Near-optimal Chiller Control for Demand Response using Input Convex Neural Networks

Guanyu Tian
Department of Marine Engineering Technology
Texas A&M University
Galveston, TX 77554, USA
Email: tiang@tamu.edu

Mucun Sun
Idaho National Laboratory
Idaho Falls, ID 83415, USA
Email: mucun.sun@inl.gov

Zheng O'Neill
Department of Mechanical Engineering
Texas A&M University
College Station, TX 77840, USA
Email: zoneill@tamu.edu

***Abstract*—Chillers, as major controllable building loads, offer strong potential for demand response (DR), yet accurate and fast control without sac rificing energy efficiency remains challenging due to their nonlinear and nonconvex behavior. This paper proposes an Input Convex Neural Network (ICNN)-based method to track the energy-efficient operation points of chillers while providing DR services. The proposed ICNN architecture guarantees convexity with respect to controllable variables, enabling global optimization and seamless integration into higher-level DR decision-making. A physics-informed training strategy is developed to ensure the learned ICNN captures both chiller thermodynamic relationships and optimal operation patterns. Case studies using a digital twin model of a real-world chiller demonstrate that the ICNN-based method achieves higher control accuracy and faster computation compared to the state-of-the-art analytical method. The results confirm that the ICNN approach provides a tractable, accurate, and scalable tool for real-time chiller control in smart grid applications.**

***Index Terms*—Demand response, HVAC system, chiller control, neural networks, convex optimization.**

I. Introduction

The integration of intermittent renewable energy sources into the power grid has led to significant challenges in maintaining grid stability and reliability. As the share of renewables such as solar and wind power increases, the grid experiences more frequent and substantial fluctuations in supply, which can result in imbalances between generation and demand [1]. To address these challenges, demand response (DR) has emerged as a crucial strategy. DR enables the adjustment of electricity demand from consumers in response to grid conditions, thereby enhancing grid flexibility and reducing the need for additional generation capacity during peak demand periods [2].

Buildings, especially large commercial buildings, play a vital role in DR programs due to their significant energy consumption and potential for flexible load management [3]. Among the electrical loads in buildings, the heating, ventilation, and air-conditioning (HVAC) systems, particularly chillers, are major contributors to electricity demand. Chillers provide the cooling capacity in buildings, and their power consumption can be adjusted flexibly for DR [4]. Quantifying chiller power to its operational condition is essential for both the grid and building perspectives. From the grid side, it provides the power flexibility information that allows grid operators to accurately estimate the available demand-side resources that can be called upon during critical grid events, such as peak load periods or when there is a sudden drop in renewable energy generation [5]. For building owners and facility managers, it guides the chiller control in response to grid signals [6].

Accurately and computationally efficiently chiller control that meets specific power consumption is a challenging task, as it is affected by multiple interacting factors, including building cooling load, ambient temperature, operating conditions, and control strategies [7]. Many existing quantification methods rely on simplified regression models, which may fail to fully capture the complex nonlinear relationships among these variables [8]. In addition to model approximation errors, control strategies exert a substantial influence on the actual demand response (DR) flexibility. Unrealistic flexibility may occur when the quantified operating point is away from the optimal efficiency curve, which deviates from the actual DR capacity that can be achieved when dispatched. Such deviated flexibility cannot be achieved in practice and may mislead grid operators into overestimating the available demand-side resources. This overestimation can result in DR events where the expected load reduction is not delivered, potentially compromising grid stability and leading to DR program failures. To address these issues, a combination of a quadratic approximation and a semidefinite programming (SDP) formulation has been proposed in [9] to quantify the minimum chiller power under varying thermal loads. However, the quadratic approximation must be recalibrated when the return water temperature changes, making the method a two-step workflow that is difficult to embed as a surrogate, tractable component within higher-level optimization problems.

In recent years, artificial intelligence and machine learning techniques have shown great potential and capability in mirroring complex energy systems with boosted computational efficiency. Neural networks, in particular, have been widely used for building energy modeling and prediction due to their ability to approximate non-linear functions [10]. However, classical neural networks have limitations when it comes to optimization-related tasks, as they do not guarantee convex input-output mappings. This lack of convexity hinders data-driven models to be widely integrated into the decision-making of critical cyber-physical systems, such as the power grid, because their near-optimal solutions are intractable and not global optimal guaranteed [11].

In this paper, we aim to present a novel approach for chiller control for DR while guaranteeing its operation efficiency using the input convex neural networks (ICNNs) [12]. Unlike

979-8-3315-8138-1

previous methods that prioritize DR capacity quantification at the expense of operational efficiency, our method inherently accounts for the operational efficiency of chillers. The ICNN is trained to track the efficient operation point of a chiller that minimizes the output power under given thermal loads. This not only enables more accurate quantification of chiller power flexibility but also ensures that the chillers operate in an energy-efficient manner, avoiding unnecessary energy waste during demand response actions. Thanks to the convex nature of the ICNN, its trained parameters can be further reformulated as convex constraints to approximate the chiller power models in higher-level decision-making in DR programs, without compromising the traceability of the higher-level optimization problem.

The contributions of this paper are threefold: 1) restructuring the chiller optimal control problem into a tractable surrogate ICNN-based model that enables efficient power flexibility quantification, 2) adjusting the network training process to the physics-informed learning that can achieve a higher accuracy, and 3) validating the accuracy and time efficiency of the proposed method using the digital twin model of a real-world chiller plant.

This paper is organized as follows. In section II, we briefly introduce the chiller plant power consumption models. The background theory of ICNN and its reformulations to the proposed ICNN-based chiller control framework are introduced in III. Section IV presents the numerical results of case studies, and section V provides conclusions and discusses future research directions.

II. Problem Formulation

Chillers are the core components of building HVAC systems, responsible for delivering the required cooling capacity by extracting heat from the building's chilled water loop and rejecting it to the external environment. The control and optimization of chiller operation are crucial not only for energy efficiency but also for enabling building participation in demand response (DR) programs.

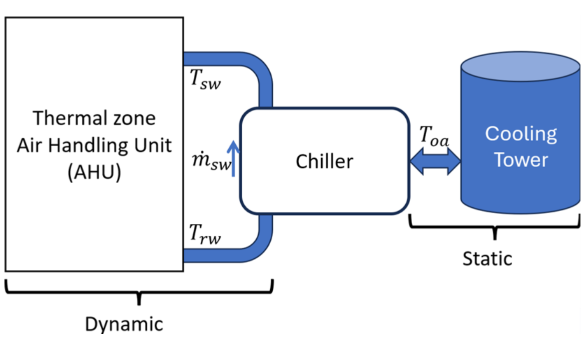

Fig. 1. Typical configuration of a water-cooled chiller with cooling tower

Fig. 1 illustrates a typical chiller plant configuration, where chilled water is supplied to the building's air handling units (AHUs) at a lower temperature (T_{sw}), absorbs heat as it circulates through the building, and returns to the chiller at a higher temperature (T_{rw}). The cooling load (Q)—the rate of heat removal—can be expressed as:

$$Q = c_w \dot{m}_{sw} \left(T_{rw} - T_{sw}\right) \tag{1}$$

where c_w is the specific heat capacity of water, and $\dot{m}_{sw}$ is the supply water mass flow rate.

The main control challenge is to determine the optimal values of the supply water temperature (T_{sw}) and mass flow rate ($\dot{m}_{sw}$) to satisfy a given cooling load at minimal total power consumption. The total power consumption of a chiller plant (P) is primarily composed of two parts: the cooling power (P_c) consumed by the chiller itself, and the pump power (P_p) required for water circulation.

$$P = P_c + P_p \tag{2}$$

A. Cooling Power Model

The cooling power depends on multiple factors, including chiller operating conditions, cooling load, and environmental variables. A widely adopted model, such as the DOE-2, expresses chiller power as a product of reference power and several correction factors:

$$P_c = P_{ref} \times CAPFT \times EIRFT \times EIRFPLR \tag{3}$$

where P_{ref} denotes the chiller power at the reference evaporator and condenser temperatures. CAPFT is the ratio curve formulated in (4a) that represents the available capacity as a function of outdoor air dry-bulb temperature T_{oa} and the chilled water supply temperature T_{sw}, EIRFT is a ratio curve formulated in (4b) that represents the full-load efficiency as a function of T_o and T_{sw}, and EIRFPLR is a ratio curve formulated in (4c) that represents the efficiency as a function of the percentage unloading (PLR), which represents the part-load operating ratio of the chiller as formulated in (4d). $a_{1\sim3}, b_{1\sim3}, c_{1\sim3}, d_{1,2}, e_{1,2}$ and $f_{1,2}$ are the regression coefficients of the ratio curves.

$$CAPFT = a_1 + b_1 T_{sw} + c_1 T_{sw}^2 + d_1 T_{oa} + e_1 T_{oa}^2 + f_1 T_{sw} T_{oa} \tag{4a}$$

$$EIRFT = a_2 + b_2 T_{sw} + c_2 T_{sw}^2 + d_2 T_{oa} + e_2 T_{oa}^2 + f_2 T_{sw} T_{oa} \tag{4b}$$

$$EIRFPLR = a_3 + b_3 PLR + c_3 PLR^2 \tag{4c}$$

$$PLR = \frac{Q}{Q_{ref} \times CAPFT} \tag{4d}$$

B. Pump Power Model

In a chiller plant, pumps play a crucial role in the fluid circulation system. They convert electrical energy into the kinetic energy of the fluid, facilitating the flow of water that is supplied to buildings for cooling purposes and subsequently returned to the chillers for re-cooling. Hence, the pump power model can be analytically derived from the fundamental laws of fluid mechanics and thermodynamics, including the work-energy theorem and the relationship between fluid flow and mechanical power. This derivation process results in a pump power model that is expressed as a third-order function of the

979-8-3315-8138-1

water mass flow rate, as presented in equations (5), where $a_4 \sim d_4$ are the coefficients of the corresponding order terms.

$$P_p = a_4 + b_4\dot{m}_{sw} + c_4\dot{m}_{sw}^2 + d_4\dot{m}_{sw}^3 \tag{5}$$

The use of a third-order function accounts for the non-linear behavior of pump operation, where small variations in the water mass flow rate can have a large impact on the required pump power, especially at higher flow rates.

C. Chiller Economic Control

The chiller plant economic control in steady states can be summarized as follows: for a given cooling load (Q) and return water temperature (T_{rw}), determine the optimal supply water temperature (T_{sw}) and mass flow rate ($\dot{m}_{sw}$) within the feasible region that minimize the total power consumption, subject to the thermal balance constraint as formulated in (6)

$$\min_{\dot{m}_{sw}, T_{sw}} P\left(\dot{m}_{sw}, T_{sw}\right) \tag{6a}$$
$$S.t. \quad Q = c_w \dot{m}_{sw}\left(T_{rw} - T_{sw}\right) \tag{6b}$$
$$\dot{m}_{sw}^{lb} \le \dot{m}_{sw} \le \dot{m}_{sw}^{ub} \tag{6c}$$
$$T_{sw}^{lb} \le T_{sw} \le T_{sw}^{ub} \tag{6d}$$

This problem is inherently non-convex due to the nonlinear relationships in the power models and the thermal constraint. Efficient and accurate convex approximations of this problem are therefore needed for real-time control and integration into higher-level DR and building management strategies.

III. Methodology

The optimal control of chillers needs to solve an inherently non-convex optimization problem, due to the nonlinear nature of chiller power consumption and the associated thermal constraints. Existing analytical methods, such as semidefinite programming (SDP), attempt to convexify the problem for tractable solutions but often introduce limitations in model fidelity or restrict applicability in broader optimization frameworks. To address these challenges, we propose a data-driven approach using Input Convex Neural Networks (ICNNs), which can efficiently learn convex approximations of complex system dynamics from data.

A. Input Convex Neural Networks

Input Convex Neural Networks (ICNNs) are a class of neural networks specifically designed to ensure that the output is a convex function of designated input variables. An ICNN comprises both "convex" and "non-convex" input pathways. By constraining the weights on the convex pathway to be non-negative and using convex activation functions (e.g., ReLU), the network output convexity with respect to the selected variables is guaranteed by the closure properties of convex functions, which states that the nonnegative sums of convex functions are also convex and the composition of a convex and convex non-decreasing function is also convex [12], [13]. This property allows for efficient global optimization of the network's output with respect to those variables, making ICNNs well-suited for control and decision-making applications.

Mathematically, the forward propagation on i^{th} layer can be expressed as:

$$u_{i+1} = ReLU\big(W_i u_i + b_i\big) \tag{7a}$$
$$\begin{aligned} z_{i+1} = ReLU\big(&W_i^z\left(z_i \circ \left(W_i^{zu} u_i + b_i^z\right)\right) \\ &+ W_i^c\left(x_c \circ \left(W_i^{cu} u_i + b_i^c\right)\right) \\ &+ W_i^u u_i + b_i^z\big) \end{aligned} \tag{7b}$$

where u and z are hidden layer vectors for the non-convex path and the convex path, respectively. x_c denotes the convex input variables, and W and b denote weights and biases.

By design, the ICNN output is a convex function of x_c, which enables the network to serve as a tractable surrogate for the chiller's power model in downstream optimization tasks.

B. ICNN-based Optimal Chiller Control

Leveraging the convexity property of ICNNs, the original non-convex chiller control problem can be reformulated as a convex optimization, with the ICNN acting as a learned surrogate model.

Fig. 2 depicts the architecture of a 3-layer Input Convex Neural Network (ICNN) employed for optimal chiller control. In this setup, we assign non-controllable variables to the non-convex path, i.e., $x_{nc} = [T_{rw}, Q]^T$, while the controllable variables, which require convexity for optimization, are provided as inputs to the convex path, i.e., $x_c = [T_{sw}, \dot{m}_{sw}]^T$.

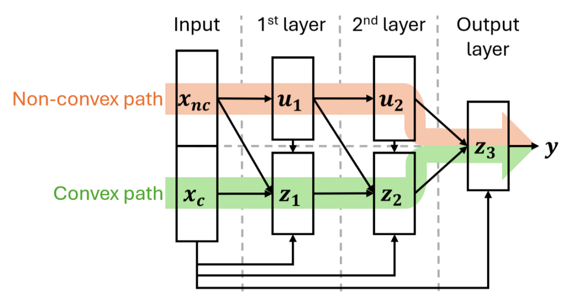

Fig. 2. 3-layer Input Convex Neural Network architecture

Here, we first assume there exists a well-trained ICNN that mirrors the features of chiller power models and thermal balance constraints with perfect parameters θ^*. Then the ICNN is parameterized by (8).

$$y\left(x_{nc}, x_c; \theta^*\right) \tag{8}$$

For any given thermal load Q and return water temperature T_{rw}, the optimal chiller control problem in (6) is equivalent to solving for the x_c^* in (9).

$$x_c^* = \arg\min_{x_c} y\left(x_{nc}, x_c; \theta^*\right) \tag{9}$$

Since $y(\cdot)$ is convex in x_c, global optimality is guaranteed and the solution can be found efficiently with off-the-shelf solvers. Note that, the objective function and constraints of this minimization problem are in the real number domain and can be analytically formulated from the forward propagation of ICNN in (7). The non-differentiability introduced by $ReLU(\cdot)$ can be eliminated through piecewise linearization or mixed integer programming to allow the use of off-the-shelf optimization solvers with global optimality guarantee [14], [15].

979-8-3315-8138-1

C. Physics-informed ICNN Training

The training of the Input Convex Neural Network (ICNN) in this work differs fundamentally from traditional neural network training methods. Conventional neural networks typically minimize the error between the network output and a set of labeled target values. However, the effectiveness of the ICNN-based chiller optimal control relies on training the network such that its convex mapping not only approximates the underlying physics-based or empirical chiller model but also ensures that the minimum point of the ICNN corresponds closely to the true optimal control action for any operating scenario.

An overview of the ICNN training workflow is illustrated in Fig. 3. This approach enables the ICNN to learn a convex mapping that directly supports optimization-based control, facilitating the deployment of the trained model in real-time demand response with guaranteed global optimality.

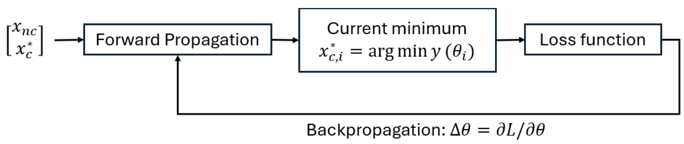

Fig. 3. Physics-informed training of ICNN for chiller DR control

During training, for each data sample, the ICNN is evaluated under the current parameters θ_i, and the corresponding optimal control variables $x_{c,i}^*$ are computed by minimizing the network output with respect to x_c. The loss function is then defined as the distance between $x_{c,i}^*$ and the ground-truth optimal solution x_c^* as formulated in (5).

$$Loss = \mid x_{c,i}^* - x_c^* \mid_2 \tag{10}$$

This loss function formulation ensures that the ICNN is trained to not only approximate the chiller power function, but also to embed the convex optimization process into the learning objective. The training procedure involves an additional physics-informed optimization step for each sample, which distinguishes it from standard supervised learning, making it a physics-informed and optimal control-aware supervised learning process. Network parameters θ are updated via gradient-based backpropagation to minimize the loss across the training set.

IV. Case Study

A. Settings

To evaluate the proposed ICNN-based chiller power flexibility quantification method, a case study is conducted using a digital twin model of a real-world chiller plant that consists of one water-cooled centrifugal chiller rated at 1,000 RT (3.52 MW cooling capacity) and two variable-speed chilled-water pumps. The simulation is performed in Python on a desktop computer with a 2.5 GHz 6-core CPU and 32 GB of memory.

The coefficients used are summarized in Table I, where the DOE-2 model parameters are derived from manufacturer data sheets and the pump power coefficients are obtained from performance testing sheets. The air loop thermal variations are directly captured by the granular sampling of thermal load Q within the feasible region, hence the outside air temperature T_{oa} is fixed at 30 °C to reduce the model dimension without loss of accuracy and generality.

TABLE I
Model parameters

	Parameter	Value	Parameter	Value
Pump power coefficients	a_4	2498	b_4	216.8
	c_4	7.037	d_4	0.393
Chiller power coefficients	a_1	0.9830684	a_2	0.6228513
	a_3	0.03295553	b_1	0.03676199
	b_2	-0.008700783	b_3	0.9114398
	c_1	5.307995e-05	c_2	0.0009719796
	c_3	0.05565265	d_1	-0.004828624
	d_2	0.005184584	e_1	-4.755963e-05
	e_2	0.0004951699	f_1	-0.0002377982
	f_2	-0.0014448555		
Other parameters	P_{ref}	91918.5 W	Q_{ref}	383300 J
	c_{water}	4184 J/(kg°C)		

B. Numerical Results

The case study considers a range of realistic operating conditions for the chiller plant to ensure the evaluation captures both typical and off-design scenarios. The thermal load (Q) varies from $10^5 J$ to $10^6 J$ in increments of $1000J$, while the return water temperature (T_{rw}) ranges from 13.5 °C to 16 °C in increments of 0.1 °C. For each combination of load and return water temperature, the supply water temperature (T_{sw}) and chilled-water mass flow rate ($\dot{m}_{sw}$) are treated as the decision variables to be optimized. The proposed ICNN-based method is evaluated against two benchmarks: the ground truth results obtained via exhaustive brute-force search over the decision variable space, and the state-of-the-art semidefinite programming (SDP) convexification method.

1) Accuracy: The accuracy of supply water temperature control is selected to be the evaluation metric for chiller plant control optimality because water mass flow rate is automatically adjusted by PI controllers in accordance with water supply temperature to satisfy thermal load Q. Hence supply water temperature control reflects the optimality of the chiller plant economic operations. The brute-force search solution indicated by the green surface in Fig. 4 denotes the ground truth optimal control for all possible scenarios within the feasible operation region. The red surface denotes the solutions given by the SDP method, and the blue surface consists of solutions from the ICNN-based method. It can be observed that the SDP method is more aggressive on supply water adjustment, while the ICNN surface has a closer slope to the ground truth surface. Although the SDP method yields a more accurate control in the high Q and low T_{rw} quadrant of the feasible region, its overall L2 distance to the ground truth is 0.9971 m^3/s. The ICNN method yields a higher control accuracy with an overall distance to the ground truth of 0.5125 m^3/s.

2) Time Efficiency: Besides the superiority in accuracy, the ICNN-based chiller control also has the advantage in terms of time efficiency, especially for a large number of scenarios. The run time of both brute-force search and SDP methods is linear to the number of scenarios because each scenarios is solved in sequence. However, the neural network can output the solution for all scenarios altogether. The increased dimension of tensors

979-8-3315-8138-1

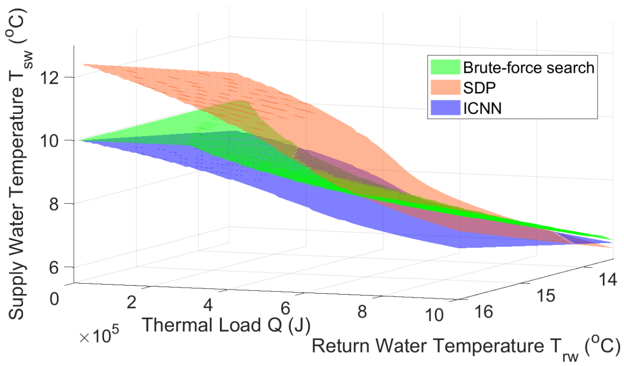

Fig. 4. Comparison of control accuracy

introduces minor run-time increases, but it is far less than the complexity of sequential methods. Table II compares the run time of the three methods over 891 random test scenarios. SDP is 10 times faster than the brute-force search due to its convexity, while ICNN is over 100 times faster than SDP as all 891 scenarios can be input to ICNN in parallel as a tensor and their control variables can be optimized independently through gradient descent methods. It is worth noting that the single-scenario runtime of SDP is lower than that of ICNN. The computational efficiency advantage of ICNN over SDP becomes more apparent as the number of scenarios increases.

TABLE II
SOLUTION TIME CONSUMPTION

	Brute-force search	SDP	ICNN
Mean (sec)	172.8888	17.5644	0.1141
Std	122.1759	2.0664	0.0390

C. Training Convergence

The ICNN model for chiller plant control was trained with 150 neurons in the first layer and 200 neurons on the second layer. The training process was performed with 100 epochs to ensure sufficient convergence and model stability. As illustrated in Fig. 5, the training loss demonstrates a clear convergence trend, steadily decreasing over the course of training. This behavior indicates that the network progressively learned the underlying mapping between the chiller operating variables and their optimal control targets while maintaining the convexity property required for optimization. The final loss reached a low and stable value ($< 10^{-5}$), confirming that the ICNN achieved good fitting accuracy and generalization capability for use in the optimal chiller control framework.

V. CONCLUSIONS

In this paper, we propose an Input Convex Neural Network (ICNN) framework for chiller demand response without loss of energy efficiency. By ensuring convexity with respect to controllable variables, the ICNN enables globally optimal and computationally efficient control decisions. Case studies using a digital-twin chiller model demonstrated that the proposed method achieves two times higher accuracy and over 100 times

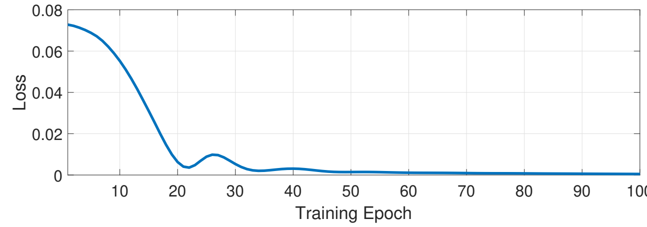

Fig. 5. Training loss convergence

faster computational efficiency than SDP-based approaches. The results highlight the ICNN's potential as a scalable and efficient tool for real-time chiller control and flexibility quantification in grid-interactive building operations. In future work, we will expand the time horizon and integrate the ICNN model into a daily demand response case under a typical design day outdoor air temperature and grid signals to further validate its comprehensive performance in practice.

REFERENCES

[1] L. G. Meegahapola, S. Bu, D. P. Wadduwage, C. Y. Chung, and X. Yu, "Review on oscillatory stability in power grids with renewable energy sources: Monitoring, analysis, and control using synchrophasor technology," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 519–531, 2021.

[2] G. Tian and Q. Z. Sun, "A stochastic controller for primary frequency regulation using on/off demand side resources," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 4141–4144, 2023.

[3] Y. Li, M. Han, Z. Yang, and G. Li, "Coordinating flexible demand response and renewable uncertainties for scheduling of community integrated energy systems with an electric vehicle charging station: A bi-level approach," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2321–2331, 2021.

[4] F. Feng and Z. O'Neill, "A real-time platform for assessment of chiller-side demand response strategies," *ASHRAE Transactions*, vol. 125, pp. 305–312, 2019.

[5] Z. Zhang, Y. Huang, H. Chen, Q. Huang, and W.-J. Lee, "A novel hierarchical demand response strategy for residential microgrid," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3262–3271, 2021.

[6] W. Wu, S. Lei, Q. Z. Sun, and J. L. Mathieu, "Energy efficiency of commercial hvac-based virtual batteries for load shifting," *IEEE Transactions on Power Systems*, pp. 1–14, 2025.

[7] W. Wang, G. Tian, Q. Z. Sun, and H. Liu, "A control framework to enable a commercial building hvac system for energy and regulation market signal tracking," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 290–301, 2023.

[8] M. M. Shepherd and S. Mohagheghi, "Demand flexibility: A review of quantification methods, models, and required data," *Applied Sciences*, vol. 14, no. 23, p. 11276, 2024.

[9] G. Tian, Q. Z. Sun, and W. Wang, "Real-time flexibility quantification of a building hvac system for peak demand reduction," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3862–3874, 2022.

[10] J. Zhu, J. Niu, Z. Tian, R. Zhou, and C. Ye, "Rapid quantification of demand response potential of building havc system via data-driven model," *Applied Energy*, vol. 325, p. 119796, 2022.

[11] M. G. Abdolrasol, S. S. Hussain, T. S. Ustun, M. R. Sarker, M. A. Hannan, R. Mohamed, J. A. Ali, S. Mekhilef, and A. Milad, "Artificial neural networks based optimization techniques: A review," *Electronics*, vol. 10, no. 21, p. 2689, 2021.

[12] B. Amos, L. Xu, and J. Z. Kolter, "Input convex neural networks," in *International conference on machine learning*. PMLR, 2017, pp. 146–155.

[13] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[14] R. Anderson, J. Huchette, W. Ma, C. Tjandraatmadja, and J. P. Vielma, "Strong mixed-integer programming formulations for trained neural networks," *Mathematical Programming*, vol. 183, no. 1, pp. 3–39, 2020.

[15] M. Fischetti and J. Jo, "Deep neural networks and mixed integer linear optimization," *Constraints*, vol. 23, no. 3, pp. 296–309, 2018.

979-8-3315-8138-1