RECA: A Multi-Task Deep Reinforcement Learning-Based Recommender System for Co-Optimizing Energy, Comfort and Air Quality in Commercial Buildings

Stephen Xia
Columbia University
New York, New York, USA

Peter Wei
Columbia University
New York, New York, USA

Yanchen Liu
Columbia University
New York, New York, USA

Andrew Sonta
Columbia University
New York, New York, USA

Xiaofan Jiang
Columbia University
New York, New York, USA

ABSTRACT

We present the design and implementation of RECA, a novel human-centric recommender system for co-optimizing energy consumption, comfort and air quality in commercial buildings. Existing works generally optimize these objectives separately, or by only controlling energy consuming resources within the building without directly engaging occupants. We develop a deep reinforcement learning architecture based on multitask learning, demonstrate how it can be used to jointly learn energy savings, comfort and air quality improvements for different actions, and build a recommender system with humans-in-the-loop. Through real deployments in multiple commercial buildings, we found that RECA has the potential to further reduce energy consumption by up to 8% in energy-focused optimization, improve all objectives by 5 − 10% in joint optimization, and improve thermal comfort by up to 21% in comfort and air quality focused optimization, over existing solutions.

CCS CONCEPTS

• **Computer systems organization → Real-time systems**; **Embedded and cyber-physical systems**; • **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

building co-optimization, deep reinforcement learning, energy savings, air quality, thermal comfort, recommender system

ACM Reference Format:
Stephen Xia, Peter Wei, Yanchen Liu, Andrew Sonta, and Xiaofan Jiang. 2023. RECA: A Multi-Task Deep Reinforcement Learning-Based Recommender System for Co-Optimizing Energy, Comfort and Air Quality in Commercial Buildings. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3600100.3623735

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11…$15.00
https://doi.org/10.1145/3600100.3623735

1 INTRODUCTION

Commercial Buildings are responsible for nearly 40% of total energy consumption in the United States [14]. To work towards future sustainability, research communities, industry, and government agencies have developed projects and policies to improve energy efficiency in buildings. However, studies have shown that these efforts can still be improved [17]. In addition to energy consumption, comfort and air quality are key targets for optimization in commercial buildings, since they can lead to many benefits such as increased productivity and occupant health [9]. Jointly optimizing energy, comfort, and air quality is challenging due to the complex and often conflicting nature of these objectives.

Prior works optimize energy, comfort, and air quality by controlling energy consuming resources. However, there is a limit in savings that can be achieved in this way. For example, optimizing thermal comfort of multiple occupants in the same room who prefer different temperatures cannot be achieved by only changing the thermostat temperature. Because most energy consuming resources in a building are used to service occupants, the improvements we can achieve without directly engaging occupants are limited.

We present and deploy RECA, a novel recommender system that generates intelligent move and setpoint recommendations to co-optimize energy, comfort, and air quality co-optimization with humans-in-the-loop and allows building managers to tune which aspects to focus on. For example, tuning RECA to aggressively reduce energy consumption may prompt more occupants to move to shared spaces to reduce heating, ventilation, and air conditioning (HVAC) service to other spaces. This decreases energy, but also decreases air quality and comfort, since more people occupy a single space and one temperature setpoint may not satisfy everyone's preferences. We present the following contributions:

1. We introduce RECA, a recommender system that co-optimizes energy savings, occupant comfort, and air quality with humans-in-the-loop in real commercial buildings. Unlike previous works, RECA is tunable, allowing building managers to prioritize energy savings, occupant comfort, and/or air quality, and engages occupants with *actionable recommendations* (*move* and *thermostat setpoint*) to improve these objectives.

2. We integrate a novel deep reinforcement learning architecture, using multi-task learning to learn the effects of actions on energy, comfort, and air quality. Our architecture utilizes an embedding

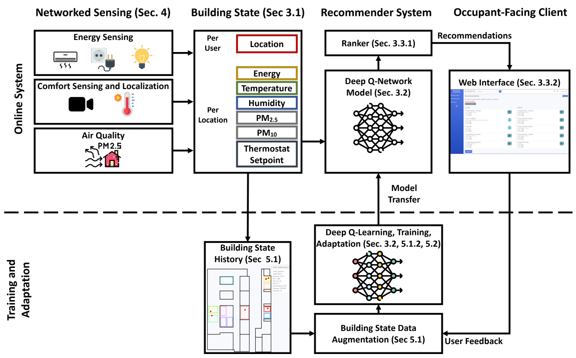


Figure 1: RECA's system architecture. To account for the challenge of cold-start, RECA leverages a simulation environment with statistical models to estimate future building states and generate more training examples from a past history of observed building states and recommendations.

to efficiently learn the location configurations of occupants and relationships between different locations.
3. Over a four week study, we evaluate RECA in two commercial office buildings and show that our system can account for a wide range of configurations that emphasize combinations of energy, comfort, and air quality.

2 RELATED WORKS

Works that optimize energy, comfort, and air quality first model a building before developing a control algorithm to perform optimization. Modeling software, such as EnergyPlus [7], have been commonly used to produce physics-based models of spaces, while data-driven models such as artificial neural networks [10] use real-world data. The building models are integrated into IoT systems [15] or paired with different controllers such as model predictive control (MPC) [8] to optimize energy consumption, comfort, and air quality. In general, these systems only control building resources, such as lighting and heating, ventilation, and air conditioning (HVAC).

Several works have proposed methods for optimizing energy and comfort by grouping occupants based on thermal preferences [18] or controlling HVAC resources with human feedback on comfort [5]. In contrast, we develop a recommender system to actively engage occupants throughout the day by delivering real-time actionable recommendations, allowing our system to adapt to changes in building resources and occupant behavior throughout the day. Moreover, our system not only uses human feedback to inform actions, but also recommends actionable steps to occupants in order to improve their own comfort, energy footprint, and/or air quality.

Reinforcement learning has become an important area for addressing dynamic environments. [1, 6] introduce reinforcement learning-based strategies for controlling windows and HVAC resources to optimize energy consumption or comfort/air quality. [24, 25] engage occupants by providing recommendations (e.g., schedule changes) to optimize energy consumption through Q-table and deep Q-network based recommender systems and demonstrate that strategies, which do not engage occupants, may pass over significant optimization opportunities. In contrast, our work co-optimizes comfort and air quality, in addition to energy consumption, by incorporating humans in the optimization process.

3 DEEP REINFORCEMENT LEARNING BASED RECOMMENDER SYSTEM

We propose RECA, a deep reinforcement learning-based (DRL) recommender system for co-optimizing energy consumption, occupant thermal comfort, and air quality, shown in Figure 1. There are several reasons why this problem is challenging. First, direct modeling of all dynamics in a commercial building that contribute to energy usage, comfort, and air quality is impractical, and the building's resources and occupants change over time, making it even more difficult. Second, the effects of different actions on these objectives are difficult to quantify, especially if changes are not realized until multiple steps in the future. Lastly, different occupants may value certain objectives and recommendations differently.

Model-free deep Q-learning can help address these challenges: first, deep Q-learning utilizes a deep neural network to approximate the state-action function, which is beneficial for large state spaces with many occupants and building resources. Secondly, deep Q-learning can learn action returns long-term, independent of the policy being followed and without requiring an explicit model of the complex effects between the environment and occupants. Lastly, returns for actions specific to each occupant can be learned separately, which allows for the model to account for different preferences among users.

We first explain how we represent the building energy, comfort, and air quality co-optimization problem in context of deep Q-learning (Section 3.1). Next, we introduce the deep Q-network model for generating actionable recommendations to occupants to co-optimize the three objectives (Section 3.2). Finally, we introduce our full recommender system that leverages the predictions provided by the Q-network to "recommend" actions to occupants to improve our objectives (Section 3.3).

3.1 Deep Reinforcement Learning Formulation

We represent the building co-optimization problem as follows. At each time step, the network uses a policy to choose an *action* a from a set of possible actions $a \in A$ based on the current state of the building, s, which constantly changes over time due to agent recommendations and external factors like outside temperature. This action is sent to the actor(s) (occupants in our case), which is then accepted or rejected. Our system then sees a *reward*, r, representing short-term changes given by the environment, which would be improvements energy, comfort, and air quality. For example, our system may recommend the only occupant in room A to move to room B. The occupant accepts and moves to room B, allowing the building to turn down the HVAC and lights in room A. As a result, the system observes a reward of energy savings from room A.

The building environment continuously changes, due to agent recommendations and external factors such as occupant location changes and environmental factors. Thus, we can formulate the problem as a finite Markov Decision Process (MDP). Each action has a transition probability of occurrence at the current state and a reward (r), representing short-term changes due to the action.

The goal of the deep Q-network is to estimate long-term changes associated with each action given the current building state s. Standard reinforcement learning aims to maximize this *return* at time t, defined as $R_t = \sum_{i=t}^{T} \gamma^i r_i$, where r_i is the reward observed at time

i, $\gamma \in (0,1)$ is the "discount factor", and T is the end time (or end of day in commercial buildings). We leverage Q-learning, a widely-used model-free reinforcement learning method, where the agent seeks an action-value function $Q(s,a)$ which represents the return of taking an action a at a given state s. If the state/action space is too large, $Q(s,a)$ is too complex to be stored in a data structure, but can be approximated effectively using a deep neural network [13].

3.1.1 State-Actions and Recommendation Types. RECA's state is comprised of the following features, which captures the three objectives we are co-optimizing.
(1) *Per Space:* energy consumption, temperature, humidity, thermostat setpoint temperature, $PM_{2.5}$, and PM_{10}.
(2) *Per Occupant:* Location of each occupant in the building at the space or room-level (e.g., the occupant is in lab space A), including a null indicator if the occupant is absent.

We divide the building into spaces based on functional purpose [21, 22] because occupants generally use and refer to spaces in this way (e.g., group A's workspace). There are three categories of *actionable recommendations* that our system recommends occupants in a building with $|O|$ occupants and $|S|$ spaces.
(1) *Move:* This recommendation suggests users to move to a different location. For example, moving multiple occupants into a single location may allow the building to turn down HVAC in other locations and reduce energy. There are $|O| \times |S|$ number of move actions that can be recommended. Suggesting occupants to move to areas that they would never reside in is unproductive. Instead, we introduce mechanisms that help RECA adapt to user location preferences (Section 5.2).
(2) *Thermostat Setpoint Changes:* This recommendation suggests users to change the thermostat setpoint at their location. We recommend changes in setpoint by $\approx \pm 2$ degrees Fahrenheit (exact values discussed in Section 5.1.1). As such, there are $2 \times |O| \times |S|$ total number of setpoint actions that can be recommended.
(3) *Temperature and Lighting Relaxation:* In empty rooms, we relax the temperature setpoint by 2 degrees Fahrenheit and turn off lights, without affecting occupants. This action is directly taken by the building when it observes empty spaces.
These categories of *actionable recommendations* allow our system to incorporate actions that existing works use to optimize energy savings and comfort, while enabling more complex action sequences, described in Section 6.3, that are otherwise not possible.

3.1.2 Reward. Since we are co-optimizing energy savings, thermal comfort, and air quality, the reward is the improvement in these objectives from one time step to the next and depend on the current building state, current action, and the building state at the beginning of the next time step. Equation 1 shows the reward at timestep n.

$$r_n = -\alpha E_n - \beta \Delta C_n - \gamma \Delta Q_n \tag{1}$$

E_n refers to the total energy consumption of all energy-consuming resources, C_n refers to the total comfort of all occupants, and Q_n is the total air quality rating experienced by all occupants in the building at time step n. Higher values of E_n, C_n, and Q_n correspond to higher energy consumption, lower overall occupant comfort, and lower overall air quality experienced by occupants. α, β, γ are used as weights for the three objectives, allowing building managers to customize and select which objective(s) to prioritize. Next, we describe how we compute E_n, C_n, and Q_n.

1. Energy consumption, E_n is computed in Equation 2, where Δ is the length of one time step and $P_d(t)$ is the power consumption of the energy consuming resource d (e.g., HVAC and lights). E_n is therefore the total energy consumed by all energy-consuming resources d in the nth time window.

$$E_n = \int_{t=n\Delta}^{(n+1)\Delta} \sum_d P_d(t)dt. \tag{2}$$

2. Thermal comfort, C_n is computed in Equation 3, where $C(R_o, t)$ is the comfort of occupant o at location R_o at time t. As such, C_n is the total comfort across all occupants in the building at time step n. Measuring comfort is challenging, and we discuss how we measure comfort in our real deployments in Section 4.

$$C_n = \frac{\sum_{o \in O} \int_{t=n\Delta}^{(n+1)\Delta} C(R_o, t)dt}{\Delta \cdot |O|} \tag{3}$$

3. Air quality, Q_n is computed as shown in Equation 4. $Q(R_o, t)$ is the air quality experienced by occupant o at location R_o at time t. Much like thermal comfort, the reward only considers the air quality of the areas where occupants are present because the air quality of empty rooms will not affect any of the occupants. We discuss how we measure air quality of different rooms, locations, and spaces in Section 4.

$$Q_n = \frac{\sum_{o \in O} \int_{t=n\Delta}^{(n+1)\Delta} Q(R_o, t)dt}{\Delta \cdot |O|} \tag{4}$$

3.2 Deep Q-Network for Generating Actionable Recommendations

There are two challenges unique to building co-optimization that prevents direct application of a standard deep Q-network. The first challenge is representing occupant locations in a building. Typically, categorical data such as occupant locations are represented using one-hot encoding. However, the number of input nodes in the neural network quickly increases with the number of occupants and rooms. We address this challenge by incorporating an embedding layer (Section 3.2.1). The second challenge is the representation of the reward; there are many actions in each state and three different objectives (energy, comfort, air quality) for each action. The network must learn the objectives for all states and actions. We address this challenge through multi-task learning (Section 3.2.2).

3.2.1 Location Embedding. There are hidden relationships in the encoding of the occupant locations. Consider an occupant who spends time in three different locations: two similar office spaces and one lab space. Let us assume that the energy consumption, setpoint temperature, humidity and air quality are similar for the two office spaces. A one-hot encoding of the locations will not uncover similarities between these spaces.

Embedding layers reduce memory and computation requirements, compared to one-hot encoding, and have been shown to learn relationships between categories [16]. In our deep Q-network, we utilize an embedding layer to better learn the occupant location configurations than standard dense layers.

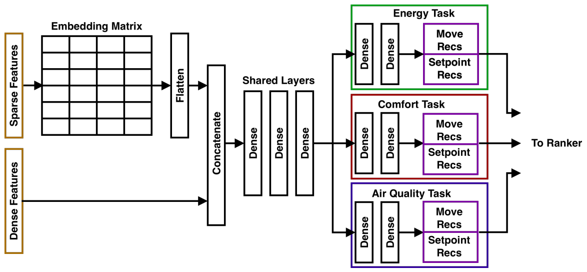


Figure 2: Our deep Q-network architecture includes an embedding layer for learning occupant locations, and has three output tasks for learning energy savings, comfort and air quality improvements for each action.

Each location corresponds to a unique row in the embedding layer. The embedding layer selects rows corresponding to the location of each occupant, which is concatenated into a vector of size $|O| \times d$, where d is the embedding dimension and $|O|$ is the number of occupants. In our deployments, we observe that $d = 3$ yielded the best tradeoff in performance vs computation.

3.2.2 Multi-Task Learning. Since there are three separate objectives (energy, comfort, air quality), there are two options for learning. The first option is to combine energy, comfort, and air quality changes into a single reward (e.g., sum all rewards in Equation 1), and use the return as the target for the Q-network. The second option is to learn the objectives separately, and then combine them at the output of the Q-network using a ranker. The advantages of the second option are that optimization emphasis can be quickly changed without retraining the network. However, this method requires learning three times the number of outputs.

A key observation is that there are hidden relationships between energy, comfort and air quality. As an example, a variable air volume system that supplies cold air turns on to supply air to a room; this will increase the energy consumption, increase air flow which may lead to improved air quality, and reduce the temperature in the room thus affecting comfort. We can take advantage of these relationships by using **multi-task learning**. In multi-task learning, the input features are fed into a number of "shared" layers which learn information about the state of the building and the individual locations. The output is then fed into "task-specific" layers, which are responsible for learning information specific to the objective. In our deep Q-network, we create task-specific layers for each of the energy, comfort, and air quality objectives in Equation 1.

3.2.3 Network Architecture. The network input consists of the sparse location features of each occupant and dense features, including temperature, humidity, setpoint temperature, air quality, and energy consumption for each location. The complete network architecture is shown in Figure 2. The sparse location features are fed into the embedding matrix, and the output vectors are flattened and concatenated with the dense features. The new vector is fed to the shared layers, then to the individual objective task layers. The output of each task are values representing the expected change in the energy consumption (E_n^a), comfort (C_n^a), and air quality (Q_n^a) objectives for all possible actions, a, at time step n.

3.3 Recommender System Design

The upper half of Figure 1 shows the full online system, consisting of the following components: the **networked sensing layer** (Section 4) that senses the **building state** (Section 3.1.1), the **deep-Q network** model (Section 3.2) and the **ranker** (Section 3.3.1) that uses the building state to generate recommendations for occupants, and the occupant-facing **web client** (Section 3.3.2) where users can view and accept/reject recommendations.

3.3.1 Ranker. Once the **Q-network** estimates the changes in energy consumption, comfort, and air quality objectives, the **ranker** arranges the recommendations for each occupant to maximize the overall energy savings, comfort, and air quality improvements. To accomplish this, the ranker generates a score for each action, a, according to Equation 5.

$$S_n^a = -\alpha E_n^a - \beta C_n^a - \gamma Q_n^a \tag{5}$$

Just like the reward (Equation 1), α, β, γ are weights that building managers can tune to prioritize certain objectives. There is a negative sign with each objective because a higher value is less desirable (Sections 3.1 and 4). In Section 6.4, we compare improvements in each objective using different weights in real deployments.

As RECA recommends two different categories of recommendations directly to users (move and setpoint), it is important to give the occupants a selection from both categories (diversity). The ranker selects two recommendations from each category that is displayed to the user. To select recommendations at time step n, the ranker samples recommendations after normalizing scores, S_n^a, for all recommendations, $a \in A$, using the softmax function (Equation 6), where τ is a temperature parameter. Sampling recommendations this way also allows RECA to incorporate exploration.

$$p(a, n) = \frac{\exp \frac{S_n^a}{\tau}}{\sum_{b \in A} \exp \frac{S_n^b}{\tau}} \tag{6}$$

3.3.2 User Feedback. To serve recommendations to occupants in real-time, we developed a web interface where occupants can browse a list of up-to-date recommendations, and select recommendations which are sent back to the system and stored in history as feedback. Additionally, we store the history of observed building states in a digital twin, similar to the one presented here [19]. In Section 5, we discuss how we use feedback and the history of building states to address several challenges for deploying RECA in the real world.

4 SYSTEM IMPLEMENTATION

To co-optimize energy savings, thermal comfort, and air quality, we need to measure these quantities per space and occupant to compute the reward and observe the building state (Section 3.1). Because we engage occupants to perform this optimization (e.g., move recommendations), we also need to estimate the location of each occupant. We make these measurements through the *networked sensing layer*, which we discuss next.

1. Energy: We measure three types of energy consumption (Equation 2), at the room level: HVAC, lighting, and individual energy devices. To monitor HVAC, we interface with the building management system over BACNet, which provides energy consumption information for large units such as fan coil units and variable air volume. Because most modern commercial buildings have a building management system (BMS) that can be accessed via BACNet, this step can be adapted to many existing and new buildings. For HVAC units that are not monitored by the BMS, we deploy wind sensors to estimate energy consumption from airflow and temperature [4]. We also deploy light sensors and plug meters to monitor lighting energy and local devices.

2. Comfort: Comfort is more challenging to measure, as it is subjective. The standard metric to measure thermal comfort is the predicted mean vote (PMV) model from ASHRAE 55 [3], where scores are generated from each occupant based purely on current environmental factors (e.g. temperature and humidity) and their own physical attributes. These scores are averaged to produce a value between -3 (cold) and 3 (hot). However, individuals may have temperature and space preferences that cannot be fully captured purely by physical attributes of the individual or the building. As such, we construct personalized thermal regression models.

We integrate the thermal comfort estimation pipeline, in [23], by deploying sensor nodes consisting of FLIR One Pro RGB-thermal cameras, Jetson Nanos, and temperature/humidity sensors into each room. We recorded thermal images of occupants over the course of two weeks to generate personalized comfort regression models. During this time, users also provide feedback or labels for their comfort levels, so we can correlate the observed thermal temperature with their perceived comfort. This pipeline estimates thermal comfort on the same scale as PMV ASHRAE 55, but more accurately by using facial temperature and feedback from users. The absolute value of this score is used as the overall comfort, $C(R_o, t)$, of each occupant o, at location R_o, at time t in Equation 3.

Though this method is more accurate, there is considerable overhead to adapt to each person. For larger deployments without thermal cameras, using the PMV ASHRAE 55 model and substituting measurements for a typical person, just like in [23], can still yield promising results, as we show in Section 6.1.

3. Air Quality: To measure air quality, we use the US Environmental Protection Agency's Air Quality Index (AQI), which incorporates $PM_{2.5}$ and PM_{10} measurements [12]. The higher the value, the more pollution is present in the air. A value of less than 50 is healthy. We deploy $PM_{2.5}$ and PM_{10} sensors at each location/room in our deployments. The air quality (Equation 4), $Q(R_o, t)$, experienced by each occupant o at time t is the AQI of the space R_o, where occupant o is residing at time t.

4. Localization: The location of each occupant is critical to determining the impacts of different actions and engaging occupants in the co-optimization process. To localize occupants, we extract head bounding boxes in the RGB domain using the *comfort estimation pipeline* we integrated [23]. We train a convolutional neural network based on VGG-16 to classify occupants by participant ID; the training data is hand-labeled using images taken over the course of one week. Because the number of occupants in our deployments is controlled, this solution is relatively simple to implement. Other methods such as wireless localization can be used in larger deployments, which we will explore in future work.

5 REAL-WORLD CONSIDERATIONS

There are several challenges that need to be addressed to ensure our system performs robustly in real environments. The first challenge is the lack of training data. Without a large amount of feedback describing the benefits of different actions, the recommender system will initially provide poor random recommendations to occupants, which inhibits useful feedback. This problem, also known as cold start, can be mitigated by providing data that is semi-realistic. We use real data, building states, and feedback from users, as discussed in Section 3.3.2 to generate more training examples (Section 5.1).

The second challenge is accounting for how a person's preferences can periodically change. For example, RECA may estimate the greatest reward if a user moves out of room A. However, if s/he needs to immediately perform lab work there, then s/he would not accept this recommendation (Section 5.2).

The third challenge is recommending conflicting actions to different users. For example, one occupant may choose to increase the setpoint temperature to reduce energy, while a second occupant at the same location and time may choose to decrease it (Section 5.3).

5.1 Data Augmentation for Exploring More Building States and Actions

We develop a simulation environment that enables the creation of a large number of potential future states, which is used to evaluate recommendations and train our recommender system. The simulation environment leverages statistical models to simulate future building states based on occupant actions and the input building state, collected from our deployment (Section 4). We first discuss how we simulate and predict key modules: HVAC energy, lighting energy, indoor comfort, indoor air quality prediction, and comfort (Section 5.1.1). Next, we introduce how we incorporate this simulator into training our deep Q-network (Section 5.1.2).

5.1.1 Predicting Future States.

1. HVAC Energy: Predicting HVAC energy is challenging due to nonlinearities associated with HVAC control and outdoor environmental conditions. As discussed in Section 2, physics-based models such as EnergyPlus have limited ability to predict HVAC energy, especially when detailed design documentation is unavailable and granular temporal scales (i.e., hourly) are required, which both apply in our setting. Data-driven surrogate energy modeling is a promising framework for addressing these limitations. However, these models are typically applied at the building level rather than room level. Here, we introduce data-driven surrogate HVAC energy modeling at the room level to support our simulation engine.

We identified three regression models that have been shown to be effective in energy prediction tasks: artificial neural networks, random forests, and gradient boosting [20]. In this task, the target is room-level HVAC energy consumption and the features are as follows (aggregated to the hourly level): number of occupants, heating degree days ($HDD = 65 - T_0$ in °F), cooling degree days ($CDD = T_0 - 65$ in °F), day of year + week + hour of day, and historical HVAC energy data (48-hr sliding window).

We used a temporal 80%–20% split to create training and testing sets. As a state-of-art comparison, we also built an EnergyPlus model based on historical building state data collected from our deployments and performed standard calibration measures, following

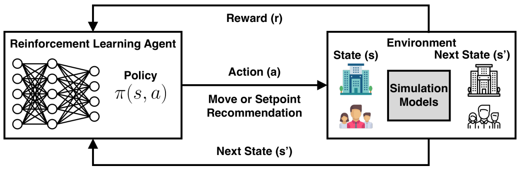


Figure 3: For a building state, the reinforcement learning agent provides an action to the simulation environment. The next state, energy savings, comfort and air quality improvements are returned to the agent to tune the policy.

the procedure in [11]. We used the coefficient of variation of the root mean squared error (CV(RMSE)) to compare the models. CV(RMSE) is commonly used to asses energy prediction performance in buildings, where a value less than 30% is considered a well-calibrated model [2]. We found that the random forest model produced the best results, with an average CV(RMSE) of 28.6% across the rooms, compared to EnergyPlus's 94.3%. We therefore implemented the random forest model for 1-hour-ahead energy prediction.

2. Lighting Energy: Based on the data from our collected building state history, we assume that lighting operation can directly follow occupancy patterns: lights are on whenever a room has at least one occupant and off whenever a room is unoccupied, since most modern office lights are controlled with motion sensors.

3. Indoor Temperature: We model the relationship between thermostat settings and indoor temperature from data collected from our networked sensing deployment. Individual rooms in the building include thermostats that give occupants the option to change the temperature. To investigate the empirical relationship between these settings and actual indoor temperature, we built linear regression models for each room that includes an indicator variable for each of the possible thermostat settings at each timestep as features and the actual temperature at each timestep as the response variable. In our deployments, the thermostats had three possible settings, "warm," "cool," and "neutral". We found that, on average, setting the thermostat to "warm" increases temperatures by 1.88°F and setting the thermostat to "cool" decreases temperatures by 1.80°F.

4. Air Quality: Our air quality prediction task closely follows that of indoor temperature. We would expect a change in the thermostat setting to increase airflow in the room because the heating and cooling system would need to supply additional air to affect air temperature. We would also expect the additional air to be cleaner, due to the filters in the HVAC system. We would therefore expect such thermostat actions to decrease PM concentrations. However, these dynamics were not clearly evident in the data—the regression models did not produce significant relationships for most rooms. For each room, we did include small factors based on the direction of the relationship between HVAC energy and PM concentrations.

5. Comfort: In real deployments, we more accurately measure thermal comfort on the PMV ASHRAE 55 scale using thermal cameras and indoor temperature (Section 4). Because it is difficult to predict future temperatures at each pixel for each thermal camera, we instead simulate comfort levels of each occupant by directly using PMV ASHRAE 55 and substituting indoor temperature and standard values, just like in this work [23].

5.1.2 Training. To train and remedy the cold start problem, we create an offline training environment by integrating our simulator, with the deep Q-network in a tightly coupled control loop, which allows for rapid data generation and learning. An illustration of the training environment is shown in Figure 3.

Training runs in episodes, with fixed $\Delta = 1$ hour time steps. In each episode, the simulation environment is instantiated with occupant locations and building states based on historical data. The state s is provided to the deep Q-network, which outputs predicted changes in energy, comfort, and air quality for each possible action from this state. An action is chosen based on softmax selection (Equation 6), just like how recommendations are displayed.

The action is sent to the simulation environment (Section 5.1) and performed in simulation, and the episode advances to the next step, producing a new building state (next state s') to be sent to the deep Q-network. Expected rewards for energy, comfort, and air quality are estimated and stored with the state in a pool of samples for experience replay.

5.2 Adapting to User Preferences

It is not productive to suggest actions that an occupant has consistently rejected. To address this, we estimate the probability that an occupant will accept a recommendation by counting the recommendations that occupants have rejected ("feedback" as described in Section 3.3.2). We add a penalty to the ranker scoring function (Equation 6) that penalizes recommendations with more rejections.

Second, a person's preferences may change over time. As such, we allow building managers to begin retraining the recommender system after an interval of time. Once started, RECA will automatically augment the data observed and stored in history using our simulation environment (Section 5.1) to retrain the deep Q-network.

5.3 Conflicting Actions

In real deployments, recommender systems serve multiple occupants concurrently, which can lead to conflicting actions (e.g., one occupant increases the setpoint temperature to reduce energy, while a second occupant decreases it to improve comfort). A second challenge is receiving multiple actions of the same type in quick succession (e.g., receiving two consecutive move recommendations).

To address these challenges, we temporarily remove any move or setpoint recommendations for an occupant at a location once a move or setpoint recommendation has been accepted, for a period $t_s = 1$ hour. Additionally, when a setpoint recommendation has been selected by an occupant for a location, no other occupants will be given setpoint recommendations for the remainder of t_s.

6 EVALUATION

We evaluate our system in two parts. First, we evaluate our recommender system using our simulation environment (Section 5.1) with different objective weights to study the ability to learn rewards from energy, comfort and air quality for different actions. Next, we deploy our system into two commercial office buildings and conduct A/B tests to evaluate the improvements our system can achieve in real settings over the course of four weeks.

We compare against two baseline strategies. The *setpoint only strategy* involves changing the setpoint temperature at each location to improve energy savings (energy emphasis), comfort (comfort emphasis), or air quality (air quality emphasis). This strategy leverages the same pipeline as RECA, except move recommendations are removed. The *service strategy* involves only relaxing HVAC services in locations that have no occupants, based on [4].

6.1 Evaluation using Simulation Environment

As the simulation environment is responsible for augmenting data for training and retraining the deep Q-networks, it is critical to evaluate the performance of the models in simulation before deploying them in the real world. We tested our recommender system with different objective weights (Equation 1) using simulated building episodes based on the building environments in our actual deployments, described in Section 6.2.

Evaluating these models in simulation involves simulating two states: one state representing that an action was performed, and the other state representing no action. We compare the energy consumption, thermal comfort, and air quality of occupants between the states. We calculate energy savings E_R as the difference in energy consumption between the two states over the nth timestep of length Δ, where $P_d(\cdot)$ and $\hat{P}_d(\cdot)$ denotes the power of resource d in the baseline state and the state with action (Equation 7).

$$E_R^n = \int_{t=n\Delta}^{(n+1)\Delta} \sum_d (P_d(t) - \hat{P}_d(t))dt. \tag{7}$$

We calculate thermal comfort and air quality improvement by taking the average PMV ($PMV(\cdot)$) and air quality ratings ($AQ(\cdot)$) for each occupant $o \in O$ in room R_o (Equations 8 and 9).

$$C_R^n = \frac{\sum_{o \in O} \int_{t=n\Delta}^{(n+1)\Delta} \widehat{PMV}(R_o, t) - PMV(R_o, t)\, dt}{\Delta \cdot |O|} \tag{8}$$

$$Q_R^n = \frac{\sum_{o \in O} \int_{t=n\Delta}^{(n+1)\Delta} \widehat{AQ}(R_o, t) - AQ\, dt}{\Delta \cdot |O|} \tag{9}$$

We simulated the performance of different models on 10000 episodes with semi-randomized occupant start locations. As shown in Table 1, we simulated episodes using the deep Q-network architecture with and without the embedding layer, and set weights, in Equation 1, for four different emphases: energy ($\alpha >> \beta = \gamma$), comfort ($\beta >> \alpha = \gamma$), air quality ($\gamma >> \alpha = \beta$), and joint optimization ($\alpha = \beta = \gamma$). We note that the architecture without the embedding layer performed worse for each emphasis in comparison to the architecture with the embedding layer and would often recommend actions that negatively impact the objectives. For example, in the comfort emphasis, RECA with embedding improved average PMV by 0.31 on a 3-point scale, compared to 0.03 without the embedding. Although the energy savings achieved for the comfort emphasis without embedding (22.2 kWh) exceeded the savings with the embedding (14.9 kWh), the system was tuned to emphasize and achieve comfort improvements at the expense of other objectives. Similar trends can be observed for all other emphases, while the joint emphasis achieves a more balanced improvement across all three objectives because it weights all three similarly in importance.

Additionally, the *service strategy* can only improve energy savings because it only turns down services in areas with no occupants. Our system also outperforms the *setpoint only strategy* across all emphases because our system can not only change setpoints, but also recommend more complex action sequences, such as grouping occupants with similar temperature preferences together.

Table 1: Comparison of our deep Q-network architecture with and without an embedding layer, against existing strategies, on simulated building episodes with four different weighting combinations to emphasize different optimizations.

Method	Emphasis	Energy Savings (kWh)	Comfort Improvement (Average PMV)	Air Quality Improvement (Average AQI)
RECA- No Embedding	Energy	27	−0.24	−2
	Comfort	22.2	+0.03	−1.8
	Air Quality	−90	−0.09	+1.0
	Joint	23.0	−0.02	+0.06
RECA- With Embedding	Energy	75	−.03	−2.5
	Comfort	14.9	+0.31	−0.73
	Air Quality	−45	+0.1	+2.0
	Joint	15	+0.01	+1.05
Setpoint Only Strategy	Energy	16	−.04	−1.3
	Comfort	5.5	+0.08	−0.7
	Air Quality	−12	+0.04	+0.8
Service Strategy	NA	10	−0.08	−1.5

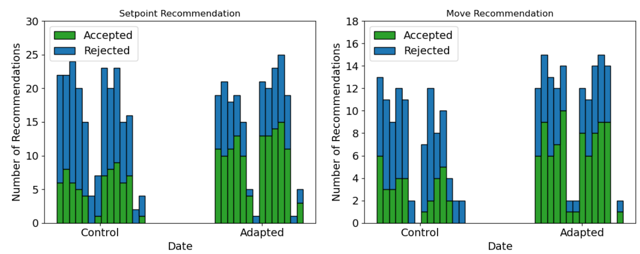


Figure 4: Acceptance rate during the first two weeks (control) and second two weeks after retraining (adapted) for the setpoint (left) and move (right) recommendations.

6.2 Recommender System Study

We deployed RECA in two commercial buildings over four weeks to evaluate energy savings, comfort improvement, and air quality improvement of different strategies and recommendation policies. In these deployments RECA ran on $\Delta = 1$ hour time steps. Building A is an office building consisting of 10 rooms, where 7 are rooms of cubicles, 3 are lab areas, and 1 break room. Building B is another office building consisting of 8 rooms, where 4 are open work areas, 2 are closed office spaces, and 2 are shared spaces. We recruited 10 occupants in Building A and 13 occupants in Building B, ranging in age from 22 to 40 from various academic disciplines, and collected recommendation feedback from participants over the course of four weeks. We obtained approval from Columbia University internal review board for all our deployments.

Since our deployment can only measure one building state sequence, we decide to use the sensed building state to measure the performance of RECA, while we simulate the energy consumption, comfort, and air quality as the baseline, using the tools from our simulation environment (Section 5.1). This is because once a user takes a recommendation, we can no longer observe the state of the building in the scenario where the occupant does not perform the action, which is required to measure improvements in energy consumption, comfort, and air quality. As such, we simulate scenarios where users do not accept any of the recommendations as the baseline. As opposed to the pure simulation environment, we

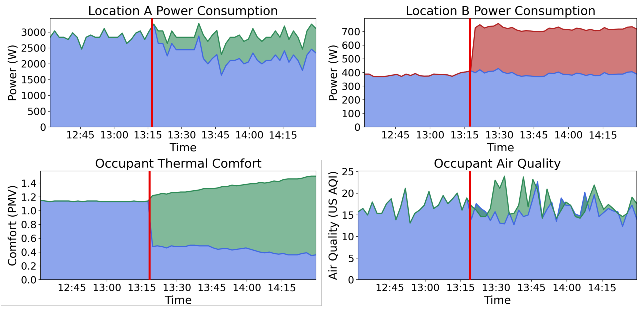


Figure 5: Location Optimization: At the red line, the occupant moves to location B, and HVAC service is reduced in location A. Due to differences in environment, the occupant's thermal comfort and air quality are improved.

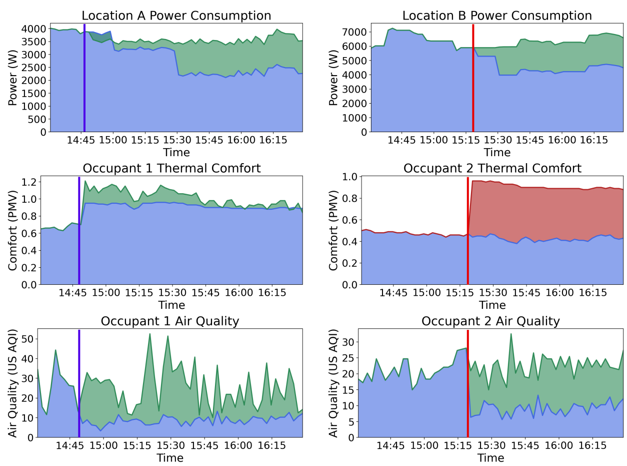


Figure 6: Group Consolidation: At the blue and red lines, occupants 1 and 2 move to location C (not shown). Locations A and B reduce HVAC and lighting service, leading to energy savings. Comfort and air quality for both occupants change due to environmental differences.

directly measure changes in energy consumption, comfort, and air quality in our real deployments as a result of RECA.

Figure 4 shows the number of accepted recommendations, broken down by recommendation type (move and setpoint). We scheduled RECA to retrain itself at two weeks (Section 5.2), using data from the first two weeks. The *control* period refers to the first two weeks, while the *adapted* period refers to the latter two weeks after retraining. There is a dramatic increase in recommendation acceptances, across both types of recommendations, in the adapted period because the retraining process allowed RECA to learn recommendations that users are more likely to accept. Across both deployments, we saw an 80% increase in accepted recommendations during the adapted period.

6.3 Learned Action Sequences

Throughout our study, we noticed and categorized three types of regularly occurring action sequences, that RECA recommended, with greater effects than individual actions: **location optimization**, **group consolidation**, and **group disbanding**.

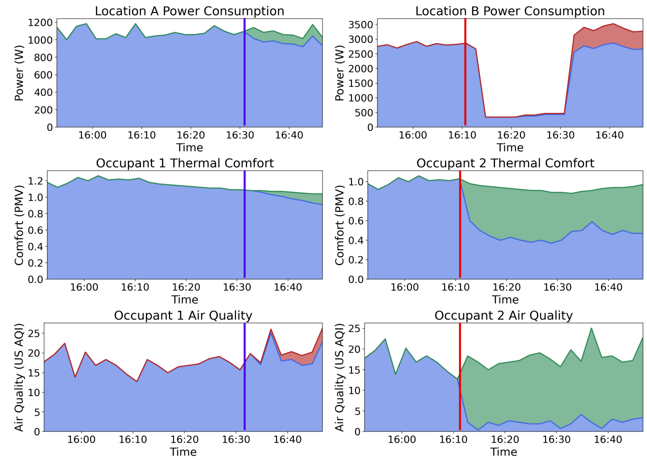


Figure 7: Group Disbanding: At the red line, occupant 2 moves from location A to location B. Since only occupant 1 remains, HVAC service is reduced in location A at the blue line, leading to a comfort improvement for occupant 1.

Location Optimization: The most common complex action sequence that we observed is location optimization. This sequence is typically made up of a setpoint change and a move recommendation. The main purpose of this sequence is, in most cases, to reduce the energy consumption of the starting location, without incurring a thermal comfort or air quality penalty by moving the occupant to another location. An example of this action sequence is shown in Figure 5. Initially, a single occupant is working in location A. At a certain time, indicated by the red line, the occupant increases the setpoint temperature of location A, and moves to location B. The normal consequence of increasing the setpoint is a reduction in thermal comfort for the occupant as the temperature gradually increases; however, moving to location B leads instead to an increase in thermal comfort due to the environment in location B being closer to the occupant's thermal preference, as shown in the bottom left plot. The green highlights the improvement in the occupant's measured PMV after the occupant moves to location B (lower value = more comfortable). Increasing the setpoint temperature in location A (upper left) reduces energy consumption of the space, as shown by the green. After the occupant moves to location B, the lights and electricity needed to service location B increases, as shown by the red (upper right). However, this increase is more than offset by the savings in location A. The air quality experienced by the occupant remains relatively stable (lower right).

Group Consolidation: Action sequences involving multiple occupants can enable even more optimization opportunities. In group consolidation, several occupants are brought to the same location often with a reduction in energy consumption in the start locations. As illustrated in Figure 6, occupants 1 and 2 are in locations A and B, respectively. At different times, denoted by the blue and red lines, occupants 1 and 2 increase the setpoint temperatures of locations A and B, and subsequently move to location C. By increasing setpoint temperatures, location A and B both experience a decrease in energy consumption as highlighted in green in the upper plots. The thermal comfort and air quality for both occupants change depending on the environmental differences at location A, B, and C. In this example, the temperature preferences of location C is more suited for occupant 1 than occupant 2. As such, we see that the thermal comfort of occupant 1 improves (middle left plot highlighted in green), while occupant 2 decreases (middle right plot highlighted in red). The air quality for both occupants are improved in location C, as compared to location A and B (lower plots highlighted in green). Note that in location optimization and group consolidation, the location that the occupants move to is critical. For example, if the destination location is too hot or too cold, the result of the action sequences may lead to an overall decrease in thermal comfort.

Group Disbanding: The final category of action sequence that we observed is *group disbanding*. The primary challenge in optimizing thermal comfort specifically, is that each occupant has a different thermal preference. Two occupants with different thermal preferences in the same location will require a compromise in setpoint temperature to prevent significant discomfort from one or both of the occupants. Group disbanding seeks to resolve this challenge by separating the occupants into different locations and group occupants with similar thermal preferences. In Figure 7, occupants 1 and 2 are in location A and have different thermal preferences. At the red line, occupant 2 is recommended to move to location B. The thermal comfort and air quality of occupant 2 increase as a result of the change in location (middle and lower right plots highlighted in green). Note that after occupant 2 leaves, occupant 1 changes the setpoint temperature higher at the blue line, leading to reduced energy consumption (upper left highlighted in green), improved thermal comfort (middle left highlighted in green), with a slight decrease in air quality (lower left highlighted in red).

6.4 Joint Optimization Results

In this section, we show the strengths of RECA against existing solutions in two real commercial buildings. Specifically, we show the *versatility* of RECA, allowing building managers to configure RECA to improve any combination of energy, comfort, and air quality. We also demonstrate how incorporating *humans-in-the-loop* allows RECA to further improve these objectives over existing solutions that only reduce services and change temperature setpoints.

Case 1: Energy Optimization: One of the most studied problems in commercial buildings is energy optimization. To study the performance of RECA on energy optimization, we select weights of the ranker to emphasize energy savings ($\alpha >> \beta$ and γ).

Figure 8a shows the percentage improvement of energy consumption, occupant thermal comfort, and occupant air quality. Since the baseline strategy only reduces service in locations with no occupants, only energy consumption is improved. In comparison, the setpoint only strategy further improves energy consumption, but reduces occupant comfort. The reason is that increasing setpoint temperature reduces HVAC service, but also increases the temperature, leading to a decrease in occupant comfort. Finally, by including move recommendations, the system achieves a further 8% and 6% increase in energy savings without sacrificing comfort in buildings A and B, respectively. As described in Section 6.3, moving occupants enables locations with high energy requirements to reduce service without incurring penalties to thermal comfort. However, we also observe a decrease in air quality in building A. There are two reasons for this. First, moving more people into the same room concentrates emissions; our system also reduces HVAC services in more rooms, which leads to a decrease in air quality because less air is being filtered. Second, in this scenario, the building manager configured our system to increase energy efficiency without considering other factors. As such, our system focused on reducing as much energy as possible, even at the cost of other factors.

Case 2: Joint Co-Optimization: In some cases, building managers may wish to save energy without sacrificing comfort or air quality. Joint co-optimization is a more complex problem, as many actions lead to tradeoffs between the three objectives. To evaluate the potential of our recommender system to jointly optimize all three objectives at once, we set the weights of the ranker to balance the three objectives ($\alpha = \beta = \gamma$). We also deployed the *setpoint and service* strategy, which combines both the *setpoint only - comfort* strategy with the *service* strategy, as a comparison

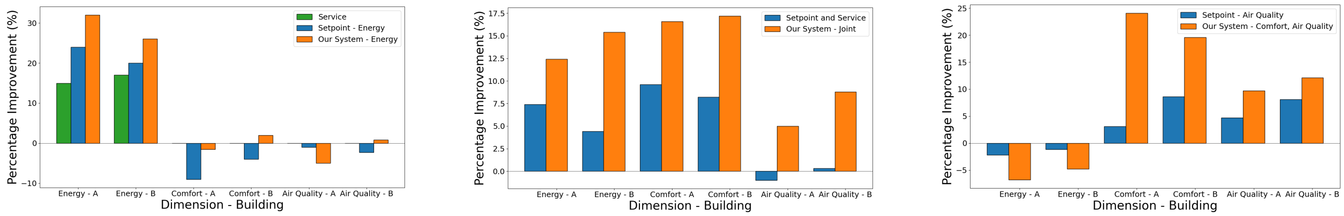
As shown in Figure 8b, the *setpoint and service* strategy improves thermal comfort and energy consumption, while minimally impacting air quality. However, the complete recommender system makes additional improvements of 5%, 7%, and 6%, for building A, and 9%, 9%, and 8%, for building B, in energy consumption, thermal comfort, and air quality. With move recommendations enabled, the recommender system can utilize more complex action sequences to find significant optimization opportunities that are not possible in the setpoint only strategy. Compared with case one, where a building manager may only be concerned about energy savings, we see that our system can jointly improve all three aspects at the same time by incorporating humans-in-the-loop.

Case 3: Comfort and Air Quality Co-Optimization: In certain cases, it may be desirable to allow increases in energy consumption to significantly increase occupant comfort and air quality. Due to the recent COVID-19 pandemic, improving air quality is a priority in work environments, while improving occupant thermal comfort can improve productivity. We deployed our recommender system with higher comfort and air quality weights ($\beta = \gamma >> \alpha$) to encourage recommendations of actions with high thermal comfort and air quality improvements. In comparison, we deployed the *setpoint only - with air quality emphasis* strategy for A/B testing.

As shown in Figure 8c, the *setpoint only baseline* is able to prioritize comfort and air quality improvements with the tradeoff of increased energy consumption. However, by treating occupants as immovable objects, we observed that many of the setpoint temperature changes were compromises for multiple occupants. In contrast, the recommender system that utilizes move recommendations shows a dramatic increase in comfort and air quality improvements of 21% and 5% for building A and 11% and 4% for building B. We noticed multiple instances of group disbanding, which allows for more personalized thermal comfort by separating groups of people with different thermal comfort preferences.

6.5 Scalability

The size of the network scales $O(|O| \times |S|)$, where $|O|$ is the number of occupants we engage and $|S|$ is the number of locations in the building. Although this is linear in both occupants and locations, it can become expensive if the building gets large. Table 2 shows the execution time and scalability of RECA as the number of occupants and spaces increases. We see that the majority of the computation



(a) (b) (c)

Figure 8: Energy savings, comfort and air quality improvements, emphasizing energy savings (a), balanced improvements (b), and comfort and air quality (c) in two deployments (A and B).

Table 2: Execution time for each component of RECA as the number of people and spaces increase.

People	Spaces	State (Section 3.1)	DQN (Section 3.2)	Ranker (Section 3.3.1)
50	25	82ms	684ms	77ms
100	50	101ms	2505ms	96ms
200	100	131ms	4322ms	101ms
400	200	213ms	8857ms	142ms
800	400	297ms	14729ms	180ms

comes from the DQN. However, even as the number of occupants and spaces increase to over 800 and 400, respectively, the total execution time is around 15 seconds. This latency is more than acceptable for our system, which updates once every hour.

In even larger deployments where computation time may exceed one time step, we can use the observation that people will typically use only a small portion of the building. As such, we can reduce the state-action space by eliminating infeasible actions and/or create multi-agent systems with smaller models to manage portions of the building. We plan to explore these avenues in future work.

7 CONCLUSION

We present RECA, a recommender system that generates real-time, human-centric, actionable recommendations for joint optimization of energy, comfort and air quality in commercial buildings. Our recommender system consists of a novel multi-task learning-based deep Q-network to jointly learn energy consumption, comfort and air quality improvement potential of different actions and is tunable to allow building managers to emphasize different dimensions. We conduct a four-week study in two real office buildings, and demonstrate the ability of this system to achieve greater energy savings, comfort improvements, and air quality improvements over prior works by incorporating occupants in the co-optimization process. Our multi-task learning based recommender system enables flexibility in optimization goals and discovers impactful actions that engage occupants in creating more energy efficient, comfortable, and healthier built environments. We envision RECA being integrated with mobile sensing and actuation platforms (e.g., drones [26]) to achieve more savings in future smart buildings.

ACKNOWLEDGMENTS

This research was partially funded by the National Science Foundation under Grant Number CNS-1943396. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

REFERENCES

[1] Yuting An, Tongling Xia, Ruoyu You, Dayi Lai, Junjie Liu, and Chun Chen. 2021. A reinforcement learning approach for control of window behavior to reduce indoor PM2. 5 concentrations in naturally ventilated buildings. *Building and Environment* (2021), 107978.

[2] ASHRAE. 2002. ASHRAE Guideline 14-2002: Measure ment of energy and demand savings.

[3] UFAD ASHRAE. 2013. Guide: Design, Construction and Operation of Underfloor Air Distribution Systems. *American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA* (2013).

[4] Bharathan Balaji, Jian Xu, Anthony Nwokafor, Rajesh Gupta, and Yuvraj Agarwal. 2013. Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. 1–14.

[5] Paulo Carreira, António Aguiar Costa, Vitor Mansur, and Artur Arsénio. 2018. Can HVAC really learn from users? A simulation-based study on the effectiveness of voting for comfort and energy use optimization. *Sustainable cities and society* 41 (2018), 275–285.

[6] Yujiao Chen, Leslie K Norford, Holly W Samuelson, and Ali Malkawi. 2018. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings* 169 (2018), 195–205.

[7] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. 2001. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings* 33, 4 (apr 2001), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

[8] Younghoon Kwak and Jung-Ho Huh. 2016. Development of a method of real-time building energy simulation for efficient predictive control. *Energy Conversion and Management* 113 (2016), 220–229.

[9] Aleksandra Lipczynska, Stefano Schiavon, and Lindsay T Graham. 2018. Thermal comfort and self-reported productivity in an office with ceiling fans in the tropics. *Building and Environment* 135 (2018), 202–212.

[10] Marcel Macarulla, Miquel Casals, Nuria Forcada, and Marta Gangolells. 2017. Implementation of predictive control in a commercial building energy management system using neural networks. *Energy and Buildings* 151 (2017), 511–519.

[11] Clayton Miller, Daren Thomas, Silvia Domingo Irigoyen, Christian Hersberger, Zoltán Nagy, Dino Rossi, and Arno Schlueter. 2014. BIM-extracted EnergyPlus model calibration for retrofit analysis of a historically listed building in Switzerland. In *Proceedings of SimBuild 2014*. 331–338. https://doi.org/10.13140/RG.2.1.1671.7285

[12] David Mintz. 2016. *Technical assistance document for the reporting of daily air quality : the air quality index (AQI).* https://stanford.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsgpr&AN=edsgpr.000990603http://www3.epa.gov/airnow/aqi-technical-assistance-document-may2016.pdfhttp://purl.fdlp.gov/GPO/gpo71010

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[14] US Department of Energy. 2015. An Assessment of Energy Technologies and Research Opportunities. Online. Accessed: 2021-05-29.

[15] Krati Rastogi, Anurag Barthwal, and Divya Lohani. 2019. AQCI: An IoT Based Air Quality and Thermal Comfort Model using Fuzzy Inference. In *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE, 1–6.

[16] Xin Rong. 2014. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738* (2014).

[17] Cherlyn Seruto. 2010. Whole-building retrofits: A gateway to climate stabilization. *ASHRAE Transactions* 116 (2010), 244.

[18] Andrew Sonta, Thomas R Dougherty, and Rishee K Jain. 2021. Data-driven optimization of building layouts for energy efficiency. *Energy and Buildings* 238 (2021), 110815.

[19] Ravi S Srinivasan, Baalaganapathy Manohar, and Raja RA Issa. 2020. Urban Building Energy CPS (UBE-CPS): Real-time demand response using digital twin. In *Cyber-Physical systems in the built environment.* Springer, 309–322.

[20] Ying Sun, Fariborz Haghighat, and Benjamin C.M. Fung. 2020. A review of the-state-of-the-art in data-driven approaches for building energy prediction. , 110022 pages. https://doi.org/10.1016/j.enbuild.2020.110022

[21] Peter Wei, Xiaoqi Chen, Jordan Vega, Stephen Xia, Rishikanth Chandrasekaran, and Xiaofan Jiang. 2017. ePrints: a real-time and scalable system for fair apportionment and tracking of personal energy footprints in commercial buildings. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. 1–10.

[22] Peter Wei, Xiaoqi Chen, Jordan Vega, Stephen Xia, Rishikanth Chandrasekaran, and Xiaofan Jiang. 2018. A Scalable System for Apportionment and Tracking of Energy Footprints in Commercial Buildings. *ACM Transactions on Sensor Networks (TOSN)* 14, 3-4 (2018), 1–25.

[23] Peter Wei, Yanchen Liu, Hengjiu Kang, Chenye Yang, and Xiaofan Jiang. 2021. A Low-Cost and Scalable Personalized Thermal Comfort Estimation System in Indoor Environments. In *Proceedings of the First International Workshop on Cyber-Physical-Human System Design and Implementation* (Nashville, TN, USA) *(CPHS21)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3458648.3460006

[24] Peter Wei, Stephen Xia, Runfeng Chen, Jingyi Qian, Chong Li, and Xiaofan Jiang. 2020. A Deep-Reinforcement-Learning-Based Recommender System for Occupant-Driven Energy Optimization in Commercial Buildings. *IEEE Internet of Things Journal* 7, 7 (2020), 6402–6413.

[25] Peter Wei, Stephen Xia, and Xiaofan Jiang. 2018. Energy saving recommendations and user location modeling in commercial buildings. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. 3–11.

[26] Stephen Xia, Rishikanth Chandrasekaran, Yanchen Liu, Chenye Yang, Tajana Simunic Rosing, and Xiaofan Jiang. 2021. A drone-based system for intelligent and autonomous homes. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 349–350.