

Chiller Plant Operation Optimization: Energy-Efficient Primary-Only and Primary–Secondary Systems

Danxu Zhang, *Student Member, IEEE*, Peter B. Luh, *Life Fellow, IEEE*, Junqiang Fan, *Senior Member, IEEE*,
and Shalabh Gupta, *Member, IEEE*

Abstract—A chiller plant consists of chiller, cooling tower, and pump subsystems. Two major configurations, primary-only and primary–secondary systems, are often used. Given the high energy costs of a plant, chiller plant operation optimization is important to save energy. For both configurations, chilled/condenser water supply temperatures are critical in improving chiller efficiency and should be considered as decision variables. However, nonlinearity of the problem is increased since chiller power consumption is a highly nonlinear function of these temperatures. Additionally, the problem is combinatorial considering the number of active units (e.g., chillers). In this paper, primary-only systems with identical units in each subsystem and primary–secondary systems with units of two sizes are studied, and both supply temperatures are optimized for energy savings. To obtain near-optimal solutions efficiently, a recent decomposition and coordination approach with little multiplier zigzagging and fast reduction of coupling constraint violations combining with sequential quadratic programming (SQP) is used. Penalties for the constraints that are difficult to be satisfied (e.g., mass balance constraints between fixed-speed pumps and variable-speed chillers) are added. After decomposition, complexity and nonlinearity of a subproblem are reduced drastically as compared with the original problem so that SQP is used. Numerical testing demonstrates that our approach is efficient in obtaining near-optimal solutions, and major energy savings are achieved as compared with benchmark strategies. The approach is scalable and can be used for chiller plant optimization and beyond.

Note to Practitioners—Traditionally, chiller plant operation is often based on rules. For example, the number of active chillers is the minimum number that satisfies cooling requirements, and chilled water supply temperature is constant. Chiller power consumption is a nonlinear function of chiller cooling load and chilled/condenser water supply temperatures. Energy is wasted since chiller efficiency is low with fixed supply temperatures, and sometimes energy consumption of two chillers is less than that of one chiller. To save energy, chiller plant optimization

with good decision variables such as the number of active units and chilled/condenser water supply temperatures is studied. The problem is challenging with high nonlinearity caused by considering such temperatures as decision variables. Furthermore, with discrete variables (e.g., the number of active chillers), the problem is combinatorial. To efficiently solve the problem for high-quality solutions, a novel decomposition and coordination approach is developed. Complexity and nonlinearity of a subproblem are reduced drastically after decomposition as compared with the original problem so that appropriate nonlinear methods are used to solve the subproblems. The results show that the solutions are near-optimal with short computational time and the approach is scalable. Additionally, major energy savings are achieved as compared with benchmark strategies. The approach provides a new and powerful way to solve chiller plant optimization problems and beyond.

Index Terms—Chiller plant optimization, sequential quadratic programming (SQP), surrogate augmented lagrangian relaxation (SALR).

I. INTRODUCTION

A CHILLER plant includes chiller, cooling tower, and pump subsystems, and is used to provide chilled water to cool buildings through heat transfer. Generally, the plant has fast response to system changes because of its fast moving processes (e.g., heat exchange in a chiller). According to the U.S. Department of Energy, heating, ventilation, and air conditioning (HVAC) systems represent 17% of total energy consumption in the United States [1]. A chiller plant represents around 40% of HVAC consumption in a typical air-conditioned building [2]. Traditionally, plant operation is often based on rules. For example, the number of active chillers is the minimum number that satisfies cooling requirements. Energy is often wasted since chillers are nonlinear and sometimes energy consumption of two chillers is less than that of one chiller. To save energy, chiller plant operation optimization is important.

There are two major chiller plant configurations: *primary-only systems* and *primary–secondary systems*. A typical primary-only system is shown in Fig. 1 with four subsystems including chiller, cooling tower, primary pump, and condenser pump subsystems, and two water loops. In the chilled water loop (orange line), chilled water is produced by chillers and supplied to buildings. Return chilled water with raised temperature is driven by primary pumps back to chillers. In the condenser water loop (blue line), condenser water

Manuscript received July 18, 2017; accepted September 1, 2017. Date of publication October 3, 2017; date of current version January 4, 2018. This work was supported in part by the United Technologies Corporation Institute for Advanced Systems Engineering at the University of Connecticut under the Supervisory Control Synthesis Project. This paper was recommended for publication by Associate Editor A. E. Smith and Editor M. P. Fanti upon evaluation of the reviewers' comments. (*Corresponding author: Danxu Zhang.*)

D. Zhang, P. B. Luh and S. Gupta are with the Department of Electrical and Computer Engineering, University of Connecticut, Mansfield, CT 06269 USA (e-mail: danxu.zhang@uconn.edu).

J. Fan is with United Technologies Climate, Controls and Security, East Hartford, CT 06108 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TASE.2017.2751605

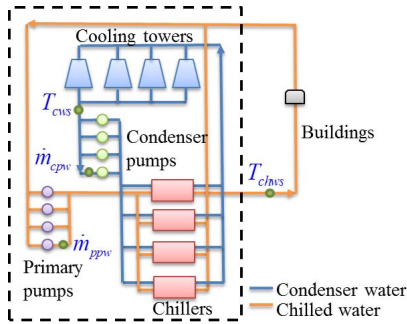


Fig. 1. Schematic of a primary-only chiller plant.

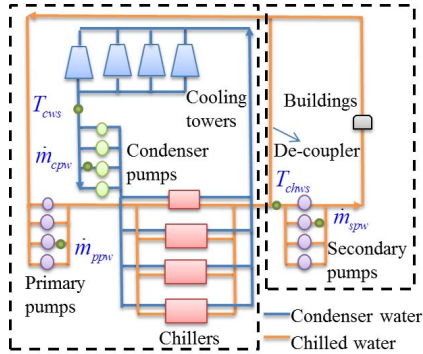


Fig. 2. Schematic of a primary-secondary chiller plant.

absorbs heat from chilled water in chillers and rejects heat to the air in cooling towers. The condenser water is driven by condenser pumps. In practice, variable-speed devices (i.e., chillers, pumps, and cooling towers) are often used and units of each subsystem are identical [3].

Different from primary-only systems, a primary-secondary system has a secondary loop with a decoupler and secondary pumps, as shown in Fig. 2. Generally, fixed-speed pumps are used in the primary loop for easy implementation and control purpose [4], and variable-speed ones are used in the secondary loop to adjust the water supplied to buildings. Primary-secondary systems are not as good as primary-only systems in terms of energy savings [4]–[6]. However, they are still preferable for some applications [7]. When the plant capacity is large enough for cooling requirements, chilled water of fixed-speed primary pumps is often more than that of secondary pumps and surplus water flows into the decoupler. When cooling requirements are too high to be satisfied, water from the decoupler and primary pumps flows into buildings together to provide chilled water. The mixed water may not be cold enough for dehumidification and we may have a problem meeting the requirement. In our study, the latter situation is discarded from the optimization point of view. Identical variable-speed cooling towers and identical variable-speed secondary pumps are often used in practice. Fixed-speed pumps (i.e., primary pumps and condenser pumps) are of two sizes and so are chillers [8].

In this paper, both configurations are studied with primary-only systems presented first and differences between primary-only and primary-secondary systems highlighted. In Section II, chiller plant optimization is reviewed.

In Section III, formulations with good decision variables which have large influences on the plant power consumption are established. Since chilled water supply temperature T_{chws} and condenser water supply temperature T_{cws} are critical in improving chiller efficiency, both T_{chws} and T_{cws} are optimized for energy savings though nonlinearity and complication of the problem will increase.

With discrete variables (e.g., the number of active chillers) and continuous variables (e.g., water temperatures), chiller plant optimization is a mixed-integer nonlinear problem. Computational requirements increase drastically as problem sizes increase. Considering the changes of external conditions (e.g., weather) and internal conditions (e.g., human activities), it is crucial to run optimization efficiently so that a chiller plant can satisfy the varying cooling load requirements quickly. Our goal, therefore, is to obtain high-quality solutions efficiently rather than global optimization. In Section IV, a novel method is developed based on a recent decomposition and coordination approach in the combination with sequential quadratic programming (SQP). For primary-secondary systems, relaxed mass balance constraints between fixed-speed pumps and variable-speed chillers are difficult to be satisfied leading to convergence issues. To resolve the issues and be consistent, surrogate augmented Lagrangian relaxation (SALR) [30] is used for both primary-only and primary-secondary systems. Quadratic penalty terms for the constraints are added to reduce constraint violations and accelerate the speed of convergence. After decomposition, complexity and nonlinearity of a sub-problem are reduced drastically as compared with those of the original problem so that the nonlinear method SQP is used.

In Section V, three examples are tested and each of them is for both configurations. Simple chiller plants are studied to explain the ideas of our methods. Medium chiller plants are tested to show the performance of the methods under different cooling load requirements, and power consumption based on a cooling load profile. Then, large chiller plants are studied to demonstrate the scalability of the methods. The results show that our approach is scalable and its performance is good with short CPU time and small gaps. In addition, significant energy savings are achieved as compared with benchmark strategies.

II. LITERATURE REVIEW

In this section, chiller plant optimization for primary-only and primary-secondary systems is reviewed with modeling presented first, followed by solution methodologies.

A. Modeling

A chiller plant has fast response to system changes because of its fast moving processes (e.g., heat exchange of chillers). Dynamics of the processes are usually not of interest to chiller plant optimization, and static models are often used. Physical models, hybrid models, and black box models are three major types of modeling techniques. Detailed physical models are generally too complex for optimization, and only few papers focusing on simulation-based optimization use such models [34]. Compared with detailed physical models, black box models based on experience are simple but without physical meanings [7]. A large amount of data are needed

to obtain such models. Different from physical models and black box models, hybrid models based on both physics and experience are often used for primary-only chiller plant optimization [10], [11]. For example, in [10], the chiller power consumption is an empirical function and chiller constraints are based on energy and mass balances. Chilled water supply temperature T_{chws} and condenser water supply temperature T_{cws} are two important factors on the plant energy consumption, and their influences are often discussed in the literature. However, to the best of our knowledge, one or two of them are often set as constant to simplify the problem since the chiller power consumption is a highly nonlinear function of the temperatures [10], [12]–[17], [33]. Objective functions include minimizing power consumption [10], energy consumption [21], and costs [33]. Since the problem is static, solutions based on above objective functions are the same. Decision variables are the number of active units, water mass flow rates of pumps, and water supply temperatures if the temperatures are not assumed fixed. With discrete variables (e.g., the number of active chillers) and continuous variables (e.g., water mass flow rates), chiller plant optimization is a mixed-integer nonlinear problem.

For primary–secondary systems, only few papers are found [22]–[25] and primary–secondary aspects considered in these papers are simple. For example, in [22], all the units were with fixed speed, and mass flow rates of the units were fixed. Decision variables were the number of active chillers, the number of cooling towers and supply temperatures. Heat exchange equations of a chiller were linear with constant water mass flow rate. The problem is simple since the relationship between the number of active pumps and the number of active chillers can be obtained easily based on the mass balance, and heat exchange constraints are linear. Similarly, a plant with fixed-speed primary pumps and fixed-speed condenser pumps was studied in [25]. Each chiller was equipped with a primary pump and a condenser pump. If a chiller is active, its corresponding pumps must be online so that there is no need to consider pumps. Decision variables were the number of active chillers and water temperatures. The problem is simplified in a major way. The decoupler was not discussed in these papers.

B. Solution Methodologies

To solve the mixed-integer nonlinear problems, intelligence algorithms such as genetic algorithm (GA) [10]–[15], simulated annealing [16], and particle swarm optimization [17] without requiring differentiable objective functions were often used. Among those algorithms, GA is the most commonly used one since it is easy to understand and implement. However, intelligence algorithms do not exploit problem structures, and solution quality is difficult to quantify. Traditional gradient-based methods, such as SQP, were also used in combination with heuristics to solve the problems [18]–[20]. The major disadvantage is that solution quality cannot be quantified. To reduce computational effort, decomposition and coordination methods were also used. Standard Lagrangian relaxation (LR) suffers from complexity difficulties, zigzagging and slow convergence, and requiring optimal dual values in the process. Such difficulties have been overcome by surrogate

Lagrangian relaxation (SLR) [9]. For problems where relaxed constraints are difficult to be satisfied, surrogate augmented Lagrangian relaxation (SALR) [30] which is an extension of SLR is used. Penalties for constraints that are difficult to be satisfied are added as used in standard augmented Lagrangian relaxation (ALR) to reduce constraint violations and accelerate convergence speed. The problem in [30] is a mixed-integer linear problem, which is different from our problem which is a mixed-integer nonlinear problem.

As mixed-integer nonlinear problems, some studies considered thermal energy storages. Such problems are dynamic and methods for dynamic optimization are used. For example, in [33], solutions for the storage were obtained via dynamic programming. Chiller power consumption was a bilinear function of chilled water return temperature. Chilled water supply temperature was assumed constant. Though on/off statuses of a chiller are considered, the model with nonlinearity reduced is much simpler than ours. After linearization, global optimization solvers were used to obtain solutions for chillers. For our problem which is highly nonlinear and with mixed integers, it is difficult to linearize the problem and solve it using existing optimization solvers. Furthermore, this paper focuses on chillers and the thermal storage rather than a chiller plant so that there are no coupling issues among subsystems. With those differences, their approach cannot be used for our problem.

Similar to primary-only systems, intelligence algorithms (e.g., GA) were used for primary–secondary systems [25]. Furthermore, nonlinear algorithms such as the generalized reduced gradient method were used [22]. All the possible combinations of active units were tested, and the final solution was the one with minimum cost. However, the method is not scalable.

III. PROBLEM FORMULATION

In this section, formulations with good decision variables are established for primary-only and primary–secondary systems. Both T_{chws} and T_{cws} are optimized for energy savings though nonlinearity of the formulations is increased. Modeling for primary-only systems is presented in Sections III-A–III-C. Coupling constraints between subsystems are presented in Section III-D, followed by the objective function and the optimization problem in Section III-E. Modeling for primary–secondary systems is presented in Section III-F with differences between the primary–secondary system and the primary-only system highlighted.

A. Chiller Model

A typical chiller includes an evaporator, a compressor, a condenser, and an expansion valve, as shown in Fig. 3. The evaporator is used to evaporate refrigerant liquid into gas by absorbing heat from chilled water. The compressor draws in the gas and compresses it by consuming electricity. The condenser takes this very hot gas and turns it into liquid by rejecting heat to condenser water. The expansion valve expands the liquid to get cold liquid. The cold liquid then flows into the evaporator. Since the working process inside a chiller is complex and the refrigerant is controlled by local controllers, a hybrid model [26], [27] focusing on water is used.

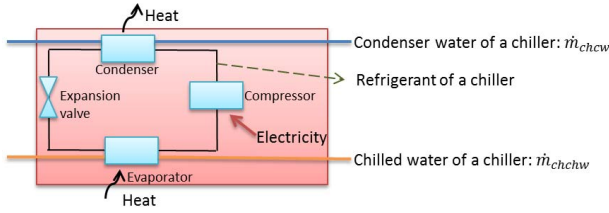


Fig. 3. Schematic of a typical chiller.

1) *Heat Exchange of Chillers*: Heat absorbed by the evaporator is denoted by Q_{ch} and formulated as

$$Q_{ch} = C_p [\dot{m}_{chchw} (T_{chwr} - T_{chws})] \quad (1)$$

where C_p is water specific heat, T_{chwr} is chilled water return temperature, and \dot{m}_{chchw} is chilled water mass flow rate of the chiller. Similarly, heat rejected by the condenser Q_{cd} is

$$Q_{cd} = C_p [\dot{m}_{chcw} (T_{cwr} - T_{cws})] \quad (2)$$

where T_{cwr} is condenser water return temperature, and \dot{m}_{chcw} is condenser water mass flow rate of the chiller. Based on energy balance of the chiller, total heat rejected by a chiller equals to the heat absorbed by the chiller plus the electricity consumed by the chiller

$$Q_{cd} = P_{ch} + Q_{ch} \quad (3)$$

where P_{ch} is chiller power consumption. Cooling load provided by chillers is assumed to be proportional to chiller cooling capacity [12] and should satisfy the building cooling requirement Q_{demand} . With identical chillers used, we have

$$N_{Ach} \times Q_{ch} = Q_{demand} \quad (4)$$

where N_{Ach} is the number of active chillers.

2) *Chiller Power Consumption*: As (1)–(3) show, P_{ch} is affected by water temperatures. Based on [26] and [27], P_{ch} is derived by power consumption at a reference condition and adjustments as

$$P_{ch}(T_{cws}, T_{chws}, Q_{ch}) = P_{ch,ref} \times CAPFT \times EIRFT \times EIRFPLR \quad (5)$$

where $P_{ch,ref}$ is power consumption at a reference condition. EIR is energy input to cooling output ratio, therefore, $P_{ch,ref} = EIR_{ref} \times Q_{ch,ref}$. $CAPFT$ is a cooling capacity adjustment which is a quadratic function of T_{chws} and T_{cws} as

$$CAPFT = a_1 + a_2 T_{chws} + a_3 T_{chws}^2 + a_4 T_{cws} + a_5 T_{cws}^2 + a_6 T_{chws} T_{cws}$$

$EIRFT$ is energy input to cooling output ratio adjustment which is a quadratic function of T_{chws} and T_{cws} as

$$EIRFT = b_1 + b_2 T_{chws} + b_3 T_{chws}^2 + b_4 T_{cws} + b_5 T_{cws}^2 + b_6 T_{chws} T_{cws}$$

$EIRFPLR$ is energy input to cooling output ratio adjustment which is a quadratic function of part load ratio (PLR) as

$$EIRFPLR = d_1 + d_2 \times PLR + d_3 \times PLR^2$$

$$PLR = \frac{Q_{ch}}{Q_{ch,ref} \times CAPFT(T_{cws}, T_{chws})}$$

and a_i , b_i , and d_i are coefficients. As shown in the model, if T_{chws} and T_{cws} are constants, P_{ch} will become a simple quadratic function of Q_{ch} . Otherwise, P_{ch} is complicated and highly nonlinear.

3) *Lower and Upper Bound Constraints of Chillers*: With safety and operation condition considered, lower and upper bounds are provided for T_{chws} , T_{chwr} , T_{cws} , and T_{cwr} . The maximum value for T_{chws} should be low enough to guarantee good humidity control capability of air handling units in buildings. Since the lowest T_{cws} can be produced by a cooling tower is the wet-bulb temperature T_{wb} , the lower bound of condenser water supply temperature is set to be 4 °C higher than the wet-bulb temperature as used in practice.

B. Cooling Tower Model

A typical cooling tower has a fan producing air by consuming electricity. Heat is removed from the condenser water by the air. Since the heat exchange process is complex, an empirical cooling tower model based on [28] is used.

1) *Heat Exchange of Cooling Towers*: In a cooling tower, heat exchange occurs between the air and the condenser water. Since the state change of the air is difficult to model, a commonly used empirical model, York cooling tower model, based on approach temperature T_{app} is used and shown as follows:

$$T_{app} = c_1 + c_2 T_{wb} + c_3 T_{wb}^2 + c_4 T_r$$

$$+ c_5 T_{wb} T_r + c_6 T_{wb}^2 T_r + c_7 T_r^2 + c_8 T_{wb} T_r^2$$

$$+ c_9 T_{wb}^2 T_r^2 + c_{10} L + c_{11} T_{wb} L + c_{12} T_{wb}^2 L + c_{13} T_r L$$

$$+ c_{14} T_{wb} T_r L + c_{15} T_{wb}^2 T_r L + c_{16} T_r^2 L + c_{17} T_{wb} T_r^2 L$$

$$+ c_{18} T_{wb}^2 T_r^2 L + c_{19} L^2 + c_{20} T_{wb} L^2 + c_{21} T_{wb}^2 L^2$$

$$+ c_{22} T_r L^2 + c_{23} T_{wb} T_r L^2 + c_{24} T_{wb}^2 T_r L^2$$

$$+ c_{25} T_r^2 L^2 + c_{26} T_{wb} T_r^2 L^2 + c_{27} T_{wb}^2 T_r^2 L^2$$

$$T_{app} = T_{cws} - T_{wb}, T_r = T_{cwr} - T_{cws}, L = \frac{(\dot{m}_{ctcw} / \dot{m}_{wdesign})}{(\dot{m}_{cta} / \dot{m}_{adesign})} \quad (6)$$

where \dot{m}_{ctcw} is condenser water mass flow rate of the cooling tower, \dot{m}_{cta} is air mass flow rate, $\dot{m}_{wdesign}$ and $\dot{m}_{adesign}$ are designed mass flow rate of water and air, respectively, T_r is range temperature, L is liquid-to-gas ratio, and c_i is a coefficient. As shown in the model, (6) is complicated and highly nonlinear, and affected by the wet-bulb temperature T_{wb} .

2) *Cooling Tower Power Consumption*: Cooling tower power consumption P_{ct} is a cubic function of the mass flow rate of the air \dot{m}_{cta} [29]

$$P_{ct} = P_{ct,nom} \left(\frac{\dot{m}_{cta}}{\dot{m}_{cta,nom}} \right)^3 \quad (7)$$

where $P_{ct,nom}$ is the nominal power consumption, and $\dot{m}_{cta,nom}$ is the nominal air mass flow rate.

3) *Lower and Upper Bound Constraints of Cooling Towers*: Lower and upper bounds of the air mass flow rate are

$$\dot{m}_{cta \min} \leq \dot{m}_{cta} \leq \dot{m}_{cta \max} \quad (8)$$

C. Variable-Speed Pump Model

Pumps consume electricity to circulate water. Therefore, pump power consumption is often modeled as a function of water mass flow rate. The pump model with varying efficiencies from [29] is used and presented as follows.

1) Primary Pump Power Consumption:

$$P_{pp} = \frac{k_{pp}X_{ppm}^3}{(1 - e^{-A_{pp}X_{ppm}})\eta_{var}(X_{ppm})}$$

$$X_{ppm} = \frac{\dot{m}_{ppw}}{\dot{m}_{ppw,nom}}$$

$$\eta_{var}(X_{ppm}) = k_1 + k_2X_{ppm} + k_3X_{ppm}^2 + k_4X_{ppm}^3 \quad (9)$$

where $\dot{m}_{ppw,nom}$ is the nominal primary pump mass flow rate, \dot{m}_{ppw} is primary pump mass flow rate, and k_{pp} , A_{pp} , and k_i are coefficients.

2) Lower and Upper Bound Constraints of Primary Pumps:

Lower and upper bounds for water mass flow rate of the primary pumps are shown as

$$\dot{m}_{ppw,min} \leq \dot{m}_{ppw} \leq \dot{m}_{ppw,max}. \quad (10)$$

Condenser pumps are similar to primary pumps, and are not presented here.

D. Coupling Constraints Between Subsystems

Based on mass balance, chilled/condenser water mass flow rates of different subsystems are the same. Therefore, the coupling constraints between different subsystems are

$$N_{App}\dot{m}_{ppw} = N_{Ach}\dot{m}_{chchw} \quad (11)$$

$$N_{Acp}\dot{m}_{cpw} = N_{Ach}\dot{m}_{chcw} \quad (12)$$

$$N_{Ach}\dot{m}_{chcw} = N_{Act}\dot{m}_{ctcw} \quad (13)$$

where N_{App} is the number of active primary pumps, N_{Acp} is the number of active condenser pumps, and N_{Act} is the number of active cooling towers.

With complicated and nonlinear heat exchange, chillers and cooling towers are highly coupled. To decouple the two subsystems, temperature constraints are involved by introducing condenser water supply temperature and return temperature for chillers and cooling towers

$$T_{cws,ch} = T_{cws,ct} \quad (14)$$

$$T_{cwr,ch} = T_{cwr,ct}. \quad (15)$$

E. Objective Function and the Optimization Problem

Since the problem is static, the solutions are the same for optimizing energy cost and power consumption. Here power consumption of a primary-only chiller plant with identical units in each subsystem is minimized. The optimization problem is

$$\min_{\left(\begin{array}{c} N_{Ach}, N_{Act}, N_{App}, N_{Acp} \\ T_{chws}, T_{cws}, \dot{m}_{ppw}, \dot{m}_{cpw} \end{array} \right)} J_{po}$$

with

$$J_{po} \equiv N_{Ach}P_{ch} + N_{Act}P_{ct} + N_{App}P_{pp} + N_{Acp}P_{cp} \quad (16)$$

subject to (1)–(4), (6), (8), and (10)–(15).

As shown in the formulation, the problem is a mixed-integer nonlinear problem with an additive objective function and additive coupling constraints. Considering the changes of external conditions (e.g., weather) and internal conditions (e.g., human activities), it is crucial to run the optimization efficiently to satisfy building load requirements. Since getting global optimization solutions of an NP-hard problem often takes a long time, our goal, therefore, is not to obtain optimal but rather near-optimal solutions in a computationally efficient manner under appropriate constraints.

F. Primary–Secondary Systems

Different from primary-only systems, a primary-secondary system usually has a primary loop with fixed-speed pumps to generate chilled water, and a secondary loop with variable-speed pumps to distribute chilled water to buildings. Therefore, we have one more coupling constraint between the primary loop and the secondary loop. Generally, water of primary pumps is more than that of secondary pumps, and surplus water flows into the decoupler. Based on the mass balance equation, we have

$$N_{App}\dot{m}_{ppw} = \dot{m}_{decoupler} + N_{Asp}\dot{m}_{sp} \quad (17)$$

where $\dot{m}_{decoupler}$ is the mass flow rate of the decoupler which is nonnegative to guarantee the flow direction, N_{Asp} is the number of active secondary pumps, and \dot{m}_{sp} is secondary pump mass flow rate. Water driven by secondary pumps is supplied to buildings to satisfy cooling requirements

$$Q_{demand} = C_p[N_{Asp}\dot{m}_{sp}(T_{chwl,b} - T_{chws})] \quad (18)$$

where $T_{chwl,b}$ is the temperature of the water leaving the buildings. Since chillers are of two sizes and cooling load of a chiller is proportional to its capacity [12], we have

$$\frac{Q_{ch,S}}{Q_{ch,ref,S}} = \frac{Q_{ch,B}}{Q_{ch,ref,B}}. \quad (19)$$

With fixed speed, power consumption and mass flow rate of a primary pump are constant. Therefore, total power consumption and total mass flow rate depend on the number of active primary pumps. Condenser pumps are similar to primary pumps. Other units are the same as those of primary-only systems and are not repeated. With the secondary loop and units of two sizes, the optimization problem of the primary–secondary system is

$$\min_{\left(\begin{array}{c} N_{Ach,S}, N_{Ach,B}, N_{Act}, N_{App,S}, N_{App,B} \\ N_{Acp,S}, N_{Acp,B}, N_{Asp}, T_{chws}, T_{cws}, \dot{m}_{sp} \end{array} \right)} J_{ps}$$

with

$$J_{ps} \equiv N_{Ach,S}P_{ch,S} + N_{Ach,B}P_{ch,B} + N_{Act}P_{ct} \\ + N_{App,S}P_{pp,S} + N_{App,B}P_{pp,B} + N_{Acp,S}P_{cp,S} \\ + N_{Acp,B}P_{cp,B} + N_{Asp}P_{sp} \quad (20)$$

subject to (1)–(4), (6), (8), (10)–(15), and (17)–(19).

IV. SOLUTION METHODOLOGY

To obtain near-optimal solutions efficiently, a novel decomposition and coordination approach with little multiplier zigzagging and fast reduction of coupling constraint violations is used. Complexity and nonlinearity of subproblems are much reduced as compared with the original problem so that SQP is used to solve the subproblems. In Section IV-A, the relaxed problem and subproblems of the primary-only system are presented. In Section IV-B, multipliers are updated after solving one subproblem, and feasible solutions for the entire problem are obtained. In Section IV-C, differences between primary-secondary and primary-only systems are highlighted.

A. Relaxed Problem and Subproblems for the Primary-Only System

Standard LR suffers from complexity difficulties, zigzagging and slow convergence, and requiring optimal dual values in the process. Such difficulties have been overcome by SLR. To reduce constraint violations and accelerate convergence speed, surrogate augmented Lagrangian relaxation (SALR) [30], which is an extension of SLR, is used. Coupling constraints (11)–(15) are relaxed by using Lagrangian multipliers $\lambda_{mb,pp}^k$, $\lambda_{mb,cp}^k$, $\lambda_{mb,ct}^k$, λ_{Tcws}^k , and λ_{Tcwr}^k , respectively. Penalties for the constraints that are difficult to be satisfied are added. The relaxed problem is as follows:

$$\min_{\left(\begin{smallmatrix} N_{Ach}, N_{Act}, N_{App}, N_{Acp} \\ T_{cws}, T_{cws}, \dot{m}_{ppw}, \dot{m}_{cpw} \end{smallmatrix} \right)} L_{po}$$

with

$$\begin{aligned} L_{po} \equiv & J_{po} + \lambda_{mb,pp}^k (N_{App} \dot{m}_{ppw} - N_{Ach} \dot{m}_{chchw}) \\ & + \lambda_{mb,cp}^k (N_{Acp} \dot{m}_{cpw} - N_{Ach} \dot{m}_{chcw}) \\ & + \lambda_{mb,ct}^k (N_{Ach} \dot{m}_{chcw} - N_{Act} \dot{m}_{ctcw}) \\ & + \lambda_{Tcws}^k (T_{cws_ch} - T_{cws_ct}) + \lambda_{Tcwr}^k (T_{cwr_ch} - T_{cwr_ct}) \\ & + 0.5c^k (T_{cws_ch} - T_{cws_ct})^2 + 0.5c^k (T_{cwr_ch} - T_{cwr_ct})^2 \end{aligned} \quad (21)$$

subject to (1)–(4), (6), and (8).

Each subsystem is formulated as one subproblem by collecting all the terms related to the subsystem. The quadratic penalty terms lead to inseparability of the augmented Lagrangian. To resolve this issue, variables of other subsystems are replaced by solutions from previous iterations.

1) *Chiller Subproblem*: By collecting all the terms related to chillers and replacing variables of cooling towers by using solutions of previous iterations, the chiller subproblem at k^{th} iteration is obtained as

$$\min_{\left(\begin{smallmatrix} N_{Ach}, \\ T_{chws}, T_{cws} \end{smallmatrix} \right)} L_{ch}$$

with

$$\begin{aligned} L_{ch} \equiv & N_{Ach} P_{ch} - \lambda_{mb,ct}^k N_{Ach} \dot{m}_{chcw} - \lambda_{mb,pp}^k N_{Ach} \dot{m}_{chchw} \\ & - \lambda_{mb,cp}^k N_{Ach} \dot{m}_{chcw} + \lambda_{Tcws}^k T_{cws_ch} + \lambda_{Tcwr}^k T_{cwr_ch} \\ & + 0.5c^k (T_{cws_ch} - T_{cws_ct}^{K-1})^2 + 0.5c^k (T_{cwr_ch} - T_{cwr_ct}^{K-1})^2 \end{aligned} \quad (22)$$

subject to (1)–(4).

Chillers are the most complex units and account for the largest portion of energy consumption in a chiller plant. Traditionally, the number of active chillers is the minimum one that satisfies the cooling requirement. Since chiller power consumption is nonlinear, more chillers may consume less energy. To strike the balance between computational effort and energy savings, the subproblem is solved by considering two cases. In Case 1, minimum number of chillers that satisfy the cooling requirement is used, and in Case 2, the number is increased by one. For each case, the problem is a nonlinear problem and a commonly used nonlinear method SQP is applied. The final solution of the subproblem is the one with lower cost. Though the solution may not be the best, the method is better than the traditional strategy, and computational effort is reduced.

2) *Cooling Tower Subproblem*: By collecting all the terms related to cooling towers and replacing variables of chillers by using solutions of previous iterations, the cooling tower subproblem at k^{th} iteration is

$$\min_{N_{Act}} L_{ct}$$

with

$$\begin{aligned} L_{ct} \equiv & N_{Act} P_{ct} - \lambda_{mb,ct}^k N_{Act} \dot{m}_{ctcw} - \lambda_{Tcws}^k T_{cws_ct} - \lambda_{Tcwr}^k T_{cwr_ct} \\ & + 0.5c^k (T_{cws_ch}^{K-1} - T_{cws_ct})^2 + 0.5c^k (T_{cwr_ch}^{K-1} - T_{cwr_ct})^2 \end{aligned} \quad (23)$$

subject to (6). With high nonlinearity, this subproblem is solved similarly as for chiller subproblems. Details are not presented.

3) *Primary Pump Subproblem*: By collecting all the terms related to primary pumps, the primary pump subproblem at k^{th} iteration is obtained as

$$\min_{(N_{App}, \dot{m}_{ppw})} L_{pp}, \text{ with } L_{pp} \equiv N_{App} P_{pp} + \lambda_{mb,pp}^k N_{App} \dot{m}_{ppw} \quad (24)$$

subject to (8). Since the pump subproblem is simple with lower and upper bound constraints, all the possible cases where the number of active pumps satisfies the cooling requirement are considered to solve the subproblem, and the solution is the one with lower cost.

The condenser pump subproblem is similar and is not repeated here.

B. Dual Problem and Feasible Solutions for the Primary-Only System

Solutions of subproblems are coordinated through the iterative updating of multipliers to maximize the dual function. The dual function is shown as

$$\max_{\lambda} q, \text{ with } q \equiv L_{ch} + L_{ct} + L_{pp} + L_{cp}. \quad (25)$$

Multipliers are updated based on (26)–(28) after solving one subproblem and then used for next subproblem until the dual function is maximized

$$\lambda^{k+1} = \lambda^k + s^k \tilde{g}(x^k) \quad (26)$$

$$s^k = \alpha_k s^{k-1} \frac{\|\tilde{g}(x^{k-1})\|}{\|\tilde{g}(x^k)\|} \quad (27)$$

$$\alpha_k = 1 - (M \cdot k^p)^{-1}, \quad 0 < p < 1, \quad M > 1, \quad k = 1, 2, \dots \quad (28)$$

where \tilde{g} is the surrogate augmented subgradient, s is the stepsize, and M and p are constants.

After convergence, a lower bound is obtained. With coupling constraints relaxed, solutions obtained above may not be feasible for the original problem. To obtain feasible solutions, the numbers of active units are checked. If no feasible solutions can be obtained, modification will be needed by increasing or decreasing the number of active units. Otherwise, the numbers will be used directly for SQP to be applied to obtain continuous variables for the entire problem. Solutions quality is quantified by the gap between the lower bound and the feasible cost.

C. Solving the Primary–Secondary System

As compared with primary-only systems, one more coupling constraint (17) is relaxed. With fixed-speed pumps and variable-speed chillers used, mass balance constraints between chillers and fixed-speed pumps are difficult to be satisfied leading to convergence issues. The reason is that the water of chillers is calculated based on water supply/return temperatures and chiller cooling load while the water of fixed-speed pumps depends on the number of the active pumps. To resolve the issues, penalties for the mass balance constraints are also added. The relaxed problem is as follows:

$$\min_{\substack{N_{Ach,S}, N_{Ach,B}, N_{Act}, N_{App,S}, N_{App,B} \\ N_{Acp,S}, N_{Acp,B}, N_{Asp}, T_{chws}, T_{cws}, \dot{m}_{sp}}} L_{ps}$$

with

$$\begin{aligned} L_{ps} \equiv & J_{ps} + \lambda_{mb,pp}^k (\dot{M}_{ppw} - \dot{M}_{chchw}) + \lambda_{mb,cp}^k (\dot{M}_{cpw} - \dot{M}_{chcw}) \\ & + \lambda_{mb,ct}^k (\dot{M}_{chcw} - N_{Act} \dot{m}_{ctcw}) \\ & + \lambda_{mb,se}^k (\dot{M}_{chchw} - \dot{m}_{decoupler} - N_{Asp} \dot{m}_{sp}) \\ & + \lambda_{Tcws}^k (T_{cws_ch} - T_{cws_ct}) + \lambda_{Tcwr}^k (T_{cwr_ch} - T_{cwr_ct}) \\ & + 0.5c^k (\dot{M}_{ppw} - \dot{M}_{chchw})^2 + 0.5c^k (\dot{M}_{chcw} - N_{Act} \dot{m}_{ctcw})^2 \\ & + 0.5c^k (\dot{M}_{chchw} - \dot{m}_{decoupler} - N_{Asp} \dot{m}_{sp})^2 \\ & + 0.5c^k (T_{cws_ch} - T_{cws_ct})^2 + 0.5c^k (T_{cwr_ch} - T_{cwr_ct})^2 \end{aligned} \quad (29)$$

subject to (1)–(4), (6), (8), (11)–(15), and (17)–(19), where \dot{M}_{ppw} is the total water mass flow rate of primary pumps, \dot{M}_{cpw} is the total water mass flow rate of condenser pumps, and \dot{M}_{chcw} is the total condenser water mass flow rate of chillers.

For the chiller subproblem, with big and small chillers used, distribution of cooling load among chillers is proportional to chiller capacity. In addition, different from primary-only systems, primary-secondary systems cannot be solved by only considering the number of active chillers. One more step is needed before the algorithm for the primary-only chiller subproblem is applied: listing the cooling capacity of the combinations of big and small chillers in an increasing order.

Then the first two combinations that satisfy the cooling requirements are used and the final solution for the subproblem is the one with lower cost. Cooling tower subproblem is the same as that of primary-only systems and is not repeated. Since primary/condenser pump subproblems are simple without constraints and only the number of active pumps needs to be determined, all the numbers of active pumps that satisfy cooling load requirements are tested and the solution is the one with minimum cost. The new subproblem, secondary pump subproblem, is similar to primary pump subproblem of the primary-only system and details are not repeated here.

D. Scalability Analysis

For primary-only systems, there are four subproblems, and iterations are counted after solving four subproblems once. Computational requirement CR_{po} depends on the computational time of subproblems and the number of iterations

$$CR_{po} = N(2t_{ch} + 2t_{ct} + n_{pp}t_{pp} + n_{cp}t_{cp}) \quad (30)$$

where N is the number of iterations, and n_{pp} and n_{cp} are the number of cases considered in primary pump and condenser pump subproblems, respectively. Generally, computational time for solving the four subproblems once is almost fixed since pump subproblems are easy to be solved, and the number of cases considered for chiller and cooling tower subproblems is constant. As a result, computational requirement for a primary-only system depends on the number of iterations. When the number of units in the system increases, CR_{po} increases almost linearly as implied in our testing results.

As mentioned earlier, with chillers of two sizes used, primary–secondary systems need one more step to generate the list for different combinations of chillers. Therefore, computational requirement CR_{ps} for a primary–secondary system is formulated as follows:

$$CR_{ps} = N(2t_{ch} + 2t_{ct} + n_{pp}t_{pp} + n_{cp}t_{cp} + n_{sp}t_{sp} + t_{list}) \quad (31)$$

where n_{sp} is the number of cases considered in secondary pump subproblem, and t_{list} is the time for generating the capacity list. Since there is no complex calculation in obtaining the list, the conclusion is similar to that of primary-only systems. As shown in our results, when the number of units in the system increases, CR_{ps} increases almost linearly.

V. NUMERICAL TESTING

Our optimization models and methods presented above have been implemented in MATLAB 2013a on a Core i7 2.8-GHz laptop with 16-GB memory. The *fmincon* function with the SQP algorithm from MATLAB is used. Three examples are tested and each example has two cases where Case 1 is for primary-only systems and Case 2 is for primary–secondary systems. In the first example, simple chiller plants are studied to explain the ideas of our methods. In the second example, chiller plants based on UTC Supervisory Control Synthesis project [31] are studied to show the performance of our methods under different cooling load requirements, and energy savings as compared with benchmark strategies. In the last example, scalability of our methods is demonstrated with the sizes of the plants increased.

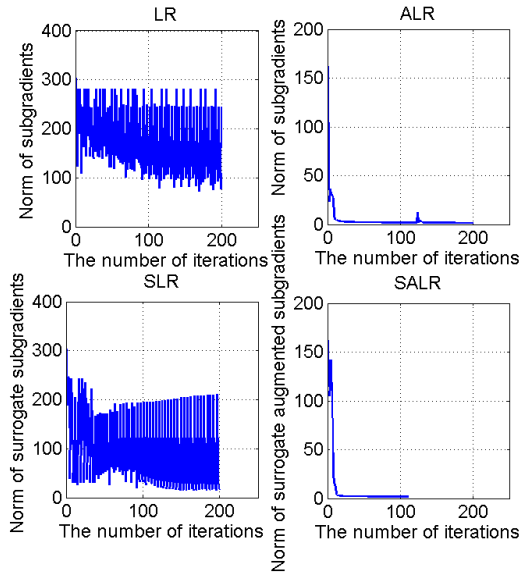


Fig. 4. Norms obtained by using LR, ALR, SLR, and SALR.

1) *Example 1 (Simple Chiller Plants)*: Simple chiller plants are used to explain the ideas of our methods. Importance of considering T_{chws} and T_{cws} as decision variables in saving energy is shown by the power consumption with and without such temperatures optimized. Convergence of our methods is implicated by the norm of surrogate augmented subgradients. Two cases are considered where Case 1 is for a primary-only chiller plant and Case 2 is for a primary–secondary chiller plant. For both plants, each subsystem has two units.

a) *Case 1 (Primary-only chiller plant)*: The plant consists of two identical chillers ($Q_{ch,ref}$: 1407 KW, $P_{ch,ref}$: 8498.28 KW), two identical primary pumps ($\dot{m}_{ppw,nom}$: 77.78 kg/s), two condenser pumps ($\dot{m}_{cpw,nom}$: 103.89 kg/s), and two identical cooling towers ($\dot{m}_{wdesign}$: 85 kg/s and $\dot{m}_{adesign}$: 125 kg/s). Model parameters are from [21] and [32], or suggested by industry partners. Wet-bulb temperature T_{wb} is 17.11 °C and load requirement Q_{demand} is 1300 KW. The penalty coefficient c is updated as $c = c_0 \times \beta^k$ with $c_0 = 0.5$ and $\beta = 1.004$. Stopping criteria is $||\lambda^k - \lambda^{k-1}|| \leq \varepsilon |\lambda^k|$ with $\varepsilon = 0.005$.

By using our method, the total power consumption obtained is 167.41 KW and the computational time is 5.09 s. Convergence is implicated by the norm of surrogate augmented subgradients in the last plot of Fig. 4. The value reduces to zero showing that relaxed coupling constraints are satisfied. A lower bound, 167.33 KW, is obtained at convergence. The gap between the lower bound and the feasible cost is 0.03%, which is small showing that the quality of the solutions is good.

To show the importance of T_{chws} and T_{cws} in saving energy, conventional rules $T_{chws} = 6.5$ °C and $T_{cws} = T_{wb} + 4$ °C are used for comparison. With T_{cws} optimized and T_{chws} fixed, the total power consumption is 170.79 KW and CPU is 3.21 s. With T_{chws} and T_{cws} fixed, the total power consumption is 185.59 KW and CPU time is 2.70 s. According to the results, power consumption is increased when more

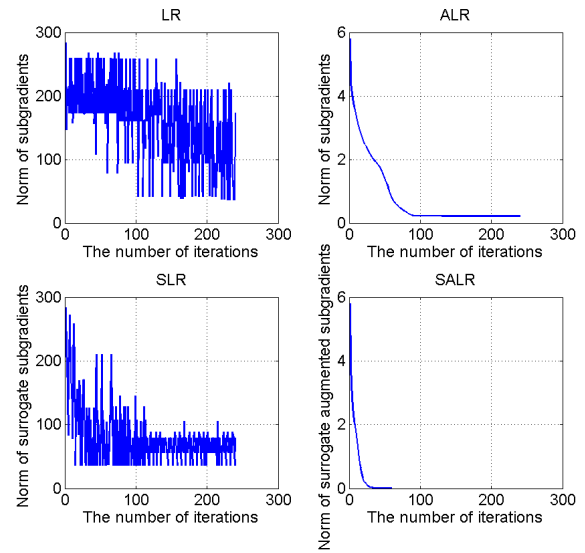


Fig. 5. Norms obtained by using LR, ALR, SLR, and SALR.

supply temperatures are fixed though CPU is decreased. The reason is that with supply temperatures fixed, complication and nonlinearity of chiller power consumption is decreased while there is limitation in improving chiller efficiency. Therefore, it is important to optimizing both T_{chws} and T_{cws} in saving energy.

To show convergence of our method, LR, ALR, and SLR mentioned in Section II are used for comparison. As shown in Fig. 5, convergence is not achieved by using LR and SLR since the norms are still oscillating. This difficulty is resolved by using ALR and SALR with penalties on coupling constraints that are difficult to be satisfied. The lower bound obtained by using ALR is 167.23 KW which is similar to that of using SALR, and CPU time is 32.97 s which is higher than that of using SALR. It takes around 120 iterations for ALR to converge, while around 60 iterations for SALR. The results show that convergence is achieved by both SALR and ALR, while SALR converges faster than ALR.

b) *Case 2 (Primary–secondary chiller plant)*: Different from Case 1, chillers, primary pumps, and condenser pumps are of two sizes. Identical secondary pumps are used. Cooling towers and the big chiller are the same as those of Case 1. For the small chiller, $P_{ch,ref}$ is 4942.09 KW and $Q_{ch,ref}$ is 703 KW. For the big primary pump, power consumption is 26 KW and mass flow rate is 81 kg/s. For the small one, the values are 13 KW and 40.5 kg/s, respectively. For the big condenser pump, power consumption is 31 KW and mass flow rate is 94 kg/s. For the small one, the values are 15.5 KW and 47 kg/s, respectively. For secondary pumps, nominal mass flow rate is 77.78 kg/s. Initial penalty coefficient c_0 is 0.1.

As mentioned earlier, mass balance constraints between variable-speed chillers and fixed-speed pumps are difficult to be satisfied leading to convergence issues. By using our method, the issues are resolved and convergence is achieved as shown in the last plot of Fig. 5. A short CPU time of 5.86 s is obtained with around 35 iterations. The lower bound is 206.95 KW and the feasible cost is 211.53 KW. The gap is 2.17% showing the quality of the solutions is good.

Similarly as in Case 1, the importance of T_{chws} and T_{cws} is studied. With T_{cws} optimized and T_{chws} fixed, the total power consumption is 255.69 KW. With T_{chws} and T_{cws} fixed, the total power consumption obtained is 257.37 KW. According to the results, power consumption reduction for the primary–secondary system with T_{chws} and T_{cws} optimized is not as good as in Case 1. This is because chiller power consumption is minimized by striking the balance between the amount of water and water temperatures. With fixed-speed pumps used, the freedom of improving chiller efficiency by adjusting T_{chws} and T_{cws} is not as good as for primary-only systems.

Similarly, LR, ALR, and SLR are used for comparison to show convergence of our method. The results are similar to primary-only systems: Convergence is not achieved by using LR and SLR but achieved by using ALR and SALR, as shown in Fig. 5. The lower bound by using ALR is 206.12 KW which is similar to that of using SALR, and CPU time is 19.01 s which is higher than that of using SALR. It takes around 100 iterations for ALR to converge while around 35 iterations for SALR. The results imply that SALR is the best among LR, SLR, and ALR.

2) *Example 2 (Performances and Energy Savings)*: Based on UTC Supervisory Control Synthesis project, chiller plants with four units in each subsystem are tested for primary-only and primary–secondary systems. With cooling load requirements from 10% to 100% of the plant capacity, the performance of our methods is demonstrated. In addition, based on a modified cooling load profile from UCONN’s chiller plant, energy savings by using our methods as compared with benchmark strategies are shown.

a) *Case 1 (Primary-only chiller plant)*: Parameters are the same as in Example 1 and plant capacity is 5628 KW. By using our method, total power consumption, the number of active units, T_{chws} and T_{cws} are obtained and shown as follows.

According to Fig. 6, chiller plant power consumption increases as the cooling load requirement increases. The numbers of active chillers are [1;1;2;2;2;3;3;4;4;4], the number of active cooling towers are [2;4;4;4;4;4;4;4;4], the number of active primary pumps are [2;4;4;4;4;4;4;4;4], and the number of active condenser pumps are [2;4;4;4;4;4;4;4;4]. Since power consumption of the units (e.g., pumps) is non-linear, more active units may consume less energy for certain cooling load requirements. Chilled water supply temperature and condenser water supply temperature are optimized and shown in Fig. 7. With high T_{chws} , chillers do not need to work hard so that energy is saved. When the cooling load requirement is approaching the plant capacity, all the units are active. To satisfy the cooling load requirement, T_{chws} decreases as (1) implies. With low T_{cws} , cooling towers need to work hard while chiller efficiency is increased. Balance between chiller power consumption and cooling tower consumption are achieved with T_{cws} optimized.

Computational time, gaps, and norms of surrogate augmented subgradients by using our method are shown in Table I for cooling load percentage Q_P from 10% to 100%. Convergence is implicated by the small norms. Short CPU time

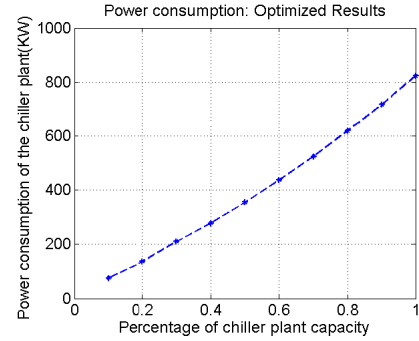


Fig. 6. Chiller plant power consumption of the primary-only system.

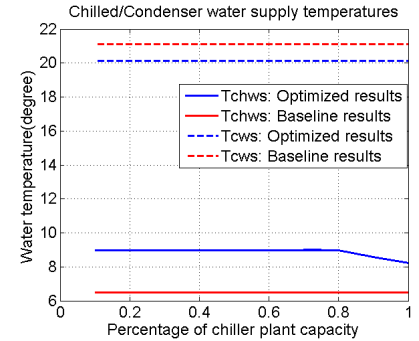


Fig. 7. Water supply temperatures of the primary-only system.

TABLE I
CPU TIME, GAPS, AND NORMS OF SURROGATE AUGMENTED SUBGRADIENTS

Q_p	0.1	0.2	0.3	0.4	0.5
CPU(s)	4.75	6.88	4.39	4.45	5.46
GAP(%)	0	0.01	0.01	0.15	0.37
NORM	0.35	1.57	1.34	1.13	1.43
Q_p	0.6	0.7	0.8	0.9	1.0
CPU(s)	5.71	6.03	6.27	8.53	10.28
GAP(%)	0.69	1.04	1.51	1.88	2.33
NORM	1.86	2.14	2.49	2.50	2.49

(e.g., maximum value 10.28 s and minimum value 4.75 s) and small gaps (e.g., maximum value 2.50% and minimum value 0) show that the performance of our method is good.

To show energy savings as compared with a baseline and the importance of considering T_{chws} and T_{cws} as decision variables, a baseline with following strategies is used: 1) T_{chws} is 6.5 °C and T_{cws} is 21.1 °C; 2) an additional chiller is turned ON when the requirement exceeds 90% of total capacity of current active chillers; and 3) the number of active chillers, cooling towers, primary pumps, and condenser pumps are the same. Results for power reduction are shown as follows.

According to Fig. 8, chillers consume much more energy than that of other subsystems. With T_{chws} and T_{cws} optimized, chillers consume less energy than that of the baseline. With the number of active primary pumps optimized, primary pump power consumption of optimized results is less than that of the baseline. To improve chiller efficiencies, more condenser

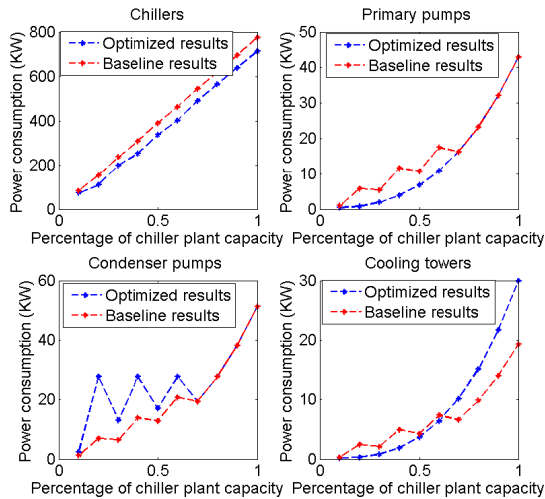


Fig. 8. Subsystem power consumption comparison.

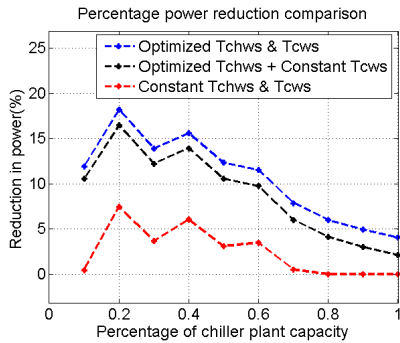


Fig. 9. Percentage power reduction comparison between our method and the baseline.

water with low temperature is provided to chillers. Therefore, condenser power consumption of optimized results is higher than that of the baseline. When cooling requirements are high, all the pumps are active for both optimized results and the baseline, therefore, power consumption of the pumps are the same. Cooling tower power consumption of optimized results is higher than that of the baseline. This is reasonable since chiller efficiency is improved with low T_{cws} while cooling towers need to work hard as mentioned earlier.

According to Figs. 9 and 10, power consumption obtained by using our method is significantly reduced as compared with the baseline, with maximum percentage power reduction around 18% and maximum power reduction around 60 KW. Since percentage power reduction is the ratio of power consumption to load requirements, the percentage is low when the requirement is high. In addition, the results show that with both T_{chws} and T_{cws} optimized, more energy is saved as compared with the results with T_{cws} fixed and results with T_{cws} and T_{chws} fixed. Therefore, it is important to consider both T_{chws} and T_{cws} as decision variables for energy savings.

b) Case 2 (Primary–secondary chiller plant): For chillers and fixed-speed pumps, three big units and one small unit are used in each subsystem. Parameters are the same as in Example 1. Results for chiller plant power consumption, chilled/condenser water supply temperature and surplus water mass flow rate are shown in Figs. 11, 12 and 13, respectively.

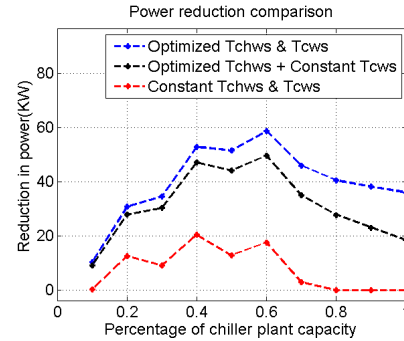


Fig. 10. Power reduction based on our method compared with the baseline.

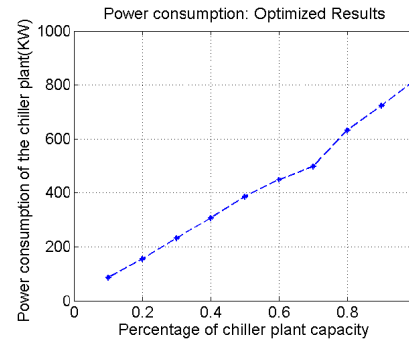


Fig. 11. Chiller plant power consumption of the medium primary–secondary system.

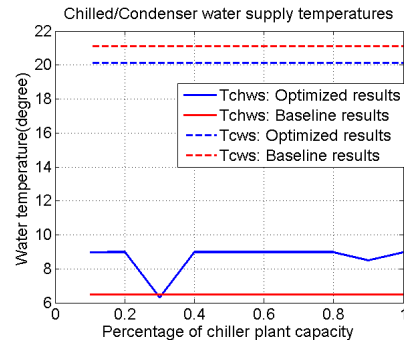


Fig. 12. Water supply temperatures of the medium primary–secondary system.

As Fig. 11 shows, power consumption increases as cooling load requirement increase.

Chilled water supply temperature T_{chws} and condenser water supply temperature T_{cws} are optimized to improve chiller efficiency. The results are similar to those of Case 1: T_{chws} is relatively high and T_{cws} is relatively low as compared with the baseline, as shown in Fig. 12. This is reasonable since a chiller does not need to work that hard with low T_{cws} and high T_{chws} .

As shown by the blue line in Fig. 13, extra water by using our method is much reduced as compared with the baseline. There are jumps in both optimized results and baseline results. For optimized results, with fixed-speed pumps used, T_{chws} decreases to satisfy the requirement as the cooling

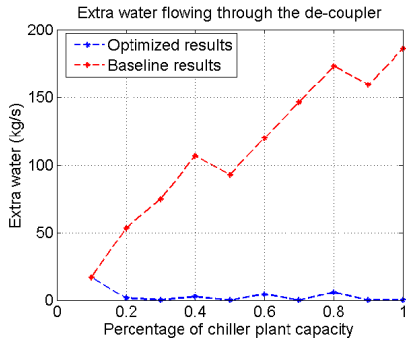


Fig. 13. Surplus water mass flow rate of the primary–secondary system.

TABLE II
CPU TIME, GAPS, AND NORMS OF SURROGATE AUGMENTED SUBGRADIENTS

Q_p	0.1	0.2	0.3	0.4	0.5
CPU(s)	11.65	9.17	8.75	9.66	9.08
GAP(%)	1.88	1.59	0.63	0.88	0.68
NORM	0.50	0.50	0.49	0.49	0.49
Q_p	0.6	0.7	0.8	0.9	1.0
CPU(s)	8.76	3.34	8.54	7.15	7.30
GAP(%)	0.90	0.04	0.70	2.30	0.74
NORM	0.68	0.73	1.04	1.05	1.04

load requirement increases. When a new pump is turned ON, chilled water is suddenly increased leading to more extra water in the decoupler. Then, T_{chws} is increased and the amount of extra water is reduced for energy savings. For the baseline, with constant T_{chws} , extra water decreases as cooling load requirements increase until a new pump is active.

By using our method, CPU, gaps, and norms of surrogate augmented subgradients are obtained and shown in Table II. Convergence is implicated by the small norms. Our method achieves short CPU time (e.g., maximum value 11.65 s and minimum value 3.34 s) with small gaps (e.g., maximum value 1.04% and minimum value 0.49%), demonstrating that the performance of our method is good.

To see energy consumption by using our method as compared to a baseline and the importance of considering T_{chws} and T_{cws} as decision variables, a baseline whose strategies are similar to those of the primary-only system is used. Since we have big and small units, the difference is that when load requirement exceeds 90% of current total capacity, if the small chiller is offline, it will be turned ON. Otherwise, it will be replaced by a big chiller. By using our method, results for subsystem power consumption and power reduction are obtained and shown as follows.

Fig. 14 shows power consumption of subsystems in the primary–secondary system. For chillers, each time when T_{chws} decreases to satisfy the increasing cooling load requirement, chiller power consumption is increased until a new primary pump is turned ON. For secondary pumps, secondary pump power consumption increases when the requirement is small. This is because with relatively high T_{chws} , more water is

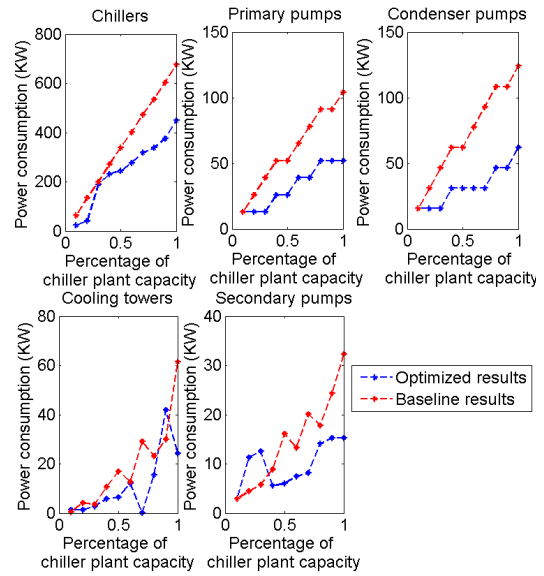


Fig. 14. Subsystem power consumption comparison.

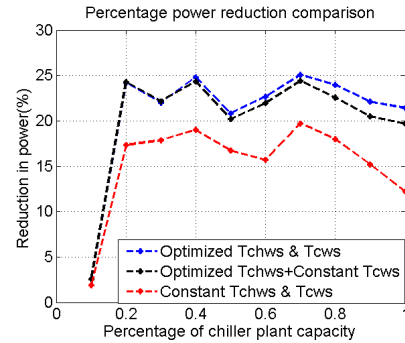


Fig. 15. Percentage power reduction for the primary–secondary chiller plant based on our method compared with the baseline.

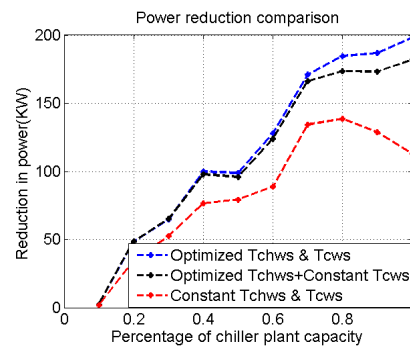


Fig. 16. Power reduction for the primary–secondary chiller plant based on our method compared with the baseline.

needed to satisfy the requirement as (18) implies. Secondary pump power consumption decreases when more secondary pumps are active since secondary pump power consumption is nonlinear.

According to Figs. 15 and 16, power consumption of the plant obtained by using our method is significantly reduced as compared with the baseline, with maximum percentage

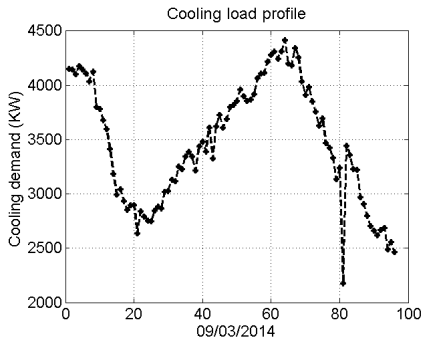


Fig. 17. Cooling load profile.

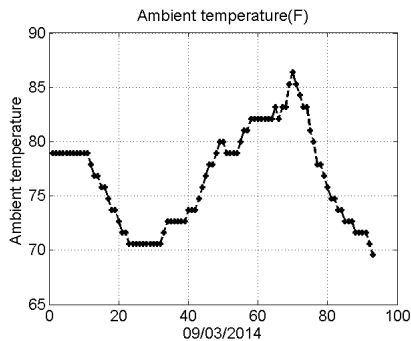


Fig. 18. Ambient temperature.

power reduction around 25% and maximum power reduction around 200 KW. The main reason is that with the number of active fixed-speed pumps optimized, surplus water, and power consumption of primary pumps and condenser pumps is reduced, as shown in Figs. 13 and 14. According to the results, power reduction of the primary–secondary system is better than that of the primary-only system by using our methods. Since with fixed-speed pumps used, chilled water provided by primary pumps is often more than that is needed resulting in extra water, and the energy used to produce the extra water is wasted. By improving chiller efficiency and reducing the extra water, total power consumption is significantly reduced. The result implies that optimizing primary–secondary systems is even more important. In addition, the results show that with both T_{chws} and T_{cws} optimized, more energy is saved as compared with the results with T_{cws} fixed and results with T_{cws} and T_{chws} fixed. However, the improvement is not as good as those for primary-only systems since with fixed-speed pumps used, there is limitation in the freedom of adjusting water temperatures.

Generally, cooling load requirements are low when it is cold and high when it is hot. To evaluate the power consumption of primary-only and primary–secondary systems by using our methods, a cooling load profile of UCONN’s chiller plant on September 3, 2014 is scaled and used. The data for the cooling loads and the ambient temperatures are recorded every 15 min and shown in Figs. 17 and 18.

Power consumption by using our methods and baseline strategies for primary-only and primary–secondary systems is as follows.

TABLE III
COMPARISON OF BASELINE AND OUR ALGORITHM

Configurations		Power consumption
Primary-only system	Baseline	60626
	Our algorithm	37761
Primary-secondary system	Baseline	78228
	Our algorithm	51957

TABLE IV
GAP AND THE NORM OF SURROGATE AUGMENTED SUBGRADIENTS

Q_p	0.1	0.2	0.3	0.4	0.5
GAP(%)	0	0.26	0.10	0.34	0.59
Norms	1.55	2.16	2.45	2.47	2.49
Q_p	0.6	0.7	0.8	0.9	1.0
GAP(%)	0.87	1.17	1.53	1.94	2.35
Norms	2.49	2.49	2.49	2.50	2.49

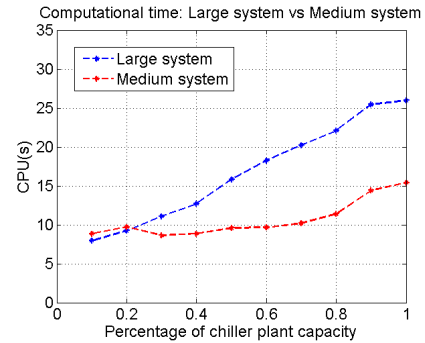


Fig. 19. CPU comparison between the large and the medium systems.

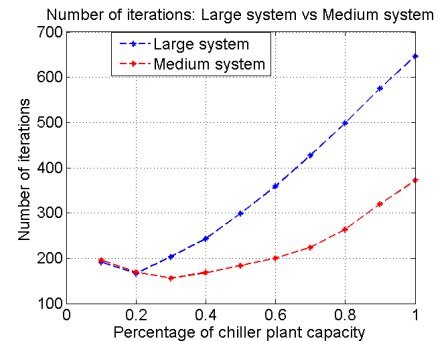


Fig. 20. Number of iterations comparison between the large and the medium systems.

As shown in Table III, power consumption is significantly reduced by using our method with 37.71% for the primary-only system and 33.58% for the primary–secondary system as compared with baseline strategies.

3) *Example 3 (Scalability Evaluation)*: In this example, a large chiller plant with the number of units in each subsystem doubled as those of Example 2 is studied to show scalability of our methods.

a) *Case 1 (Primary-only chiller plant)*: Parameters of units, penalty coefficient, and stopping criteria are the same as those of Example 1. By using our method, gaps and norms of surrogate augmented subgradients are shown in Table IV. The small values of norms show that convergence is achieved

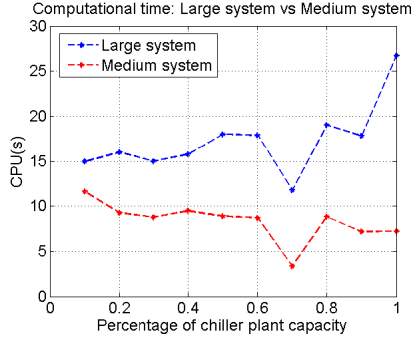


Fig. 21. CPU comparison between the large and the medium systems.

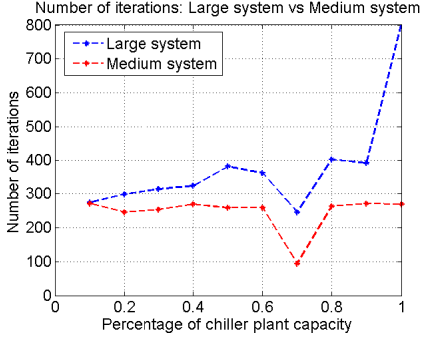


Fig. 22. Number of iterations comparison between the large and the medium systems.

TABLE V

GAP AND THE NORM OF SURROGATE AUGMENTED SUBGRADIENTS

Q_p	0.1	0.2	0.3	0.4	0.5
GAP(%)	1.83	0.06	0.10	0.23	1.55
Norms	0.09	0.09	1.98	0.99	2.98
Q_p	0.6	0.7	0.8	0.9	1.0
GAP(%)	0.66	0.01	2.06	1.44	0
Norms	1.99	1.39	2.59	1.99	2.99

by using our methods and the small values of gaps show that the quality of the solutions is good.

To show the scalability of our method, computational time of the plants is shown in Fig. 19 and the numbers of iterations are shown in Fig. 20. According to the results, as the size of the problem increases, both CPU time and the number of iterations increase almost linearly. Additionally, the results imply that the time of solving the four subproblems is almost fixed. Scalability of our method is demonstrated.

b) Case 2 (Primary–secondary chiller plant): Parameters of units, penalty coefficient, and stopping criteria are the same as those of Example 1. By using our method, gaps and norms of surrogate augmented subgradients in Table V. The small values of norms show that convergence is achieved by using our methods and the small values of gaps show that the quality of the solutions is good.

To show scalability of our method, CPU and the number of iterations for primary–secondary systems are as follows.

Similar as for primary-only systems, the relationship of CPU time and the number of iterations of the large and the medium systems is almost linear. Since computational effort of solving the five subproblems is almost fixed as implied in Figs. 21 and 22, computational time of the system depends

on the number of iterations and increases almost linearly from the medium system. Table V shows that the performance of our algorithm is good with small gaps. Scalability of our method is demonstrated.

VI. CONCLUSION

In this paper, chiller plant optimization with good decision variables such as chilled/condenser water supply temperatures is studied for energy savings. To solve the mixed-integer nonlinear problem efficiently, a novel decomposition and coordination approach, SALR, combined with SQP is developed. Numerical testing demonstrates that our approach is scalable and its performance is good with short CPU time and small gaps. Additionally, major energy savings are achieved as compared with benchmark strategies. The approach provides a powerful way to solve chiller plant optimization problems and beyond.

APPENDIX

Coefficients of units used in Example 1 Case 1 are as follows.

1) Coefficients of big chillers

$$\mathbf{a} = [1.042261, 2.644821e-3, -1.468026e-3, \\ \times 1.366256e-2, -8.302334e-4, 1.573579e-3]$$

$$\mathbf{b} = [1.026340, -1.612819e-2, -1.092591e-3, \\ -1.784393e-2, 7.961842e-4, -9.586049e-5]$$

$$\mathbf{d} = [1.188880e-1, 6.723542e-1, 2.068754e-1].$$

2) Coefficients of primary pumps $k_{pp} = 70.69$, $A_{pp} = 9.04$, and $k = [0.507, 1.280, -1.420, 0.584]$.

3) Coefficients of condenser pumps $k_{cp} = 76.62$, $A_{cp} = 8.54$, and $k = [0.507, 1.280, -1.420, 0.584]$.

4) Coefficients of cooling towers

$$\mathbf{c} = [-3.597412e-1, -5.505361e-2, 2.385043e-3, \\ 1.739269e-1, -2.484738e-2, 4.843020e-4, \\ -5.589849e-3, 5.770071e-4, -1.342427e-5, \\ \times 2.847658, -1.217651e-1, 1.459924e-3, \\ \times 1.680420, -1.669208e-2, -7.190532e-4, \\ -2.548519e-2, 4.874917e-5, 2.719234e-5, \\ -6.537663e-2, -2.278167e-3, 2.500254e-4, \\ -9.105654e-2, 3.181763e-3, 3.862177e-5, \\ -3.428538e-3, 8.565899e-6, -1.516821e-6].$$

Coefficients of units used in the Example 1 Case 2 are

1) Coefficients of small chillers

$$\mathbf{a} = [1.042261, 2.644821e-3, -1.468026e-3, \\ 1.366256e-2, -8.302334e-4, 1.573579e-3]$$

$$\mathbf{b} = [1.026340, -1.612819e-2, x - 1.092591e-3, \\ -1.784393e-2, 7.961842e-4, -9.586049e-5]$$

$$\mathbf{d} = [1.188880e-1, 6.723542e-1, 2.068754e-1].$$

2) Coefficients of secondary pumps

$k_{sp} = 70.69$, $A_{sp} = 9.04$, and $k = [0.507, 1.280, -1.420, 0.584]$.

ACKNOWLEDGMENT

The authors would like to thank Dr. T. Bailey and C. Netter from United Technologies for their guidance and constructive comments on the research. Any opinions expressed herein are those of the authors and do not represent those of the sponsor.

REFERENCES

- [1] *Armstrong Fluid Technology Sees HVAC-Industry Future in Design Envelope*. Accessed: Sep. 10, 2016. [Online]. Available: <http://hvac.com/piping-pumping/armstrong-fluid-technology-sees-hvac-industry-future-design-envelope>
- [2] *Air-Conditioning*. Accessed: Sep. 21, 2016. [Online]. Available: <http://psgfacilityservices.com/services/air-conditioning>
- [3] *Variable Primary Chilled Water Systems*. Accessed: Sep. 10, 2016. [Online]. Available: http://www.tranehk.com/en/tech_info/doc/2_vpf%20hong%20kong%20rv1.pdf
- [4] T. B. Hartman, "Design issues of variable chilled-water flow through chillers," *ASHRAE Trans.*, vol. 102, no. 2, pp. 679–683, 1996.
- [5] W. Kirsner, "The demise of the primary-secondary pumping paradigm for chilled water plant design," *HPAC Eng.*, vol. 68, no. 11, pp. 73–78, 1996.
- [6] T. Hartman, "All-variable speed centrifugal chiller plants," *ASHRAE J.*, vol. 43, no. 9, pp. 43–53, 2001.
- [7] S. T. Taylor, "Primary-only vs. Primary-secondary variable flow systems," *ASHRAE J.*, vol. 44, no. 2, pp. 25–29, 2002.
- [8] Q. Shi, Y. Kuang, and W. K. Kwok, "Energy-saving control strategies for Shanghai International Financial Center (IFC) chiller plant system," *Building Energy Efficiency*, vol. 6, pp. 24–31, 2013.
- [9] M. A. Bragin, P. B. Luh, J. H. Yan, N. Yu, and G. A. Stern, "Convergence of the surrogate Lagrangian relaxation method," *J. Optim. Theory Appl.*, vol. 164, no. 1, pp. 173–201, 2015.
- [10] L. Lu, W. J. Cai, Y. C. Soh, and L. H. Xie, "Global optimization for overall HVAC systems—Part II problem solution and simulations," *Energy Convers. Manage.*, vol. 46, nos. 7–8, pp. 1015–1028, 2005.
- [11] L. Lu, W. J. Cai, Y. C. Soh, L. H. Xie, and S. J. Li, "HVAC system optimization—Condenser water loop," *Energy Convers. Manage.*, vol. 45, no. 4, pp. 613–630, 2004.
- [12] A. R. Simpson, G. C. Dandy, and L. J. Murphy, "Genetic algorithms compared to other techniques for pipe optimization," *J. Water Resour. Planning Manage.*, vol. 120, no. 4, pp. 423–443, 1994.
- [13] W. Huang and H. N. Lam, "Using genetic algorithms to optimize controller parameters for HVAC systems," *Energy Buildings*, vol. 26, no. 3, pp. 277–282, 1997.
- [14] Y. C. Chang, "Genetic algorithm based optimal chiller loading for energy conservation," *Appl. Thermal Eng.*, vol. 25, no. 17, pp. 2800–2815, 2005.
- [15] Y.-C. Chang, J.-K. Lin, and M.-H. Chuang, "Optimal chiller loading by genetic algorithm for reducing energy consumption," *Energy Buildings*, vol. 37, no. 2, pp. 147–155, 2005.
- [16] Y.-C. Chang, "An innovative approach for demand side management—Optimal chiller loading by simulated annealing," *Energy*, vol. 31, no. 12, pp. 1883–1896, 2006.
- [17] A. J. Ardakani, F. F. Ardakani, and S. H. Hosseini, "A novel approach for optimal chiller loading using particle swarm optimization," *Energy Buildings*, vol. 40, no. 12, pp. 2177–2187, 2008.
- [18] J. Sun and A. Reddy, "Optimal control of building HVAC&R systems using complete simulation-based sequential quadratic programming (CSB-SQP)," *Building environment*, vol. 40, no. 5, pp. 657–669, 2005.
- [19] R. T. Olson and J. S. Liebman, "Optimization of a chilled water plant using sequential quadratic programming," *Eng. Optim.*, vol. 15, pp. 171–191, Apr. 1990.
- [20] Z. Liu, F. T. Song, Z. Y. Jiang, X. Chen, and X. H. Guan, "Optimization based integrated control of building HVAC system," *Building Simul.*, vol. 7, no. 4, pp. 375–387, 2014.
- [21] Y. Yao and J. Chen, "Global optimization of a central air-conditioning system using decomposition-coordination method," *Energy Buildings*, vol. 42, no. 5, pp. 570–583, 2010.
- [22] Z. Q. Zhang, W. D. Turner, Q. Chen, C. Xu, and S. Deng, "Optimization of a chilled water plant without storage using a forward plant model," in *Proc. Enhanced Building Oper. Conf.*, 2010. [Online]. Available: <http://hdl.handle.net/1969.1/94064>
- [23] S. T. Taylor, "Optimizing design and control of chilled water plants," *ASHRAE J.*, vol. 54, no. 3, p. 60, 2011.
- [24] Z. J. Ma and S. W. Wang, "An optimal control strategy for complex building central chilled water systems for practical and real-time applications," *Building Environ.*, vol. 44, no. 6, pp. 1188–1198, 2009.
- [25] Z. J. Ma and S. W. Wang, "Supervisory and optimal control of central chiller plants using simplified adaptive models and genetic algorithm," *Appl. Energy*, vol. 88, no. 1, pp. 198–211, 2011.
- [26] *DOE-2 Reference Manual, Part 1, Version 2.1*, Berkeley Lawrence Berkeley National Laboratories, CA, USA, 1980.
- [27] M. Hydeman and K. L. Gillespie, "Tools and techniques to calibrate electric chiller component models/discussion," *ASHRAE Trans.*, vol. 108, p. 733, Jan. 2002.
- [28] M. Wetter, W. D. Zuo, T. S. Noudui, and X. F. Pang, "Modelica buildings library," *J. Building Perform. Simul.*, vol. 7, no. 4, pp. 989–999, 2014.
- [29] Y. Yao, Z. Lian, Z. Hou, and X. Zhou, "Optimal operation of a large cooling system based on an empirical model," *Appl. Thermal Eng.*, vol. 24, no. 16, pp. 2303–2321, 2004.
- [30] X. Sun, P. B. Luh, M. A. Bragin, Y. Chen, J. Wan, and F. Wang, "A decomposition and coordination approach for large-scale security constrained unit commitment problems with combined cycle units," in *Proc. IEEE Power Energy Soc. Gen. Meet.*, 2017, pp. 1–5.
- [31] *Supervisory Control Synthesis for Energy Efficient BIS Chiller Plants*. Accessed: Dec. 10, 2016. [Online]. Available: http://utciase.uconn.edu/research/utci-projects/gupta_utci/
- [32] *Buildings.Fluid.Chillers.Data.ElectricEIR*. Accessed: May 4, 2016. [Online]. Available: https://simulationresearch.lbl.gov/modelica/releases/latest/help/Buildings_Fluid_Chillers_Data_ElectricEIR.html#Buildings.Fluid.Chillers.Data.ElectricEIR
- [33] K. Deng *et al.*, "Model predictive control of central chiller plant with thermal energy storage via dynamic programming and mixed-integer linear programming," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 2, pp. 565–579, Apr. 2015.
- [34] X. Li, Y. Li, J. E. Seem, and P. F. Li, "Dynamic modeling and self-optimizing operation of chilled water systems using extremum seeking control," *Energy Buildings*, vol. 58, pp. 172–182, Mar. 2013.



Danxu Zhang (S'17) received the B.E. degree in automation from the Harbin Institute of Technology, Harbin, China, in 2013, and the M.S. degree in electrical engineering from the University of Connecticut, Mansfield, CT, USA, in 2016, where she is currently pursuing the Ph.D. degree with the Electrical Engineering Department.

Her current research interests include optimization and energy-efficient chiller plant systems.



Peter B. Luh (S'76–M'80–SM'91–F'95–LF'16) received the B.S. degree from National Taiwan University, Taipei, Taiwan, the M.S. degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, and the Ph.D. degree from Harvard University, Cambridge, MA, USA.

He has been with the University of Connecticut, Mansfield, CT, USA, since 1980, where he is currently the SNET Professor of Communications and Information Technologies. He is also a member of the Chair Professors Group, Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing, China; and a member of the Thousand-Talent Program, the State Key Laboratory of Synthetical Automation for Process Industry, Northeastern University, Shenyang, China. His current research interests include smart and green buildings, smart power systems, and intelligent manufacturing.

Prof. Luh received the Robotics and Automation Society (RAS) 2013 Pioneer Award and the 2017 George Saridis Leadership Award. He is the Chair of the IEEE TAB Periodicals Committee during 2018–2019. He was the Vice President of Publications of the IEEE RAS during 2008–2011, the founding Editor-in-Chief of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING during 2003–2007, and the Editor-in-Chief of the IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION during 1999–2003.



Junqiang (James) Fan (SM'07) received the B.Eng. degree from Zhejiang University, Hangzhou, China, the M.Eng. degree from the Shenyang University of Technology, Shenyang, China, and the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada.

He has 24 years of research and development experience in industrial automation systems and control systems, and his control designs reside on numerous industrial installations around the world. He was with United Technologies Research Center (UTRC), East Hartford, CT, USA; Honeywell Autom. & Control Solutions, Vancouver, BC, Canada; Carnegie Mellon University, Pittsburgh, PA, USA; and the Chinese Academy of Sciences, Beijing, China. He is currently a fellow with United Technologies Climate, Controls and Security, and is responsible for model-based tools and methods in controls and diagnostics. He has published his research findings in 16 publications, and holds 12 granted and three pending U.S. patents.

Dr. Fan received three Outstanding Achievement Awards from UTRC, two Outstanding Technical Achievement Awards from Honeywell, and one Industrial Research Fellowship from the Natural Sciences and Engineering Research Council of Canada.



Shalabh Gupta (M'07) received the B.E. degree in mechanical engineering from IIT Roorkee, Roorkee, India, in 2001, the M.S. degrees in mechanical engineering and electrical engineering, and the Ph.D. degree in mechanical engineering from Pennsylvania State University, University Park, PA, USA, in 2004, 2005, and 2006, respectively.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Connecticut, Mansfield, CT, USA. His current research interests include distributed autonomy, cyber-physical systems, robotics, network intelligence, data analytics, information fusion, and fault diagnosis in complex systems.

Dr. Gupta is currently serving as the Specialty Chief Editor of the *Frontiers in Robotics and AI: Sensor Fusion and Machine Intelligence* and an Associate Editor of the *Structural Health Monitoring: An International Journal*. He is a member of the American Society of Mechanical Engineers.