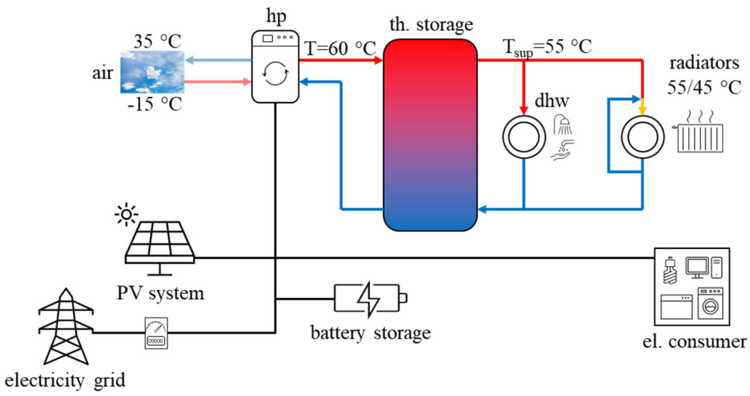

Fig. 1: Investigated building energy system configuration

The multi-family building itself was modeled with the "House simple" block, which uses a single zone building model with radiators as the heat transfer system. The parameters (geometry, U-values, etc.) were defined according to the data given in Loga et al. (2015) and supplementary data.

For the internal heat gains caused by the occupants, the occupancy profile given in SIA (2021) was used and upscaled to 60 persons.

The test reference year (TRY) of *Ingolstadt* provided by the German Weather Service (DWD, 2017) was used as input weather data for the simulations as well as for the load profile generation described below.

The method described in the VDI guideline 4655 (VDI, 2021) was used to generate the household electricity consumption profile, which is also used for internal heat gains caused by electrical appliances, and the domestic hot water (DHW) consumption profile. The guideline defines 12 typical days according to ambient temperature (winter, transition, summer), irradiation (clear or cloudy), and day of the week (workday or Sunday/holiday). For each typical day, normalized daily standard load profiles for electricity, DHW and heat consumption and PV generation are given in hourly time resolution for existing and low-energy buildings. The sequence of typical days and the annual normalized load profiles were determined from the DWD TRY data. The normalized profiles were scaled by an annual consumption of 96,000 kWh/a (3,000 kWh/a per household, VDI, 2021) for electricity consumption and 61,500 kWh/a (approx. 2,000 kWh/a per household, Loga et al., 2015) for DHW consumption.

The DHW hydraulic circuit is modeled in a simplified manner using a freshwater station. The supply water mass flow is controlled to match the given DHW load profile at a fixed return water temperature.

A separate hydraulic circuit is defined for the space heating with radiators. The supply water temperature is calculated by a heating curve according to the ambient temperature. The mass flow is controlled by a thermostatic valve to maintain the room temperature at 21 °C.

The ventilation system is included in the building model and ensures a hygienic air exchange with a ventilation rate of 0.4 1/h and prevents an overheating of the building in summer with additional ventilation at night.

To decouple the thermal energy consumption from the heat production and to provide some flexibility in the operation of the heat pump, a stratified thermal storage tank is modeled as a hydraulic separator between the heat pump hydraulic circuit and the consumer hydraulic circuit. For this purpose, the block "Buffer storage tank - type 1" has been modified to provide two consumer hydraulic circuits. Its volume is 8.5 m³, which corresponds to approximately 50 l/kW of the heat pump's thermal power.

For the heat supply, an air source heat pump with a thermal capacity of approximately 170 kW was modeled by a modified version of the "Heat pump" block. The modification removes the (unknown) thermal inertia and allows direct control of the supply water temperature. Real data from a heat pump manufacturer was used for the parameter set. To provide the DHW preparation, the setpoint for the supply water temperature of the heat pump was set to 60 °C.

To partially cover the electricity demand of the households and the heat pump, a PV system is modeled on the south-facing roof of the building. The usable roof area was examined from an existing multi-family building in *Ingolstadt* and is about 215 m², which results in a 65 kW_p PV system producing approximately 65,000 kWh_{el}/a.

In an extended version for the simulations, a battery storage with a usable capacity of 65 kWh (1 $kWh_{el}/kW_{el,PV}$) is modeled. It is located between the consumers/producers and the electricity grid. The (dis)charging power is limited according to the specifications of the battery storage. The control strategy (described in Section 3.2) is implemented in the model. The outputs are the state of charge (SOC) and an energy balance showing the electricity consumed from and fed into the grid.

Due to the high complexity of the model described, a variable time step solver was required to solve the problem. As a result, the time steps for the model calculation were much shorter than the hourly time resolution used in the MILP-based approaches.

3.2. Typical control strategies

For the control of the heat pump and the battery storage, typical control strategies have been defined and implemented in the simulation model. They are used as a reference.

Heat pumps in combination with thermal storages are typically operated in an on/off mode depending on the temperatures in the thermal storage (basic control strategy) as shown in Figure 2. In this study 55 °C are used as setpoint for the heat pump control.

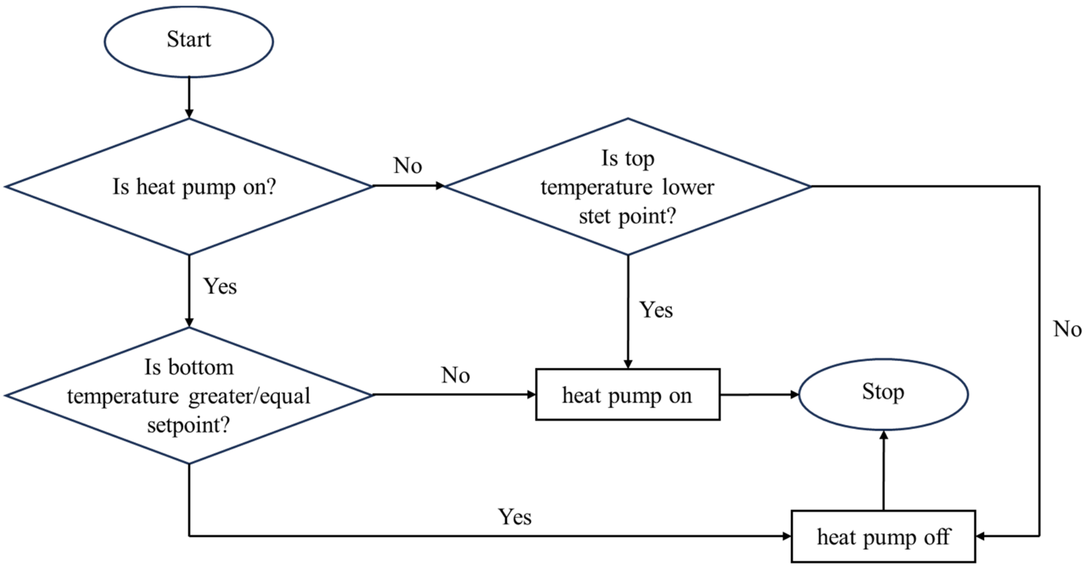

Fig. 2: Typical control strategy of heat pump

Sometimes simple rule-based approaches are used to better match the heat pump electricity consumption and the PV generation. For example, they increase the set point for the heat pump's supply water temperature by 5 K when the PV power exceeds the household's electricity demand (rule-based PV optimized strategy). This results in the heat pump being turned on when it was off, and at the same time, when it is turned on, the electricity demand increases due to the increased temperature difference between the heat source and heat sink. In the case shown here, the supply water temperature set point is increased when the PV surplus exceeds 25 kW_{el}, which is about half of the maximum simulated PV power.

When a battery storage is part of the building energy system, it is typically charged when not all of the electricity generated by the PV system is consumed directly by the households and the heat pump, and is discharged when the electrical loads exceed PV power.

3.3. Development of reinforcement learning based building energy system control

An RL agent learns an optimal policy by interacting with an unknown environment. In the study presented here, the policy is the control strategy for the heat pump to be learned, and the environment is the building energy system modeled in MATLAB/Simulink. Interacting in this case means that at each time step, the agent receives some information from the building energy system (observations) and decides how to operate the heat pump (action) based on the observations and the learned policy. For each action, the RL agent receives a reward depending on the defined reward function that evaluates the chosen action. In the learning phase, the policy is optimized to obtain the maximum reward over a defined period of time (in this case over the total episode length) to optimally perform the required task. The challenge in designing an RL agent is the goal-oriented definition of the reward function, the selection of an appropriate RL agent/method with its policy (e.g., deep neural network) and learning algorithm (policy optimization), and the parameterization of the hyper-parameters e.g., sample time, discount factor or learning rate.

The workflow for setting up an RL-based controller is shown in Figure 3.

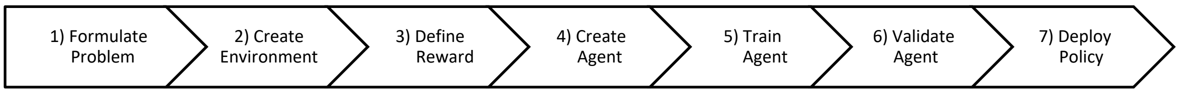

Fig. 3: Workflow of setting up an RL-based controller according to The MathWorks, Inc. (2023)

For steps 3) through 6) the Reinforcement Learning Designer application provided by the Reinforcement Learning Toolbox (The MathWorks, Inc., 2023) was used.

1) Formulate Problem

At the beginning of the RL agent design process, it is necessary to define the task, the way of interaction and the objectives. In this study, the objective of the RL agent is to control the heat pump in such a way that the thermal storage always delivers temperatures of at least 55 °C to the consumers, as the typical controller (Section 3.2) does. The continuous action space (0.2 to 1) is the control signal of the modulating heat pump. It is switched off when the control signal is less than the minimum modulation level of 0.3, or switched on when it is between 0.3 and 1. It is chosen according to the observation of the top T_{top} and bottom storage temperatures T_{bottom} and, in an advanced version, the thermal load for space heating and domestic hot water preparation.

2) Create Environment

The environment the agent interacts with is created as a MATLAB/Simulink subsystem and describes the building energy system shown in Section 3.1. This subsystem is connected to the RL agent block by the input of the control signal for the heat pump and the outputs of the top and bottom temperatures of the thermal storage, the thermal load, and the reward.

3) Define Reward Function

The reward function is based on the temperatures of the thermal storage. It gives a negative reward r i) if the top and bottom temperatures are too low to maintain the heat supply temperature and ii) if the bottom temperature is too high to prevent the thermal storage from overheating. In all other cases, the reward r is zero or positive. The definition of this region-based reward is shown in the following equations.

$$r_1 = -4 \quad if\ T_{top} < 55\ °C \qquad \text{(eq. 1)}$$
$$r_2 = 0 \quad if\ 55\ °C <= T_{top} < 57\ °C \quad else\ 3 \qquad \text{(eq. 2)}$$
$$r_3 = -16 \quad if\ T_{bottom} > 58\ °C \qquad \text{(eq. 3)}$$
$$r_4 = -6 \quad if\ 58\ °C >= T_{bottom} > 53\ °C \qquad \text{(eq. 4)}$$
$$r_5 = 2 \quad if\ 53\ °C >= T_{bottom} >= 45\ °C \quad else\ -5 \qquad \text{(eq. 5)}$$

At each time step t the reward r_t is calculated:

$$r_t = r_1+r_2+r_3+r_4+r_5 \qquad \text{(eq. 6)}$$

This reward function is implemented in the MATLAB/Simulink subsystem that contains the building energy system model.

4) Create Agent

An RL agent contains two components: the policy, represented by a function approximator (e.g., artificial neural networks), and a learning algorithm that optimizes the policy based on the chosen actions, observations from the environment, and rewards. Both depend on the type of agent.

The RL Designer application offers six types of agents:

- Deep Q-Network (DQN)
- Deep Deterministic Policy Gradient (DDPG)
- Twin-Delayed Deep Deterministic (TD3)
- Proximal Policy Optimization (PPO)
- Trust Region Policy Optimization (TRPO)
- Soft Actor-Critic (SAC)

When creating an RL agent, the RL Designer application only suggests agent types that are compatible with the environment being created. In this case, the Soft Actor-Critic (SAC) is chosen because it is model-free and can handle continuous action spaces. Values between 0.2 and 1 are defined as the action space for the control signal of the heat pump. In contrast to the time variable step size of the model solver, the RL agent requires a fixed time step size for its actions. As a trade-off between high temporal resolution of the RL agent actions and an acceptable computation time, a sample time of 600 seconds was chosen.

Two agents were created to explore the influence of the number of observations on the performance of the learned policy. The first receives only the two temperatures of the thermal storage as observations, while the second also observes the thermal load for space heating and domestic hot water preparation.

5) Train Agent

The agents were trained for a maximum episode length of three winter days. The rewards are cumulated over a whole episode. The stopping criterion for the leaning process was the average reward of the episodes. So, when the average reward over a defined number of episodes is not improving significantly the training will be stopped.

6) Validate Agent

After training the agent, its learned policy was evaluated in an annual simulation. The results are presented in Section 4.

7) Deploy Policy

As a final step in the RL agent design process, the learned policy can be transferred to real controllers by generating appropriate code. This paper focuses on a simulation-based investigation; therefore, this step was skipped.

4. Results and Discussion

The building energy system model described in Section 3.1 was used to run simulations with different control strategies for the heat pump. The simulation results are presented in the following subsections.

4.1. Simulation of typical control strategies

As a reference, the typical on/off strategy based on the temperatures of the thermal storage described in Section 3.2 (basic control strategy) was simulated with and without stationary battery storage. The calculated annual electricity consumption of the heat pump is 90,250 kWh_{el}/a. It provides 230,500 kWh_{th}/a of heat with an annual coefficient of performance (COP) of 2.555. The monthly distribution for the electricity consumption of the heat pump and households is shown in Figure 4.

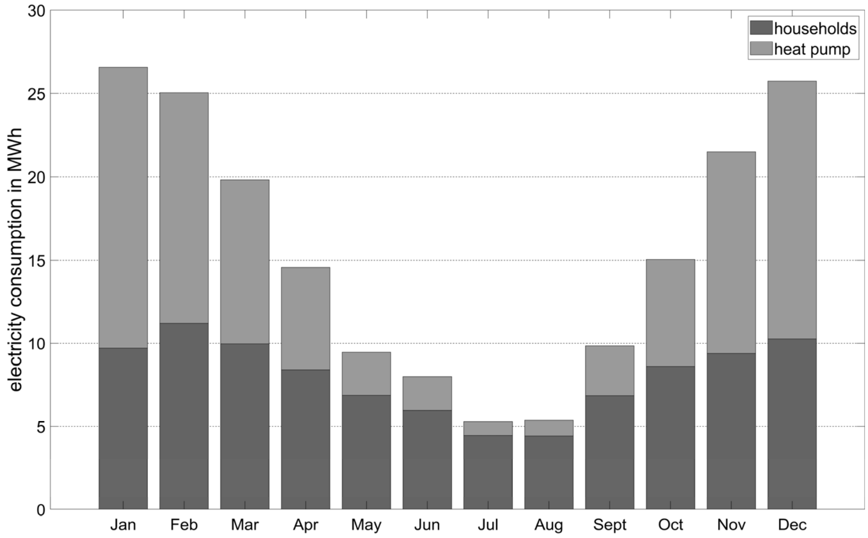

Fig. 4: Monthly distribution of total electricity consumed, broken down by heat pump and households

The monthly over all electricity consumption varies between 5.3 MWh in July and 26.6 MWh in January. Comparing the monthly distribution of the electricity consumption with the PV generation (Figure 5), it is clear that there is a PV surplus in summer and a shortage of self-generated electricity in winter.

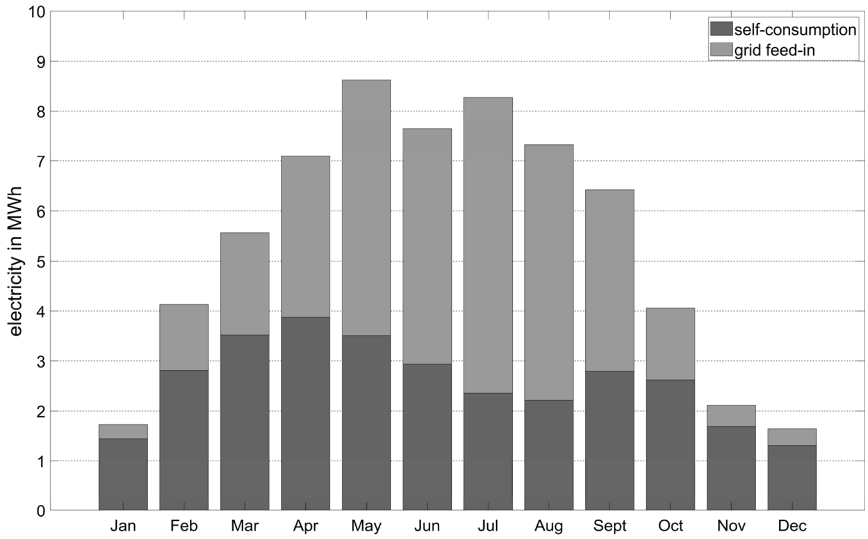

Fig. 5: Monthly distribution of PV electricity generation, broken down by self-consumption and grid feed-in for basic control strategy

For the variant without battery storage, the simulation results (Figure 5 and Table 1) show a higher annual self-consumption rate (48.1 %) and a lower annual self-sufficiency (16.7 %) compared to conventional single-family buildings (Quaschning, 2022) or a net-zero energy single-family building (both approx. 37 %, Milan et al., 2012). This is due to the higher electricity demand compared to the size of the PV system. In the winter months, up to 80 % of the monthly generated PV electricity is self-consumed, while in summer months up to 70 % has to be fed into the grid (Figure 5). Even in an ideal case (infinite capacity of the battery storage), only about 1/3 of the total electricity consumption can be covered by the PV system.

By implementing battery storage, the annual self-consumption rate and self-sufficiency can be increased by a factor of about 1.5, and the grid consumption and feed-in can be reduced. The monthly self-consumption rate can be increased up to 100 % in winter.

Tab. 1: Simulation results

Simulation Variant / Performance Indicator	Basic Control without Battery Storage	Basic Control with Battery Storage	PV Optimized without Battery Storage
Self-consumption rate in %	48.1	72.2	49.3
Self-sufficiency in %	16.7	25.1	17.1
Electricity demand of heat pump in kWh	90,250		90,511
Annual coefficient of performance	2.555		2.547
Electricity consumption from grid in kWh	155,192	139,564	154,620
Electricity feed-in into grid in kWh	33,577	17,949	32,744

Instead of implementing battery storage, the rule-based PV optimized strategy described in Section 3.2 was tested. The simulation shows only very small improvements of about 2.5 % for the self-consumption rate and self-sufficiency compared to the basic control strategy. One reason for this is that, in the on/off mode, the electric power consumed by the heat pump is always higher than the surplus of PV power, which leads to electricity consumption from the grid. Therefore, a modulating operating mode is necessary to improve the match between the surplus and the heat pump's electricity demand and should be considered in the development of RL approaches. The second reason for only very small improvements is that the high PV electricity surplus in summer cannot be shifted to cover the household's electricity consumption in the evening and at night, as a battery storage can do. Therefore, even with more advanced control strategies for the heat pump, battery storage should be used.

The simulation results shown here will be used as a reference for comparison with more advanced RL-based strategies that will be developed based on the RL agent introduced in Sections 4.2 and 4.3.

4.2. Design and training of reinforcement learning based heat pump control

During the initial learning process, high temperatures in the bottom storage tank caused an infinite mass flow between the heat pump and the thermal storage, resulting in a training termination. To ensure that the agent can explore actions that lead to high bottom storage tank temperatures and high negative rewards without terminating the learning process, an additional controller was implemented in the environment. It turns off the heat pump when the bottom storage tank temperature exceeds 58 °C, even if the action chosen by the RL agent is the opposite. This approach allows the agent to learn a policy that avoids too high temperatures in the bottom storage tank.

Two RL agents were trained with different numbers of observations. Figure 6 shows the episode reward (light blue line) and the average of 5 episodes (dark blue line) for both trainings. The increasing rewards from episode to episode show a good improvement of the policy. In both cases, the stopping criterion is met after only eleven episodes with an episode reward of about 1,880. So, the learning process itself is obviously not improved by an additional observation.

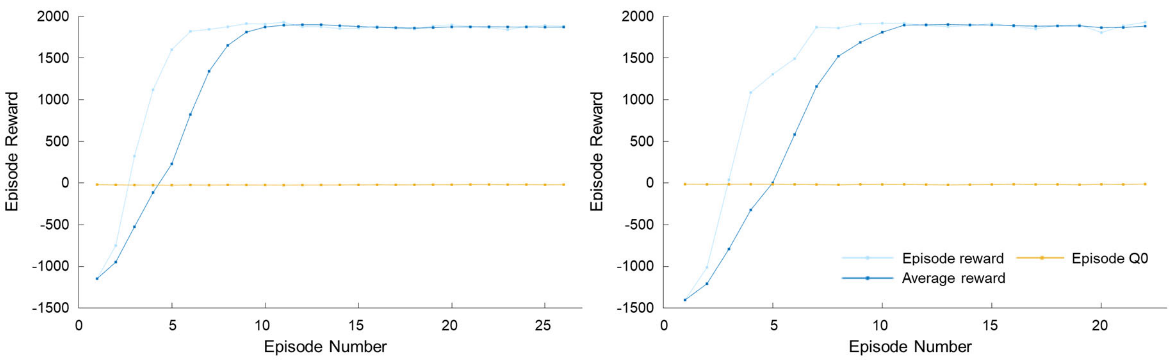

Fig. 6: Episode reward during learning process of RL agent with two observations (left) and three observations (right)

Due to the different thermal loads, especially for space heating, in winter and summer, it was tested to train the RL agent for an additional summer period of three days. However, this resulted in decreasing rewards, so the policy learned for winter only was used for the annual simulations presented in Section 4.3.

4.3. Simulation of reinforcement learning based heat pump control

The annual simulations show similar electricity consumption from and feed-in into the electricity grid compared to the basic control strategy, which was expected since there is no PV optimization implemented yet.

Both RL agents can always provide the required supply water temperatures. However, the analysis of the annual positive and negative rewards earned by the two agents (Table 2) shows that in an identical environment and with an identical reward function, adding one observation can improve the operation of the heat pump to better meet the temperature constraints implemented in the reward function, even though the training itself did not show any improvement.

Tab. 2: Comparison of gained annual rewards

Reward	2 Observations	3 Observations	Improvement absolute	Improvement relative in %
Positive	2,436,545	3,361,755	925,210	38.0
Negative	-1,530,056	-1,011,139	518,917	33.9
Sum	906,489	2,350,616	1,444,127	159.3

Including an additional observation can increase the positive annual reward by 38.0 % and reduce the negative annual reward by 33.9 %. In total the reward increases by 159.3 %.

The intervention of the additional controller, which switches off the heat pump when the bottom storage tank temperature exceeds 58 °C, was also analyzed. The total intervention time of the controller and the number of interventions were calculated for both agents (Table 3).

Tab. 3: Analyses of the intervention of the additional controller

	2 Observations	3 Observations
Total intervention time in h	53.1	74.4
Number of interventions	80	105

The additional controller only needs to intervene for up to approximately 75 hours, or 105 times per year. The average intervention time is less than one hour. Most of the interventions take place in summer when the thermal load is low. This is consistent with the results of the additional summer period training (Section 4.2). Thus, there is a potential for further improvement in learning the agent switching of the heat pump at times of high bottom storage temperatures.

One solution is to modify the heat pump model. The fixed heat pump output temperature of 60 °C used in this simulation causes the infinite mass flow when the thermal storage is fully charged at 60 °C (Section 4.2). Therefore, using a variable output temperature depending on the return water temperature may be a promising approach to allow the RL agent to better learn to switch off the heat pump at high storage temperatures even without the intervention of the additional controller.

5. Conclusions and outlook

A review of the literature showed that most of the publications investigating RL approaches for building energy (sub)systems have been carried out for only one or two (combined) systems and have the reduction of energy costs or consumption as their primary objective. Only two publications use system configurations like the one investigated here. However, the models were developed for energy efficient single-family homes and are based on simplified linear models with time steps of one hour, which is a coarse time resolution for building energy system control and not suitable for detailed models. Also, mass flows and temperatures in hydraulic circuits and thermal storages are not considered in detail. However, it is very important to consider the temperature distribution in the thermal storage because its thermal capacity is highly dependent on it, and the efficiency of the heat pump is influenced by the output temperature. Therefore, the detailed model used here provides more realistic results.

A detailed model of a multi-family building energy system was developed in MATLAB/Simulink using the CARNOT toolbox (Solar Institut Jülich, 2022). Typical control strategies were used for an annual simulation, which will serve as a reference for future RL agent developments implementing PV optimization.

The Reinforcement Learning Designer application of the Reinforcement Learning Toolbox (The MathWorks, Inc., 2023) was used to design and train two RL agents for the heat pump control. One agent receives only two storage tank temperatures as observations, while the other agent also observes the thermal load of the building. The RL agents were trained to provide the required supply water temperature of at least 55 °C to the consumers, as the typical controller does.

Annual simulations show that both RL agents can always provide the required supply water temperature. The annual reward of the RL agent using three observations instead of two increases by about 160 %, so the control strategy can be significantly improved by adding one observation.

The analysis of the additional controller that prevents too high bottom storage temperatures shows that it only intervenes up to 75 hours per year, mostly in summer, which is a satisfactory result. However, there is potential to improve learning the agent switching of the heat pump at times of high bottom storage temperatures e.g., by modifying the heat pump model.

With a well-trained reinforcement learning agent that satisfies the temperature constraints in the thermal storage, the next step is to design an agent that takes into account the PV electricity generation to control the heat pump to increase the self-consumption rate. Therefore, PV electricity generation and household electricity consumption need to be added to the observations, and electricity exchange with the grid needs to be implemented as a criterion in the reward function. The challenge in designing the reward function will be the weighting of the two objectives, the required temperatures in the thermal storage, and the reduced electricity exchange with the grid.

Acknowledgments

The work presented in this paper was funded by the Bavarian Research Foundation within the research network "Energy – Sector Coupling and Micro Grids (STROM)" (Project ID: AZ 1473-20) within subproject 7 "Decentralized Energy Management". The objective of subproject 7 is to develop forecast and/or AI-based control strategies for energy systems in multi-family buildings that couple the heat, electricity, and e-mobility sectors.

References

Bongs, C., Wapler, W., Dinkel, A. Miara, M., Auerswald, S., Lämmle, M., Hess, S., Kropp, M., Eberle, R., Rodenbücher, B., Schmidt, F., Ruppert, M., Carbonare, N., Wagner, A., 2023. LowEx-Konzepte für die Wärmeversorgung von Mehrfamilien-Bestandsgebäuden, final report. Available online: http://www.lowex-bestand.de/wp-content/uploads/2023/03/Abschlussbericht_LiB.pdf. Accessed 8 August 2023

Dounis, A.I., Caraiscos, C., 2009. Advanced control systems engineering for energy and comfort management in a building environment - A review. Renewable and Sustainable Energy Reviews, vol. 13, pp. 1246-1261. DOI: 10.1016/j.rser.2008.09.015

DWD Deutscher Wetterdienst, 2017. Testreferenzjahre. Available online: https://www.dwd.de/DE/leistungen/testreferenzjahre/testreferenzjahre.html. Accessed 4 September 2023

Kuboth, S., Heberle, F., König-Haagen, A., Brüggemann, D., 2019. Economic model predictive control of combined thermal and electric residential building energy systems. Applied Energy, vol. 240, p. 372-385. DOI: 10.1016/j.apenergy.2019.01.097

Langer, L., Volling, T., 2022. A reinforcement learning approach to home energy management for modulating heat pumps and photovoltaic systems. Applied Energy, vol. 327, p. 120020. DOI: 10.1016/j.apenergy.2022.120020

Loga, T., Stein, B., Diefenbach, N., Born, R., 2015. Deutsche Wohngebäudetypologie: Beispielhafte Maßnahmen zur Verbesserung der Energieeffizienz von typischen Wohngebäuden, 2nd ed. IWU, Darmstadt. ISBN 9783941140479. Available online: http://www.building-typology.eu/downloads/public/docs/brochure/DE_TABULA_TypologyBrochure_IWU.pdf. Accessed 8 August 2023

Milan, C., Bojesen, C., Nielsen, M. P., 2012. A cost optimization model for 100% renewable residential energy supply systems. Energy. vol. 48, pp. 118-127. DOI: 10.1016/j.energy.2012.05.034

Prognos, Öko-Institut., Wuppertal Institut, 2021. Towards a Climate-Neutral Germany by 2045. How Germany can reach its climate targets before 2050. Executive Summary conducted for Stiftung Klimaneutralität, Agora Energiewende and Agora Verkehrswende. Available online: https://www.agora-energiewende.de/en/publications/towards-a-climate-neutral-germany-2045-executive-summary. Accessed 9 October 2023

Quaschning, V., 2022. Regenerative Energiesysteme -Technologie, Berechnung, Klimaschutz, 11th ed. Hanser, München. ISBN: 978-3-446-47163-4

Solar Institut Jülich. CARNOT Toolbox Ver. 7.3, 2022 für Matlab/Simulink R2018b, Available online: https://www.fh-aachen.de/forschung/solar-institut-juelich/carnot. Accessed 8 August 2023

SIA Schweizerischer Ingenieur- und Architektenverein, 2021. SIA 2024 – Raumnutzungsdaten für die Energie- und Gebäudetechnik, Zürich.

Sutton, R. S., Barto, A. G., 2018. Reinforcement Learning : An Introduction, second edition. MIT Press, Cambridge, ISBN 9780262039246

The MathWorks, Inc., 2023. Reinforcement Learning Toolbox™ – User's Guide. The MathWorks, Inc.

Vázquez-Canteli, J. R., Nagy, Z., 2019. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy, vol. 235, pp. 1072-1089. DOI: 10.1016/j.apenergy.2018.11.002

VDI Verein Deutscher Ingenieure e.V., 2021. VDI 4655 - Referenzlastprofile von Wohngebäuden für Strom, Heizung und Trinkwarmwasser sowie Referenzerzeugungsprofile für Fotovoltaikanlagen, Düsseldorf.

Wu, D., Rabusseau, G., François-lavet, V., Precup, D., Boulet, B., 2018. Optimizing Home Energy Management and Electric Vehicle Charging with Reinforcement Learning. *ALA 2018: Adaptive Learning Agents Workshop at the Federated AI Meeting 2018*, Stockholm, 14-15 July. Available online: https://ala2018.cs.universityofgalway.ie/papers/ALA_2018_paper_37.pdf. Accessed 30 August 2023

Ye, Y., Qiu, D., Wu, X., Strbac, G., Ward, J., 2020. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3068-3082. DOI: 10.1109/TSG.2020.2976771

Yu, L., Xie, W., Xie, Di, Zou, Y., Zhang, D., Sun, Z., Zhang, L., Zhang, Y., Jiang, T., 2020. Deep reinforcement learning for smart home energy management. IEEE Internet of Things Journal, vol. 7, no. 4, pp. 2751-2762. DOI: 10.1109/JIOT.2019.2957289

Yu, L., Qin, S., Zhang, M., Shen, C, Jiang, T., Guan, X., 2021. A review of deep reinforcement learning for smart building energy management. IEEE Internet of Things Journal, vol. 8, no. 15, pp. 12 046-12 063. DOI: 10.1109/JIOT. 2021.3078462

Zator, S., Skomudek, W., 2020. Impact of DSM on Energy Management in a Single-Family House with a Heat Pump and Photovoltaic Installation. Energies, vol. 13, no. 13, 5476. DOI: 10.3390/en13205476