Control Optimization of Residential Air-Source Heat Pumps Using Reinforcement Learning

Ben Anton Goff
University of Applied Sciences Augsburg
Email: bengoff16@gmail.com

I. Motivation and Background

Heating and cooling account for a major share of residential energy use, and in many European countries air-source heat pumps (ASHPs) are being installed at record rates to replace fossil-fuel boilers. While these systems are efficient, they are typically operated with fixed rule-based control parameters that cannot adapt to the building's changing thermal dynamics, occupancy, or weather conditions. This results in suboptimal performance, excessive compressor cycling, and unnecessary energy use.

Every dwelling exhibits unique thermal inertia, insulation, and occupant behavior, meaning that a single static control strategy cannot achieve optimal operation across diverse scenarios. Reinforcement learning (RL) provides a framework in which a control policy can learn from interaction with the environment and improve its performance autonomously. In contrast to classical model predictive control (MPC), which relies on explicit optimization at each timestep, RL can learn directly from data to approximate long-term cost-to-go functions.

However, applying RL to building systems presents challenges: data collection is slow, exploration can damage equipment, and purely model-free algorithms often require millions of interactions to converge. The thermal dynamics of heat pumps are continuous and sluggish, with state changes occurring over minutes rather than milliseconds. These characteristics make model-based reinforcement learning (MBRL) especially suitable. By incorporating a simplified physical model of heat transfer and compressor behavior, the agent can simulate trajectories internally, improving sample efficiency and stability. In this project, MBRL is proposed as a method to enable adaptive, safe, and interpretable control of residential ASHPs, bridging the gap between physical modeling and data-driven decision making.

II. Related Work

A. RL for HVAC and Building Energy Control

Wei [1] applied deep Q-learning to HVAC control in commercial buildings and demonstrated significant energy savings without compromising comfort. Yang [2] developed an RL-based approach for bi-directional energy management in smart buildings, enabling adaptive control under dynamic conditions. De Ridder [3] combined physical building models with deep RL to improve control stability and interpretability. A survey by Li [4] reviewed model-based reinforcement learning methods for building energy optimization, highlighting hybrid approaches as the most promising.

B. Heat Pump Control and Demand Response

Kazmi [5] investigated reinforcement learning for optimizing residential heat pump operation, achieving better efficiency under variable loads. Ruelens [6] introduced batch RL for thermostatically controlled loads, providing a foundation for residential demand response applications. The Dynamic framework proposed by Sutton [7] remains a key conceptual basis for integrating simulated and real experience in model-based learning.

C. Model-Based and Hybrid Methods

Mohammadi [8] developed a hybrid MBRL controller for energy-efficient building operation, balancing physical modeling with policy learning. Schlüter [9] implemented model-based RL for HVAC control in smart buildings, demonstrating improved adaptability and robustness. Zhang [10] explored combining predictive models with reinforcement learning for residential energy systems, achieving higher sample efficiency and safer control.

III. Problem Formulation

The control of an air-source heat pump can be formulated as a Markov Decision Process (MDP):

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$$

where $\mathcal{S}$ represents the set of measurable system states, $\mathcal{A}$ the available control actions, P the transition model, R the reward function, and γ the discount factor. The aim is to derive an optimal policy $\pi^*(a|s)$ that minimizes long-term operational cost and energy use while maintaining indoor comfort and extending system lifespan.

A. State and Action Definitions

The state vector s_t contains environmental and operational variables:

$$s_t = [T_{in}, T_{out}, T_{set}, P_{hp}, \text{mode}, \text{time}, \text{price}]$$

where T_{in} is indoor temperature, T_{out} is outdoor temperature, T_{set} is the target setpoint, P_{hp} is the compressor power, and mode denotes heating, idle, or defrost operation. The action space includes compressor power modulation and discrete mode switching.

B. Physical Transition Model

System transitions are described by a simplified heat balance equation:

$$T_{in,t+1} = T_{in,t} + \frac{\eta P_{hp,t} - U(T_{in,t} - T_{out,t})}{C_{th}} \Delta t$$

where C_{th} is the thermal capacity of the building, U is the overall heat loss coefficient, η is the coefficient of performance (COP) of the heat pump, and Δt is the time step. This deterministic model forms the basis for the internal predictive model $\hat{P}(s_{t+1}|s_t, a_t)$ used for planning and policy updates.

C. Model-Based Learning Process

Unlike model-free RL, the MBRL agent employs both real and simulated experiences. Using the learned or analytic model $\hat{P}$, the agent generates synthetic trajectories to estimate future returns and refine the value function:

$$\theta \leftarrow \theta + \alpha \nabla_\theta \mathbb{E}_{s,a \sim \hat{P}}[R(s,a) + \gamma V_{\theta'}(s') - V_\theta(s)]$$

This enables faster learning from limited data and constrains the policy search to physically realistic state transitions.

D. Compressor Dynamics and Longevity

A key secondary objective is to minimize compressor wear by reducing the number of start–stop cycles. Frequent switching causes mechanical stress and shortens compressor lifespan, directly affecting the reliability and sustainability of the system. To incorporate this into the optimization, an additional penalty term C_{cycles} is added to the reward function:

$$r_t = -\alpha |T_{in,t} - T_{set}| - \beta P_{consumed,t} - \lambda C_{cycles,t}$$

where λ balances comfort, energy efficiency, and longevity. By discouraging rapid state changes, the RL agent learns smoother control actions that maintain temperature stability while extending component life.

E. Objectives and Expected Outcomes

The objective is to learn a policy π that minimizes long-term cumulative cost:

$$\min_\pi \mathbb{E}\left[\sum_t \left(\alpha |T_{in,t} - T_{set}| + \beta P_{consumed,t} + \lambda C_{cycles,t}\right)\right]$$

Expected outcomes include:

- Improved sample efficiency compared to model-free RL.
- Reduced energy consumption while maintaining comfort.
- Fewer compressor start–stop cycles and smoother operation.
- Increased interpretability and safety through physically constrained transitions.

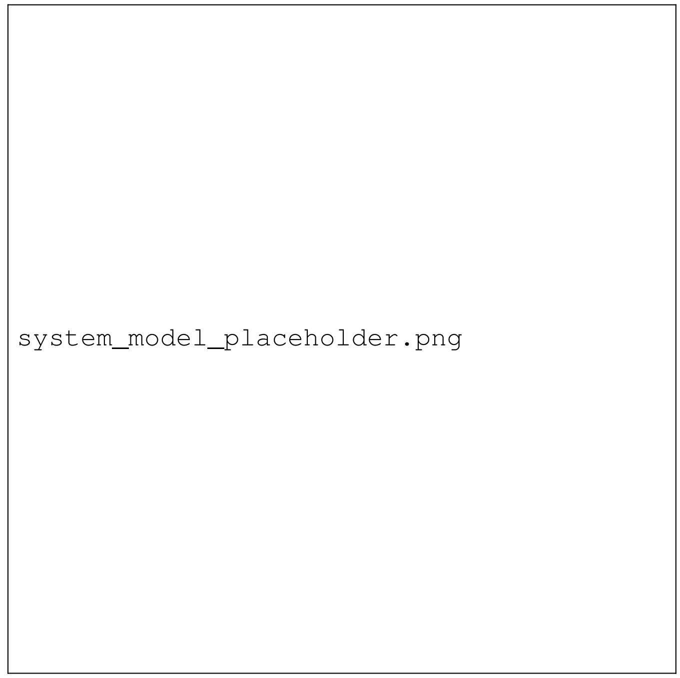

Fig. 1. Conceptual structure of the proposed model-based RL controller for a residential air-source heat pump.

References

[1] T. Wei, Y. Wang, and Q. Zhu, "Deep Reinforcement Learning for Building HVAC Control," in *Proc. 54th Design Automation Conf.*, 2017, pp. 1–6.

[2] R. Yang, T. Zhang, and L. Wang, "Reinforcement Learning for Building Energy Management through Bi-directional Coordination," in *IEEE ISGT*, 2015, pp. 1–5.

[3] D. De Ridder, S. Verhelst, and L. Helsen, "Combining Physical Building Models with Deep Reinforcement Learning," *Energy and Buildings*, vol. 224, 2020.

[4] J. Li, K. Li, and Z. Wang, "Model-Based Reinforcement Learning for Building Energy Optimization: A Survey," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 53, no. 1, pp. 312–325, 2023.

[5] H. Kazmi, J. Driesen, and B. Minne, "Reinforcement Learning Based Control of Domestic Heat Pumps," in *IEEE ENERGYCON*, 2018, pp. 1–6.

[6] F. Ruelens, B. Claessens, and S. Vandael, "Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4921–4932, 2018.

[7] R. S. Sutton, "Dyna, an Integrated Architecture for Learning, Planning, and Reacting," *ACM SIGART Bulletin*, vol. 2, no. 4, pp. 160–163, 1991.

[8] S. Mohammadi, A. Sadeghi, and M. Behrangrad, "Model-Based Reinforcement Learning Control for Energy-Efficient Building Operation," in *IEEE SmartGridComm*, 2021, pp. 1–6.

[9] A. Schlüter, T. van Treeck, and D. Müller, "Reinforcement Learning for Model-Based HVAC Control in Smart Buildings," *Energy and Buildings*, vol. 214, 2020.

[10] Y. Zhang, X. Xu, and K. Li, "Hybrid Reinforcement Learning with Predictive Models for Residential Energy Systems," *IEEE Trans. Ind. Informatics*, vol. 18, no. 7, pp. 4820–4832, 2022.