Multi-Agent DRL Framework for Coordinated HVAC Control in Large-Scale Buildings

Wenhan Yu[a], **Jun Zhao**[a]

[a] *Nanyang Technological University, Singapore* junzhao@ntu.edu.sg

Abstract

This paper presents a proposal for addressing the complexities of coordinating multiple Air Handling Units (AHUs) within large indoor environments to enhance energy efficiency and occupant comfort. Traditional Reinforcement Learning (RL) methods typically focus on scenarios with a single AHU, often neglecting critical variables such as humidity. We propose a detailed system model that incorporates temperature, humidity, and CO_2 dynamics within a multi-objective optimization framework. Our proposed method leverages a multi-agent Deep Reinforcement Learning (DRL) framework, where each AHU operates as an autonomous agent. This conceptual approach aims to improve the scalability and adaptability of HVAC systems in extensive building infrastructures, aiming to surpass the constraints associated with conventional single-AHU strategies.

1. Introduction

Modern commercial buildings require advanced HVAC control strategies that balance energy efficiency with stringent comfort demands, particularly in configurations involving multiple AHUs. Traditional model predictive control (MPC) approaches struggle with the complex dynamics of these systems [1]. Recent DRL applications show potential but tend to simplify system dynamics or the control structure [2–4]. Moreover, most DRL studies on HVAC control focus on single-AHU setups and fail to address the scalability needed for multiple AHUs that manage extensive spaces [5]. This oversight is significant in large-scale buildings where several AHUs must work in unison to maintain optimal conditions.

Addressing these challenges, this paper introduces a more detailed system model that considers the interplay among temperature, humidity, and air quality across multiple AHUs. We propose a multi-agent DRL approach where each AHU acts as a learning agent with localized zone control but operates within a coordinated system-wide strategy. This method aims to reduce energy consumption while improving comfort in complex building environments, leveraging decentralized control to enhance system responsiveness and efficiency.

2. System Model

Consider a building floor or large commercial space partitioned into N zones. Let $\mathcal{N} = \{1, 2, \ldots, N\}$ denote the set of zones. These zones are grouped under M AHUs, so that each AHU $m \in \{1, \ldots, M\}$ controls a subset $\mathcal{K}_m$ of zones ($\sum_m |\mathcal{K}_m| = N$). The time horizon is divided into L discrete slots, $t \in \{1, \ldots, L\}$, each of duration Δt.

Control Variables:

Zone-Level Damper Angles, θ_n^t: Each zone n is served by a VAV box with a damper controlling airflow. The angle θ_n^t influences $f_n^t(\theta_n^t)$, the airflow delivered to zone n.

AHU Air Exchange Rates, x_m^t: Each AHU m determines how much fresh air is mixed with return air, represented by $x_m^t \in [0, 1]$. A higher x_m^t corresponds to drawing in a larger fraction of outside air.

Mixed-Air States: For AHU m, let C_m^t, T_m^t, and H_m^t be the CO_2, temperature, and humidity of the mixed air. If C_{out}^t, T_{out}^t, H_{out}^t denote outside conditions and $\{C_n^t, T_n^t, H_n^t\}$ are zone returns, then:

$$C_{\text{mix},m}^t = x_m^t \, C_{\text{out}}^t \; + \; (1 - x_m^t) \, \frac{\sum_{n \in \mathcal{K}_m} C_n^t \, f_n^t}{\sum_{n \in \mathcal{K}_m} f_n^t}, \quad (1)$$

The same for $T_{\text{mix},m}^t, H_{\text{mix},m}^t$. This captures how fresh outside air and return air from zones are blended.

Thermal Dynamics: Let T_n^t be the temperature in zone n at time t. We adopt a lumped heat-balance approach, where the net heat flow in each zone depends on walls, adjacent zones, internal loads, and HVAC-supplied airflow:

$$\begin{aligned} T_n^{t+1} = T_n^t \; + \; \Delta t \cdot \Big(& h_{\text{wall}}(T_n^t, \{T_{n'}^t\}) + h_{\text{internal}}(O_n^t) \\ & + h_{\text{AHU}}\big(f_n^t(\theta_n^t), \, T_m^*, \, T_n^t\big)\Big), \end{aligned} \quad (2)$$

where T_m^* is the supply-air setpoint, and O_n^t captures occupancy-based heat gains. Functions $h_{\text{wall}}(\cdot)$ and $h_{\text{internal}}(\cdot)$ can be derived from standard building thermal models [5]. The term $h_{\text{AHU}}(\cdot)$ encapsulates the cooling effect from air delivered to zone n.

Humidity Dynamics: Let H_n^t denote the humidity in zone n. Similar to temperature, humidity changes arise from (i) moisture flow through walls or adjacent zones, (ii) occupant-generated moisture, and (iii) AHU-supplied air. We write:

$$\begin{aligned} H_n^{t+1} = H_n^t \; + \; \Delta t \cdot \Big(& g_{\text{zone}}(H_n^t, \{H_{n'}^t\}) \\ & + g_{\text{occupant}}(O_n^t) + g_{\text{AHU}}\big(f_n^t(\theta_n^t), \, H_m^*, \, H_n^t\big)\Big), \end{aligned} \quad (3)$$

where H_m^* is the target humidity set by AHU m. Functions $g_{\text{zone}}(\cdot)$ and $g_{\text{occupant}}(\cdot)$ follow standard building moisture exchange principles.

CO_2 Dynamics: Maintaining healthy CO_2 levels is critical for occupant comfort and safety. Let C_n^t be the CO_2 concentration in zone n. The next state depends on how much of the zone air is replaced by

HVAC air (carrying mixed CO_2 level C_m^t) and on occupant respiration:

$$C_n^{t+1} = \left(1 - \alpha_n^t\right) C_n^t + \alpha_n^t \, C_m^t + r_n^t, \tag{4}$$

where $\alpha_n^t \propto \frac{f_n^t(\theta_n^t)\,\Delta t}{V_n}$ (a ratio of incoming airflow to zone volume V_n), C_m^t merges outside air with return air (scaled by x_m^t), and r_n^t is occupant-generated CO_2.

Energy Costs and Constraints: The AHU must cool/dehumidify the mixed air to maintain setpoints (T_m^*, H_m^*). This incurs a cost that depends on how much outside air (x_m^t) is used and how much total airflow $\sum_{n\in\mathcal{K}_m} f_n^t(\theta_n^t)$ is needed:

$$E^t(\boldsymbol{\theta}^t, \boldsymbol{x}^t) = \sum_{m=1}^{M} \frac{p^t}{\eta COP} \Bigg\{ \sum_{n\in\mathcal{K}_m} \rho\, c_p\, f_n^t\big(\theta_n^t\big) \Big(T_{\mathrm{mix},m}^t - T_m^*\Big) + \sum_{n\in\mathcal{K}_m} \rho\, f_n^t\big(\theta_n^t\big) \Big(H_{\mathrm{mix},m}^t - H_m^*\Big)\, L_v \Bigg\},$$

where p^t is the electricity price at time t, η is an efficiency factor, and COP is the Coefficient of Performance for the cooling/heating coil. ρ is the density of air, c_p is its specific heat capacity, and L_v is the latent heat of vaporization, relevant for dehumidification processes.

Feasibility Constraints:

$$T_{\min} \le T_n^t \le T_{\max}, H_{\min} \le H_n^t \le H_{\max}, C_n^t \le C_{\max}.$$

The above enforces safe ranges for all physical and comfort-related quantities.

3. Proposed Multi-Agent DRL Framework

Our control framework adopts a multi-agent Deep Reinforcement Learning (DRL) structure to manage the joint optimization of multiple AHUs and their respective zone-level dampers. Each AHU acts as a system-level agent controlling outside-air exchange, and each zone functions as a local agent modulating damper angles, designed to handle the complexity of large-scale HVAC systems.

Agents and Action Spaces Control is decomposed into two layers: system-level agents (AHU agents) output continuous actions representing the fraction of outside air mixed with recirculated air, affecting mixed-air states and subsequently influencing energy usage and indoor air quality. Zone-level agents (damper agents) modulate airflow through damper angles, directly impacting local temperature, humidity, and CO_2 dynamics.

State Space Each agent observes both local measurements and relevant global signals such as temperature, humidity, CO_2 levels, occupancy, and potentially a weather forecast or other exogenous factors.

Reward Design The global reward structure reflects energy efficiency and comfort objectives, penalizing deviations from desired setpoints for temperature and humidity and violations of CO_2 levels, balancing the trade-offs among these objectives.

Multi-Agent Learning Architecture A centralized training with decentralized execution (CTDE) scheme allows individualized policy learning for each agent while promoting coordinated behaviors through a shared global critic. This includes employing a multi-agent variant of Proximal Policy Optimization (PPO), handling policy updates and constraint enforcement to maintain system stability and ensure occupant safety.

Hierarchical Extensions and Implementation Considerations While MAPPO handles multiple agents, a hierarchical RL structure can simplify credit assignment and scaling. Implementation includes fully connected or recurrent neural networks, parameter sharing, and transfer learning across similar buildings to enhance training efficiency and system adaptability.

Conclusion

This paper proposes an initial framework for a DRL system aimed at managing multiple AHUs in expansive commercial settings while adhering to physical system constraints for optimization. Given the preliminary nature of this work, future efforts will focus on refining the system model and algorithms based on the formulated problem structure. Detailed crafting of these algorithms will be essential for advancing the framework's broad applicability in complex HVAC control scenarios.

References

[1] Roger Kwadzogah, Mengchu Zhou, and Sisi Li. Model predictive control for hvac systems—a review. In *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 442–447. IEEE, 2013.

[2] Liang Yu, Yi Sun, Zhanbo Xu, Chao Shen, Dong Yue, Tao Jiang, and Xiaohong Guan. Multi-agent deep reinforcement learning for hvac control in commercial buildings. *IEEE Transactions on Smart Grid*, 12(1):407–419, 2020.

[3] Liang Yu, Shuqi Qin, Meng Zhang, Chao Shen, Tao Jiang, and Xiaohong Guan. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal*, 8(15):12046–12063, 2021.

[4] Matthias Hutsebaut-Buysse, Kevin Mets, and Steven Latré. Hierarchical reinforcement learning: A survey and open research challenges. *Machine Learning and Knowledge Extraction*, 4(1):172–221, 2022.

[5] Srinarayana Nagarathinam, Arunchandar Vasan, Venkata Ramakrishna P, Shiva R Iyer, Venkatesh Sarangan, and Anand Sivasubramaniam. Centralized management of hvac energy in large multi-ahu zones. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 157–166, 2015.