Science and Technology for the Built Environment (2016) **22**, 733–750

ISSN: 2374-4731 print / 2374-474X online
DOI: 10.1080/23744731.2016.1197716

Taylor & Francis
Taylor & Francis Group

General approaches for determining the savings potential of optimal control for cooling in commercial buildings having both energy and demand charges

JIE CAI[1,*], JAMES E. BRAUN[1], DONGHUN KIM[1], and JIANGHAI HU[2]

[1]*Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN 47907, USA*
[2]*School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA*

This article presents a general approach for determining maximum monthly energy cost savings associated with optimal supervisory control for cooling in commercial buildings in the presence of utility rates that include both demand and time-of-use energy charges. The resulting tool has a month-long time horizon because of the nature of demand changes and is only useful for benchmarking the performance of simpler and shorter-term demand response and optimal control approaches. Attempts to solve this optimization problem using a centralized formulation failed and, therefore, the benchmarking problem was formulated as a dynamic optimization problem within a multi-agent control framework so that the monthly optimization problem is segmented into several sub-problems where each sub-problem involves system optimization for a shorter period of time, for example, a 1-day period. The daily-scale optimization involves determination of trajectories of zone set-point temperatures that minimize an integral cost function with a demand cost constraint determined by a demand agent. In order to further simplify the daily optimization, dynamics associated with the cooling system are neglected and optimal control of the cooling system is assumed to be based on heuristics determined through upfront analysis. This approach leads to significantly reduced computational requirements and more importantly, it provides guaranteed convergence in the multi-agent optimization. Results for a single-zone building case study are presented to illustrate the potential cost savings. In addition, a simpler and more practical short-term optimization approach with a demand-limiting heuristic is proposed and evaluated in comparison to the benchmarking optimization results for this case study and achieves most of the potential savings.

Introduction

Demand charges and time-of-use (TOU) pricing exist in most utility markets as an incentive to reduce the peak power demand and thus, the capital investment on the power plants in a grid. Buildings account for more than 70% of the total electricity consumption in the United States (U.S. EPA 2009), thus control strategies for buildings that consider both energy and peak demand costs are beneficial for both the grid and building end users.

For existing buildings, one of the primary opportunities for reducing peak demand and shifting energy use to lower cost periods is through the use of building thermal mass. This involves adjustment of building zone temperature set-points where the building is pre-cooled during off-peak and low-demand periods and then the temperature set-points are adjusted upwards during the on-peak period to reduce cooling power. The range of set-points is limited by comfort bounds but significant opportunities exist because of the large and distributed thermal mass inherent in commercial buildings. Wang et al. (2014) provided a relatively comprehensive review of building demand response strategies that could be employed within a smart grid scenario including a number of building load shifting strategies. For example, Keeney and Braun (1997) demonstrated a 25% peak cooling load reduction by using a pre-cooling strategy in a large office building with two identical wings. With the reduced cooling load, it was found that a chiller that had failed did not need to be replaced leading to a potential $500,000 cost savings. In addition, simulation results showed a 15% reduction in the chiller electricity energy cost with a simple pre-cooling strategy. Xu et al. (2004) carried out a pre-cooling test within an office building where the chiller power was reduced by 80–100% for a 3-hour peak period under mild weather conditions and the savings were dependent on the amount of pre-cooling (limited and extended pre-cooling) and weather conditions. Xu and Haves (2006) performed an extended test during hot summer days where the peak demand reduction only lasted for 2 hours and

Received January 8, 2016; accepted April 11, 2016
Jie Cai, PhD, is a Postdoctoral Research Associate. **James E. Braun, PhD, PE,** Fellow ASHRAE, is a Professor. **Donghun Kim, PhD,** is a Postdoctoral Research Associate. **Jianghai Hu, PhD,** is an Associate Professor.
*Corresponding author e-mail: cai40@purdue.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhvc.

a significant rebound was observed during the third hour of the peak period. To avoid the power rebound, Lee and Braun (2008) proposed a model-based demand-limiting control strategy where a data-driven model is used to predict the building thermal behavior in the demand-limiting period. Also, an optimal space temperature set-point trajectory is sought using the model so that the cooling power is maintained at the lowest constant level in the demand-limiting period. As demonstrated in Lee and Braun (2006), the obtained optimal set-point trajectory led to larger peak load reduction compared to a step or linear setup trajectory. Xue et al. (2015) proposed a fast demand response control approach relying on a pre-cooling strategy that controlled the cooling plant directly. Compared to the space temperature set-point control strategies, the proposed approach is able to provide fast demand response to grid requests and may be better suited for building demand management in a smart grid scenario.

These previous studies did not explicitly consider the tradeoff between cooling energy and demand costs. Peak load reduction comes at a price of increased cooling energy since the lower zone temperatures in the pre-cooling periods induce larger heat gains from the ambient through heat conduction and ventilation. Therefore, a poorly devised demand-limiting strategy could lead to total cooling energy costs that are actually greater than those for a non-demand-limiting strategy. Although extensive research work has been done on reducing demand (or energy) cost, there is little work on reducing the total electricity cost. Braun (1990) formulated two dynamic optimization problems for building energy cost and peak electrical demand minimizations, respectively, both evaluated over individual days with the beginning and end states of the building forced to be the same. The performance of the optimal solution was assessed in terms of energy cost savings and peak demand reduction with respect to different weather conditions, HVAC system characteristics, and electricity rate structures. Although the parametric study results highlighted some trade offs between energy and demand costs, this study did not consider the total cost reduction explicitly. Ma et al. (2012) proposed an economic model predictive control (MPC) formulation that considered the sum of energy and demand costs explicitly in the optimization cost function. With simplified building and equipment models, the MPC problem was formulated as a linear programming problem and the optimal solution was implemented in a shrinking horizon scheme. However, it only considered daily system optimization which posed unrealistic weighting between energy and demand costs.

The challenge of studying the tradeoff between energy and demand costs lies in the month-long time scale associated with assessing demand charges, since they are based on peak demands over an entire monthly billing period. In a practical sense, this makes optimal operation impossible when considering both energy and demand costs. However, it is still useful to understand the maximum savings potential associated with optimal control over monthly billing periods in order to provide a benchmark for evaluating simpler control approaches. Toward this goal, Henze et al. (2008) proposed a hierarchical optimization solution where an outer loop was used to enumerate the target monthly demand threshold, while the inner loop carried out sequential daily optimizations constrained by the target demand threshold for a month period. The daily optimization concerned minimizing the energy cost plus a demand penalty with artificially imposed weighting. Thus, the obtained daily solutions are energy-cost-prioritized and could be sub-optimal due to the unrealistic weighting. Applying a similar idea, Sun et al. (2010) performed multiple simulations for a demand limiting strategy that was implemented for each day of the month with a given monthly peak threshold where the thresholds were adjusted from one simulation to the next to identify the best threshold value. However, the daily control followed a demand limiting strategy and not a cost minimization strategy, which could be sub-optimal. Note that the two approaches by Henze et al. (2008) and Sun et al. (2010) both involved unrealistic cost formulations in the daily optimization sub-problems due to difficulties in assessing demand cost on a daily time scale. In addition, solutions obtained by a series of daily optimizations could be very sub-optimal for the overall monthly operation.

This article presents a general approach for determining zone temperature set-point trajectories over a month that minimize total monthly utility costs in the presence of general utility rate structures that include demand and TOU energy charges. In order to achieve a solution to this complex problem, a multi-agent based approach is proposed where the original optimization problem is fragmented into multiple sub-problems with individual agents solving an optimization for a shorter period of time (e.g., a day) with demand constraints and the demand cost optimization assigned to another agent. Solving the decentralized problem is much more tractable than a centralized solution and can be implemented in parallel while a coordination mechanism is needed to ensure the interfacing conditions match between neighboring agents or sub-problems. Although solving a month-long optimization problem is not practical for real-time implementation because of forecasting and computational limitations, the proposed approach is useful in understanding cost savings opportunities and as a benchmarking tool for evaluating the performance of simpler and more practical control strategies. In this study, benchmarking results obtained with this tool were used to evaluate a more practical short-term MPC formulation that considers the tradeoff between the energy and demand costs by utilizing a simple peak demand cost limit heuristic along with a daily energy cost minimization. The short time-horizon MPC was tested with different configurations and compared to the benchmark. This simpler approach achieves most of the cost savings and could be the basis for a practical control implementation or as part of a tool for evaluating additional simplifications.

Case study description and the component models

A single-zone building case study was used to test the proposed approaches and the system schematic is shown in Figure 1. The building zone is actually a graduate student office at Purdue University that is served by an individual air-handling unit that utilizes campus chilled water. In order to consider both the air-handling unit and cooling plant, a dedicated air-cooled chiller (a 10-ton unit with a rated energy efficiency ratio (EER) of 10) was assumed and modeled. The chiller capacity

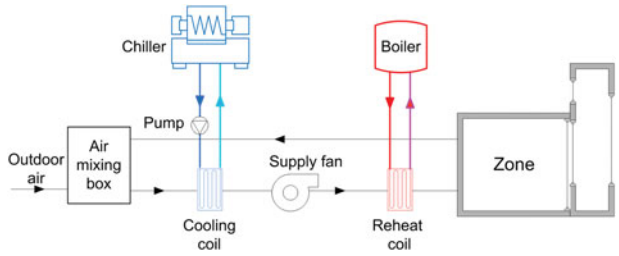

Fig. 1. Case study schematic.

was scaled down by a factor of 4 to match the building loads. A variable speed pump is used to deliver variable chilled water flow (nominal flow of 30 GPM and total head of 23 ft) to the cooling coil in order to maintain a set-point supply air temperature (T_{sa}). The inlet chilled water temperature ($T_{w,in}$) to the cooling coil is controlled to a fixed value via chiller capacity modulation. Zone air temperature is regulated by varying the supply fan speed and the entering airflow (nominal flow of 2600 cubic feet per minute (CFM) and external static pressure of 1.1 in.W.C.). However, there is a minimum airflow due to ventilation and a lower speed limit for the variable frequency drive. When the controlled airflow reaches the minimum level and the zone cooling capacity is larger than the load, then hot-water reheat (Q_{ht}) is enabled to maintain the room temperature set-point. A simple boiler model with fixed efficiency is used to characterize the reheat energy cost. Return air from the building space is partially circulated back to a mixed-air chamber where it is mixed with outdoor fresh air. The mixed air then passes through the cooling coil to exchange heat with chilled water.

The concerned office space has a floor area of 1000 sq ft (93 m^2) and houses a maximum of 20 students. Based on average occupancy, the peak internal heat gains are about 2.5 kW (8530 Btu/h) with 1.2 kW (4095 Btu/h) from lighting and 1.3 kW (4436 Btu/h) contributed by occupants and their electrical device usage. Internal heat gain variations due to plug loads and lighting, were obtained from measurements available from the building management system (BMS). Since the space serves as a student office, a typical occupancy profile has occupied periods starting from 10 am and ending at 10 pm and the internal heat gain profiles closely match the occupancy pattern. The building envelope has good insulation in both the external and internal walls so its thermal couplings to the ambient and other offices are weak. However, the space is configured with a south-facing double-skin facade that introduces significant sensitivity of the indoor space temperature to solar radiation. During evening times, the double facade also induces some heat exchange between the office space and the ambient. The building thermal mass is mainly from the concrete floor and inside furniture. Some step tests have shown that the building construction has a significant thermal buffering effect that has good potential for load shifting.

Building envelope model

A simplified thermal network model has been developed for the building envelope where the model parameters were estimated from on-site measurements. The model incorporates two three-resistance-two-capacitance wall branches representing a floor and an external wall, respectively. The double-skin facade space is modeled as an individual zone with two single-resistance branches representing the thermal interactions through windows to the office space and to the ambient, respectively. The model details and validation can be found in Appendix C of Cai (2015) or Cai and Braun (2015) and the obtained model can be formulated under a discrete-time state-space representation:

$$\mathbf{x}[i+1] = \mathbf{A}\mathbf{x}[i] + \mathbf{B}_w\mathbf{w}[i] + \mathbf{B}_u\, Q_z[i]$$
$$y[i] = \mathbf{C}\mathbf{x}[i] = x_z[i]$$

where $\mathbf{w}$ is a vector of uncontrollable inputs or disturbances including outdoor conditions and internal heat gains due to occupants and equipment, Q_z is the sensible cooling or heating provided to the space by the HVAC system and is the only controllable input, and y or x_z is the zone air temperature.

Chiller model

An empirical model was constructed for the air-cooled chiller where a quadratic correlation is used to calculate the chiller cooling capacity based on the leaving water temperature (T_{lw}) and outdoor air temperature (T_{oa}):

$$Cap_{rate} = a_1 + a_2 T_{oa} + a_3 T_{oa}^2 + a_4 T_{lw} + a_5 T_{lw}^2 + a_6 T_{lw} T_{oa}$$

The chiller power, Pow_{rate}, is calculated using the same empirical form, though with a different set of parameters a_1 to a_6. Under part load conditions, another quadratic correlation is used to scale the power based on the load ratio, LR, which is defined as the ratio of the actual load ($Q_{cl,tot}$) to the rated capacity (Cap_{rate}):

$$Pow_{PL} = \left(b_1 + b_2 LR + b_3 LR^2\right) Pow_{rate}.$$

Parameters a_1 to a_6 (two parameter sets, one for the capacity correlation and the other for the power correlation) and b_1 to b_3 were estimated via linear regression applied to catalog data. Good model fits were obtained with a maximum error of 3.4% for the power prediction.

Supply air fan and chilled water pump model

A cubic correlation to the airflow/water flow is used to calculate the supply fan/chilled-water pump power. Actual measurements were used to train the correlation parameters. Performances of the estimated models can be found in Appendix C of Cai (2015) where two sets of parameters c_0 to c_3 were obtained for the pump and fan, respectively.

$$Pow_{pump/fan} = c_0 + c_1 m_{w/a} + c_2 m_{w/a}^2 + c_3 m_{w/a}^3$$

Cooling coil model

A quasi-steady-state model was developed for the cooling coil from on-site measurements. A moving boundary modeling

approach, adapted from Braun (1988), is adopted where the transition point of the coil from dry to wet is determined iteratively with air and chilled-water energy balances. Dry and wet coil heat transfer coefficients are calculated based on correlations in terms of air and water mass flow rates where the correlation parameters were estimated from measurements. The model details can be found in Appendix C of Cai (2015) and the obtained model is of the form:

$$[Q_{cl,sen},\ Q_{cl,tot}] = \mathrm{ClCoil}\,(T_{ma},\ RH_{ma},\ T_{w,in}, m_a, m_w)$$

where $Q_{cl,sen}$ and $Q_{cl,tot}$ are the sensible and total capacities of the cooling coil; T_{ma} and RH_{ma} are the coil inlet air temperature and relative humidity; $T_{w,in}$ is the coil inlet water temperature which is the same as the chiller leaving water temperature T_{lw}; m_a and m_w are the air and water flow rates, respectively.

Cooling plant control optimization and near-optimal heuristics

Integration of the HVAC component models provides an overall HVAC system model for performance evaluation. Dynamics of the HVAC system components are neglected and perfect feedback control is assumed in response to supervisory set-points. The primary dynamics are associated with energy storage within the building structural materials. For this simple case study system, the free supervisory control variables are set-point temperatures for the zone air (thermostat), supply air from the cooling coil, and chilled water supply and perfect feedback control is implemented as follows: (1) zone supply air flow (when greater than the minimum) or reheat (when zone supply air is at the minimum) is adjusted to maintain the specified zone temperature set-point; (2) chilled water flow rate is adjusted to maintain the specified supply air temperature set-point; and (3) chiller cooling rate is adjusted to maintain the specified chilled water supply temperature set-point. In the optimization framework presented in the next section, zone air temperature set-points are free variables in a dynamic optimization that yields a trajectory of values whereas the cooling plant control variables are assumed to be determined using a static optimization or from heuristics determined from an upfront analysis. For this simple case study, the chilled water set-point was assumed to be constant, whereas a heuristic for resetting the supply air temperature set-point was determined using the cooling plant model as described in the remainder of this section. Note that the chilled water temperature set-point was not considered as an optimization variable here because the case study building receives chilled water at a constant temperature of 8.5°C from the Purdue central cooling plant and the obtained results in the present study with the same settings can be used as a fair benchmark for a planned experimental test of the proposed short-term MPC strategy.

Define $Q_{sen,net}$ as the cooling coil net capacity which equals the coil sensible capacity minus the heat dissipated by the fan. This net capacity is the effective cooling rate of the AC system. Figure 2 shows total HVAC power variations with respect to airflow under four example operating conditions. The chilled water set-point, which is the cooling coil inlet chilled water temperature, was assumed to be a fixed value of 8.5°C (47.3°F). The airflow can vary between 1200 (0.67 kg/s) to 2600 CFM (1.44 kg/s).

Figure 2c shows a case under dry coil conditions. To achieve a specified net capacity (2 kW or 6824 Btu/hr in the plotted case), higher airflow requires more fan power and thus, requires more chiller power to compensate for the heat dissipated from the fan. This can also be observed in the coil heat exchange rate variations shown in Figure 2c (the sensible and total rate curves are overlapped). The chilled water pump power is small compared to both chiller and supply fan powers. As a consequence, the total power increases monotonically with airflow for this particular system.

Under wet coil conditions, shown in Figure 2a, coil sensible capacity increases with increasing airflow to offset the fan heat. However, less dehumidification (latent capacity) occurs with higher airflows due to higher coil surface temperature and this latent capacity decrease is more significant than the sensible capacity increase. As a consequence, the coil total capacity and chiller power decrease with increasing airflow. A slight decrease can also be observed in the pump power because less chilled water is needed. However, the increase in fan power with air flow is more significant than the decrease in chiller power so that total power increases with airflow although the curve is relatively flat for low air flows.

Similar trends are seen in the other two subplots for cases where the coil changes from dry to wet at some intermediate airflow. As a result of these trends, a near-optimal control heuristic for this particular system is to maintain the airflow at the minimum level and vary the chilled water flow for capacity modulation. It should be noted that this particular heuristic might be unique to this system. Other systems that have a lower ratio of fan to chiller power could have a somewhat different tradeoff and optimization result.

The cooling plant control heuristic was used to determining a simpler model for the plant power consumption in terms of load and ambient temperature that was then utilized within the longer-term cost optimization formulation presented in the next section. By virtue of the simple heuristic, the coil sensible cooling rate is formulated as:

$$Q_{cl,sen} = LR \cdot Cap_{rate}(\underline{T_{oa}}) \cdot SHR.$$

where SHR is ratio of the sensible to the total cooling rate. Note that only sensible loads were considered in this study because the simulation periods in the case study had very dry weather conditions and moisture balance model was not considered. Thus, SHR was assumed to be unity within this study. A building moisture model could be readily incorporated within the model to handle more humid climate conditions. Since the coil inlet chilled water temperature is fixed, the chiller capacity is a function of the outdoor temperature T_{oa} only. Defining Pow_{ch} as the total power consumed by the chiller and chilled water pump, 4th-order convex polynomial fits were obtained that correlate Pow_{ch} to the total load ratio LR at different outdoor air temperature:

$$Pow_{ch} = \mathrm{Pow}_{ch}\left(LR, \underline{T_{oa}}\right).$$

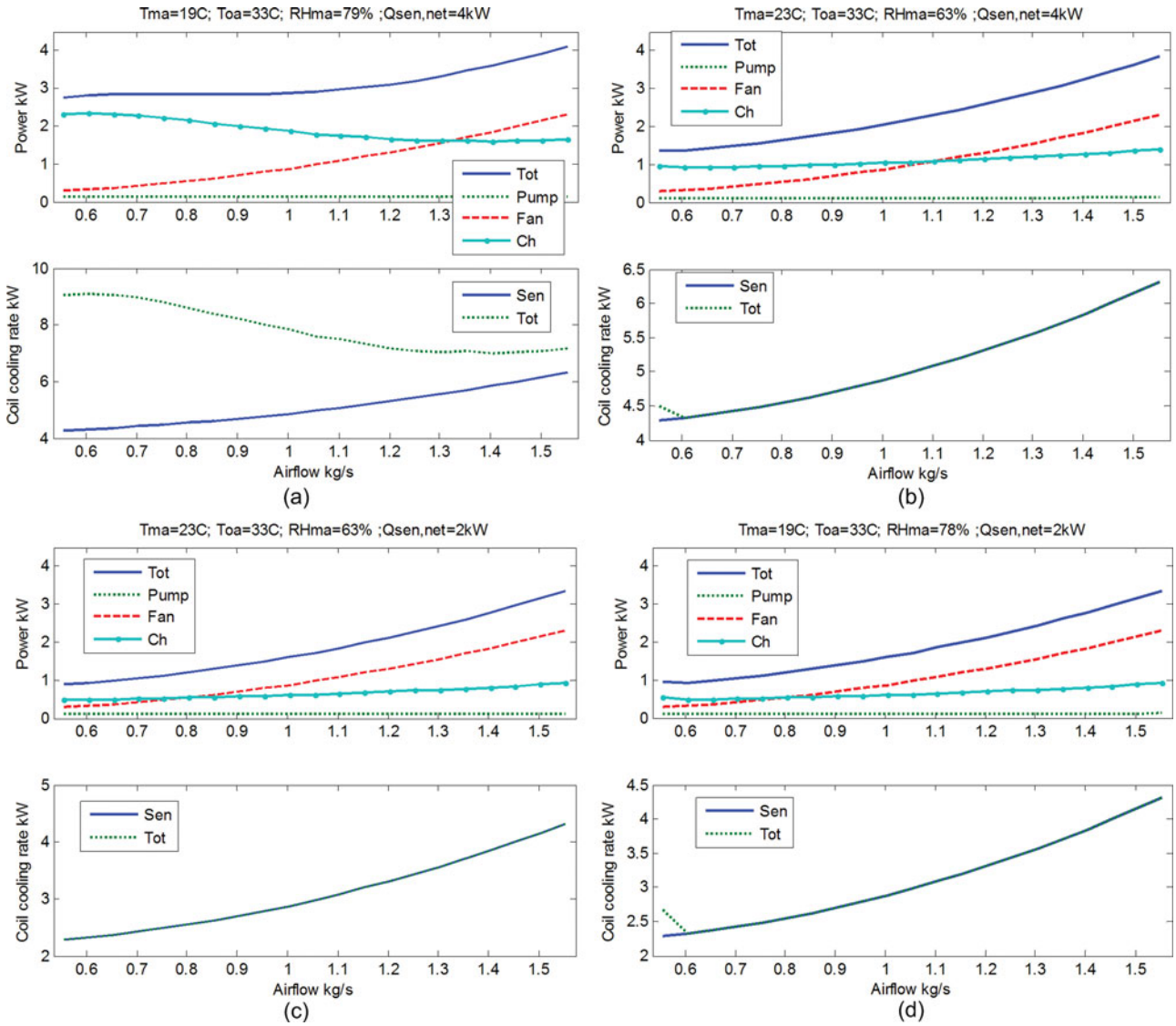

Fig. 2. HVAC model integrated performance under four example operating conditions.

Different polynomial fit coefficients were determined for different outdoor air temperatures, T_{oa}, with values ranging from 10°C to 45°C with 0.5°C increments, using an offline analysis. Note that a two-dimensional fit of Pow_{ch} with respect to LR and T_{oa} did not work well. The building control optimization then uses the model that is closest to the outdoor temperature at any given time to determine the plant power consumption with the load determined from the building and coil models in terms of the current and previous zone temperature set-points and other external driving conditions. Since the outdoor air temperature is a boundary condition, this approach does not introduce any numerical complexity in solving the optimization problem. Cai et al. (2016b) showed that obtaining 4th-order convex polynomial fit in terms of load is itself a convex problem that can be solved with a convex programming package. Figure 3 shows convex shape associated with the variation of Pow_{ch} with respect to LR at an example outdoor air temperature (T_{oa}) of 33°C (91.4°F). This convex shape can also be observed under other tested values for T_{oa} and because of this, good fits were obtained with $R^2 > 0.99$ for most of the tested outdoor air temperatures. Because of the convexity in the proposed model formulation, the dynamic optimization problem formulated in the following section is also convex, which guarantees convergence in the multi-agent optimizations. Note that the calculation of the pump power requires coil air side conditions. However, since the pump power is small compared to the chiller power, pump power calculations are only performed under a nominal air condition.

Utility rates

The electricity tariffs shown in Table 1 are used to analyze the energy cost savings potential under different control strategies. The considered tariffs include different electricity energy rates for three rating periods: on-peak, mid-peak, and off-peak periods. While the proposed approach in the subsequent section is able to naturally handle multiple demand pe-

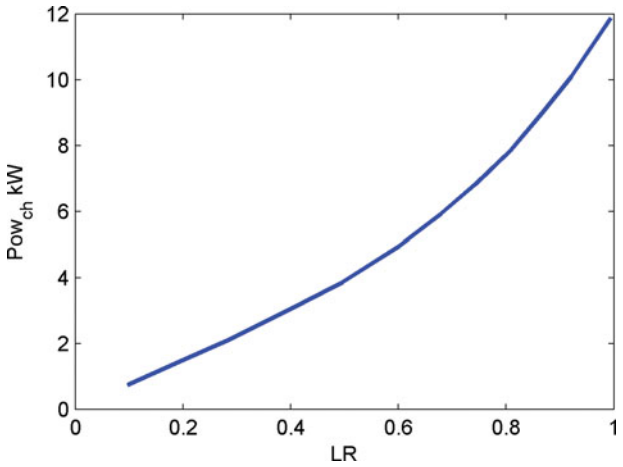

Fig. 3. Example variation of total power of chiller and pump with respect to load ratio.

riods, only one single-demand rate is considered here, i.e., $Nd = 1$, that is charged by the anytime monthly peak. Including multiple demand periods would not increase the computation complexity much but would complicate the overall analysis. The natural gas price is assumed to be a constant value of 0.03$/kWh-heating.

Mathematical formulation for the monthly cost optimization problem

Centralized formulation

A centralized optimization problem is first presented to illustrate the complexity associated with a long time-scale optimization in the presence of complicated utility rate structures that may include multiple TOU and demand rate periods. The optimization problem is formulated for each billing cycle, for example, 1 month, as shown in Equations 1 and 2. The cost function is shown in Equation 1 which corresponds to the monthly utility bill including gas and electricity costs. Np is the number of the optimization time steps ($Np = 720$ for a 30 day optimization with 1-h time steps). r_e, $r_{DC,l}$, and r_{gas} are the electricity energy rate ($/kWh), electricity demand charge rate ($/kW), and gas price ($/kWh), respectively. The electricity energy rate can vary with the time of the day (e.g., on-peak and off-peak rates) and is time indexed. Demand charges under different demand periods (e.g., on-peak, mid-peak, off-peak, anytime) are considered where P_l is the set of the time indices within demand period l. The number of periods having different demand charges is denoted by Nd. However, for ease of analysis, only an anytime-peak demand rate was considered in the case study so $Nd = 1$ and $P_1 = [1, \ldots, Np]$. The gas price is assumed constant with respect to time. Non-HVAC energy consumption is denoted by Pow_{oth}.

$$\min_{\left\{\begin{array}{l}\mathbf{x}[i], LR[i-1], Q_{ht}[i-1], m_{OA}[i-1]\ldots \\ |i=2,\ldots,Np+1\end{array}\right\}} \left\{\begin{array}{l} \sum_{i=1}^{Np}\left\{\left(\mathrm{Pow}_{ch}\left(LR[i], \underline{T_{oa}[i]}\right) + Pow_{oth}[i]\right)\cdot r_e[i] \right. \\ \qquad \left. + Q_{ht}[i]\cdot r_{gas}\right\} \\ + \sum_{l=1}^{Nd}\left\{\max_{i\in P_l}\left(Pow_{ch}[i] + Pow_{oth}[i]\right)\cdot r_{DC,l}\right\} \end{array}\right\} \tag{1}$$

subject to:

$$\left.\begin{array}{rl} \mathbf{x}[i+1] &= \mathbf{A}\mathbf{x}[i] + \mathbf{B}_w\mathbf{w}[i] + \mathbf{B}_u Q_z[i] \\ Q_z[i] &= -LR[i]\cdot \underline{Cap_{ch}\left(T_{oa}[i]\right)}\cdot SHR \\ & \quad + Q_{ht}[i] + m_{OA}[i]\left(T_{amb}[i] - T_{RA,nom}\right) \\ T_{z,lb}[i+1] &\le x_z[i+1] \le T_{z,ub}[i+1] \\ m_{OA,\min} &\le m_{OA}[i] \le m_{OA,max} \\ 0 &\le LR[i] \le 1 \\ 0 &\le Q_{ht}[i] \le Q_{ht,max} \end{array}\right\} i = 1, \ldots, Np \tag{2}$$

Equation 2 lists all the optimization constraints. The first constraint comes from the discrete-time dynamic model for building envelope presented in the component model section. The second constraint calculates the net sensible cooling rate by considering the different sources of energy from cooling coil, hot-water reheat, and ventilation. Again, the SHR is assumed to be unity since only the sensible performance is of concern in this study. Note that this constraint is originally bilinear since the return air temperature T_{RA} is in the design variable $\mathbf{x}$. However, this is simplified to a linear constraint by using a fixed nominal return air temperature $T_{RA,nom}$. This is a reasonable assumption since the zone air temperature is typically regulated within a narrow range for the sake of occupant comfort. This treatment preserves convexity in the problem formulation without compromising much accuracy. Simulation results have shown that this simplification leads to less than 1% difference in the daily energy consumption. The third time-varying interval constraint is to ensure thermal comfort for the occupants and x_z is the element of the state vector corresponding to zone air temperature. The upper and lower bounds ($T_{z,lb}$ and $T_{z,ub}$) can vary depending on the occupancy of the room. In the fourth constraint, a minimum outdoor air intake is

Table 1. Summer TOU tariffs with demand charge.

Rate periods	Electricity price ($/kWh)	Hours	Demand charge
On-peak period	0.108	Noon–6 PM	$19.2/kW*
Mid-peak period	0.089	8 AM–noon; 6 PM–11 PM	$19.2/kW*
Off-peak period	0.064	All other hours	$19.2/kW*

*anytime peak demand.

imposed due to the ventilation requirement and the maximum outdoor air intake is set to the total airflow entering the room. The remaining constraints are due to capacities of the specific equipment.

Distributed formulation

In the centralized formulation, there are nearly 7000 optimization variables along with a large number of constraints which is computationally intractable due to the high dimensionality. Attempts to solve this centralized problem were not successful on a workstation computer with an Intel i5 CPU and 6GB RAM due to the large memory requirement. In order to find a practical solution, a distributed problem is formulated under a multi-agent scenario where a "day-based" agent is assigned for energy cost minimization for each day and one "demand" agent is used for the total demand cost reduction of the whole month. A day-based agent coordinates with its neighbor as well as the demand agent through information exchange.

Day-based agent

For day j, the energy cost minimization sub-problem is formulated as:

$$\min_{\left\{\begin{array}{l}\mathbf{x}^j[1],\,\mathbf{x}^j[i],\,LR^j[i-1],\,Q^j_{ht}[i-1],\,m^j_{OA}[i-1],\,...\\ Pow^j_{max,l}|i=2,...,Np_j+1,\,l=1,...,Nd\end{array}\right\}} \sum_{i=1}^{Np_j}\left\{\left(\mathrm{Pow_{ch}}\left(LR^j[i],\,\underline{T^j_{oa}[i]}\right)+Pow^j_{oth}[i]\right)\cdot r^j_e[i] + Q^j_{ht}[i]\cdot r_{gas}\right\} \tag{3}$$

subject to:

$$\left.\begin{array}{rl}\mathbf{x}^j[i+1] &= \mathbf{A}\mathbf{x}^j[i]+\mathbf{B}_w\mathbf{w}^j[i]+\mathbf{B}_u\,Q^j_z[i]\\ Q^j_z[i] &= -LR^j[i]\cdot\underline{Cap_{ch}(T^j_{oa}[i])}\cdot SHR+\,Q^j_{ht}[i]...\\ &\quad +m^j_{OA}[i]\left(T^j_{oa}[i]-T_{RA,nom}\right)\\ T_{z,lb}[i+1]\le x^j_z[i+1]&\le T_{z,ub}[i+1]\\ m_{OA,min}\le m^j_{OA}[i]&\le m_{OA,max}\\ 0\le LR^j[i]&\le 1\\ 0\le Q^j_{ht}[i]&\le Q_{ht,max}\\ Pow^j_{max,l}\ge\max_{i\in P_l}&\left(Pow_{ch}\left(LR^j[i],\,\underline{T^j_{oa}[i]}\right)+Pow^j_{oth}[i]\right)\end{array}\right\} \begin{array}{l} i=1,...,Np_j,\\ j\in\{1,...,Nday\},\\ l=1,...,Nd\end{array} \tag{4}$$

The superscript j denotes the association with day j. Np_j is the number of optimization time steps for each day. For example, a 1-h time step was used in the case study so $Np_j = 24$. Compared to the centralized formulation, a couple of new variables are introduced: $\mathbf{x}^j[1]$ is the initial state for day j; $Pow^j_{max,l}$ is the expected monthly peak power within the l-th demand period (e.g., on-peak, mid-peak, off-peak, anytime demand period) for the jth day. A new constraint is also introduced as shown in the last inequality in Equation 4 where the power consumption in each demand period is bounded by the corresponding peak. Note that this reformulation does not compromise convexity of the original problem since the newly introduced constraint corresponds to an epigragh of the convex power function (Pow_{ch}) and thus, the new feasible region is still convex. All other variables correspond to the sub-vectors of the corresponding centralized design variable vectors, but with the time indices shifted to the range of 1 to 24. For the first day sub-problem, i.e., $j = 1$, there is an additional constraint:

$$\mathbf{x}^1[1]=\mathbf{x}_0$$

where $\mathbf{x}_o$ is the initial state at the beginning of the month. Note that the design variable vector consists of local variables that are only involved in the sub-problem, such as LR, Q_{ht}, m_{OA}, and those intermediate state variables; and shared variables which are connected to other sub-problems, such as the interfacing state variables and the expected peak power $Pow^j_{max,l}$.

Demand agent

The demand agent, i.e., the ($Nday$+1)-th agent, is responsible for the overall demand reduction of the whole month. The associated optimization problem is simply to minimize the demand cost for the month, which is the sum of the demand costs over all of the demand periods:

$$\min_{[Pow^{Nday+1}_{max,l}|l=1,...,Nd]}\sum_{l=1}^{Nd}\left\{Pow^{Nday+1}_{max,l}\cdot r_{DC,l}\right\}$$

So the day-based agent only minimizes the cooling energy and reheat gas costs for the corresponding day subject to demand constraints for each demand period while the demand agent is responsible for demand cost reduction of the whole month.

Consensus constraints

Day-based agents form a unidirectional row in the sense that the state variable at the end of a day should be the same as the initial state variable for the following day, i.e.,

$$\mathbf{x}^{j+1}[1]=\mathbf{x}^j[24]=\mathbf{x}[24\cdot j+1]\ \text{ for }\ j=1,...,Nday \tag{5}$$

These relations form one set of consensus constraints for the distributed sub-problems. In addition, the expected monthly peak power for each demand period, $Pow^j_{max,l}$, should be equal among all agents, i.e.,

$$Pow^j_{max,l}=Pow_{max,l}\ \text{for}\ j=1,...,Nday+1,\ \ l=1,...,Nd \tag{6}$$

Note that the variables that do not carry a superscript are global shared variables which should equal the corresponding local shared variables among different agents once the con-

sensus constraints are satisfied. The global consensus variable vector is defined as:

$$\mathbf{Z} = \left[(\mathbf{x}[25])^T, (\mathbf{x}[24 \cdot 2 + 1])^T, \ldots, (\mathbf{x}[24 \cdot Nday + 1])^T, Pow_{max,l}\right]^T,$$

which consists of the interfacing state variables and the monthly peak powers for each demand period. Since the consensus constraints are linear, they can be reformulated as:

$$\mathbf{X}^j = \mathbf{F}^j \mathbf{Z}$$

where $\mathbf{X}^j$ is the vector of the shared variables in sub-problem j or agent j and $\mathbf{F}^j$ is a matrix of appropriate size that bridges $\mathbf{X}^j$ and $\mathbf{Z}$. Denoting $\mathbf{X}$ as the vector stacking all the shared variable vectors $\mathbf{X}^j$ and $\mathbf{F}$ as the stacked matrix of $\mathbf{F}^j$, the consensus constraints can be reformulated in a compact way:

$$\mathbf{X} = \mathbf{F}\mathbf{Z}. \tag{7}$$

Each sub-problem can be written in a general form:

$$\begin{aligned} &\min \left\{ f^j\left(\mathbf{X}^j\right)\right\} \\ &\text{s.t. } \mathbf{X}^j \in \mathbf{C}^j \end{aligned}$$

where f^j represents the cost function and $\mathbf{C}^j$ is the feasible region for the shared variables of agent j. Note that the local variables and their feasible regions are omitted in the formulation above for ease of notation.

Solution algorithm

The alternating direction multiplier method (ADMM; Bertsekas and Tsitsiklis, 1989; Boyd 2011) is utilized in this study to solve the distributed optimization problem formulated in the preceding section. Another alternative is the subgradient method (Bertsekas and Tsitsiklis 1989; Nedic and Ozdaglar 2010) which differs from the ADMM in that a non-augmented Lagrangian is considered. Cai et al. (2016a) compared these two methods for application to a HVAC system multi-agent control problem. The subgradient method is slightly easier to implement and the associated primal-dual problem is intrinsically decomposable. Due to the existence of the quadratic penalty term, the ADMM is not directly decomposable and an alternating direction procedure is needed for a distributed implementation. However, the ADMM has superiority over the subgradient method in two aspects: (1) the subgradient method requires strict convexity for convergence while the ADMM does not; (2) the step size for the dual update needs to be chosen appropriately for the subgradient method to converge (Bertsekas, 1995) but the ADMM can use a constant step size (equal to the penalty multiplier). So the ADMM is used in this article and the method is briefly discussed here with more details provided by Cai et al. (2016a) or Boyd (2011).

An augmented Lagrangian shown in Equation 8 is considered in the ADMM algorithm:

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) &= \sum_{i=1}^{n} f^i(\mathbf{X}^i) + \mathbf{Y}^T(\mathbf{X} - \mathbf{F}\mathbf{Z}) + (\sigma/2)||\mathbf{X} - \mathbf{F}\mathbf{Z}||_2^2 \\ &= \sum_{i=1}^{n} f^i(\mathbf{X}^i) + \frac{\sigma}{2}\left\|\mathbf{X} - \mathbf{F}\mathbf{Z} + \frac{\mathbf{Y}}{\sigma}\right\|_2^2 - \frac{1}{2\sigma}||\mathbf{Y}||_2^2 \end{aligned} \tag{8}$$

where $\mathbf{Y}$ is the Lagrange multiplier vector and σ is the penalty multiplier. The first two terms on the right hand side of the top equality form a Lagrangian (which is used directly in the subgradient method) and the third term adds a quadratic penalty. Let $\mathbf{Y}^i$ denote the sub-vector of $\mathbf{Y}$ corresponding to the sub-problem i. Then the augmented Lagrangian can be reformulated as:

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) = \sum_{i=1}^{n} \left\{ f^i(\mathbf{X}^i) + \frac{\sigma}{2}\left\|\mathbf{X}^i - \mathbf{F}^i\mathbf{Z} + \frac{\mathbf{Y}^i}{\sigma}\right\|_2^2 \right\} \\ - \frac{1}{2\sigma}||\mathbf{Y}||_2^2. \end{aligned} \tag{9}$$

It can be noticed that the problem in Equation 9 is not decomposable between $\mathbf{X}^i$ and $\mathbf{Z}$ due to the existence of the quadratic term. An alternating direction procedure is taken which first solves the $\mathbf{X}^i$ problem while fixing $\mathbf{Z}$, and then solves for $\mathbf{Z}$ with fixed $\mathbf{X}$. It is trivial from Equation 8 to see that the optimal value $\mathbf{Z}^*$ satisfies:

$$\mathbf{X} - \mathbf{F}\mathbf{Z}^* + \frac{\mathbf{Y}}{\sigma} = \mathbf{0},$$

which gives an estimate $\mathbf{Z}^* = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\left(\mathbf{X} + \frac{\mathbf{Y}}{\sigma}\right)$. With this alternating direction procedure, the minimization of Equation 9 in the primal problem can be distributed to the previously mentioned agents running in parallel. It is obvious from Equation 8 that the dual ascent direction is **X-FZ** so the dual problem update is:

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \sigma\left(\mathbf{X}_{k+1} - \mathbf{F}\mathbf{Z}_{k+1}\right) \tag{10}$$

where the subscript k represents the iteration number. This primal-dual scheme is carried out iteratively until some stopping criteria are met. The primal problems shown in Equation 9 are solved in parallel by the corresponding agents with the convex programming package CVX (Grant et al. 2008) and with the SDPT3 solver (Toh et al. 1999). The coordination or dual problem performs a simple multiplier variable update as shown in Equation 10.

Convergence and stopping criterion

Two criteria are used to determine if convergence is reached and if the iterative optimization process can be terminated, which are defined as:

$$\varepsilon_k^1 = \|\mathbf{X}_k - \mathbf{F}\mathbf{Z}_k\|_2^2 \tag{11}$$

$$\varepsilon_k^2 = \|\mathbf{X}_k - \mathbf{X}_{k-1}\|_2^2 \tag{12}$$

As explained in Boyd (2011), the optimal solution of the distributed optimization problem in the preceding section needs

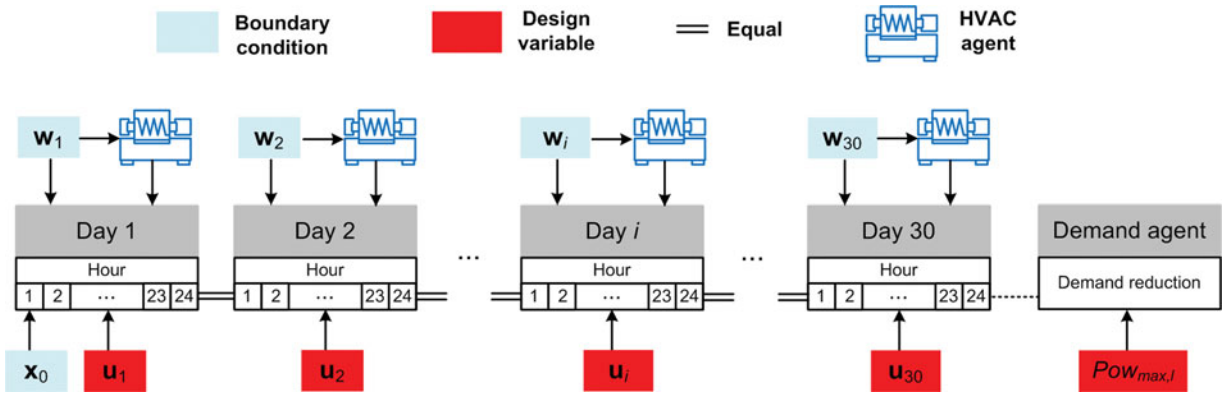

Fig. 4. Monthly utility cost optimization problem diagram in the multi-agent framework.

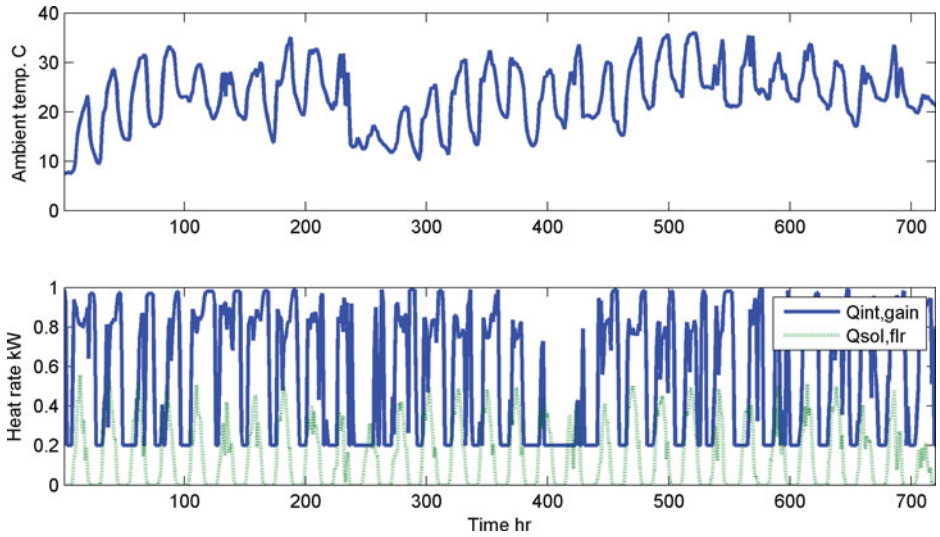

Fig. 5. Ambient temperature and estimated heat gains from May 22, 2015∼June 21, 2015.

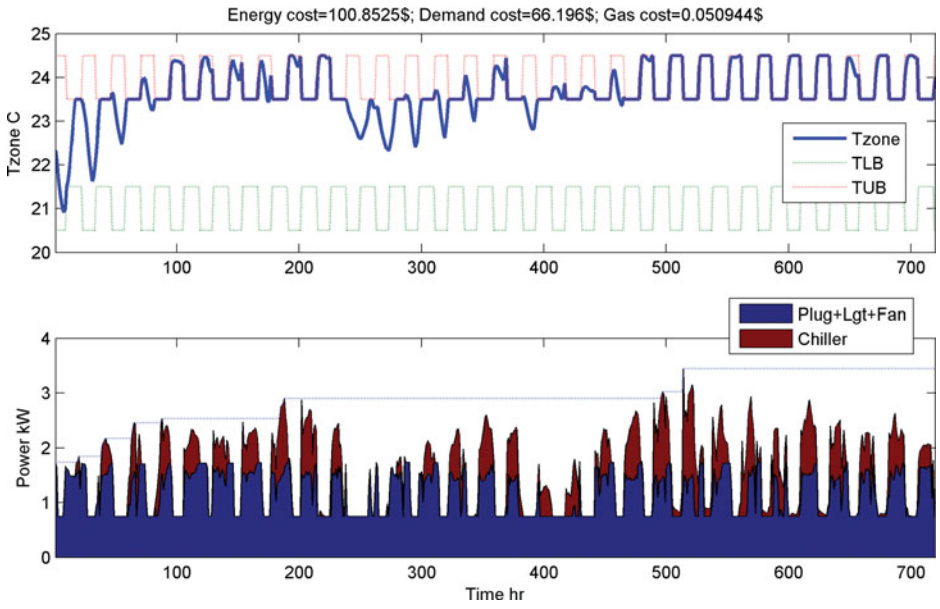

Fig. 6. Baseline control results for the billing cycle May 22, 2015 to June 21, 2015. Peak days occur on June 12th and June 13th (hours from 490 to 538).

to satisfy primal and dual feasibilities. The first criterion defined in Equation 11 is the Euclidean norm of the primal residual, which corresponds to violations of the consensus constraints shown in Equation 7. The second criterion defined in Equation 12 is the Euclidean norm of the dual residual that represents the difference in the shared variables between current and previous steps. So users specify a stopping threshold ε such that when both of the criteria ε^1 and ε^2 are below the threshold, the iterative process stops and the final iteration point is used as the optimal solution.

A multi-agent control framework

A general multi-agent control software framework was developed by Cai et al. (2016a) to facilitate distributed controller design for HVAC systems. Figure 4 depicts implementation of the monthly optimal control problem for minimizing utility costs within this framework. A user only needs to drag the day-based agent, the demand agent and the corresponding HVAC agents from a predefined library and drop them into a project canvas. After simple configurations, such as specifying the day-based agent parameters (state-space matrices), defining the inter-agent connections and loading the boundary conditions, the framework automatically composes the optimization problem and implements the algorithm described in the previous section. This framework reduces the engineering effort required for setting up this type of optimization problem.

Simulation was carried out with measured excitation data collected from May 12 to July 11, 2015 in the case study building. The excitation data include weather information (ambient temperature, humidity, and solar radiation) along with internal heat gains. Figure 5 shows the measured outdoor air temperature, estimated internal heat gains, and solar radiation transmitted through the window for a month period from May 22, 2015 to June 21, 2015. The outdoor airflow rates (m_{OA}), cooling relative to capacity (LR), and reheat (Q_{rh}) rates are the independent optimization variables. The case study building serves as a graduate student office so the imposed occupancy schedule differs from a typical office building: the occupied periods are from 10 am to 10 pm while for the rest of the time, the building is assumed to be unoccupied. Indoor temperature lower/upper bounds are set to 20.5/24.5°C (68.9/76.1°F) during unoccupied periods and 21.5/23.5°C (70.7/74.3°F) during occupied periods. The minimum outdoor air intake ($m_{OA,min}$) is set to 250 CFM (0.14 kg/s) for ventilation and the maximum ($m_{OA,max}$) is set to 1200 CFM (0.67 kg/s) which is the total airflow entering the office space. Since the analysis is only carried out during cooling season, reheat is barely needed and the maximum reheat capacity ($Q_{rh,max}$) is set to 5 kW (17,061 Btu/hr). The nominal return air temperature ($T_{RA,nom}$) is set to 22.5°C (72.5°F) which is the median of the comfort band.

Simulation Results

Baseline

A baseline case was considered with a conventional control strategy: When the space temperature tends to rise above the space temperature upper bound, mechanical cooling is enabled to maintain the temperature at the upper bound; when the space temperature drops below the comfort lower bound, reheat is utilized to keep the space temperature at the lower bound; when the space temperature is within the comfortable range, both mechanical cooling and reheat are turned off to save energy and the space temperature floats. Fan is assumed to operate continuously providing the minimum airflow in all control strategies.

Figure 6 shows monthly simulation results where the top plot shows variation of the space temperature along with its upper and lower bounds. The bottom plot shows the nonsheddable power consumption as well as the sheddable chiller power. The nonsheddable power includes space plug loads, lighting, and the supply air fan. The plug load and lighting powers are nonsheddable due to the occupant activities and they are available as measurements from the BMS. The fan operates at a constant speed to maintain the minimum airflow based on the derived heuristic, so fan power is constant with respect to time. As can be seen from Figure 5, the ambient temperature is high for most of the days and cooling is needed to prevent the space temperature from going beyond the comfort upper bound. The peak power occurs during the days of June 12th and June 13th (hours from 490 to 538) and the monthly energy and demand costs are $100.9 and $66.2 as listed in Table 2.

Maximum savings potential under different strategies

This section considers some different scenarios and evaluates the maximum savings potential relative to the baseline case from the previous section. The different scenarios involve solving optimization problems with different cost functions and constraints:

- Scenario 1: Only considers energy cost (no demand cost) with constant outdoor air intake.
- Scenario 2: Considers both energy and demand costs with constant outdoor air intake.
- Scenario 3: Considers both energy and demand costs with optimally scheduled outdoor air intake (economizer optimization).

Although different cost functions are considered in different scenarios, the actual electricity costs are still evaluated based on the rates shown in Table 1. The optimization results from the different scenarios are helpful in understanding the sources of cost savings among different control strategies. Electricity energy and demand costs are analyzed separately and Table 2 summarizes the costs and savings for the different scenarios as determined using the distributed optimization algorithm. The results are discussed below for the different scenarios, but first it is instructive to consider the behavior of the optimization algorithm. Figure 7 plots variations of the convergence indices and the total electricity cost with iterations for scenario 1. In each iteration, the agents optimize their own sub-problems in parallel and then feed the optimal solutions to a central data hub where the data is fused and broadcast back to the local agents. In the next iteration, the agents repeat the

Table 2. Cost savings potential under different control scenarios.

Scenarios	Description	Energy cost ($)	Energy cost saving	Demand cost ($)	Demand cost saving	Total cost ($)	Total cost saving
Baseline	—	100.9	—	66.2	—	167.1	—
Scenario 1	Without economizer; no demand	96.6	4.3%	50.4	23.9%	147	12.3%
Scenario 2	Without economizer; with demand	97.6	3.3%	42.8	35.4%	140.4	16%
Scenario 3	With economizer; with demand	93.2	7.7%	42.9	35.2%	136.1	18.6%

optimization based on the updated information. The iterations move forward until the convergence criteria are met. Recall that ε^1 indicates the violation level of the consensus constraints and ε^2 indicates the relative change of the shared variables between two consecutive iterations. It can be seen from Figure 7 that the two indices approach zero in an oscillatory manner. The total electricity cost could be lower than the optimal value in some intermediate iterations because constraints are not satisfied. But as constraint violations approach zero, the electricity cost approaches to the optimal value. As a consequence of the convexity in the problem formulation, the distributed optimization converged under all tested scenarios. The stopping criterion used in this study was for both ε^1 and ε^2 to be below 0.02. For the case demonstrated in Figure 7, the distributed optimization stopped at the 91st iteration and for all other tested cases, the optimization stopped within 120 iterations.

Scenario 1

In this scenario, the control optimization was solved in the absence of demand charges and with the outdoor air intake always set at the minimum level which ruled out opportunities for economizer operation. This scenario was implemented by setting $m_{oa,max} = m_{oa,min}$ and $r_{DC} = 0.01$ \$/kW in the problem formulation in Equations 1 and 2. The demand charge rate was set to small but nonzero value to avoid an ill-conditioned optimization problem. With this setting, the optimization only tries to reduce energy costs without trying to reduce the peak demand. Figure 8 shows optimization results for this scenario. Significant pre-cooling occurs for most days. Toward the end of the month, less or even no pre-cooling exists because the zone has a significant time constant and the stored "cooling energy" tends to be released before the end of the month. In this scenario, Table 2 shows that the optimal solution leads

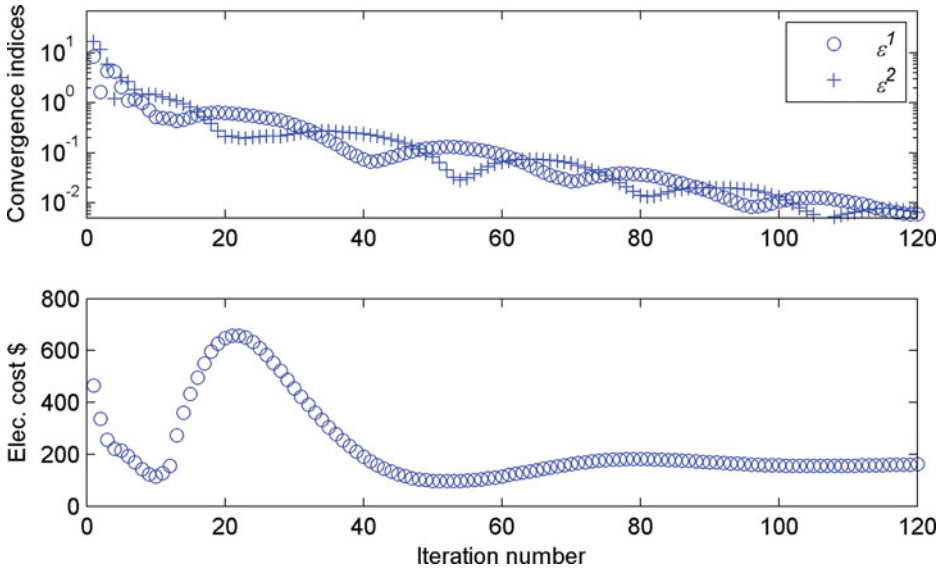

Fig. 7. Convergence results for an example scenario.

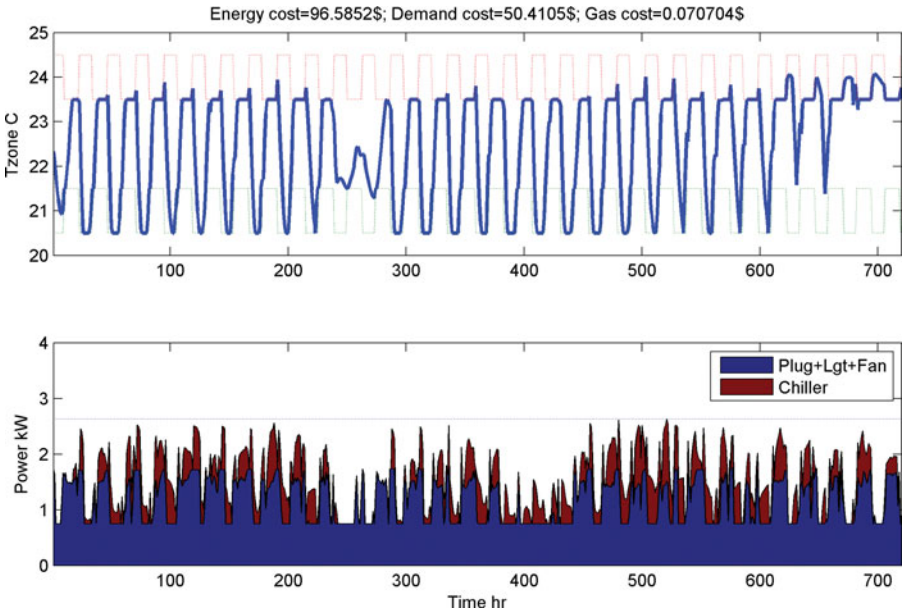

Fig. 8. Simulation results under scenario 1.

to a 4.3% energy cost savings and would reduce the demand cost by 23.9% if the actual costs included the demand rate from Table 1, even though demand cost was not considered in the optimization cost function. That is because the optimal solution shifts a portion of on-peak loads to off-peak periods and the TOU on-peak period coincides with peak building loads.

Scenario 2

In this scenario, the actual demand cost is considered in the optimization cost function with no opportunity for economizer operation (outdoor air intake still maintained at the minimum level). The optimization results are shown in Figure 9. The temperature profile pattern is similar to that in scenario 1 except for the period of 390 to 460 h. This is because the monthly peak occurs within the period of 490 to 520 h and the space temperature is maintained at the lower bound even during the on-peak hours for the days prior to the monthly peak to enable deep pre-cooling. Because of that, overall energy cost savings were reduced by 1% compared to scenario 1, but significantly greater demand cost savings were achieved as shown in Table 2. Another observation is that the total power profile is very flat during on-peak periods while in scenario 1, many spikes exist.

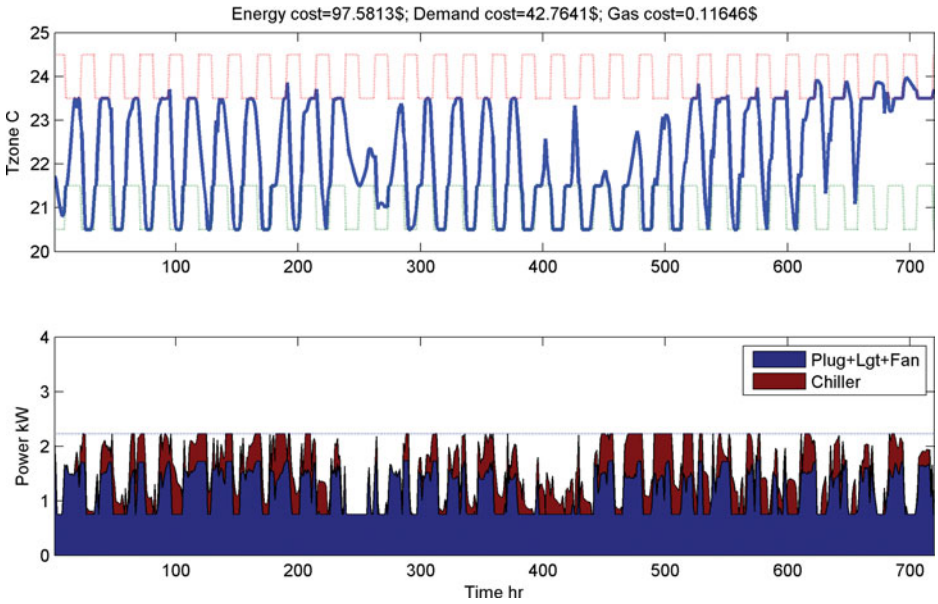

Fig. 9. Simulation result under scenario 2.

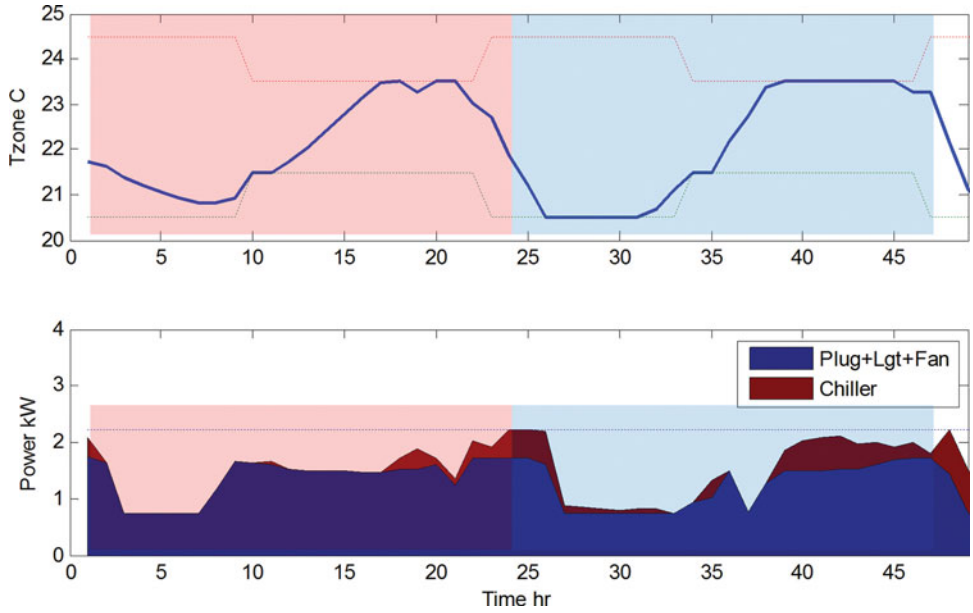

Fig. 10. Zoomed results for the first 2 days of scenario 2.

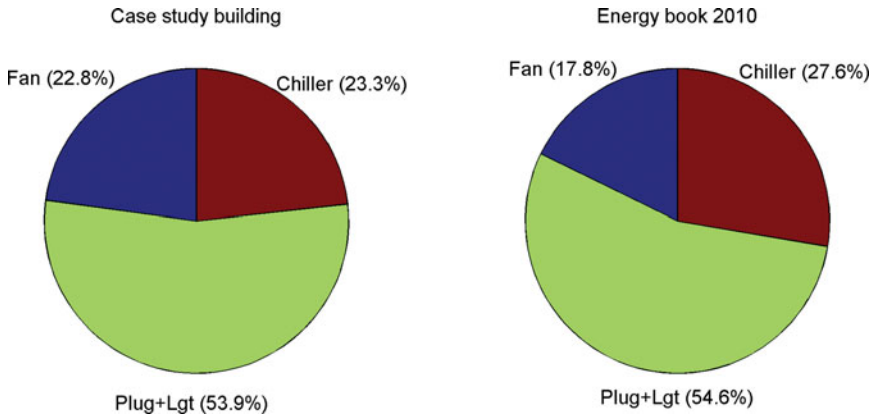

Fig. 11. Energy splits for the case study building and a typical commercial building from the Energy Book (DOE 2010).

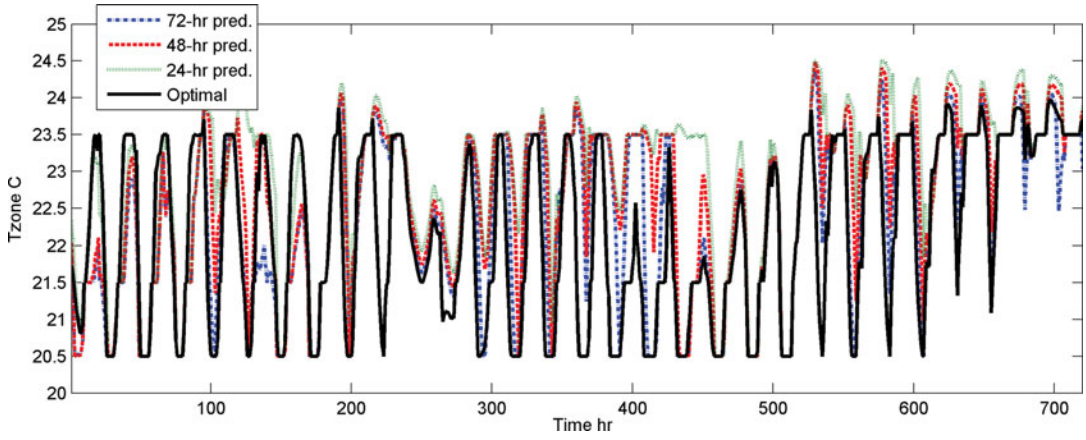

Fig. 12. Comparison of MPC temperature profiles with different look ahead horizons.

Figure 10 shows simulation results for the first 2 days of scenario 2 with shaded blocks highlighting each day period. It can be noticed that significant cooling exists at the end of the first day with the temperature far below the upper bound. This is not an optimal solution for the first day agent itself. However, the information for the second day is propagated to its neighboring days via inter-agent coordination and pushes the first day agent to cool down the room at the end of the first day to store "cooling energy" to be used by the second day agent. This plot shows how the inter-agent coordination makes it possible to fragment the original problem into multiple subproblems that are solved in parallel.

Scenario 3

In this case, both demand costs and economizer operation are considered in the optimization. The optimal temperature profile is very similar to the result for scenario 2 but larger energy cost savings are achieved as shown in Table 2 since some mechanical cooling energy is eliminated by utilizing "free" economizer cooling when the outdoor air is cool enough. However, the demand cost savings are not enhanced since during high load periods, outdoor air temperature is typically high and no "free" cooling is available.

The simulation results indicate that energy cost savings are much smaller than demand cost savings for the considered case study. The validity of this conclusion is highly dependent on the ratio of sheddable energy (chiller energy consumption) to the total energy consumption of the building and the utility rate structure. Figure 11 shows that the energy splits for the case study building are similar to those for a typical commercial building. Also, the utility rate structure (Table 1) is representative of structures found in California and elsewhere around the country.

Short horizon MPC strategies

The monthly optimization tool is useful in understanding the potential magnitude and sources for cost savings associated with optimal control under different scenarios and as a benchmarking tool for evaluating the performance of simpler and more practical control strategies. In this section, a more practical short horizon MPC strategy is proposed that accounts for the tradeoff between the energy cost and demand cost within a smaller time window. For each decision step, a solution is sought for the following optimization problem,

$$\min_{\left\{\substack{x[i],LR[i-1],Q_{ht}[i-1],m_{OA}[i-1]\ldots \\ |i=k+1,\ldots,k+Np}\right\}} \left\{\begin{array}{l} \sum_{i=k}^{k+Np-1}\left\{\mathrm{Pow_{ch}}\left(LR[i],\underline{T_{oa}[i]}\right)\cdot r_e[i]+Q_{ht}[i]\cdot r_{gas}\right\}+ \\ \sum_{l=1}^{Nd}\left\{\max\left(\max_{i\in P_l}\left(Pow_{ch}[i]+Pow_{oth}[i]\right)\right.\right. \\ \qquad \left.\left.-Pow_{pk,l}[k],0\right)\cdot r_{DC,l}\right\} \end{array}\right\} \tag{13}$$

subject to constraints similar to those shown in Equation 2. The only difference in this formulation compared to Equation 1 is that it introduces a peak load threshold $Pow_{pk,l}[k]$ for each demand period, which is equal to the peak load that occurs within the lth demand period of the past portion of current billing cycle where k is the current time step. In addition, Np is a smaller number representing a short and predictable horizon (e.g., 24 h). In addition to the energy cost over the prediction horizon, this formulation considers the incremental demand cost for each demand period associated with the amount of demand beyond the peak for that period that has already occurred. Choosing the parameters $Pow_{pk,l}[1]$ is non-trivial. A simple yet intuitive strategy would be to set $Pow_{pk,l}[1]=0$. This strategy might pose unrealistic weighting of demand cost relative to energy cost: Demand cost would be over-weighted during the days prior to the monthly peak and the energy cost savings might be compromised. However, results in the next section indicate that this simple heuristic works well.

Effect of prediction horizon

Different look-ahead horizons were tested within the short-horizon MPC formulated in Equation 13 with $Pow_{pk,l}[1]=0$, where simulations were carried out for a one-month billing

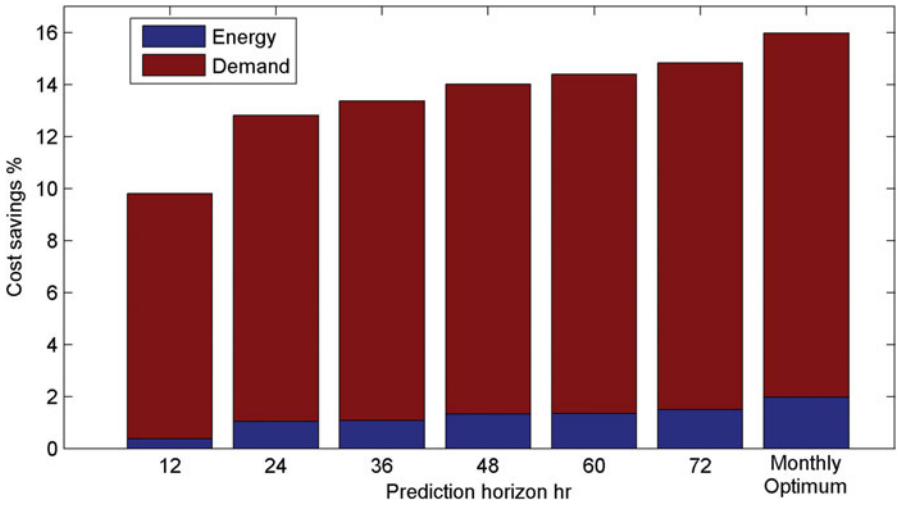

Fig. 13. Energy and demand cost savings with respect to MPC prediction horizons.

Table 3. MPC cost savings within different simulation periods.

Test #	Periods	Energy cost ($)			Demand cost ($)		
		Baseline	Optimal	MPC	Baseline	Optimal	MPC
1	05/12/15–06/11/15	94.2	93.4	93.9	56	41.8	45.4
2	05/22/15–06/21/15	100.9	97.6	99.1	66.2	42.8	46.5
3	06/01/15–07/01/15	98.8	95.1	97	65.4	42.7	45.7
4	06/11/15–07/11/15	100.3	95.7	99	61.1	46	46.1

period that used data from May 22, 2015 to June 21, 2015. Figure 12 compares the MPC temperature profiles along with the monthly optimal profile and Figure 13 shows the energy and demand cost savings with different look-ahead horizons in comparison to the monthly optimal benchmarking results. It can be noticed from Figure 12 that a longer prediction horizon leads to a temperature profile that is closer to the monthly optimum while the short prediction horizon leads to higher space temperatures than the optimum. That is because with longer prediction horizons, the MPC is able to foresee a longer "cooling energy" release and thus, favors lower temperatures. The shorter look-ahead MPC is not able to see the payback of deeper pre-cooling and is only concerned with the benefit within the predictable period. Recall that the monthly peak occurs within the period 490 to 520 h. The 72-h look-ahead MPC is able to see this peak early and starts deep pre-cooling three days ahead, while the 48 and 24-h look-ahead MPCs perform significantly less pre-cooling. As a consequence, both demand and energy cost savings approach their optimal levels as the prediction horizon increases, as can be seen in Figure 13. Note that the 72-h look-ahead MPC still performs pre-cooling at the end of the month which would not be necessary when only considering the utility cost of the demonstrated month. However, this would help in reducing costs for the following month's cycle. Another observation is that the MPCs carry out significant deep cooling for the first couple of days because the peak load threshold $Pow_{pk,l}[1]$ starts from zero and peak load reduction is prioritized for the first few days. An unnecessarily high energy cost should be expected for the first couple of days. However, since the deep cooling only lasts for a couple of days and a portion of the "cooling energy" stored within this period can be utilized later, the energy cost penalty is small. In addition, the energy cost savings potential is much smaller than the demand cost savings, so the overall performance deterioration due to this zero peak start is very minor.

Effect of peak load timing within billing period on 24-hour MPC cost savings

The time for the occurrence of the peak load could have some impact on the performance of the proposed short-horizon MPC. It is expected that if the peak occurs at the beginning of the billing cycle, then the heuristic of setting $Pow_{pk,l}[1] = 0$ should work better. To study this effect, short-horizon MPC simulations were carried out for a 1-month summer period with different assumptions for the start and end of the billing cycle with respect to the weather driving the building loads as listed in Table 3. The start times were shifted by 10 days for each 1-month simulation test. The preceding results show that

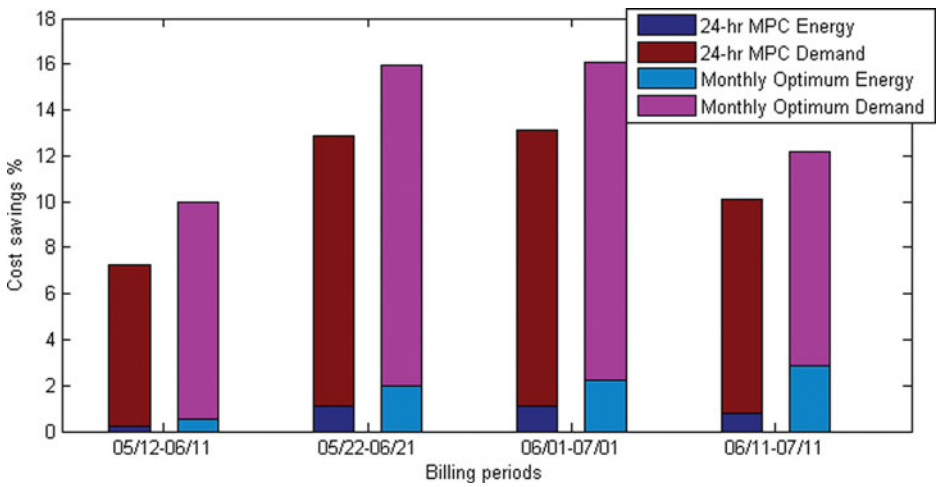

Fig. 14. Energy and demand cost savings with different billing cycles.

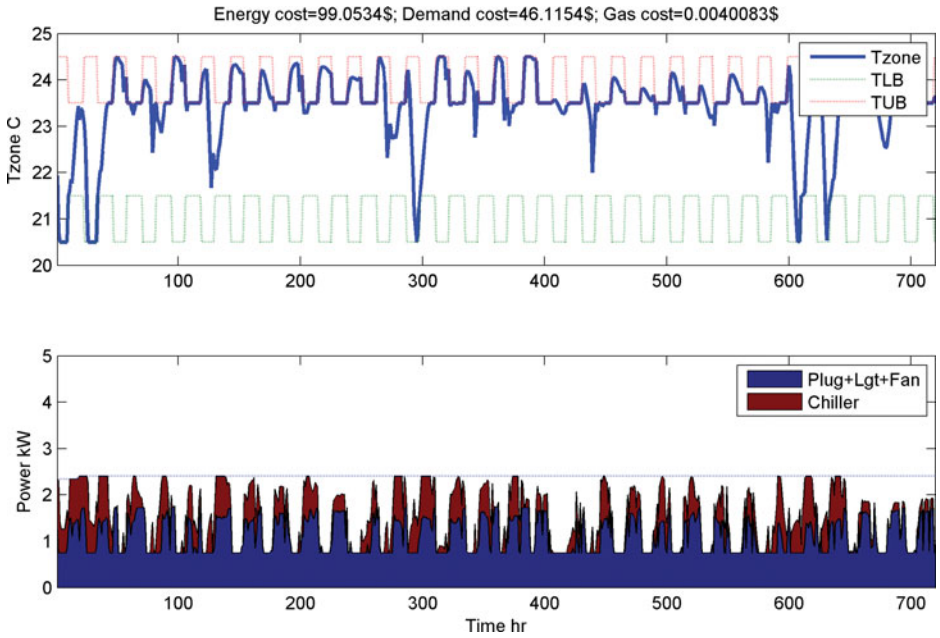

Fig. 15. MPC result with billing cycle from June 11, 2015 to July 11, 2015.

the 24-h look-ahead MPC is able to recover reasonable savings for both demand and energy costs. In addition, longer look-ahead MPC is not practical due to the difficulties in predicting future weather. So the following results were obtained with a 24-h look-ahead MPC.

Figure 14 summarizes overall energy and demand cost savings for the different test cases. The MPC was able to recover most of the cost savings regardless of when the peak occurs within the billing period. It should be noted that the primary peaks occur on June 12th and June 13th which is at the later half, the middle and the beginning of the billing cycles in tests 2, 3, and 4, respectively. Figure 15 shows the MPC results for test 4 where the peak occurs near the beginning of the billing period. Since there is only 1 day prior to this peak that could be utilized for deep pre-cooling, the 24-h look-ahead MPC leads to a demand cost savings that is very close to the optimal savings. The zero-start MPC strategy best leverages the energy and demand costs within this billing cycle and it was expected that the energy savings would also be close to the optimal level. However, this was not the case and the MPC only

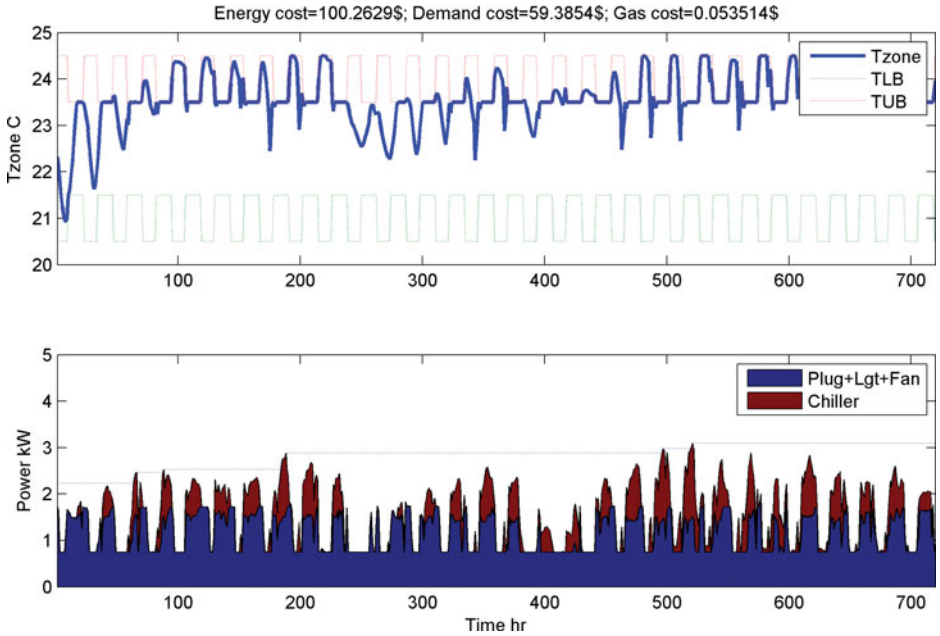

Fig. 16. MPC result with billing cycle from May 22, 2015 to June 21, 2015 with no demand term in the cost formulation.

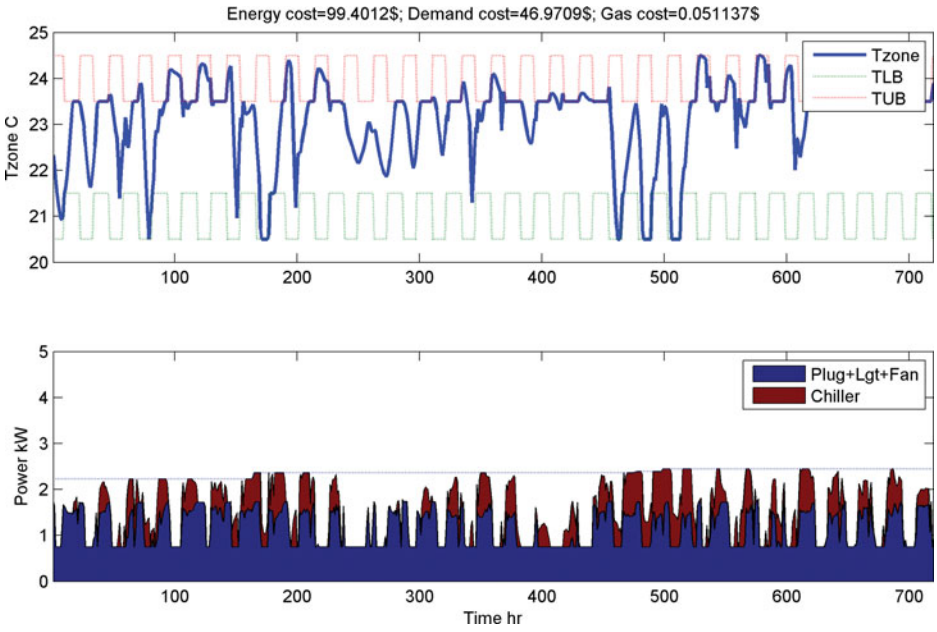

Fig. 17. MPC result with billing cycle from May 22, 2015 to June 21, 2015 with optimal peak threshold start, i.e., $Pow_{pk,l}[1] = 2.227$ (kW).

achieved 29% of the optimal energy cost savings. The reason is explained in the next subsection.

Effect of peak demand charge and initial target on energy cost savings for 24-hour MPC

To better understand the unrecovered cost savings, two extreme cases were simulated. Figure 16 shows MPC results with no demand term in the cost formulation, i.e., $r_{DC,l} = 0.01$ \$/kW. In this formulation, the MPC only tries to minimize energy cost while neglecting the demand cost. However, this energy-priority strategy achieved only 0.6% savings compared to the baseline, which is smaller than the savings determined when the demand charge was considered. Comparing its temperature profile with the monthly optimum shown in Figure 8, the energy-priority MPC leads to much less pre-cooling than the monthly optimal solution because it cannot foresee the utilization of the "pre-cooling energy" beyond its prediction period. As a result, an optimal 24-h MPC leads to a very sub-optimal solution in the scope of a whole month. Considering the demand cost in the MPC leads to larger energy cost savings than a pure energy-priority strategy since the demand reduction requires deeper pre-cooling that helps to reduce energy costs in the long run for this case study.

Another test used the optimal monthly demand obtained from Figure 9 as the peak start value, i.e., $Pow_{pk,l}[1] = 2.227$ kW (7599 Btu/h), which should lead to the best tradeoff between energy and demand costs. The simulation results are plotted in Figure 17. However, the obtained energy and demand cost savings are both smaller than those shown in Table 3 for $Pow_{pk,l}[1] = 0$. The reason is that less pre-cooling is required for the days prior to the monthly peak and thus, less "cooling energy" is available to help to reduce the peak day load.

These two limiting case simulation results indicate that the choice of value for the peak start, $Pow_{pk,l}[1]$, does not have a big affect on the performance of the short-term MPC. This is an important result because it is difficult to identify a more accurate method for estimating a peak demand target for any month. However, additional work is needed to evaluate the generality of this result for different buildings, locations, and utility rates.

Conclusions and Discussion

This article proposed and demonstrated a multi-agent control optimization approach that can determine maximum cost savings opportunities for adjusting supervisory control setpoints over time when considering both energy and demand costs. The method considers the entire monthly billing period since that is the time scale associated with imposing demand charges and decomposes the problem into multiple sub-problems where each sub-problem only deals with a daily optimization or monthly demand reduction. The daily optimizations are focused on determining the optimal trajectory of zone temperature setpoints that minimize daily energy costs subject to demand constraints and rely on the assumption of quasi-steady cooling plant performance and simple heuristics for setting plant supervisory control variables. Example results determined with the distributed optimization approach were presented for different scenarios that were not possible to obtain using a centralized formulation because of computational and memory problems. The uniqueness and significance of the present study is in the proposal of the multi-agent solution scheme as a scalable approach to determine optimal control strategies of building energy systems in a long time horizon. As

a specific application, this article treated the proposed methodology as a general solution method to obtain optimal demand response strategies for building cooling systems.

The overall multi-agent-based approach provides a benchmarking tool for identifying maximum savings potential and for evaluating the performance of more practical supervisory control algorithms. In this respect, this article proposed and evaluated a short-term MPC formulation that accounts for the incremental energy and demand costs within a short and predictable future horizon. The short-term MPC approach imposes a heuristic weighting between the energy and demand costs in the predictable horizon to enable an implementable demand response control. This is another main contribution of the present study. Assuming perfect weather and internal heat gain predictions, it was shown that the performance of the MPC approaches the monthly optimal solution as the length of the prediction horizon increases. However, look-ahead horizons longer than 24 h would not be practical due to the uncertainties in the weather predictions. Fortunately, the 24-h look-ahead MPC achieved over 78% of the maximum possible cost savings potential for all tested scenarios. The effects of different look-ahead horizons, timing of peak demand, and initial peak demand target on the MPC performance were considered. It was found that setting the initial peak demand to zero works well as a heuristic for the 24-h MPC approach. However, further study is needed to generalize this result for different building types, locations, HVAC systems and utility rates.

References

Bertsekas, D.P. 1995. *Nonlinear Programming*. Belmont, MA: Athena Scientific.

Bertsekas, D.P., and J.N. Tsitsiklis. 1989. *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice Hall.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1):1–122.

Braun, J.E. 1988. Methodologies for the design and control of central cooling plants. Ph.D. dissertation, University of Wisconsin, Madison, Wisconsin, USA.

Braun, J.E. 1990. Reducing energy costs and peak electrical demand through optimal control of building thermal storage. *ASHRAE Transactions* 96(2):876–88.

Cai, J. 2015. A low-cost multi-agent control approach for building energy system management. Ph.D. dissertation, Purdue University, West Lafayette, Indiana, USA.

Cai, J., and J.E. Braun. 2015. An inverse hygrothermal model for multi-zone buildings. *Journal of Building Performance Simulation* 1–19.

Cai, J., D. Kim, R. Jaramillo, J.E. Braun, and J. Hu. 2016a. A general multi-agent control approach for building energy system optimization. *Energy and Buildings* 127(1):337–351.

Cai, J., J.E. Braun, D. Kim, and J. Hu. 2016b. A multi-agent control based demand response strategy for multi-zone buildings. *American Control Conference, Boston, July 6-8*.

DOE. 2010. Buildings energy data book. http://buildingsdatabook.eren.doe.gov/docs%5CDataBooks%5C2010_BEDB.pdf.

Grant, M., S. Boyd, and Y. Ye. 2008. CVX: Matlab software for disciplined convex programming.

Henze, G.P., M. J. Brandemuehl, C. Felsmann, H. Cheng, A. R. Florita, and C. E. Waters. 2008. Optimization of Building Thermal Mass Control in the Presence of Energy and Demand Charges. *ASHRAE Transactions* 114(2):75–84.

Keeney, K.R., and J.E. Braun. 1997. Application of building precooling to reduce peak cooling requirements. *ASHRAE Transactions* 103(1):463–9.

Lee, K-H., and J.E. Braun. 2006. An experimental evaluation of demand limiting using building thermal mass in a small commercial building. *ASHRAE Transactions* 112(1):559.

Lee, K.-H., and J.E. Braun. 2008. Model-based demand-limiting control of building thermal mass. *Building and Environment* 43(10):1633–46.

Ma, J., J. Qin, T. Salsbury, and P. Xu. 2012. Demand reduction in building energy systems based on economic model predictive control. *Chemical Engineering Science* 67(1):92–100.

Nedic, A., and A. Ozdaglar. 2010. Cooperative distributed multi-agent optimization. *Convex Optimization in Signal Processing and Communications*. Cambridge, UK: Cambridge University Press.

Sun, Y., S. Wang, and G. Huang. 2010. A demand limiting strategy for maximizing monthly cost savings of commercial buildings. *Energy and Buildings* 42(11):2219–30.

Toh, K.C., M.J. Todd, and R.H. Tutuncu. 1999. SDPT3—a Matlab software package for semidefinite programming. *Optimization Methods and Software* 11(1-4):545–581.

U.S. EPA. 2009. Buildings and their impact on the environment: a statistical summary. http://www.epa.gov/greenbuilding/pubs/gbstats.pdf.

Wang, S., X. Xue, and C. Yan. 2014. Building power demand response methods toward smart grid. *HVAC&R Research* 20(6):665–87.

Xu, P., P. Haves, M.A. Piette, and J. Braun. 2004. *Peak Demand Reduction from Pre-Cooling with Zone Temperature Reset in an Office Building*. Technical Report. Lawrence Berkeley National Laboratory, Berkeley, CA.

Xu, P., and P. Haves. 2006. Case study of demand shifting with thermal mass in two large commercial buildings. *ASHRAE Transactions* 112(1):572.

Xue, X., S. Wang, C. Yan, and B. Cui. 2015. A fast chiller power demand response control strategy for buildings connected to smart grid. *Applied Energy* 137:77–87.