Optimal HVAC System Operation Using Online Learning of Interconnected Neural Networks

Ye-Eun Jang, Student Member, IEEE, Young-Jin Kim, Senior Member, IEEE, and João P. S. Catalão, Senior Member, IEEE

Abstract—Optimizing the operation of heating, ventilation, and air-conditioning (HVAC) systems is a challenging task that requires the modeling of complex nonlinear relationships among the HVAC load, indoor temperature, and outdoor environment. This paper proposes a new strategy for optimal operation of an HVAC system in a commercial building. The system for indoor temperature control is divided into three sub-systems, each of which is modeled using an artificial neural network (ANN). The ANNs are then interconnected and integrated into an optimization problem for temperature set-point scheduling. The problem is reformulated to determine the optimal set-points using a deterministic search algorithm. After the optimal scheduling has been initiated, the ANNs undergo online learning repeatedly, mitigating overfitting. Case studies are conducted to analyze the performance of the proposed strategy, compared to strategies with a pre-determined temperature set-point, an ideal physics-based building model, and other types of machine learning-based modeling and scheduling methods. The case study results confirm that the proposed strategy is effective in terms of the HVAC energy cost, practical applicability, and training data requirements.

Index Terms—Artificial neural networks (ANNs); deterministic search; heating, ventilation, and air-conditioning (HVAC); online learning; temperature set-point scheduling.

NOMENCLATURE

The main notations used in this paper are summarized here.

A. Sets and Indices:
- $d, t$: indices for day and time
- $m, n$: indices for neural networks and linear power blocks
- $\max, \min$: subscripts for maximum, minimum, reference, and set-point values
- $\text{ref}, \text{set}$: normalizosed root mean square error between $\cdot$ and $\cdot'$
- $L_1, L_2, L_3$: neural networks to model building thermal dynamics
- $S_1, S_2, S_3$: original and reformulated optimization problems

B. Parameters:
- $C$: retail electricity price at time $t$
- $D$: dead-band between $T_{\text{ref}}$ and $T$
- $E$: building thermal environments at time $t$
- $F_{\text{seg}}$: linear gradient of $T_{\text{seg}}$ at time $t$
- $L_{\text{d}, t}$: maximum time delays of input data for neural networks
- $N_{\text{ep}}, N_{\text{et}}, N_{\text{EO}}$: numbers of epochs for network training and optimal scheduling

I. INTRODUCTION

COMMERCIAL buildings accounted for more than 36% of total energy consumption in the United States in 2019 [1]. Heating, ventilation, and air conditioning (HVAC) units represent approximately 40% of the electricity used in commercial buildings [2]. Therefore, significant attention has
been given to the modeling and optimal operation of HVAC systems to improve energy efficiency and reduce electricity bills in commercial buildings.

Physics-based modeling of HVAC units requires numerous parameters to reflect the complex nonlinear relationships among the HVAC load, indoor temperature, and outdoor environment [3]. Most of the physics-based modeling parameters are unknown and need to be extracted using sophisticated estimation techniques. Therefore, previous studies using simple RC circuit models need further analysis to reflect the thermal dynamics of buildings accurately [4]. Moreover, the types, sizes, and operating characteristics of HVAC units vary by manufacturer and by the building in which they are installed [5]. This prevents application of physics-based modeling and optimal operation to various buildings with different types of HVAC system.

To overcome these challenges, machine learning (ML) and artificial neural networks (ANNs) have increasingly been considered in recent studies on building energy management systems (BEMSs). This paper proposes a new ML-based strategy for an HVAC system in a commercial building, wherein the optimal temperature set-points are deterministically scheduled using the online supervised learning (SL) of interconnected ANNs. Specifically, the system for a building’s temperature control is divided into three sub-systems: a thermostat controller, an HVAC unit, and a building envelope. Long short-term memory (LSTM) networks are implemented and trained to model each sub-system. The LSTM networks are then interconnected to establish a complete model of the temperature control system. Using the LSTM-based model, an optimization problem is formulated to schedule the optimal temperature set-points, given day-ahead forecasts of the electricity price and the thermal environment. The problem is then reformulated, so that the optimal solution can be deterministically searched for using a gradient descent (GD) algorithm. After the optimal scheduling has been initiated, the LSTM-based model continues to undergo online SL, as new data on the building’s operation are collected. This gradually improves the accuracy of the LSTM-based model and hence the performance of the optimal scheduling. The results of sensitivity analyses and comparative case studies confirm that the proposed strategy ensures cost-effective operation of the HVAC system and the thermal comfort of occupants.

The main contributions of this paper are summarized below:

- To the best of our knowledge, this study is the first to develop and interconnect the ANN models of the sub-systems that are required for building temperature control, mitigating the complexity of the ANNs and improving the modeling accuracy of the building thermal dynamics.
- The interconnected ANNs are directly integrated into the optimization problem for temperature set-point scheduling. The problem is then reformulated to apply a deterministic search algorithm and find the optimal schedule within a reasonable computation time.
- The online SL is incorporated into the optimal scheduling of HVAC system operation, reducing the requirement for initial training data and hence facilitating the application of ML-based modeling and control in practice.
- The proposed strategy is comprehensively evaluated, both using sensitivity analyses and via comparison with strategies that use a traditional temperature setting rule, an ideal physics-based building model, and other types of ML algorithm.

II. RELATED WORKS

In recent years, Internet of things (IoT) technologies have been widely used to facilitate interactions between BEMSs and in-building infrastructures [6], [7], as significant attention has been given to improving building energy efficiency. The costs of IoT sensors and data analytics tools continue to decrease and they have become more widely and immediately available. Consequently, labeled datasets of HVAC unit operations and building thermal conditions have become increasingly available to BEMSs [7]–[9], enabling data-driven modeling and operation of building temperature control systems in practice.

Given this data availability, various ML algorithms have been used in recent studies on optimal control of indoor temperatures. For example, in [8]–[10], an ANN was trained offline via SL to model building thermal dynamics. Given the ANN model, the solution to the problem for the optimal HVAC system operation was searched for using heuristic algorithms, such as GA, PSO, and firefly algorithms. However, in a heuristic search, the solution is highly likely to fall into one of numerous local optima. Therefore, the optimization problem should be iteratively solved to find the best solution closer to the global optimum, increasing the computation time [8]. Moreover, in [8]–[10], only a single ANN was implemented to reflect the highly nonlinear characteristics of the building thermal dynamics. In practice, this risks compromising the modeling accuracy and hence the scheduling performance, even for the case of an ANN with deep hidden layers.

In [11]–[16], reinforcement learning (RL) was adopted to take advantage of the fact that it requires little knowledge of HVAC system operations and building thermal dynamics. For example, in [11] and [12], the optimal operation of air conditioners was explored using Deep Q-Network (DQN) and Deep Policy Gradient (DPG) algorithms. In [13]–[15], deep DPG (DPG) and A3C algorithms were adopted to minimize HVAC energy consumption, considering the high-dimensional action spaces. In the RL algorithms, for each episode, an RL agent chooses an action based on the exploration-and-exploitation mechanism [16], where the agent explores untried actions to gain more experience and combines this with exploitation of the already known successful actions to obtain high long-term reward. In other words, the optimal HVAC load schedule still needs to be iteratively searched for using random variables. For the heuristic search, the number of learning episodes should be set to an arbitrarily high value, increasing the computation time.

In addition, the exploration-and-exploitation mechanism is highly likely to include the risk that a slight change in hyperparameters for the RL agent’s training can lead to unstable and poor control of the HVAC system, particularly in the initial learning episodes [17], [18]. When the spaces of states and actions are discretized, a large step size can also lead to poor capability of the RL agent to learn the problem characteristics.
and failure to ensure the thermal comfort of occupants. This implies difficulties in directly applying RL-based control algorithms to real buildings where HVAC systems are currently in daily service. Therefore, in recent studies (e.g., [13] and [19]), data-driven models of HVAC systems and building envelopes were developed first, so that the RL agent was trained using the input and output datasets obtained from the models, as in the case of model predictive control (MPC) [20]. However, once the models are implemented, it can be more stable and time-efficient to apply SL-based control strategies using deterministic optimal solvers, rather than RL-based strategies.

The application of SL requires historical data on HVAC system operations under various building thermal environments. When the size of the historical dataset is small and the variability is limited, the ANNs are likely to be over-fitted [21]: i.e., too closely fitted to only a limited set of data points. The requirement for historical data needs to be mitigated for wide application of SL-based modeling and optimal operation. For example, in new buildings, insufficient historical data may have been collected. In traditional energy-inefficient buildings, a rule-based strategy is often adopted to operate HVAC systems with pre-determined temperature set-points. To reduce the data requirement, recent studies have been conducted on online SL. For example, in [22], the optimal operation of an air-conditioning system was achieved online, although variations in the ambient temperature and electricity price were not considered. In [23], hyper-parameters for optimal HVAC system operation were updated online; however, the temperature set-point was chosen from only a limited set of discrete values and was fixed during a day.

III. MODELING OF BUILDING THERMAL DYNAMICS

A. ANN-based Modeling of Sub-systems

![Fig. 1. A schematic diagram of a common system for building temperature control, consisting of a thermostat control loop, an HVAC unit, and a building envelope.](image)

Fig. 1 shows a common system for indoor temperature control in a commercial building. It consists of three sub-systems: a thermostat-control loop, an HVAC unit, and a building envelope. Specifically, in the thermostat loop, a proportional-integral (PI) controller is adopted to adjust the reference power input \( P_{ref} \) of the HVAC unit, based on the difference between the set-point and actual values of the indoor temperature: i.e., \( T_{set} \) and \( T_i \), respectively. In practice, the PI controller is accompanied by nonlinear signal processing functions, such as saturations and ramp rate limits, to ensure reliable system operation. The HVAC unit receives \( P_{ref} \) as an input signal and provides thermal energy \( Q' \) to the envelope, given the ambient temperature \( T_e \) and the evaporator-side air or water temperature \( T_e' \). In this paper, a variable-speed heat pump is considered as an example of an HVAC unit [8]. The time response of the variable speed drive is fast and, consequently, the actual power input \( P_t \) is almost the same as \( P_{ref} \) (i.e., \( P_t \approx P_{ref} \)), particularly in the scheduling time horizon. In the building envelope, the profile of \( T_e' \) is determined by the HVAC system operation (i.e., \( Q' \)) and the building thermal environments \( E' \), such as \( T_i', T_e', \) and indoor thermal load \( Q' \).

Each sub-system is modeled using an ANN, as shown in Fig. 2. The ANNs are then linked together, based on the inter-connections of the sub-systems, as discussed above. The operating characteristics of each sub-system can successfully be reflected into an ANN with a rather simple architecture. This mitigates the overall complexity of the ANN model that represents the complete system for the building temperature control, shown in Fig. 1. By contrast, the conventional modeling methods often consider only a single ANN [8]–[10]. The ANN then needs to be significantly complicated and deep to reflect the operation of the complete system accurately, requiring a large amount of building operation data. This implies the risk of compromising modeling accuracy and hence the temperature control performance for a practical case with data of limited size and variability.

B. ANN Architecture and Training

For the sub-systems, the ANNs are implemented in the form of an LSTM network, which is widely used for time-series data learning and system identification. Note that the proposed strategy can readily be achieved using different types of ANNs, as discussed in Section V-D. Specifically, the LSTMs consist of multiple hidden layers, each of which includes multiple hidden nodes with self-loops. Furthermore, Fig. 2 shows that the LSTM L1 has an inner feedback loop between the output and input neurons for \( P_t \), which is indicated by the red circles.
Similarly, \( L_1 \) has an inner feedback loop of \( T'_i \), marked by the yellow circles. An outer feedback loop of \( T'_i \) also exists between the output neuron of \( L_1 \) and the input neuron of \( L_1 \), which is represented by the blue circles. Moreover, \( L_{1-3} \) have pre- and post-processors to normalize the input data and recover the output data with their original units, respectively, preventing the training speed from dropping too low.

In addition, each LSTM has a single output neuron and multiple input neurons. As shown in Fig. 2, the outputs of \( L_{1-3} \) are defined as \( P_i, Q_i \), and \( T'_i \), respectively. The inputs of \( L_1 \) are the current and time-delayed values of \( T_{sett} \) and the time-delayed values of \( P_i \) and \( T'_i \). For \( L_2 \), the inputs are the current and time-delayed values of \( P'_i, T'_i \), and \( T'_i \). The inputs of \( L_3 \) are set to the current and time-delayed values of \( Q'_i \) and \( E'_i \) and the time-delayed \( T'_i \). In this study, the time-delayed inputs of \( L_{1-3} \) are explicitly considered to achieve better accuracy in modeling the building thermal dynamics by reflecting the effects of the integral controller in the thermostat loop, the heat exchanger in the HVAC system, and the thermal energy storage inherent in the building envelope, respectively. Specifically, the search range for the hyper-parameters of \( L_{1-3} \) is established by the minimum and maximum values of the time delays of input neurons, the numbers of hidden layers and neurons, and the learning rates, considering the trade-off between the modeling accuracy and the computational burden. While examining all possible combinations, one is selected that leads to good training and testing results for historical datasets. Through this procedure, the maximum time delays of the inputs of \( L_{1-3} \) are set to \( L_{P1} = 24 \) h, \( L_{P2} = 4 \) h, and \( L_{P3} = 4 \) h, respectively. For brevity, each LSTM has the same values of \( L_P \) for its inputs, and the selected hyper-parameters of \( L_{1-3} \) are fixed during the online SL, as discussed in Section IV-B.

The individual LSTMs are trained separately using the database of a BEMS to determine the weighting coefficients and biases for all the input, hidden, and output neurons. The separate training can reduce the structural complexity of the LSTMs, facilitating modeling of the temperature control system. The feedback loops for each LSTM are also open, so that the actual time-delayed data can be fed into the input neurons and hence an SL algorithm can be applied for the LSTM training. The training data are obtained during the actual, normal operation of the temperature control system, ensuring the modeling convergence of \( L_{1-3} \). Moreover, the physics-based modeling parameters of the HVAC system and building envelope are not required to train the LSTMs and hence formulate the optimization problem, discussed in Section V, wherein the LSTMs are integrated for optimal scheduling of the set-point temperatures. This enables wide application of the proposed strategy in practical BEMSs. After the training, the LSTMs are then interconnected and tested with closed feedback loops, so that the outputs estimated at the current time step can be used as the time-delayed inputs at the next step. This also enables the interconnected LSTMs to reflect the interactions among the sub-systems and hence the operating characteristics of the completed system.

### IV. OPTIMAL SCHEDULING INTEGRATED WITH ONLINE SL

In the proposed strategy, the optimal operation of the HVAC system is scheduled for the next 24 hours, based on day-ahead forecasts of the electricity price and building thermal conditions. The scheduling is consistent with current practices for demand response (DR) [24] and existing strategies for scheduling of power system operation [25], [26]. Numerous forecasting algorithms have been discussed, for example, in [27]–[29] and, therefore, appropriate algorithms can readily be selected and incorporated into the proposed strategy. In this study, the forecast data are assumed to be already available in the BEMS database for brevity, as in [8] and [9]; integration has been left for future research.

#### A. Optimization Problem Formulation

Using the trained \( L_{1-3} \), the optimal schedule for \( T_{sett} \) can be determined by solving \( S_1 \) as:

\[
S_1: \text{Problem for optimal HVAC system operation}
\]

\[
\text{arg} \min_{T_{sett}} C_E = \sum_{i=1}^{N_i} C_i P_i,
\]

subject to

\[
T_{sett, min} \leq T_{sett} \leq T_{sett, max}, \quad \forall t,
\]

\[
T'_i \leq T'_i \leq T'_{i, max}, \quad \forall t,
\]

\[
P_{min} \leq P_i \leq P_{max}, \quad \forall t,
\]

\[
R_g \leq \left( P' - P^{\min} \right) / \Delta t \leq R_g, \quad \forall t,
\]

where

\[
P' = L_1 \left( T'_{sett}, ..., T'_{sett-L_i+1}, P^{\min}, ..., P^{\min-L_i+1}, T^{\min}, ..., T^{\min-L_i+1} \right), \quad \forall t,
\]

\[
Q' = L_2 \left( P'_i, ..., P'_{i-L_i+1}, T'_i, ..., T'_{i-L_i+1}, T^{\min-L_i+1}, ..., T^{\min-L_i+1} \right), \quad \forall t,
\]

\[
T'_i = L_3 \left( Q'_i, ..., Q'_{i-L_i+1}, E'_i, ..., E'_{i-L_i+1}, T'_i, ..., T'_{i-L_i+1} \right), \quad \forall t.
\]

The objective function (1) aims to minimize the energy cost \( C_E \) of the HVAC system: i.e., the 24-h sum of the hourly-varying retail electricity price \( C \) multiplied by the power input \( P' \) of the HVAC system. Note that \( C \) can be negative, for example, when there is an excess of renewable generation [30].

In the set of constraints, (2) shows the limits of the operating range of the thermostat controller (i.e., from \( T_{sett, min} = 15^\circ \text{C} \) to \( T_{sett, max} = 35^\circ \text{C} \)) to secure reliable operation of the HVAC system. Moreover, (3) represents that \( T'_i \) should be maintained within an acceptable range from \( T'_{i, min} \) to \( T'_{i, max} \) to ensure the thermal comfort of occupants. Note that \( T'_{sett} \) and \( T'_{i} \) can differ under normal operating conditions of the HVAC system, mainly due to the large thermal capacity of the building envelope. The constraints (4) require \( P_i \) to be maintained between \( P_{max} \) and \( P_{min} \); in this paper, these are set to the rated power input and zero, respectively. Furthermore, (5) specifies the limits on the upward and downward ramp rates of \( P \) for the time period \( \Delta t = 1 \) h. In (5), \( P' \) at \( t = 0 \) h is set to zero, assuming that the HVAC system is turned off at night (after 7 pm to midnight) when the commercial building has low occupancy.

In (6)–(8), the LSTM-based sub-system models, discussed in Section III, are parameterized as the functions \( L_{1-3}(\cdot) \), in which the current and time-delayed inputs and the output are specified. In other words, the operating characteristics of the sub-systems are integrated as nonlinear equality constraints in \( S_1 \), so that the optimal solution of \( S_1 \) reflects the relationships between the controllable variable \( T_{sett} \) and the dependent variables \( P', Q', \)
and $T_l$, given the constant vector $E'$ for the current and delayed time steps. Specifically, in (6), $L_1(\cdot)$ specifies the relationships between the input variables $T_{sett}$ and $T_l$ and the output variable $P$ of the thermostat controller, thus establishing the links of (2) and (3) with (4) and (5). Similarly, in (7) and (8), $L_2(\cdot)$ and $L_3(\cdot)$ connect the variable $P'$ in (4) and (5) with the variables $Q'$ and hence $T_l$ in (3).

The optimization problem $S_1$ [i.e., (1)–(8)] can be equivalently expressed in a compact form using the simple expressions of $T_{sett}$, $P'$, $Q'$, $T_l$, and $E'$, as well as of $L_1(\cdot)$, as:

**S2: Compact form of the original problem $S_1$**

$$\arg\min_{\nu'} C_E = \sum_{t=1}^{N} y'_t,$$

subject to

1. $u_{\text{min}} \leq u' \leq u_{\text{max}}, \quad \forall t,$
2. $s'_{\text{min}} \leq s' \leq s'_{\text{max}}, \quad \forall t,$
3. $s' = g(s'^{-1}, v'), \quad \forall t,$
4. $y' = f(s', v'), \quad \forall t.$

In $S_2$, $y'$ and $u'$ are defined as the output $C_p'P'$ and the controllable input $T_{sett}'$, respectively, of the system for optimal building temperature control. Moreover, for notational simplicity, a vector $v'$ is used to represent the system inputs $[u', w]^T$, including the system disturbances $w' = E'$. Similarly, $s'$ is used to indicate the system states $[T_l', Q', P', \Delta P]_t$ that characterize the operating condition of the temperature control system at time $t$. Then, (1) can be equivalently represented as (9). Moreover, (2)–(5) can be simply expressed as (10) and (11), and (6)–(8) correspond to (12), where $g(\cdot)$ represents a set of nonlinear functions. Therefore, $S_1$ and $S_2$ are the same as each other. To complete the $S_2$ formulation using the simplified notation, (13) is added to connect $y'$ with $s'$ and $v'$, considering the relationship of $P'$ with $T_l$, $T_{sett}'$, and $E'$.

As clearly shown in (9)–(13), optimal operation of the HVAC system is achieved by solving a constrained nonlinear optimization problem. To find the optimal $u'$, $S_2$ is relaxed to an unconstrained problem using the continuous, quadratic penalty functions of (10) and (11) as:

**S3: Reformulated problem of $S_2$**

$$\arg\min_{\nu'} \sum_{t=1}^{N} \left[ \lambda_1 \left( y'^2 \right) \cdot \text{sgn}(y') + \lambda_2 \left( k' \right) + \lambda_3 \left( h' \right) \cdot h' \right],$$

where

1. $u' = \begin{cases} u_{\text{max}} & \text{for } u' > u_{\text{max}} \\ u_{\text{min}} & \text{for } u' < u_{\text{min}}, \forall t, \\ 0 & \text{otherwise} \end{cases}$
2. $s' = \begin{cases} s_{\text{max}} & \text{for } s' > s'_{\text{max}}, \forall t, \\ s_{\text{min}} & \text{for } s' < s'_{\text{min}}, \forall t, \\ 0 & \text{otherwise} \end{cases}$

In (14), $\lambda_1$ and $\lambda_2$ are the corresponding penalty factors. Large values of the penalty factors lead to good consistency between the optimal solutions of $S_1$ and $S_3$; see Appendix A. Note that penalty factors that are too large are likely to create steep valleys on the constraint boundaries, rendering it difficult to solve $S_3$ within a reasonable computation time. Therefore, it is common to apply penalty factors with small values and gradually increase them [31].

A GD algorithm [32], [33] is adopted to search for the minimum of the continuous, nonlinear function $S_3$, where the next step is determined proportional to the negative gradient of $S_3$ at the current step. Since this requires only the first derivative, the GD solver can readily be implemented in the BEMS, facilitating the optimal operation of the HVAC system in practice. Moreover, unlike heuristic, RL-based algorithms, the GD solver is deterministic and hence ensures that the optimal solution of $S_1$ leads to stable, reliable system operation.

**B. Online Supervised Learning**

After optimal scheduling of $T_{sett}$ has been initiated, $L_{1-3}$ undergo repeated online SL, as new data of $T_{sett}$, $P'$, $Q'$, and $T_l'$ are obtained for various profiles of $C'$ and $E'$. This gradually mitigates the overfitting of $L_{1-3}$. In other words, $L_{1-3}$ become well adapted to changes in the operating conditions of the building, further improving the accuracy of modeling the building thermal dynamics. Specifically, Fig. 3 shows a flowchart for the online SL of $L_{1-3}$. In Step 1, the optimal day-ahead scheduling of $T_{sett}$ is initiated, after $L_{1-3}$ are trained with the initial historical data of the BEMS. Due to the small size and variability of the data, $L_{1-3}$ are likely to be rather inaccurate, limiting the performance of the optimal scheduling. In Step 2, the HVAC system operates according to the optimal schedule of $T_{sett}$ on day $d$, and the BEMS collects the corresponding dataset $[T_{sett}, P', Q', E', T_l']$ for $1 \leq t \leq N_t$. The profiles of the dataset are likely to differ from those of the historical BEMS datasets before the optimal scheduling is initiated. This increases the variability in the training data, improving the accuracy of $L_{1-3}$ when they are re-trained using the newly collected dataset in Step 3. The re-training is
conducted using the historical and online datasets for a number of epochs, and stops when the modeling accuracy at the current epoch is marginally improved, compared to that at previous epochs. In Step 4, the optimization problem \( S_3 \) is updated using the retrained \( L_{1-3} \) and solved for the forecasts of \( E' \) on day \( d+1 \). The improved accuracy of \( L_{1-3} \) will lead to expansion of the feasible solution area of \( S_3 \), enhancing the performance of the optimal HVAC system operation. Step 2 is then repeated on day \( d+1 \) with the new optimal schedule of \( T_{\text{set}} \). In this paper, Steps 2–4 are performed on each scheduling day during the period from day \( d = 1 \) to \( N_d \) to achieve continual improvement of the modeling accuracy and the scheduling performance. In practice, Steps 2–4 can be conducted once every several days and repeated continuously until the results are satisfactory.

V. CASE STUDIES AND SIMULATION RESULTS

A. Test Conditions

The proposed strategy was tested for an experimental setup of an office building with an HVAC system, as shown in Fig. 4. Briefly, the experimental setup is divided into test and climate rooms, both of which are within a larger laboratory room with a temperature of \( T_i \). The test room has lights and heat sources to emulate the internal thermal load \( Q_i \) of a common office. The walls and floor consist of multiple layers of different building materials. For the case studies, the power rating of the HVAC system was set to \( P_{\text{max}} = 50 \text{ kW} \), and a scaled fraction of the corresponding \( Q_i \) was used to control \( T_i \) in the test room.

The climate room contains a separate heating unit to emulate the building thermal environments \( E' \). For the experimental setup, a building simulator was implemented in [34] to estimate \( T_i \) for \( P_i \), given \( E'=\{T_i', T_{\text{d}}', T_{\text{t}}', Q_i'\} \). In this study, the simulator was further extended by integrating the thermostat control loop with the HVAC system, as shown in Fig. 1; this enabled indirect control of the HVAC unit, as is common in real buildings.

To establish the initial training datasets, the building simulator was run using the data of \( Q_i' \) estimated from a real building [34], [35] and of \( T_i' \) measured in Boston from June 1 to August 31, 2017. As shown in Fig. 5(a) and (b), respectively. Note that \( Q_i' \) can also be surveyed and measured for benchmark buildings [37]. Given \( Q_i' \) and \( T_i' \), the simulator was run with the pre-determined profiles of \( T_{\text{set}}' \), such that \( T_i' \) was controlled within an acceptable range under the conditions of traditional HVAC system operation. Fig. 5(c) shows the corresponding profiles of \( P_i' \) obtained from the simulation runs. Moreover, Fig. 5(d) shows the profiles of \( C' \) [38] for the same time period as when the \( Q_i' \) and \( T_i' \) data were acquired. Note that on several days, \( C' \) decreased below zero in the early morning. The sizes of the initial datasets \( \{T_{\text{set}}', P_i', Q_i', E_i', T_c'\} \) were 1,200 (i.e., 50 days) and 8 with respect to time and objects, respectively. The size with respect to time continued to increase, further extended by integrating the thermostat control loop with the HVAC system. Fig. 5(c) shows the profiles of \( T_i' \), and solving \( S_3 \) during the period from \( d = 1 \) to \( N_d \). Note that the time-delayed data for the objects were not considered in the size estimation. The datasets were then randomly shuffled and divided into three parts with the ratios of 0.8:0.1:0.1 for the training, validation, and testing, respectively.

Table I lists the parameter values used for the modeling and optimal operation of the HVAC system in the case studies. The parameter values were determined mainly based on [8]–[10] and considering the current practices for DR, the sampling rates

![Fig. 4. Experimental setup for the data acquisition and comparative case studies](image)

![Fig. 5. Case study conditions from June 1 to August 31, 2017: (a) \( Q_i' \), (b) \( T_i' \), (c) \( P_i' \), and (d) \( C' \). The profiles of \( Q_i' \), \( T_i' \), \( P_i' \), and \( C' \) in 2018 and 2019 were similar.](image)
of the BEMS datasets, and the convergence rates of the solution of $S_t$. In particular, the learning rate for the LSTM training was set to be a small value of $4 \times 10^{-3}$, and the learning rate for the GD solver was reduced from $R_0 = 10^{-3}$ to $R_d = 10^{-4}$ when the epoch number increased to greater than two thirds of the total number of epochs. This aimed to achieve high accuracy of the LSTM models $L_{1-3}$ and the optimal solution to $S_t$. Moreover, $\lambda_y$ and $\lambda_e$ were set to relatively large values to reduce the HVAC energy cost while ensuring the occupants’ thermal comfort. By contrast, $\lambda_d$ did not have to be set to a large value, because $T_{set}$ and $P'$ varied within the acceptable ranges before the proposed strategy was applied. In other words, $L_1$ and $L_2$ were trained with the historical datasets of $T_{set}$ and $P'$ ranging only between $T_{set,min}$ and $T_{set,max}$ and between $P_{min}$ and $P_{max}$, respectively. For simplicity, $\lambda_1$, $\lambda_2$, and $\lambda_3$ were fixed during $1 \leq d \leq N_d$.

The HVAC system operations were compared for three cases: the proposed SL-based strategy (Case 1), an ideal physics-based strategy (Case 2), and a traditional rule-based strategy (Case 3). Table II lists the main features of Cases 1–3. In Case 2, the piecewise linear equations for variations in $T_{set}$ and $P'$ were employed, and then applying mixed-integer linear programming (MILP). Note that Case 2 is referred to as the ideal case, because most of the information is not available in practice. In Case 3, $T_{set,i}$ was fixed at 23°C, regardless of the variation in $C_i$. For fair comparison of Cases 1–3, the HVAC system was assumed to be capable of operating from $t = 1$ h in the case studies. This also allowed the building to take advantage of pre-cooling for all cases 1–3; late start of HVAC system and building envelope [39]; see (B1)–(B5) in Appendix B. The optimal schedule of $T_{set,i}$ was then obtained by replacing (6)–(8) with (B1)–(B5) and then applying mixed-integer linear programming (MILP). Note that Case 2 is referred to as the ideal case, because most of the information is not available in practice. In Case 3, $T_{set,i}$ was fixed at 23°C, regardless of the variation in $C_i$. For fair comparison of Cases 1–3, the HVAC system was assumed to be capable of operating from $t = 1$ h in the case studies. This also allowed the building to take advantage of pre-cooling for all cases 1–3; late start of HVAC operation had the risk of causing an increase in $C_E$ and a deviation of $T_i$ from the acceptable range.

### B. Improvement via Online Supervised Learning

The accuracy of the LSTM-based building model was verified by comparing the actual values of $P'$, $Q'$, and $T_i$ in the testing datasets (discussed in Section V-A) with the corresponding estimates obtained from $L_{1-3}$. Note that the estimates were acquired after training and interconnecting $L_{1-3}$. Fig. 6 shows the results of the comparisons for $d = 1$ and $N_d$, where the x- and y-axes represent the actual values and the estimates, respectively. For $d = 1$, the normalized root mean square errors (nRMSEs) of $L_{1-3}$ were estimated to be rather considerable: i.e., $1.1 \times 10^{-3}$, $9.5 \times 10^{-3}$, and $5.8 \times 10^{-3}$, respectively. As the online SL and optimal scheduling continued, the nRMSEs for $d = N_d$ were reduced to low levels of $9.1 \times 10^{-5}$, $3.1 \times 10^{-5}$, and $4.8 \times 10^{-4}$, respectively. Fig. 7 shows the variations in the nRMSEs over the period from $d = 1$ to $N_d$. For all $L_{1-3}$, the nRMSEs were reduced rapidly during the initial period and decreased gradually for the remaining period. The results of the case studies confirmed that the online SL integrated with the optimal scheduling is effective in improving the accuracy of LSTM-based models of sub-systems (and hence the complete system) for building temperature control. In particular, the reduction of the nRMSEs for the testing datasets verified not only the improvement of the modeling accuracy of $L_{1-3}$ but also

![Image](https://via.placeholder.com/150)

Fig. 6. Comparisons of the actual and estimated values of (a) $P'$, (b) $Q'$, and (c) $T_i$. The red and blue dots indicate the test results for $d = 1$ and $N_d$, respectively.

![Image](https://via.placeholder.com/150)

Fig. 7. Variations in the nRMSEs of (a)–(c) $L_{1-3}$ during the online SL.

### TABLE III. AVERAGE VALUES OF THE COST REDUCTION RATES FOR CASE 1

<table>
<thead>
<tr>
<th>Periods</th>
<th>$0 \leq d \leq N_d/4$</th>
<th>$N_d/4 &lt; d \leq N_d/2$</th>
<th>$N_d/2 &lt; d \leq 3N_d/4$</th>
<th>$3N_d/4 &lt; d \leq N_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{PCR}^{\text{avg}}$ [%]</td>
<td>20.96</td>
<td>21.66</td>
<td>22.62</td>
<td>23.36</td>
</tr>
</tbody>
</table>

![Image](https://via.placeholder.com/150)

Fig. 8. Optimal scheduling results for Cases 1–3 over the time period from $d = 1$ to $N_d$: (a) $C_E$ and (b) the average of $\text{w}$. The enhancement of their generalization capability, because the testing datasets were not used to train $L_{1-3}$, as discussed in Section V-A. In other words, the online SL continued, $L_{1-3}$ became less over-fitted and hence more capable of accurately predicting the outputs of the sub-systems for the unseen inputs.

Figs. 6 and 7 show that the nRMSEs of $L_1$ were estimated to be higher than those of $L_2$ and $L_3$. This was mainly because the output of $L_1$ (i.e., $P' \approx P_{opt}$) changed faster and with larger magnitudes than the outputs of $L_2$ and $L_3$ (i.e., $Q'$ and $T_i'$, respectively) due to the thermal capacity inherent in the HVAC refrigerant loop and building envelope. Note that in the case studies, the cooling energy supplied by the HVAC system was assumed to be equally divided into $Q'$ and used to control $T_i'$ in the test building room, as discussed in Section V-A.

In addition, Table III and Fig. 8 show the optimal scheduling results for the proposed strategy (i.e., Case 1) in comparison with those for the ideal and traditional strategies (i.e., Cases 2 and 3). For Cases 1 and 3, $C_E$ was calculated as $\Sigma C_i P_i$ and $\Sigma C_i P_{i,avg}$, respectively, during $1 \leq t \leq N_d$. The cost reduction rate

---

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2021.3051564, IEEE Transactions on Smart Grid

Fig. 9. Comparisons of the 24-h schedules for Cases 1–3: (a) $C$, (b) $T_i$, (c) $Q_i$, (d) $T_{set}$, (e) $P$, and (f) $T_f$.

Fig. 10. Comparisons of the 24-h schedules for Cases 1–3 for different profiles of $C$ and $E$: (a) $C$, (b) $T_i$, (c) $Q_i$, (d) $T_{set}$, (e) $P$, and (f) $T_f$. In Fig. 10(a), the $y$-axis is broken to accommodate the peak of $C$ for $t = 15$ h.

**TABLE IV. COMPARISONS OF THE PROPOSED STRATEGY WITH THE IDEAL AND RULE-BASED STRATEGIES**

<table>
<thead>
<tr>
<th>Profiles of $C$ and $E'$</th>
<th>Proposed (Case 1)</th>
<th>Ideal (Case 2)</th>
<th>Rule-based (Case 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_0$ [S]</td>
<td>6.75</td>
<td>6.74</td>
<td>9.39</td>
</tr>
<tr>
<td>$r_{CR}$ [%]</td>
<td>28.1</td>
<td>28.2</td>
<td>-</td>
</tr>
<tr>
<td>$T_{set}$ [°C]</td>
<td>0.18</td>
<td>0.18</td>
<td>0</td>
</tr>
<tr>
<td>$C_0$ [S]</td>
<td>4.90</td>
<td>4.84</td>
<td>8.41</td>
</tr>
<tr>
<td>$r_{CR}$ [%]</td>
<td>41.7</td>
<td>42.4</td>
<td>-</td>
</tr>
<tr>
<td>$T_{set}$ [°C]</td>
<td>0.18</td>
<td>0.08</td>
<td>0</td>
</tr>
</tbody>
</table>

$r_{CR}^d$ for day $d$ is then estimated as:

$$r_{CR}^d := \frac{\sum_{i=1}^{N_d} C_i^d (P_i^d - P_i')}{\sum_{i=1}^{N_d} C_i^d P_i'}$$

In Table III, the average value of $r_{CR}^d$ during the period of every $N_d/4$ (i.e., 50) days is calculated as:

$$r_{CR}^{avg} := \frac{1}{N_d/4} \sum_{d=1}^{N_d} r_{CR}^d$$

where $T_i = \{i-1\}N_d/4 + 1, \ldots , iN_d/4\}$ for $i = 1, \ldots , 4$. As the online SL continued, the average reduction rate $r_{CR}^{avg}$ gradually increased from 20.96% to 23.36%. Fig. 8(a) shows the comparisons of $C_E$ for Cases 1–3 for each day $d$. For Case 1, $C_E$ was only slightly larger than for Case 2 but considerably smaller than for Case 3. Fig. 8(b) shows the average of the accumulated deviations in $T_i$ during the period of every $N_d/4$ days, given by:

$$v_{TC} := \frac{\sum_{i=1}^{N_d} \left[ \max \left( T_i - T_{i,\text{max}} \right) + \max \left( T_{i,\text{min}} - T_i \right) \right]}{T_{i,\text{max}} - T_{i,\text{min}}}$$

which results from the penalty function of (3). For brevity, (19) is expressed using a linear form, rather than a quadratic form, because $v_{TC}$ can be directly calculated from the optimal profile of $T_i$. In Fig. 8(b), the average of $v_{TC}$ for Case 1 was gradually reduced and became comparable to that for Case 2. As $\Delta t$ in (14) increases, $v_{TC}$ can be reduced more rapidly and maintained further lower, although $C_E$ is likely to increase. The case study results verify that the proposed strategy is effective in reducing the HVAC energy cost, while ensuring the thermal comfort.

C. Comparisons of Operating Schedules of HVAC System

Fig. 9 represents the 24-h schedules of $T_{set}$ and the corresponding variations in $P$ and $T_f$ for Cases 1, 2, and 3, given the forecasts of $C$ and $E'$. Specifically, for Case 1, $T_{set}$ was scheduled at relatively low levels in the early morning due to the low values of $C$, whereas $T_f$ and $Q_i$ were maintained high during $7 \leq t \leq 19$ h. As $C'$ began to increase, $P$ for Case 1 then became lower than that for Case 3. In other words, the proposed strategy achieved the HVAC load shift from on-peak hours to off-peak hours, leading to the pre-cooling operation and hence the reduction of the HVAC energy cost. Table IV shows that $C_E$ for Case 1 was estimated as $6.75$, which is 28.1% less than $9.39$ for Case 3. Fig. 9(f) shows that in Case 1, $T_f$ was still successfully controlled within the acceptable range.

Fig. 10 shows the scheduling results for different profiles of $C$ and $E'$. Specifically, $C$ differed more between the off- and on-peak hours. Fig. 10(a) shows that $C'$ was negative at $t = 3$ h and 5 h and increased up to 12.1 C/kWh at $t = 15$ h; note that the $y$-axis was broken to better display the variation in $C'$. Moreover, $T_f$ and $Q_i$ were estimated to be lower, compared to the cases for Fig. 9(c) and (d), respectively. Therefore, the shift in $P$ became larger than for the case of Fig. 9(e), leading to a larger reduction in $C_E$: i.e., from $r_{CR} = 28.1\%$ to 41.7%. In other words, a larger amount of the demand-side flexibility was provided due to the larger difference between $C'$ for the on- and off-peak hours and the more favorable operating conditions of the HVAC system during the on-peak hours, compared to the case shown in Fig. 9. This confirms that the proposed strategy...
could successfully reflect the load shifting capabilities of the HVAC system in response to the different profiles of time-varying electricity prices and building thermal conditions.

In Figs. 9 and 10, the optimal schedules for the proposed and ideal strategies (i.e., Cases 1 and 2) were considerably similar, confirming the accuracy of $L_{1-3}$ and the convergence of the solution of $S_3$ to that of $S_1$, and further to that of the ideal strategy. The small difference arose mainly because the proposed strategy was developed using the actual operating data of the temperature control system, whereas the ideal strategy was achieved using complete information on the system modeling parameters. It was also attributable to the difference between the GD and MILP solvers.

D. Comparisons with Other SL- and RL-based Strategies

The case studies discussed in Sections V-B and V-C were repeated to further evaluate the performance of the proposed SL-based strategy. In particular, as shown in Table V, the proposed strategy was evaluated by comparison with the conventional SL-based strategies, in which a single LSTM was trained offline and online to model the temperature control system. The case with two online-trained LSTMs was also considered, the first of which modeled the thermostat control loop, and the second corresponded to the HVAC system and the building envelope. The comparative study results confirm that the proposed strategy is more effective in improving the building modeling accuracy and temperature control performance, while maintaining the computation time within reasonable limits. The computation time was estimated on a computer with a six-core 4.3-GHz CPU and 32 GB of RAM.

The proposed strategy was also developed using different types of ML models: e.g., one RNN for $L_1$ and two ARMAXs for $L_{2,3}$ (Case 4) and two RNNs for $L_{1,2}$ and one GRU for $L_3$ (Case 5). The ML models were simpler than the LSTMs. Fig. 11 shows the nRMSEs of $L_{1-3}$ for the testing datasets in Cases 4 and 5 over the period from $d = 1$ to $N_d$. As the online SL continued, the nRMSEs of all $L_{1-3}$ were still reduced to low levels in both Cases 4 and 5. After the online SL had finished, the optimal schedules of $P^*$ and the corresponding variations in $T_i^*$ were obtained in Cases 4 and 5 for the profiles of $C'$ and $E'$ shown in Fig. 9. Fig. 12 shows that for Cases 4 and 5, the proposed strategy still achieved the HVAC load shift from on-peak hours to off-peak hours, while leading to small deviations in $T_i^*$ from the acceptable range. This led to a reduction in $C_E$, compared to Case 3, as shown in Table VI. The case study results confirmed that the proposed strategy can be widely and adaptively applied in real buildings with different temperature control systems and corresponding operating datasets. Moreover, for Cases 4 and 5, the computation times were lower than for Case 1, whereas the HVAC energy costs were higher than for Case 1, revealing the trade-off between the computational burden and the modeling accuracy and scheduling performance.

Furthermore, the proposed strategy was compared with an RL-based strategy using a DDPG algorithm (Case 6) [13], [14]. After initially trained with the historical datasets, the critic and actor networks were further trained for 200 episodes, as in the proposed strategy, each of which was characterized by the

<p>| TABLE V. COMPARISONS BETWEEN THE PROPOSED AND CONVENTIONAL SL-BASED STRATEGIES |</p>
<table>
<thead>
<tr>
<th>Profiles of $C'$ and $E'$</th>
<th>Online SL</th>
<th>3 LSTMs</th>
<th>2 LSTMs</th>
<th>1 LSTM</th>
<th>Offline SL</th>
<th>1 LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e(T_i, T_{i-1})$</td>
<td>$4.02 \times 10^5$</td>
<td>$1.53 \times 10^3$</td>
<td>$3.71 \times 10^4$</td>
<td>$3.86 \times 10^5$</td>
<td>$4.02 \times 10^5$</td>
<td></td>
</tr>
<tr>
<td>$C_E$ [S]</td>
<td>6.75</td>
<td>7.11</td>
<td>8.11</td>
<td>10.6</td>
<td>6.75</td>
<td></td>
</tr>
<tr>
<td>$r_{cr}$ [%]</td>
<td>28.1</td>
<td>31.6</td>
<td>22.0</td>
<td>-1.92</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>$v_{rc}$ [°C]</td>
<td>1.03</td>
<td>1.20</td>
<td>1.34</td>
<td>0.55</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>comp. time [s]</td>
<td>1,618</td>
<td>1,377</td>
<td>1,052</td>
<td>1,073</td>
<td>1,618</td>
<td></td>
</tr>
<tr>
<td>$e(T_i, T_{i-1})$</td>
<td>$3.97 \times 10^5$</td>
<td>$1.24 \times 10^3$</td>
<td>$7.09 \times 10^4$</td>
<td>$3.82 \times 10^5$</td>
<td>$3.97 \times 10^5$</td>
<td></td>
</tr>
<tr>
<td>$C_E$ [S]</td>
<td>4.90</td>
<td>5.77</td>
<td>6.07</td>
<td>8.18</td>
<td>4.90</td>
<td></td>
</tr>
<tr>
<td>$r_{cr}$ [%]</td>
<td>41.7</td>
<td>31.4</td>
<td>27.8</td>
<td>2.73</td>
<td>41.7</td>
<td></td>
</tr>
<tr>
<td>$v_{rc}$ [°C]</td>
<td>0.18</td>
<td>0.60</td>
<td>0.72</td>
<td>0</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>comp. time [s]</td>
<td>1,616</td>
<td>1,442</td>
<td>1,056</td>
<td>1,078</td>
<td>1,616</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 11. Variations in the nRMSEs for Cases 4 and 5 during the online SL.
of (a) $L_1$, (b) $L_2$, and (c) $L_3$ on the nRMSEs of the other LSTMs.

![Effects of $\alpha(P, P')$](image)

![Effects of $\alpha(Q', Q')$](image)

![Effects of $\alpha(T, T')$](image)

Fig. 13. Effects of the nRMSE of (a) $L_4$, (b) $L_5$, and (c) $L_6$ on the nRMSEs of the other LSTMs.

![Effects of $\lfloor L_4 \rceil$ on the nRMSE of $\lfloor L_4 \rceil$, $\alpha(P, P')$](image)

![Effects of $\lfloor L_5 \rceil$ on the nRMSE of $\lfloor L_5 \rceil$, $\alpha(Q', Q')$](image)

![Effects of $\lfloor L_6 \rceil$ on the nRMSE of $\lfloor L_6 \rceil$, $\alpha(T, T')$](image)

Fig. 14. Effects of the maximum time delays $L_{P1}$, $L_{P2}$, and $L_{P3}$ on the nRMSEs of (a)–(c) $L_4$, (d)–(f) $L_5$, and (g)–(i) $L_6$ for $d = 1$, $N_d/2$, and $N_d$.

![Effects of $\lambda_y$ and $\lambda_y$ on $C_{E'}$ and $v_{TC}$](image)

![Effects of $\lambda_y$ and $\lambda_y$ on $C_t$ and $v_{TC}$](image)

Fig. 15. Effects of the penalty factors (a) $\lambda_y$ and (b) $\lambda_y$ on $C_{E'}$ and $v_{TC}$.

E. Sensitivity Analyses

For the proposed strategy, the effect of the modeling error of each LSTM was analyzed on the modeling accuracy of the other LSTMs. Fig. 13(a) shows the variations in the nRMSEs of $L_4$ and $L_5$ for an increase in the nRMSE of $L_1$ approximately from $4.40 \times 10^{-3}$ to $1.45 \times 10^{-1}$. When $\alpha(P, P')$ was smaller than about $2.42 \times 10^{-2}$, both $\alpha(Q', Q')$ and $\alpha(T, T')$ marginally increased to low levels. When it became greater than $2.42 \times 10^{-2}$, $\alpha(Q', Q')$ and $\alpha(T, T')$ increased rather rapidly until they were saturated at high levels. This is also the case for the nRMSE variations shown in Fig. 13(b) and (c). Note that in Fig. 13(c), the nonlinearity of the building sub-systems led to the sharp variations in $\alpha(P, P')$ when all the nRMSEs were at high levels. Given the analysis, the permissible error margins of $L_4$–$L_6$ can be specified as $2.42 \times 10^{-2}$, $6.79 \times 10^{-2}$, and $1.41 \times 10^{-2}$, respectively. For all $L_1$–$L_3$, the nRMSEs were smaller than the margins, particularly as the online SL started and continued (see Fig. 7).

In addition, Fig. 14 shows the variations in the nRMSEs of $L_1$–$L_3$ for gradual increases in the maximum time delays of the network input data (i.e., $L_{P1}$ in (6)–(8)). Specifically, Fig. 14(a)–(c) show the variation in $\alpha(P, P')$ with respect to an increase in $L_{P1}$ for $d = 1$, $N_d/2$, and $N_d$, respectively, while $L_{P2}$ and $L_{P3}$ were fixed at 4 h. It can be seen that $L_{P1} = 24$ h led to the smallest value of $\alpha(P, P')$. When $L_{P1}$ was too small, $L_1$ could not accurately reflect the thermostat controller operation. Moreover, the operation of the thermostat controller at the current time step was marginally affected by the operations at previous time steps long before the current step. This was also the case for the nRMSE variations in $L_2$ and $L_3$ for changes in $L_{P2}$ and $L_{P3}$, respectively. Fig. 14(d)–(i) show that $L_{P2} = 4$ h and $L_{P3} = 4$ h led to the smallest values of $\alpha(Q', Q')$ and $\alpha(T, T')$, respectively, for $d = 1$, $N_d/2$, and $N_d$. Large values of $L_{P2}$ and $L_{P3}$ did not noticeably improve the accuracy of $L_2$ and $L_3$ due to the limited thermal capacity of the test building room.

The case studies, discussed in Section V-C, were also repeated while increasing $\lambda_y$ and $\lambda_y$ to 4.0 and 15.0, respectively. Fig. 15(a) represents that an increase in $\lambda_y$ led to a decrease in $C_{E'}$ and an increase in $v_{TC}$ for both profiles of $C$ and $E^r$ shown in Figs. 9 and 10. When $\lambda_y$ increased to greater than 2.0, $C_{E'}$ was marginally reduced, whereas $v_{TC}$ was rapidly increased to an inadmissible level. Similarly, Fig. 15(b) shows the case for an increase in $\lambda_y$. It can be seen that $v_{TC}$ was slightly reduced when $\lambda_y$ increased higher than 7.5. Note that the nonlinearity of the sub-systems led to sudden variations in $C_{E'}$ particularly when $\lambda_y$ varied from 7.5 to 15.0 for the profile of $C$ and $E^r$ shown in Fig. 10. Moreover, for $\lambda_y = 0.5$, $T_{set}$ and $P'$ were successfully maintained within the acceptable ranges for all $\lambda_y$ and $\lambda_y$.

VI. CONCLUSIONS

This paper proposed a new SL-based strategy for optimal operation of an HVAC system in a commercial building. The system for indoor temperature control was divided into three sub-systems, each of which was modeled using an LSTM. The LSTMs were then interconnected and integrated directly into the optimization problem for temperature set-point scheduling. The optimization problem was reformulated and solved using a deterministic search algorithm within reasonable computation time limits. After optimal scheduling was initiated, the interconnected LSTMs went through the online SL repeatedly, gradually improving the modeling accuracy and the scheduling performance. Case studies were conducted to validate the performance of the proposed strategy in comparison with other strategies using a rule-based temperature set-point, an ideal physics-based building model, and other types of ML-based modeling and scheduling methods. The case study results confirmed that the proposed strategy accurately reflects the...
load-shifting capability of the HVAC system in response to the
time-varying electricity prices and building thermal environ-
ments, successfully reducing the HVAC energy cost. The
results also verified that the proposed strategy effectively
mitigates the requirement for historical building data and the
risk of unstable operation of the HVAC system and thermal
discomfort of occupants in the initial learning period, which is
of utmost importance for practical application.

**APPENDIX**

**A. Consistency Between the Optimal Solutions of S2 and S3**

The consistency between the solutions of S2 and S3 is proved
considering a general case [40] as:

Minimize \( \{ f(x) : x \in D \} \) , \( \text{(A1)} \)

where \( f \) is a nonlinear continuous function on \( \mathbb{R}^n \) and \( D \) is a
constraint set in \( \mathbb{R}^n \). Then, (A1) is reformulated to:

Minimize \( \{ J(x, \lambda) = f(x) + \lambda P(x) \} \) , \( \text{(A2)} \)

where \( \lambda \) is a positive constant and \( P(\cdot) \) is a penalty function on
\( \mathbb{R}^n \) that satisfies \( P(x) \geq 0 \) for all \( x \in \mathbb{R}^n \) and
\( P(x) = 0 \) if and only if \( x \in D \). Lemma 1 then gives a set of
inequalities that follows from the definition of \( x_0 = \arg \min_x J(x, \lambda) \),
and the inequality \( \lambda_{k+1} > \lambda_k \).

**Lemma 1:** \( J(x_k, \lambda_k) \leq J(x_{k+1}, \lambda_{k+1}) \); \( P(x_k) \geq P(x_{k+1}) \); and \( f(x_k) \leq f(x_{k+1}) \)

**Proof:** \( J(x_{k+1}, \lambda_{k+1}) = J(x_{k+1}, \lambda_{k+1}) + \lambda_{k+1} P(x_{k+1}) \geq J(x_{k+1}, \lambda_{k+1}) + \lambda_{k+1} P(x_{k+1}) \)

which proves the first inequality. Moreover, (A3) leads to:

\( f(x_k) + \lambda_{k+1} P(x_k) \leq f(x_{k+1}) + \lambda_{k+1} P(x_{k+1}) \)

(A4) and (A5) and rearranging the terms yield:

\( (\lambda_{k+1} - \lambda_k) P(x_{k+1}) \leq (\lambda_{k+1} - \lambda_k) P(x_k) \)

which proves the second inequality. In conjunction with (A6),
the definition of \( x_k \) gives:

\( f(x_{k+1}) + \lambda_{k+1} P(x_{k+1}) \geq f(x_k) + \lambda_{k+1} P(x_k) \geq f(x_k) + \lambda_{k+1} P(x_{k+1}) \)

which proves the third inequality.

**Lemma 2:** Let \( x^* \) be a solution of \( (A1) \). Then, \( f(x^*) = J(x_k, \lambda_k) \geq f(x_k) \) for each \( k \).

**Proof:** \( f(x^*) = f(x^*) + \lambda_{k+1} P(x^*) \geq f(x_k) + \lambda_{k+1} P(x_k) \geq f(x_k) \).

The two lemmas supports the proof of the theorem on the
convergence of the solution of \( (A2) \) to that of \( (A1) \).

**Theorem 1:** Let \( 0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < \lambda_{k+1} < \cdots \rightarrow \infty \). Let \( \bar{x} \) be an
arbitrary limit point of \( \{ x_k \} \) \( k \rightarrow \infty \). Then, \( \bar{x} \) solves \( (A1) \).

**Proof:** The limit point is defined as \( \bar{x} = \lim x_k \). Since \( f \) is continuous, \( \lim f(x_k) = f(\bar{x}) \). Then,

\( J^* = \lim_{k \rightarrow \infty} J(x_k, \lambda_k) = f(\bar{x}) \).

\( J^* = \lim_{k \rightarrow \infty} f(x_k) + \lambda \lim_{k \rightarrow \infty} \lambda_{k+1} P(x_k) = f(\bar{x}) \).

\( J^* = f(\bar{x}) + \lambda \lim_{k \rightarrow \infty} \lambda_{k+1} P(x_k) = f(\bar{x}) \).

\( J^* = \lim_{k \rightarrow \infty} f(\bar{x}) + \lambda \lim_{k \rightarrow \infty} \lambda_{k+1} P(x_k) \leq f(\bar{x}) \).

Since \( J^* - f(\bar{x}) \) and \( f(\bar{x}) \) are finite, \( \lim \lambda_{k+1} P(x_k) \) is a finite quantity.

For \( \lambda \rightarrow \infty \), \( P(x_k) \) converges to zero, verifying \( P(\bar{x}) = 0 \).

**B. Optimization Problem Constraints for the Ideal Strategy**

For the comparative case studies, a physics-based model of
the system for the temperature control was implemented as:

\[ e^t = T^d - T^c, \quad \forall t, \]

(B1)

\[ P^e = P^0 - k_e e^t - \sum_{i=1}^{n} G_i e^i, \quad \forall t, \]

(B2)

\[ \sum_{i=1}^{n} \delta^t = P^c, \quad \forall t, \]

(B3)

\[ T^c = T^d + \sum_{i=1}^{n} F_i \delta^t, \quad \forall t, \]

(B4)

\[ \forall \delta^t, \{ 0, 1 \}, \forall n, \forall t. \]

The constraints (B1) and (B2) represent the operation of the PI
controller in the thermostat feedback loop. Moreover, (B3)–
(B5) correspond to the piecewise linear approximation of
the nonlinear variation in \( T^c \) for a change in \( P^e \) [39].
Specifically, in (B3), \( P^e \) is divided into \( N_t \) linear blocks.
In (B4), the variation from \( T^c \) in \( T^c \) is calculated as the sum of the temperature
variations that are led by the incremental HVAC loads assigned
in the linear blocks. This is possible because in (B4), \( F_i \)
contains the complete information on the inter-time thermal
response of the building to the HVAC system operation.
Moreover, (B5) represents the boundaries of the linear blocks
to complete the piecewise linear approximation.

**REFERENCES**


www.eia.gov/consumption/commercial/data/2012/pdf/user_guide_public
use_may2016.pdf.

2018.

Sep. 2012.

techopen.com/chapter/pdf-preview/62059.

scheduling considering dynamic electricity pricing in smart home,” IEEE Trans.


control (MPC) and optimization of HVAC systems: A state of the art
review and case study of a residential HVAC system,” Energy Build., vol.

between predicted and actual energy consumption in buildings,” IEEE Trans.

depth reinforcement learning,” IEEE Syst. J., vol. 214, no. 4, pp. 5362-5372,


[38] Ye-Eun Jang (S’20) received the B.S. and M.S. degree in electrical engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2016 and 2020, respectively, where she is currently pursuing the Ph.D. degree in electrical engineering. Her research interests include power system reliability, smart grid, and smart buildings.

[39] Young-Jin Kim (Senior Member, IEEE) received the B.S. and M.S. degree in Electrical Engineering from Seoul National University in 2007 and 2010, respectively, and his Ph.D. degree in Electrical Engineering from Massachusetts Institute of Technology in 2015.

He worked for Korea Electric Power Corporation as a power transmission and distribution system engineer from 2007 to 2011. He was also a visiting scholar at the Catalonia Institute for Energy Research in 2014, and a postdoctoral researcher for Center for Energy, Environmental, and Economic Systems Analysis at the Energy Systems Division at Argonne National Laboratory from 2015 to 2016. He joined the faculty of the Pohang University of Science and Technology (POSTECH), where he is currently an Associate Professor in the Department of Electrical Engineering. His research fields of interest include distributed generators, renewable energy resources, and smart buildings.

[40] João P. S. Catalão (Senior Member, IEEE) received the M.Sc. degree from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2003, and the Ph.D. degree and Habilitation for Full Professor ("Agregação") from the University of Beira Interior (UBI), Covilhã, Portugal, in 2007 and 2013, respectively. Currently, he is a Professor at the Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal, and Research Coordinator at INESC TEC. He was also appointed as Visiting Professor by North China Electric Power University (NCEPU), Beijing, China. His research interests include power system operations and planning, distributed renewable generation, power system economics and electricity markets, demand response and smart grid.