PHOTOGRAPH OF THE PEPSICO SUSTAINABILITY CENTER IN CHICAGO. COURTESY OF SHANNON MASH.

Predictive Control for Energy Efficient Buildings with Thermal Storage

MODELING, SIMULATION, AND EXPERIMENTS

YUDONG MA, ANTHONY KELMAN, ALLAN DALY,
and FRANCESCO BORRELLI

The building sector is the largest energy consumer in the world. Therefore, it is economically, socially, and environmentally significant to reduce the energy consumption of buildings. Achieving substantial energy reduction in buildings may require rethinking the whole processes of design, construction, and operation of a building. This article focuses on the specific

Digital Object Identifier 10.1109/MCS.2011.2172532
Date of publication: 13 January 2012

1066-033X/12/$31.00©2012IEEE

issue of advanced control system design for energy efficient buildings.

Building controls design becomes especially challenging as practitioners move beyond standard heuristic controls approaches and seek to incorporate predictions of weather, occupancy, renewable energy availability, and energy price signals. [1], [2]. Model predictive control [3] (MPC) naturally enters the picture as a control methodology that can systematically use all the aforementioned predictions to improve building thermal comfort, decrease peak demand, and reduce total energy costs.

In buildings, performance improvement using forecasted information is possible due to two basic mechanisms. The first mechanism is referred to as *load shifting* or *active storage*. Load shifting consists of shaping the energy profile delivered to a building, exploiting the possibility of storing energy for later use. Thermal storage is inherent to a building's structure and can be increased by including additional external energy storage devices. The optimal profile of delivered energy depends on various factors which include time varying utility prices, availability of renewable energy, ambient temperature variation, and load shedding signals received from the utility grid. The second mechanism is component optimization. Buildings can be large systems with many control variables and degrees of freedom. Predictive models of building thermal dynamics and energy costs of control actuators allow computation of the optimal inputs to each actuator in order to deliver the desired energy profile. The two mechanisms are coupled in an actual MPC control strategy. It is challenging to isolate the contribution of each mechanism to the total performance improvement, especially under model mismatch and uncertain forecasts.

This article has two main objectives. In the first part we show the basic mechanism of active thermal storage and how this mechanism naturally emerges in a predictive control scheme. We use a thermal mass model in order to demonstrate a fundamental tradeoff between savings, losses, and uncertainty. This basic tradeoff also exists when realistic building models and performance indices are used. In the second part we present a more complex predictive control scheme with reduced order models for the building components and for thermal storage. Simulation and experimental tests are used to show the effectiveness of MPC. In particular, we demonstrate that MPC is able to systematically reproduce and coordinate a variety of established commercial solutions for energy savings, including demand response, economizer-mode, and precooling/preheating.

This article focuses on building heating, ventilation, and air conditioning (HVAC) systems where cooling and heating rely on centralized chilled water and hot water generation, respectively. The two mechanisms of load shifting and optimal component operation can be implemented during both energy conversion and energy distribution to the building spaces. We present the predictive control design for both systems. The ultimate goal is to provide the main ingredients of a predictive control framework needed for an actual implementation. At the same time we want to explain, through simple examples, the benefits of active thermal storage and the potential advantages and disadvantages of MPC over conventional building control sequences.

MODEL PREDICTIVE CONTROL AND THERMAL STORAGE: A SIMPLE EXAMPLE

To create thermally comfortable indoor environments, energy is converted and delivered to occupied spaces in buildings. In this article, occupant comfort is measured by the air temperature of a given space. The objective of this section is to use a simple thermal mass model to show the basic principles of active thermal storage, to demonstrate that active storage naturally emerges as a closed-loop behavior of an MPC scheme, and to discuss the main tradeoffs that arise in active thermal storage.

The temperature dynamics of a given space can be modeled using a resistance-capacitance (RC) circuit analogy

$$C\dot{T} = u + P_{\mathrm{d}} + (T_{\mathrm{oa}} - T)/R, \tag{1}$$

where T is the temperature of the space, P_{d} is the external disturbance load generated by occupants, direct sunlight, and electrical devices, T_{oa} is the temperature of outside air, and u is the heating and cooling power input to the space. The space is cooled when $u \leq 0$ and heated when $u \geq 0$.

In an energy conversion problem, the simplified thermal mass model (1) can be viewed as the abstraction of an entire building and the temperature T is an average temperature of all the building spaces. In an energy distribution problem, the simplified thermal mass model (1) can be viewed as the abstraction of one zone within a building, and T is the temperature measurement of the zone. In this case, the lumped parameter R describes the thermal resistance of walls and windows isolating the zone from the outside environment, and the parameter C lumps up the thermal capacitance of the zone. The active storage mechanism shares the same properties for both control problems.

The representation of system (1) in discrete time is obtained using Euler discretization with a sampling time of Δt,

$$T(k+1) = AT(k) + Bu(k) + d(k), \tag{2}$$

where

$$A = 1 - \frac{\Delta t}{RC},\ B = \frac{\Delta t}{C},\ d = \frac{P_{\mathrm{d}}\Delta t}{C} + \frac{T_{\mathrm{oa}}\Delta t}{RC}.$$

A simple MPC problem is formulated with the objective of minimizing total heating and cooling energy

consumption, minimizing the peak power consumption, and maintaining zones within a desired temperature range despite predicted load changes. At each time step, the predictive controller solves the following problem:

$$\min_{\mathbf{U}_t,\, \underline{\varepsilon},\, \overline{\varepsilon}} \sum_{k=0}^{N-1} |u_{t+k|t}|\Delta t + \kappa \max\{|u_{t|t}|, \ldots, |u_{t+N-1|t}|\} + \rho \sum_{k=1}^{N} (|\overline{\varepsilon}_{t+k|t}| + |\underline{\varepsilon}_{t+k|t}|) \tag{3}$$

subject to

$$T_{t+k+1|t} = AT_{t+k|t} + Bu_{t+k|t} + d_{t+k|t}, \tag{4}$$

$$\underline{T} - \underline{\varepsilon}_{t+k|t} \le T_{t+k|t} \le \overline{T} + \overline{\varepsilon}_{t+k|t}, \tag{5}$$

$$\underline{\varepsilon}_{t+k|t}, \overline{\varepsilon}_{t+k|t} \ge 0, \tag{6}$$

where the symbol $v_{t+k|t}$ is read as "the variable v at time $t+k$ predicted at time t." For instance, $T_{3|1}$ represents the predicted temperature at time 3 when the prediction is made at time $t = 1$ starting from the current temperature $T(1)$. It is, in general, different from $T_{3|2}$, which is the predicted temperature at time 3 when the prediction is made at time $t = 2$ starting from the current temperature $T(2)$. With this notation in place, $\mathbf{U}_t = [u_{t|t}, u_{t+1|t}, \ldots, u_{t+N-1|t}]$ is the vector of energy control inputs, $\underline{\varepsilon} = [\underline{\varepsilon}_{t+1|t}, \ldots, \underline{\varepsilon}_{t+N|t}]$ is the vector of temperature violations below the lower bound, $\overline{\varepsilon}$ collects the temperature violation above the upper bound, $T_{t+k|t}$ is the thermal zone temperature, and $d_{t+k|t}$ is the load prediction. $\underline{T}$ and $\overline{T}$ are the lower and upper bounds on the zone temperature, respectively. ρ is the penalty on the comfort constraint violation, κ is the penalty on peak power consumption, and N is the length of the prediction horizon.

Let $\mathbf{U}_t^* = \{u_{t|t}^*, \ldots, u_{t+N-1|t}^*\}$ be the optimal solution to (3)–(6) at time t. The first element of $\mathbf{U}_t^*$ is applied to system (2)

$$u(t) = u_{t|t}^*. \tag{7}$$

The optimization problem (3)–(6) is repeated at the next time step $t+1$ based on the new measured temperature $T_{t+1|t+1} = T(t+1)$, yielding a moving or receding horizon control strategy.

The following parameters are used in the simulations presented next. Thermal capacitance $C = 9.2 \times 10^3$ kJ/°C, thermal resistance $R = 50$ °C/kW, sampling time $\Delta t = 1$ hour, prediction horizon $N = 24$ hours, and thermal comfort interval $[\underline{T}, \overline{T}] = [21, 26]$ °C. Note that the plant model (2) and the model used in the MPC scheme (4) are the same. It is assumed that weather and load are periodic with a period of one day, and that their predictions are perfect without mismatch between predictions and actual measurements. The outside air temperature profile $T_{\text{oa}}(t)$ used in (1) is depicted in Figure 1(a), and the disturbance load profile $P_{\text{d}}(t)$ used in (1) is depicted in Figure 1(b). The load $P_{\text{d}}(t)$ resembles a 3-kW thermal load of a conference room with one meeting from 9:00 to 11:00.

Two controllers are considered. Controller C1 is a proportional controller designed to reject the load without predictive information. Controller C1 inputs zero power when the space temperature is within the comfort range, otherwise the proportional control law

$$u(t) = \begin{cases} K(\overline{T} - T(t)) & T(t) \ge \overline{T}, \\ 0 & \underline{T} < T(t) < \overline{T}, \\ K(\underline{T} - T(t)) & T(t) \le \underline{T} \end{cases} \tag{8}$$

is applied to system (2). Controller C2 implements the MPC problem (3)–(7).

Simulations of system (2) in closed-loop with C1 and C2 are performed until the system settles to steady-periodic behavior. The performance of the controllers is measured by the closed-loop total energy consumption

$$J^u = \sum_{k=0}^{N-1} |u^*(k)|\Delta t, \tag{9}$$

the peak power consumption

$$J^p = \max\{|u^*(0)|, \ldots, |u^*(N-1)|\}, \tag{10}$$

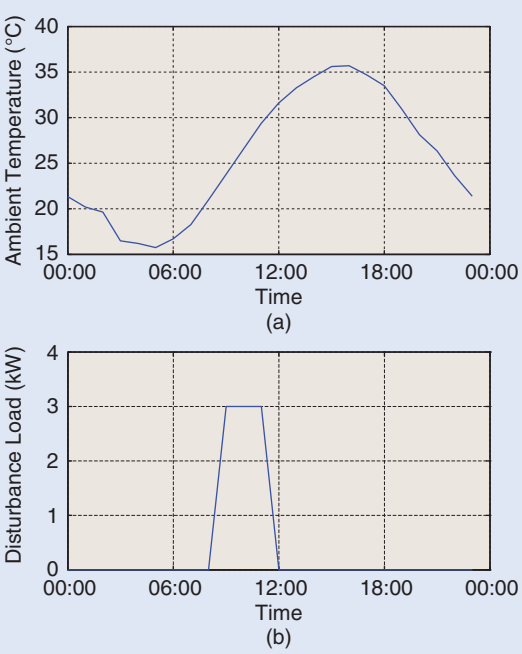

FIGURE 1 Predicted ambient temperature and load profiles used for the simple model predictive control example. The ambient temperature profile is measured from a sensor located at the University of California, Merced, on July 3, 2009. The load profile corresponds to a group meeting scheduled between 8:00 and 11:00 every day with a constant 3-kW occupancy load. (a) Ambient temperature (°C) and (b) disturbance load (kW).

and the total comfort violation

$$J^{\varepsilon} = \sum_{k=0}^{N} (|\overline{\varepsilon}^{*}(k)| + |\underline{\varepsilon}^{*}(k)|)\Delta t. \tag{11}$$

Closed-loop simulations are performed with various gains K of the controller C1 and various tuning ρ of MPC with no penalty on peak power consumption, that is $\kappa = 0$. Figure 2 shows the tradeoff between comfort violation and total energy consumption. It is observed that C1 and C2 use the same energy for the same amount of constraint violation. As expected, increased comfort violation corresponds to a lower energy use for both controllers. Note that the results above are valid for the model (2) and the performance indices defined by (9)–(11).

When $\kappa \neq 0$, a different behavior is observed. Simulation results for C1 with $K = 400$ and C2 with $\rho = 1000$ and $\kappa = 2$ are plotted in Figure 3. The peak power consumption J^{p} is reduced by 89% relative to the proportional controller C1 when MPC C2 is used. For both controllers the comfort violation index J^{ε} is zero. This behavior is obtained by taking advantage of the predictive information of the disturbance and using the space thermal storage. In fact, MPC precools the space temperature to 22.8 °C before the occupancy load begins. This *precooling* behavior reduces the peak power consumption of the system and flattens the control profile.

In the previous simulation, the total energy consumption J^{p} of MPC is increased by 6.3% compared to the proportional controller. The increase in energy consumption is due to energy losses through the resistance R while precooling. The tradeoff between the total energy losses $(J^{u}_{C2} - J^{u}_{C1})/J^{u}_{C1}$ and the peak power reduction $(J^{p}_{C1} - J^{p}_{C2})/J^{p}_{C1}$ is further explored in Figure 4. The tradeoff lines are generated for different tunings of κ taken from the interval $[0, 5]$ and three different thermal resistances with values $R_0 = 50$ °C/kW, $R_0/5$, and $5R_0$. The MPC achieves lower peak power consumption at the cost of higher energy consumption. MPC energy losses relative to the proportional controller decrease as the thermal resistance R in (1)

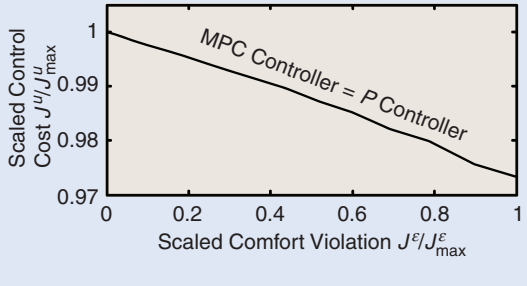

FIGURE 2 Comparing model predictive control (MPC) (C2) and proportional controller (C1). Observe that C1 and C2 use the same energy $J^{u}_{C1} = J^{u}_{C2}$ for the same amount of constraint violation $J^{\varepsilon}_{C1} = J^{\varepsilon}_{C2}$. As expected, increased comfort violation corresponds to a lower energy use for both controllers.

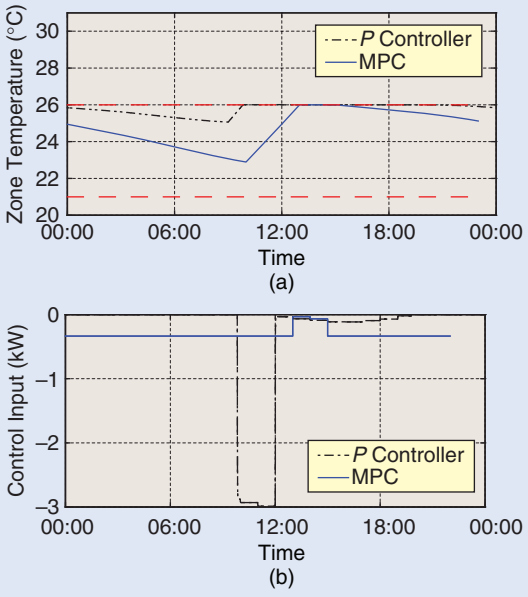

FIGURE 3 Closed-Loop simulations with model predictive control (MPC) and proportional control. The peak power consumption J^{p} is reduced by 89% relative to the proportional controller when the MPC is used. For both controllers the comfort violation index J^{ε} is zero. This behavior is obtained by taking advantage of the predictive knowledge of the disturbance and using the space thermal storage. In fact, the MPC precools the space temperature to 22.8 °C before the occupancy load begins. This *precooling* behavior reduces the peak power consumption of the system and flattens the control profile. (a) Zone temperature (°C) and (b) cooling input (kW).

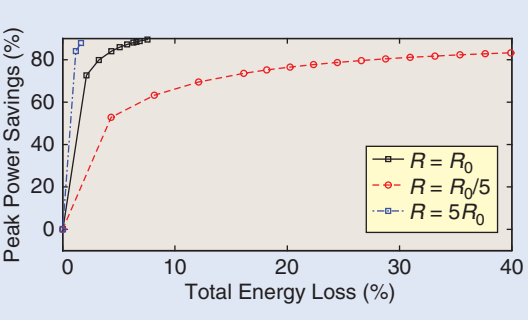

FIGURE 4 Comparison between model predictive control (MPC) and proportional control in terms of closed-loop total energy losses $((J^{u}_{C2} - J^{u}_{C1})/J^{u}_{C1})$ and peak power reduction $((J^{p}_{C1} - J^{p}_{C2})/J^{p}_{C1})$. Simulations are performed for different MPC tunings and three different thermal resistances with $R_0 = 50$ °C/kW. The total energy consumption of MPC (C2) is larger than that of the proportional controller (C1). Larger peak power savings of the MPC correspond to higher energy losses. The increased energy consumption is due to energy lost though the resistance R while precooling. MPC can achieve lower peak power consumption at the cost of higher energy consumption. MPC energy losses with respect to the proportional controller decrease as the thermal resistance R in (1) increases.

The Problem Complexity

Building control design is not straightforward for a long list of reasons. Here we highlight some of the main features of the problem that make it so complex.

Building heating, ventilation, and air conditioning (HVAC) systems convert and transport energy through working fluids, primarily air and water. The flow dynamics of air and water through distribution networks are described by nonlinear partial differential equations, specifically the Navier-Stokes equations. The computational fluid dynamics technique is computationally intensive and requires a complete geometry description at all length scales. This level of detail is rarely available for a real building. A more common approach is to approximate the velocity, temperature, and pressure distributions with reduced order lumped nodal models. The nodes in a nodal HVAC system model are defined by components. Each component has distinctive and potentially complicated behavior often described by nonlinear equipment performance curves.

HVAC components are arranged into HVAC systems in a variety of different configurations as a result of evolving design practices. A handful of standard configuration types are more common than others, but virtually every building is unique. Therefore, the spatial locations, type of components, and methods used to implement a control action are highly dependent on the specific HVAC system. For instance, overhead air distribution systems use a different set of actuators from under-floor air distribution systems, and both differ from systems that use water-based radiators for conditioning.

The control objective of an HVAC system is occupant thermal comfort. In this article, we treat comfort as being equivalent to a specific range of spatial air temperatures. A large body of ASHRAE and other literature [32] have investigated more complicated representations of occupant comfort. These more detailed comfort models take into account metabolism and biological factors, air velocity, humidity, heat transfer through radiation, free convection, and other effects [33].

In addition to maintaining comfort and temperature regulation, HVAC controllers can have additional requirements on humidity regulation, proportions of fresh versus recirculated air for indoor air quality, flow rates for ventilation, and pressurization of spaces.

increases. Thorough studies on load shifting have appeared in [4] and [5].

The above example shows the benefits and tradeoffs resulting from the use of MPC. However, three important elements are not captured by the example. The first element is the complexity of the real problem. Buildings are more complicated than simple RC systems. For details, see "The Problem Complexity." It is therefore necessary to derive descriptive models that are simple enough for real-time optimization in MPC and bring substantial savings to the real world. In the next sections we address this issue by providing a brief description of how building cooling and heating systems work, describing abstraction models at different hierarchical levels, and presenting MPC algorithms for each hierarchical level. The second element not captured by the example is the complexity of the cost function. The cost function of an MPC scheme in practice includes more detailed energy consumption functions for system components and external signals such as time varying utility price, availability of renewable energy, and load shedding signals received from the utility grid. The resulting MPC logic combines load shifting with additional features such as peak electrical power reduction and free cooling. We provide details on this issue in the next sections. The third element not captured by the example is the uncertainty in predictions. In reality, the model and predictions of load and weather are uncertain. The last part of this article briefly discusses this issue and addresses future research directions.

For the sake of brevity, we concentrate on the cooling side. Despite different central equipment for heating conversion, the heating architecture is similar to the cooling one. As a result, the abstraction levels and control methodology in this article can be extended to heating systems with minimal effort.

COOLING CONTROL SYSTEMS

This section describes a building HVAC system by dividing it into two parts, namely, energy conversion and energy distribution. We refer to the energy conversion system as *high level* and to the energy distribution system as *low level*. At the high level system, a centralized chilled water generation system produces the required cooling energy. It is assumed that an energy storage device is available. This article focuses on thermal energy stored in a stratified water tank. At the low-level system, subsystems known as air handling units (AHUs) transfer energy from distributed chilled water into localized air flows. These air flows are transported to buildings' spaces, delivering cooling energy where required. The next sections describe the high-level chilled water generation, storage element, and low-level AHU system.

High Level: Components Outside a Building

Figure 5 depicts the main components of a cooling system based on chilled water generation, storage, and distribution. The system can serve either a single building or multiple buildings. Chillers and cooling towers are

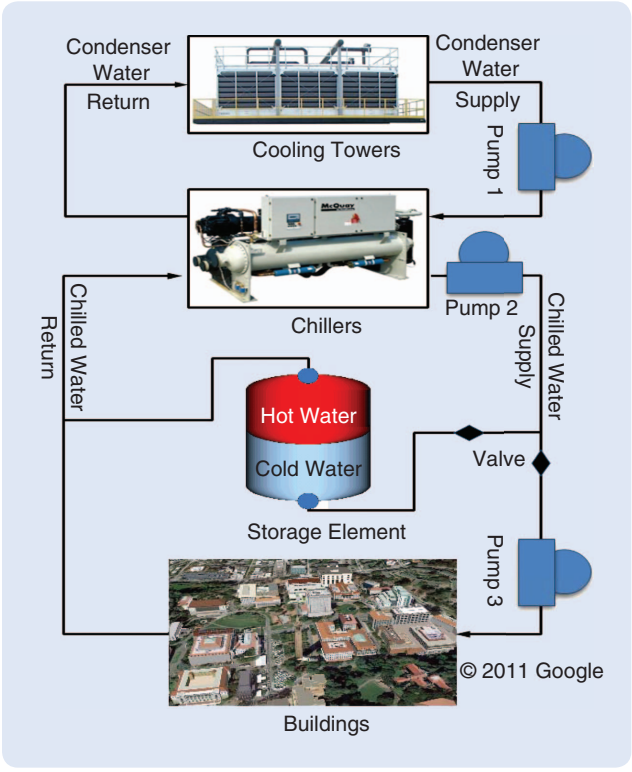

FIGURE 5 Schematic of a cooling system based on chilled water generation, storage, and distribution. The system can serve either a single building or multiple buildings. The chillers and the cooling towers are responsible for generating the chilled water. The chilled water is distributed through insulated piping using hydraulic pumps. A valve controls the chilled water flow to either fill the storage element or serve the buildings with a desired mass flow rate of chilled water. The chilled water storage element can shift the peak cooling demand so that the chillers and cooling towers can run only when it is most efficient to do so. The logic behind the load shifting can depend on various factors, which include time varying utility prices, availability of renewable energy, lower ambient temperature, and load shedding signals received from the utility grid. The thermal storage tank can operate in two modes. When the chilled water supply flow is greater than the water flow demanded by the buildings, the excess flow charges the tank. When the chilled water supply flow is less than the water flow demanded by the buildings, the tank is discharged to compensate for the insufficient chilled water supply.

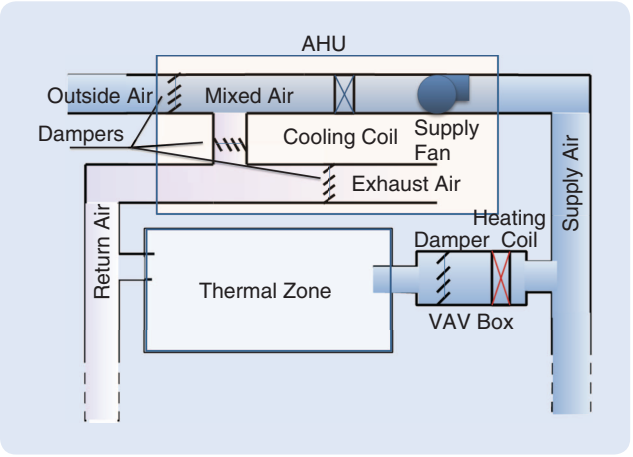

FIGURE 6 Schematic of the air distribution system. Air handling units (AHUs), supply fans, and variable air volume boxes constitute the main equipment used to produce and distribute cool air in a building. The AHU recirculates return air from building spaces, and mixes it with fresh outside air. The ratio of return air flow to outside air flow is controlled by damper positions in the AHU. The mixed air is then cooled by the cooling coil that transfers the cooling energy from the chilled water that is generated or stored by the high-level system.

responsible for generating chilled water. The chillers remove heat from the chilled water loop by means of a vapor-compression or absorption refrigeration cycle. Cooling towers reject heat from the chiller cycle to the ambient environment by evaporation and forced convection using electric fans.

The chilled water storage element in Figure 5 can shift the peak cooling thermal energy load so that the chillers and cooling towers can run only when it is most efficient. The logic behind the load shifting depends on various factors including time varying utility prices, availability of renewable energy, lower ambient temperature, and load shedding signals received from the utility grid. The chilled water is distributed through insulated piping using hydraulic pumps. The valve in Figure 5 controls the chilled water flow to either fill the storage element or serve the buildings with a desired mass flow rate of chilled water.

Several thermal storage elements are available in the building industry. This article focuses on a stratified chilled water tank that can store chilled water, such as the one installed on the campus of the University of California at Merced [6], [7]. Other storage devices are available using ice balls or concrete slabs to store thermal energy.

Low Level: Components Inside a Building

The main components used to produce and distribute cool air in a building are depicted in Figure 6. They are AHUs and variable air volume (VAV) boxes. The AHU recirculates return air from building spaces, and mixes it with fresh outside air. The ratio of return air flow to outside air flow is controlled by dampers located inside the AHU. The mixed air is cooled by a cooling coil that transfers cooling energy from the chilled water that is generated or stored by the high-level system.

The air temperature downstream of the cooling coil depends on the mass flow rate of chilled water through the cooling coil, the temperature of the chilled water, the temperature of mixed air entering the cooling coil, the mass flow rate of the mixed air, and the physical characteristics as well as thermal effectiveness of the cooling coil. Cool air is delivered to the building spaces by electric supply fans. Before reaching the building spaces, the air goes through VAV boxes. At each VAV box, the air flow

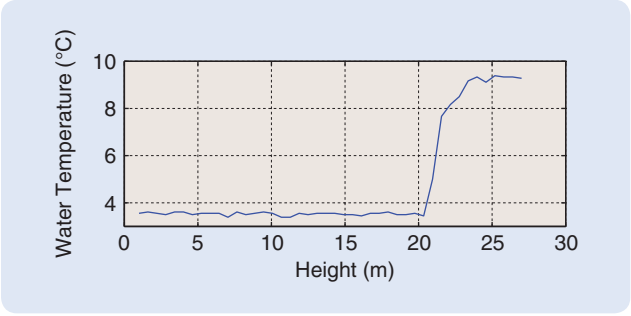

FIGURE 7 The water in a stratified storage tank is subject to negligible mixing. This figure depicts the temperature of water measured inside the stratified tank of University of California, Merced, at different heights. The presence of a thin layer of water that has a steep temperature gradient over the height of the tank can be observed. This layer is called a thermocline.

rate supplied to the space is regulated by a damper. In addition, air temperature can be increased using a reheat coil installed in the VAV box when needed. A space served by one VAV box is referred to as a *thermal zone*. The delivered air enters a thermal zone through diffusers that are designed to fully mix the incoming air with the air in the thermal zone.

HIERARCHICAL LEVELS AND SYSTEM MODELING

We follow the structure of the previous section and present reduced order models for both the high level and low level of buildings. For each level we describe the associated dynamics, components that use energy, and thermal energy load. When a multitude of buildings is controlled, the relevant model dynamics at the high level are the dynamics of thermal energy storage devices. The buildings' cooling or heating thermal energy loads are modeled as a lumped heat flux, and the main components consuming energy are chillers, cooling towers, and pumps. At a low level the relevant model dynamics are the dynamics of thermal zones. The thermal energy load combines the loads generated by occupants, solar radiation, and electrical devices. The components consuming energy are a supply fan, a cooling coil, and heating coils.

High Level: Modeling a Building from the Outside

Storage Dynamics

A model of a water tank used for actively storing chilled water is presented and validated with data collected from the campus of the University of California, Merced. The water in the tank is subject to negligible mixing, thus the tank can be modeled as a stratified system with layers of warmer water at the top and cooler water at the bottom. Figure 7 depicts the temperature of water measured inside the tank at different heights at 8:30 on November 29, 2007. A thin layer of water, known as a thermocline, that has a steep temperature gradient over the height of the tank can be observed. Warmer water above the thermocline and cooler water below the thermocline are lumped up to obtain a four-state system describing the height and temperature of the warmer and cooler water, respectively.

The tank is assumed to be part of a closed hydraulic loop, thus the mass flow rate entering the tank is equal to the mass flow rate exiting the tank. Subsequently, the total height of water in the tank z_{tank} is the sum of the height of warm water z_{a} and the height of cool water z_{b} in the tank. The tank dynamics are governed by mass and internal energy conservation laws

$$\dot{z}_{\text{b}} = \frac{\dot{m}_{\text{CHWS}} - \dot{m}_{\text{cmp}}}{\rho \pi r_{\text{tank}}^2}, \quad \dot{z}_{\text{a}} + \dot{z}_{\text{b}} = 0, \tag{12}$$

$$\dot{U}_{\text{a}} = \dot{H}_{\text{a}} + \dot{Q}_{\text{b}>\text{a}} + \dot{Q}_{\text{oa}>\text{a}}, \ \dot{U}_{\text{b}} = \dot{H}_{\text{b}} + \dot{Q}_{\text{a}>\text{b}} + \dot{Q}_{\text{oa}>\text{b}}, \tag{13}$$

where ρ is the density of the water, r_{tank} is the inner radius of the tank, $\dot{m}_{\text{CHWS}}$ is the mass flow rate of water supplied to the buildings, and $\dot{m}_{\text{cmp}}$ is the mass flow rate of water returning from the buildings. $U_* = \rho \pi r_{\text{tank}}^2 z_* c_p T_*$ is the internal energy of the water in the tank, where $* = a$ denotes variables for warmer water, and $* = b$ denotes variables for cooler water. $\dot{Q}_{\text{oa}>\text{a}}$ is the heat transferred from ambient to the warmer water in the tank, $\dot{Q}_{\text{oa}>\text{b}}$ is the heat transferred from ambient to the cooler water in the tank

$$\begin{aligned} \dot{Q}_{\text{oa}>\text{a}} &= (T_{\text{oa}} - T_{\text{a}})(2\pi r_{\text{tank}} z_{\text{a}}) k_1, \\ \dot{Q}_{\text{oa}>\text{b}} &= (T_{\text{oa}} - T_{\text{b}})(2\pi r_{\text{tank}} z_{\text{b}}) k_1. \end{aligned} \tag{14}$$

$\dot{Q}_{\text{a}>\text{b}}$ is the heat conducted from warmer water to cooler water in the tank, and $\dot{Q}_{\text{a}>\text{b}}$ is the heat conducted from cooler water to warmer water in the tank.

$$\dot{Q}_{\text{a}>\text{b}} = (T_{\text{a}} - T_{\text{b}})(\pi r_{\text{tank}}^2) k_2, \ \dot{Q}_{\text{b}>\text{a}} = -\dot{Q}_{\text{a}>\text{b}}, \tag{15}$$

where k_1 and k_2 are heat transfer coefficients, $\dot{H}_{\text{a}}$ is the enthalpy rate for the warm water in the tank contributed by water flow, and similarly $\dot{H}_{\text{b}}$ for the cooler water.

The thermal storage tank can operate in two modes. When the chilled water flow $\dot{m}_{\text{CHWS}}$ is greater than the water flow demanded by the buildings $\dot{m}_{\text{cmp}}$, the excess flow fills the tank. Hence the water flow enthalpy rates are calculated as

$$\dot{H}_{\text{a}} = -(\dot{m}_{\text{CHWS}} - \dot{m}_{\text{cmp}}) c_p T_{\text{a}}, \ \dot{H}_{\text{b}} = (\dot{m}_{\text{CHWS}} - \dot{m}_{\text{cmp}}) c_p T_{\text{CHWS}}. \tag{16}$$

When the chilled water flow $\dot{m}_{\text{CHWS}}$ is less than the water flow demanded by the buildings, the water in the tank compensates for the insufficient chilled water supply. Hence the water flow enthalpy rates are calculated as

$$\dot{H}_{\text{a}} = -(\dot{m}_{\text{CHWS}} - \dot{m}_{\text{cmp}}) c_p T_{cmp,r}, \ \dot{H}_{\text{b}} = (\dot{m}_{\text{CHWS}} - \dot{m}_{\text{cmp}}) c_p T_{\text{b}}, \tag{17}$$

where $T_{cmp,r}$ is the temperature of return water from buildings.

The simplified model (12)–(17) is validated using measured data from May 22–29, 2007. The measured inputs $[\dot{m}_{CHWS}, T_{CHWS}, \dot{m}_{cmp}, T_{cmp,r}]$ are applied to the tank model, and the output of the model $[z_a, z_b, T_a, T_b]$ is compared with the measurements in Figure 8. Figure 8(a) shows the tank water temperature validation results. The solid lines are the temperature measurements of the top layer water T_a and bottom layer water T_b in the tank, and the dotted lines show the predicted cooler and warmer water temperature. The tank model captures the temperature dynamics of the top and bottom layers of the tank water as well as the dynamics of the cool water height [Figure 8(b)]. However, the second peak of the top water temperature during the day is not captured due to the formation of a second thermocline (note in Figure 8(a) the bumps above 15.2 °C every afternoon). A higher order model overcomes this limitation. We preferred to not increase the model order in order to avoid real-time implementation issues. More details can be found in [8].

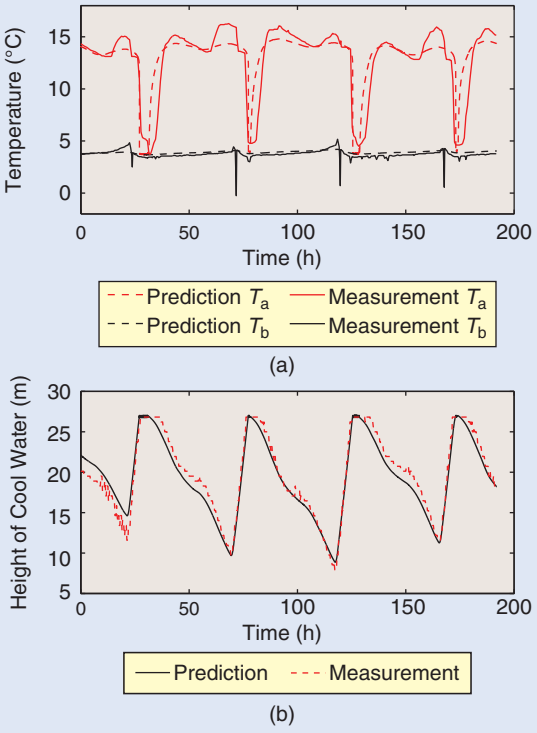

FIGURE 8 The tank model (12)–(17) is validated using measured data collected from the tank at University of California, Merced. We applied the measured inputs to the tank model and the output of the model $[z_a, z_b, T_a, T_b]$ is compared with the measurements. Part (a) shows the tank water temperature validation results. The solid lines are the temperature measurements of top layer water T_a and bottom layer water T_b in the tank, and the dotted lines show the predicted cool and warm water temperature. The simplified tank model captures the temperature dynamics of the top and bottom water layers as well as the height of the cool water in the tank. (a) Water temperature validation and (b) water height validation.

Load Modeling

At a high level, buildings can be modeled as *load demand* elements. A lumped load model predicts the total energy requested by a building based on date, time, occupancy, and weather. In the approach studied in [9], the building load model has two subcomponents, namely, the solar and internal load predictor and the building thermal load predictor [34]. Both components are depicted in Figure 9. The solar and internal load predictor uses time, date, and cloud coverage as its inputs and calculates inside and outside solar loads as well as internal loads. The outside solar load reflects the solar energy on the outer surface of the building, while the inside solar load is the solar radiation into the building through windows. The internal load includes the heat from people, lights, and equipment.

The building thermal load predictor predicts the cooling load of buildings, which are conventionally modeled by RC circuit analogy [10], [11]. The building thermal load model is sketched in Figure 10. R_1 represents the thermal resistance of windows. The walls are separated into two layers, where C_{in} and C_{out} capture the thermal capacitance of the wall when influenced by outside and inside solar loads, respectively. The thermal resistance between C_{in} and C_{out} is modeled by R_3, while R_2 and R_4 capture the thermal resistance associated with heat convection. The interconnection of the thermal components is shown in Figure 10. The model inputs are outside air temperature T_{oa}, outside solar load $\dot{Q}_{Solar,out}$, inside solar load $\dot{Q}_{Solar,in}$, internal load $Q_{internal}$, and the indoor temperature setpoint T_{sp}. The model internal states are the temperatures of the thermal masses T_{in} and T_{out}. The model output is the cooling load demand $\dot{Q}_{Load}$.

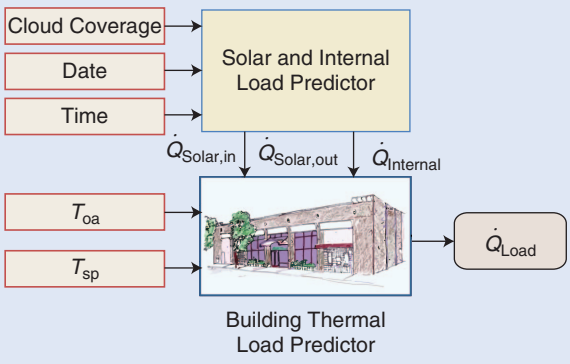

FIGURE 9 The building load model has two subcomponents, namely, the solar and internal load predictor and the building thermal load predictor. The solar and internal load predictor uses time, date, and cloud coverage as its inputs and calculates inside and outside solar loads and internal loads. The building thermal load predictor predicts the cooling load of buildings.

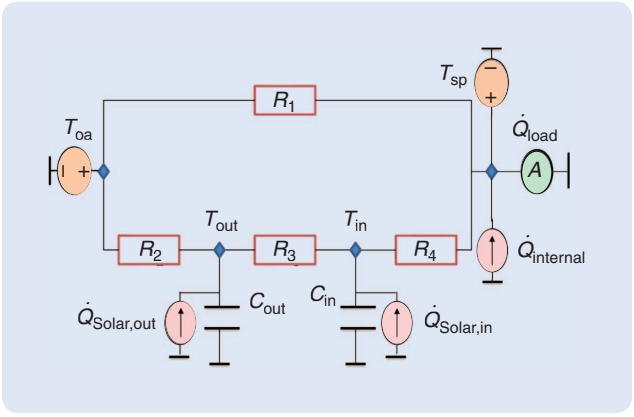

FIGURE 10 A building thermal load model. R_1 represents the thermal resistance of windows. The walls are separated into two layers. C_{in} and C_{out} capture the wall thermal capacitance. The thermal resistance between C_{in} and C_{out} is modeled by R_3, while R_2 and R_4 capture the thermal resistance associated with heat convection. The model inputs are outside air temperature T_{oa}, outside solar load $\dot{Q}_{Solar,out}$, inside solar load $\dot{Q}_{Solar,in}$, internal load $Q_{internal}$, and the indoor temperature setpoint T_{sp}. The model output is the cooling load demand $\dot{Q}_{Load}$.

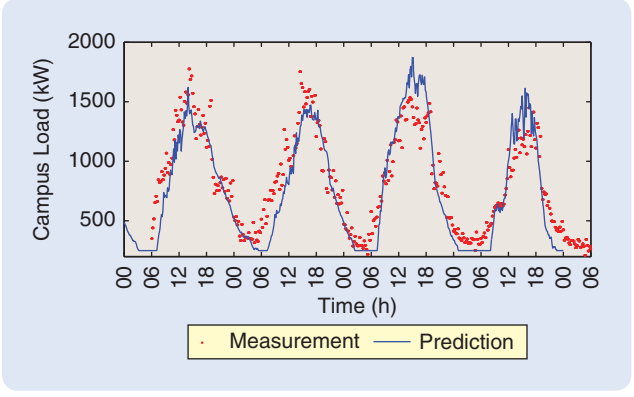

FIGURE 11 Campus load validation. The measured building load of the University of California, Merced, is depicted with a dotted line. The time frame is June 1–4, 2009. The solid line shows the building load predicted by the building load model in Figure 10.

The detailed equations describing the model in Figure 10 can be found in [8].

The load model in Figure 10 is used to model all the buildings on the University of California, Merced, campus where the parameters R_1, R_2, R_3, R_4, C_{in}, C_{out} are estimated using historical data. A nonlinear regression problem is solved to minimize the error between the actual cooling load demand and the predicted load demand. The predicted load demand is obtained by simulating the model in Figure 10 with measured inputs. The load model with estimated parameters is validated using load measurements from June 1 to June 4, 2009 at University of California, Merced. Figure 11 presents the validation results. The measured building load is depicted as the dotted line and the solid line shows the building load predicted by the model in Figure 10. When the prediction mismatch exceeds desired tolerances, the system parameters need to be reidentified based on a new set of measured data. For other types of building load models, see [12] and [13].

Main Components Modeling

At a high level, the main components consuming energy are chillers, cooling towers, and pumps. It is assumed that the setpoints sent to these components are tracked instantly and without errors. The equations estimating the power consumption of each component are listed in Table 1, and Table 2 explains the notation used in Table 1.

Chillers

A simple regression-based centrifugal chiller model is used [6]. In Table 1(1), the chillers' electrical power is modeled as a static function of the chilled water supply temperature, condenser water supply temperature, mass flow rate of the chilled water supply, and the chilled water return temperature. The static chiller model is used to describe the chillers performance at University of California, Merced. Figure 12 presents the validation results based on the data collected from June 1 to June 3, 2009. The predicted chiller power in solid line captures the measurements in dotted line.

Cooling Towers

The cooling towers use variable speed fans to track a setpoint for the condenser water supply temperature by rejecting the condenser water heat to the ambient environment through evaporation and convection. In Table 1(2), the simple cooling tower model [7] employs a regression equation to compute the fan speed required to produce condenser water with supply temperature T_{CWS} and mass flow rate $\dot{m}_{CWS}$ when the wet bulb temperature is T_{wb} and the condenser water return temperature is T_{CWR}. Then Table 1(3) approximates the

TABLE 1 Component models outside a building. This table lists the equations describing the components outside a building including chillers, cooling towers, and pumps. The descriptions of variables used in (1)–(6) are listed in Table 2.

Equation Number	Component	Equation
(1)	Chillers	$P_{CH} = \text{Chiller}(T_{CHWS}, T_{CWS}, \dot{m}_{CHWS}, T_{wb}, T_{CHWR})$.
(2)	Cooling Towers	$f_{APP}(T_{CWS}, T_{CWR}, T_{wb}, \dot{m}_{CWS}, \omega_{fan}) = 0$.
(3)	—	$P_{CT} = c(\omega_{fan})^3$.
(4)	Pumps	$\frac{q}{q^0} = \frac{\omega_{pump}}{\omega^0_{pump}}, \frac{\Delta p}{\Delta p^0} = \left(\frac{\omega_{pump}}{\omega^0_{pump}}\right)^2$.
(5)	—	$\Delta p^0 = f_p(q^0), \eta^0 = f_\eta(q^0)$.
(6)	—	$P_{pump} = \Delta p q / \eta^0 (q\omega^0/\omega)$.

TABLE 2 Notation for the models used in Table 1.

Variable	Description
P_{CH}	Chiller power consumption
T_{CHWS}	Chilled water supply temperature
T_{CWS}	Condenser water supply temperature
$\dot{m}_{CHWS}$	Mass flow rate of the chilled water supply
T_{CHWR}	Chilled water return temperature
P_{CT}	Cooling tower power consumption
T_{wb}	Wet bulb temperature
ω_{fan}	Fan speed
T_{CWR}	Condenser water return temperature
q	Pump volumetric water flow
ω_{pump}	Pump angular speed
Δp	Pressure difference across the pump
η	Pump efficiency
P_{pump}	Pump power consumption
q^0	Pump nominal volumetric water flow
ω^0_{pump}	Pump nominal angular speed
Δp^0	Nominal pressure difference across the pump
η^0	Pump nominal efficiency

cooling tower power consumption as a cubic function of the fan speed. This model is applied to model the cooling towers at University of California, Merced, and the validation results are reported in Figure 13.

Pumps

When modeling the energy consumption of the pumps, it is assumed that the enthalpy change through a pumps is constant. It is also assumed that the pump volumetric water flow, the pump speed, the pressure difference across the pump, and their corresponding nominal values denoted with the superscript ^{0} satisfy the affinity laws in Table 1(4). In Table 1(5), the pressure difference and efficiency under nominal operating conditions are approximated as polynomial functions of the nominal volumetric water flow. The polynomial coefficients are obtained by fitting historical pump performance data. For a given pressure difference Δp and volumetric water flow rate q, the pump power is calculated using Table 1(6). The pump model is validated using measured data from the chilled water supply pump at University of California, Merced. The validation results are shown in Figure 14.

Low Level: Modeling a Building from the Inside

Thermal Zones Temperature Dynamics

The objective of this section is to develop a simplified control oriented model for building HVAC systems suitable for

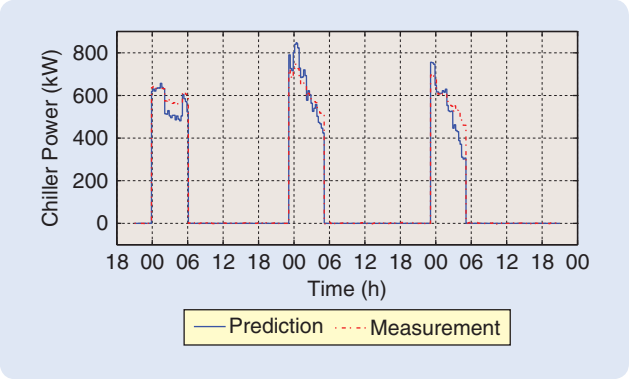

FIGURE 12 Chiller model validation. The dotted line shows the electric power consumption of the two chillers installed at University of California, Merced, and the solid line is the power consumption of the chillers predicted by the simplified model Table 1(1). Data was collected from June 1–3, 2009.

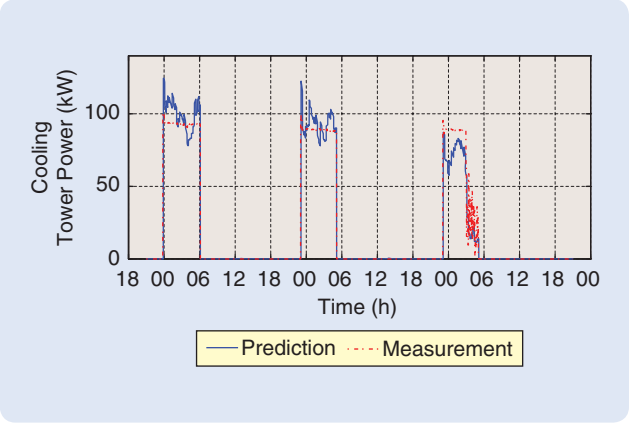

FIGURE 13 Cooling tower model validation. The dotted line depicts the measured electric power consumption of cooling towers located at University of California, Merced, and the solid line shows the power consumption of the cooling towers predicted by model Table 1(2) and (3). Data was collected from June 1–3, 2009.

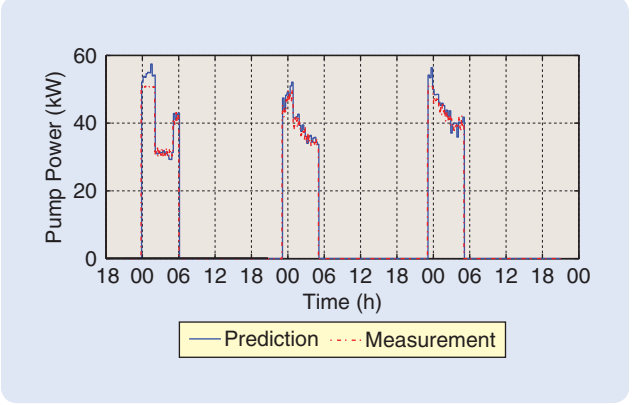

FIGURE 14 Pump model validation. The dotted line is the measured electric power consumption of the hydraulic pump 2 in Figure 5 located at University of California, Merced. The solid line is the pump power consumption predicted by model Table 1(4)–(6). Data was collected from June 1–3, 2009.

real-time optimization. We focus on a dynamic model of zone temperatures. In order to develop a simplified yet descriptive model, it is assumed that the pressure is constant throughout the system and the amount of air exiting the zones is the same as the amount of air entering.

Table 3 collects the static component models, and Table 4 clarifies the notation used in Table 3. Table 3(1) calculates the supply air temperature to each zone, where T_{oa} is the outside air temperature, δ is the AHU damper position, and T_r is the return air temperature. In Table 3(2), T_r is calculated as the flow-weighted average temperature of return air from each thermal zone. The return air is not recirculated if $\delta = 0$ and no outside fresh air is used if $\delta = 1$. δ has to be strictly less than one to guarantee minimum ventilation required by building codes.

For a zone j in a system of N_v thermal zones, the temperature dynamics can be modeled as a two-mass system

$$C_1^j \dot{T}_1^j = \dot{m}_s^j c_p (T_s^j - T_1^j) + \frac{T_2^j - T_1^j}{R^j} + \sum_{i \in \mathcal{N}^j} \frac{T_1^i - T_1^j}{R_{ij}} + \frac{T_{\text{oa}} - T_1^j}{R_{\text{oa}}^j} + P_{\text{d}}^j, \tag{18}$$

$$C_2^j \dot{T}_2^j = \frac{T_1^j - T_2^j}{R^j}, \tag{19}$$

for all $j = 1, 2, \ldots, N_v$.

The thermal capacitance C_1^j is associated with the fast-dynamic masses such as the air around VAV diffusers. The thermal capacitance C_2^j is associated with the slow-dynamic masses such as the floors, walls, and furniture. The appearance of fast and slow temperature dynamics is described in [14] for residential buildings. In (18) and (19), T_1^j and T_2^j are the system states representing the temperatures of the lumped masses of C_1^j and C_2^j, respectively. The air temperature of thermal zone j is the temperature of the fast dynamic mass $T^j = T_1^j$. $\mathcal{N}^j$ is the set of neighboring thermal zones of zone j, R_{oa}^j is the thermal resistance between zone j and ambient temperature, and c_p is the specific heat capacity of zone air. R^j models the thermal resistance between C_1^j and C_2^j, $R_{ij} = R_{ji}$ models the thermal resistance between zone i and neighboring zone j, and P_{d}^j is an unmeasured thermal load that is imposed by occupancy and solar radiation. Figure 15 depicts a representative example RC network.

The two-mass approach is used to model the temperature dynamics of thermal zones in the Bancroft Library located on the campus of the University of California, Berkeley. Next we show the results by focusing on a conference room without windows, indexed by $j = 1$, with one adjacent thermal zone indexed by $j = 2$. The parameters $p = [C_1^1, C_2^1, R^1, R_{12}, R_{\text{oa}}^1]$ are identified using a nonlinear regression algorithm using measured data during weekends when the conference room has no occupants, that is $P_{\text{d}}^1 = 0$. Figure 16 compares the measured room temperature to the room temperature predicted by model (18) and (19) when driven by the measured inputs.

Load

The load prediction $P_{\text{d}}^j(t)$ for each thermal zone j is important for designing predictive feedback controllers and assessing potential energy savings. Various approaches are available to estimate occupancy load. For instance, the authors of [15] develop an agent-based model to simulate the occupants' behavior in a building, and the work in [16] and [17] focuses on occupants' behavior and mobility patterns using a wireless camera sensor network.

Time varying bounds on the disturbance load $P_{\text{d}}(t)$ can be computed from the mismatch between a nominal model and historical data, and correlating the load bounds with shared calendars, weather predictions, as well as predicted cloud coverage. This concept can be illustrated using the conference room discussed previously. The conference room calendar contained two regularly scheduled group meetings at 10:00 and 14:00 every Wednesday. The same meetings can be identified by inspecting the model mismatch between nominal model predictions and historical data. Figure 17 depicts

TABLE 3 Component models inside a building. Equation (1) evaluates the supply air temperature T_s^j to zone j when the return air temperature is T_r, the outside air temperature is T_{oa}, the damper position in air handling units (AHU) is δ, and the temperature difference across the cooling coil and heating coils are ΔT_c and ΔT_h, respectively. The return air temperature is calculated by (2) as airflow rate weighted average of zone temperatures. Equations (3)–(6) collect formulas modeling the power consumption of the supply fan in the AHU, the cooling coil in the AHU, and the heating coils in variable air volume boxes. The descriptions of variables used in (1)–(6) are listed in Table 4.

Equation Number	Component	Equation
(1)	Supply air temperature	$T_s^j = \delta T_r + (1-\delta) T_{\text{oa}} - \Delta T_c + \Delta T_h^j$
(2)	Return air temperature	$T_r = \sum_i \dot{m}_s^i T^i / \sum_i \dot{m}_s^i$
(3)	Fan power	$P_f = c_0 + c_1 \dot{m}_s + c_2 \dot{m}_s^2$
(4)	Cooling coil power	$P_c = \dfrac{\dot{m}_{\text{sc}} c_p \Delta T_c}{\eta_c \text{COP}_c}$
(5)	Heating coil power	$P_h^j = \dfrac{\dot{m}_{\text{sh}}^j c_p \Delta T_h^j}{\eta_h^j \text{COP}_h}$
(6)	Coefficient of performance	$\text{COP} = \dfrac{E_{\text{thermal}}}{E_{\text{input}}}$

TABLE 4 Notation for the models used in Table 3.

Variable	Description
$\dot{m}_s^j$	Supply air mass flow rate to thermal zone j
T_s^j	Supply air temperature to thermal zone j
ΔT_c	Change in air temperature across AHU cooling coil
ΔT_h^j	Change in air temperature across reheat coil in VAV box at zone j
δ	AHU recirculation damper position
T_{oa}	Outside air temperature
T_r	Return air temperature
C_1^j	Thermal capacitance associated with fast-dynamic masses in zone j
C_2^j	Thermal capacitance associated with slow-dynamic masses in zone j
T_1^j	Temperature state representing lumped mass of C_1^j (air temperature of zone j)
T_2^j	Temperature state representing lumped mass of C_2^j
R_{oa}^j	Thermal resistance between zone j and ambient temperature
c_p	Specific heat capacity of zone air
R^j	Thermal resistance between C_1^j and C_2^j
$R_{ij} = R_{ji}$	Thermal resistance between zone i and adjacent zone j
P_d^j	Unmeasured load imposed by occupancy and solar radiation
c_0, c_1, c_2	Fan power fitting parameters
η_c	Cooling coil efficiency
η_h^j	Efficiency of the heating coil in VAV box j
P_c	Electric power used by the cooling coil
P_h^j	Fuel energy consumption rate associated with the heating coil in VAV box j.
$\dot{m}_{sc}$	Airflow through the cooling coil
$\dot{m}_{sh}^j$	Airflow through the heating coil in VAV box j
COP_c	Cold water coefficient of performance
COP_h	Hot water coefficient of performance

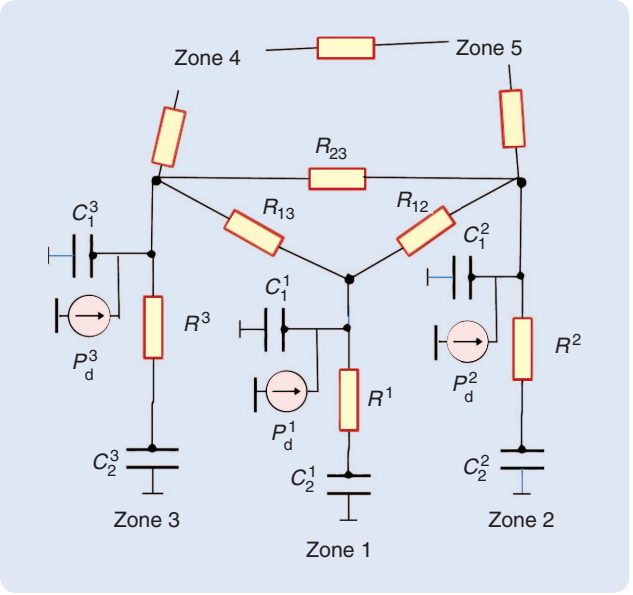

FIGURE 15 Thermal zone network model. The thermal capacitance C_1^j is associated with the fast-dynamic masses such as the air around variable air volume diffusers. The thermal capacitance C_2^j is associated with the slow-dynamic masses such as the floors, walls, and furniture. R^j models the thermal resistance between C_1^j and $R_{ij} = R_{ji}$ models the thermal resistance between zone i and neighboring zone j, and P_d^j is a current source modeling an unmeasured load imposed by occupancy and solar radiation.

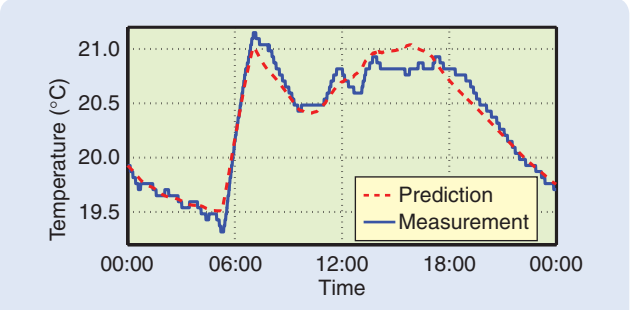

FIGURE 16 Identification results of the thermal zone model (18) and (19). The solid line is the measured temperature collected from a conference room in the Bancroft Library located on the campus of the University of California, Berkeley, USA, on July 4, 2010. The dashed line is the temperature predicted by the simplified thermal zone model (18) and (19) when driven by the measured inputs. The result shows that the model captures the thermal dynamics of the conference room without occupants or solar load.

the envelope-bounded disturbance load during all Wednesdays in July 2010. The envelope is computed as the point-wise minimum and maximum difference between measured data and the nominal model described by (18) and (19) with $P_d^1 = 0$. The two peaks in the disturbance load envelope in Figure 17 correspond to two regularly scheduled group meetings. The off-peak prediction errors can be attributed to unmodeled dynamics and external disturbances.

Main Components

The components at the lower level of the architecture that use energy include dampers, supply fans, heating coils, and cooling coils. They are depicted in Figure 6. The supply fan uses electric power to distribute air to the zones, while the cooling and heating coils use the energy of the chilled and hot water. It is assumed that the power to drive the dampers is negligible. A simple energy consumption model for each component is presented next.

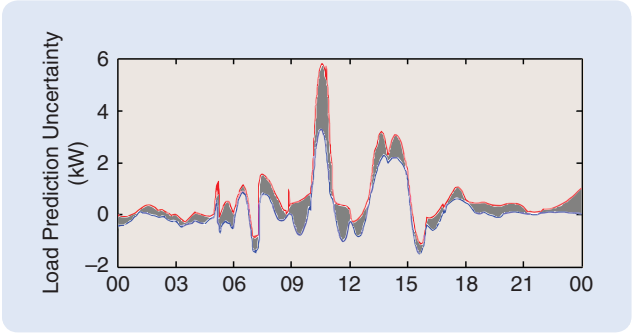

FIGURE 17 Envelope-bounded disturbance load during all Wednesdays in July 2010 in the conference room of the Bancroft Library at the University of California, Berkeley. The envelope is computed as the point-wise min and max difference between measured data and the nominal model. The two peaks in the disturbance load envelope correspond to two regularly scheduled group meetings.

Fan Power

In Table 3(3), the fan power is approximated as a second order polynomial function where c_0, c_1, and c_2 are parameters identified by fitting recorded data, and $\dot{m}_s$ is the mass flow rate of supply air through the supply fan. The simplified fan model predictions are compared in Figure 18 to the data recorded at the University of California, Berkeley, Bancroft Library from October 1–10, 2010.

Cooling and Heating Coils

Cooling coils and heating coils are air-water heat exchangers. There are many examples of coil models in the literature [18]–[20]. In this article we use a simple coil model with constant efficiency (η_c for cooling coils and η_h for heating coils). With this simplification the energy consumption model is a static function of the load on the air side. In Table 3(4) and (5), the electric power associated with the chilled water used by the cooling coil P_c and the thermal power associated with the hot water used by the heating coils P_h are computed as the thermal power delivered to the air side divided by the energy conversion efficiency (η) and the coefficient of performance (COP). The COP defined in Table 3(6) captures the efficiency of the chilling and heating systems, that is the amount of thermal energy E_{thermal} (J) generated by the central plant with one Joule of total energy consumed E_{input} by the plant.

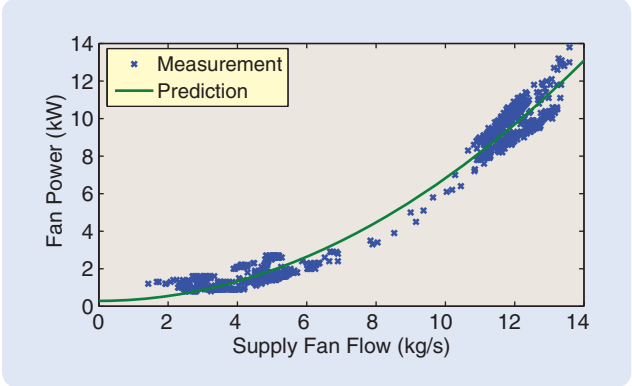

FIGURE 18 Fan identification results. The crosses are the measured power consumption of an air handling unit supply fan at the Bancroft Library, and the solid line is the power predicted using the simplified model Table 3(3).

PREDICTIVE CONTROL DESIGN WITH ACTIVE THERMAL STORAGE

As buildings become smarter the large number of decoupled local controllers, together with coordination strategies based on heuristic rules-of-thumb, become difficult to design, tune, maintain, and upgrade. For more details on current approaches for building control design, see "Current Building Operation and Control Logic." MPC naturally enters the picture as a systematic control methodology that can handle large-scale multi-input, multi-output (MIMO) dynamically coupled systems, with performance guarantees and the unique capability to explicitly handle system constraints. The models presented in the previous section can be used to design MPC algorithms at both the high and low level of building control architectures.

The MPC architecture presented in this article is depicted in Figure 19. A high-level MPC (HMPC) is deployed to optimize the operation and schedule of the cooling and heating systems with active thermal storage. A low-level MPC (LMPC) controls the VAV boxes and the AHUs in each building to guarantee occupant thermal comfort constraints. At both levels, various predictions can be included in the constraints and in the cost function to control the system in an efficient and effective way. These predictions include building loads, load shedding signals from the power grid, utility prices, weather, occupancy, and solar radiation. In addition, HMPC and LMPC can exchange information to achieve a better performance. For example, the occupancy load prediction from LMPC can help HMPC achieve a better accuracy of building load predictions. Also, the chilled and hot water temperature predictions from HMPC impose constraints for the LMPC on achievable supply air temperature downstream of the cooling and heating coils, respectively.

The following optimization problem is used to describe both the HMPC and LMPC

$$J^{\star}(x(t), t) = \min_{u_{t|t}, \ldots, u_{t+N-1|t}} \sum_{k=0}^{N-1} J(x_{t+k|t}, u_{t+k-1|t}, k) + J_N(x_{t+N|t}) \tag{20}$$

subject to

$$x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}, d_{t+k|t}, k), \text{ for all } k = 0, 1, \ldots, N-1, \tag{21}$$

Current Building Operation and Control Logic

Most modern buildings employ some level of automated control. In certain cases the control logic may be complex and optimized, but in the majority of cases building systems are controlled by basic control logic that errs on the side of simplicity over subtlety. This simple control logic is implemented with distinct but interconnected proportional-integral-derivative (PID) control loops and switching logic. This logic responds to setpoints and schedules for building components such as chillers and cooling towers. Controlled parameters include the chilled water supply mass flow rate, chilled water temperature, condenser water mass flow rate, condenser water temperature, operation schedule of the chilling system, and the storage level of the thermal storage element.

Advanced decision systems are available on the market to optimize the high-level system based on component modeling, feedback, and forecasts. A variety of proprietary control sequences for chillers, boilers, and cooling towers are available in the building industry. However, to the best of the authors' knowledge, their implementation is not widespread and often limited to specific configurations and components of the cooling and heating systems.

At the low level, the current practice is to use single-input, single-output cascaded controllers to achieve stable behavior and modest disturbance rejection. To this aim, controllers are designed locally and independently, with the goal of decoupling the multiple components involved.

A classical example for control logic architectures used for air handling units (AHUs) and variable air volume (VAV) boxes is depicted in Figure S1. In each thermal zone, the control system generates a cooling or heating request when the zone temperature is higher or lower than the setpoint, respectively. Each individual request is used to adjust setpoints for the corresponding VAV box. The sum of all requests is used to adjust setpoints for the AHU.

The setpoint adjustment usually follows proportional laws, as depicted in Figure S2, for a VAV box. When the zone temperature is lower than the comfort range, denoted by the dashed lines in Figure S2, the heating coil valve position command is proportional to the difference between zone temperature and the lower-bound on comfort level, and the air mass flow rate is set to the minimum ventilation level. When the zone temperature is within the comfort range, the VAV box maintains the minimum ventilation level. When the zone temperature is higher than the comfort range, the supply airflow rate increases proportionally with the difference between zone temperature and the upper bound on comfort level.

The total number of cooling requests is used to determine the total cooling energy required to guarantee the thermal comfort. The AHU supply fan is often controlled to a pressure setpoint which varies in time depending on the

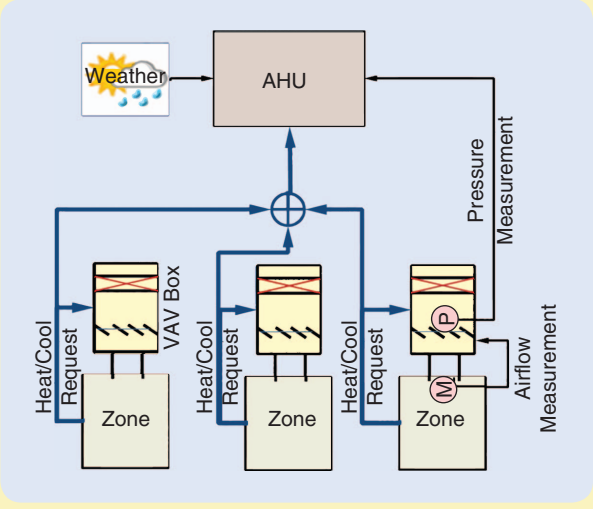

FIGURE S1 Current control logic schematic. In each thermal zone, the control system generates a cooling or heating request if the zone temperature is higher or lower than the set point, respectively. Each individual request is used to adjust set points for the corresponding variable air volume (VAV) box. The total number of cooling and heating requests is used to determine the total cooling and heating energy required to guarantee the thermal comfort. The air handling unit (AHU) supply fan is operated to track an air pressure setpoint to ensure that enough air flow can be provided by the VAV box to each zone.

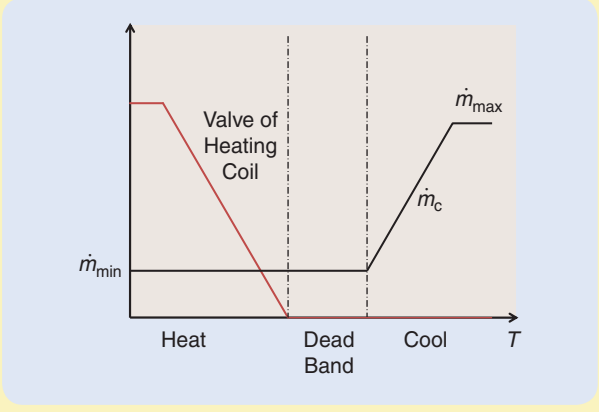

FIGURE S2 Simple proportional control logic. When the zone temperature is lower than the comfort range (dashed lines) the heating coil valve position command is proportional to the difference between zone temperature and the lower-bound on comfort level, and the air mass flow rate is set to the minimum ventilation level. When the zone temperature is within the comfort range, the variable air volume box maintains the minimum ventilation level. When the zone temperature is higher than the comfort range, the supply airflow rate increases proportionally to the difference between zone temperature and the upper bound on comfort level.

total number of cooling requests. This scheme, known as supply pressure reset, ensures that the air pressure at each VAV box is high enough to provide the requested air flow to all zones.

In addition to basic control logic, a myriad of advanced heuristics are used to reduce the overall energy consumption. For instance in certain combinations of ambient conditions and zone thermal demands, energy use can be lowered by increasing the proportion of outside air flow. This control logic is known as *economizer* operation. *Supply temperature reset* is another energy-saving control strategy and is implemented to reduce the chilled water consumption. When the zone temperatures are within the comfort constraints, the AHU supply temperature setpoint is increased slowly until one of the zones flags a cooling request. This modification to the supply air temperature setpoint enables the cooling coil to consume less chilled water, thus reducing the energy consumption of the chillers and cooling towers. However, when supply temperature is raised, zones in cooling mode require a higher flow rate to maintain the same zone temperature. Therefore, when a zone is in cooling mode the chiller energy saved is counterbalanced by increased fan energy. In this case, energy prediction models need to be used to assess the validity of supply temperature reset strategies.

$$y_{t+k|t} = g(x_{t+k|t}, u_{t+k-1|t}, d_{t+k|t}, k), \text{ for all } k = 0, 2, \ldots, N, \tag{22}$$

$$y_{t+k|t} \in \mathbb{Y}, \text{ for all } k = 1, 2, \ldots, N, \tag{23}$$

$$u_{t+k|t} \in \mathbb{U}, \text{ for all } k = 0, 1, \ldots, N-1, \tag{24}$$

$$d_{t+k|t} \in \mathbb{D}(t+k), \text{ for all } k = 1, 2, \ldots, N, \tag{25}$$

where $\mathbb{Y}$ is set of feasible system outputs y, $\mathbb{U}$ is the feasible set of control inputs u, $J_N(x)$ is the terminal cost function, $f(x, u, d, k)$ is the time varying state update equation, d is the disturbance, and $\mathbb{D}$ is the set of possible disturbances. Disturbances d_k can be predicted by a dynamic model such as building load model in Figure 10. An alternative approach is to obtain the future admissible set of disturbances $\mathbb{D}(t+k)$ by external modules such as the occupancy model in Figure 17.

Let $U^{\star}_{t\to t+N-1|t} = \{u^{\star}_{t|t}, \ldots, u^{\star}_{t+N-1|t}\}$ be the optimal solution to problems (20)–(25) at time t. Then the first element of $U^{\star}_{t\to t+N-1|t}$ is applied to the system, $u(t) = u^{\star}_{t|t}$. The optimization problems (20)–(25) are repeated at $t + \Delta t$, with the updated state $x_{t+\Delta t|t+\Delta t} = x(t+\Delta t)$, yielding a moving or receding horizon control strategy.

The cost function, model dynamics, constraints, and disturbances depend on the abstraction level and the specific problem of interest. Two detailed implementations of HMPC and LMPC schemes are presented in the following sections. We remark that when a nominal model of the disturbances d_k is replaced by a set valued model, that is $d_k \in \mathbb{D}(k)$ with a given probability distribution function, then robust or stochastic MPC formulations [21]–[25] need to be used in place of (20)–(25).

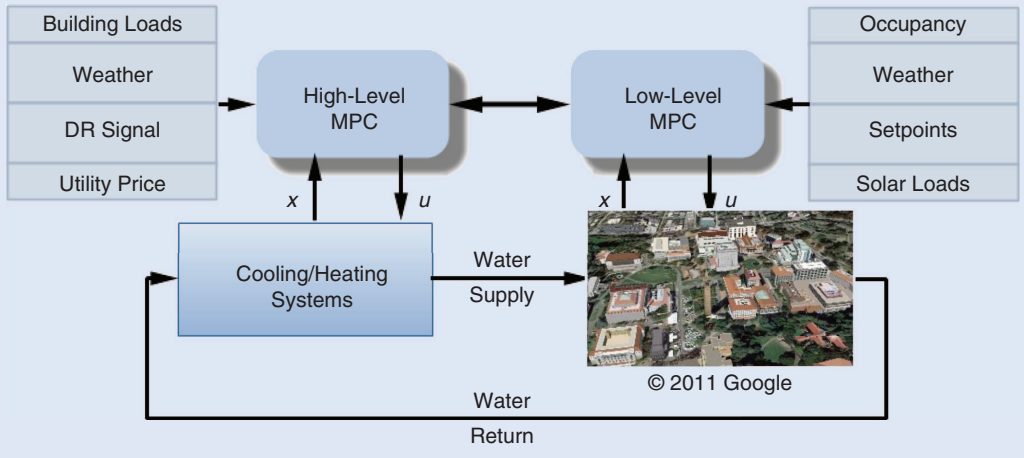

FIGURE 19 Hierarchical model predictive control (MPC) structure for a building control system. A high-level MPC (HMPC) is deployed to optimize the operation and schedule the cooling and heating systems with active thermal storage. A low-level MPC (LMPC) controls the variable air volume boxes and the air handling units by considering thermal comfort constraints of the occupants. At both levels a variety of predictions can be included in the models and in the cost function to control the system in an efficient and effective way. These predictions include building loads, load shedding signals from the power grid, utility prices, weather, occupancy, and solar loads.

MPC for High-Level Control

The objective of HMPC is to minimize the electrical energy consumption while generating enough chilled and hot water. A typical cost function of the HMPC in (20) penalizes total electricity cost and the deviation from the building thermal energy demand satisfaction. The cost can be further extended to include the peak load requests, time varying utility prices, and time varying availability of renewable energy. The control variables to be optimized by MPC include the chilled water supply temperature T_{CHWS}, condenser water supply temperature T_{CWS}, chilled water supply flow rate $\dot{m}_{\mathrm{CHWS}}$, chilling system start time t_s, and chilling system end time t_{end}. The dynamic system $f(x, u, d, k)$ includes the storage dynamics in (12)–(13), and the disturbance d includes weather and building load demand.

University of California at Merced Experimental Testing

A version of the HMPC controller is implemented at University of California, Merced. The detailed experimental setup can be found in [9] and is briefly described next. MPC computes the setpoints for cooling towers, chillers, and the thermal storage tank at the central plant. The MPC algorithm is implemented in Matlab and runs in real time on a Pentium 4 Intel processor. The MPC algorithm receives and sends data to the campus through the building automation system "Automated Logic Web Control."

Two experiments are executed in order to evaluate the benefits of MPC compared to conventional controllers. Experiment 1 (E1) is the baseline performance. During E1, the plant is operated manually using the policy defined by the plant managers. The control policy is based on the operators' experience. The data for E1 was collected from May 27–31, 2009. Experiment 2 (E2) implements the HMPC controller described in the previous section. The data for E2 was collected from October 6–10, 2009. The quantity of chilled water stored in the tank at the end of the experiment is required to be equal to the quantity available at the beginning of the experiment. Despite the difference in time, the weather conditions during E1 and E2 are similar as shown in Figure 20(a). For this reason, we can fairly compare the HMPC performance to the baseline control logic. The performance of the cooling system is measured through the COP defined by Table 3(6).

The results of real-time experiments indicate that the performance of the central plant controlled by HMPC is improved by 19% in terms of COP, which corresponds to a

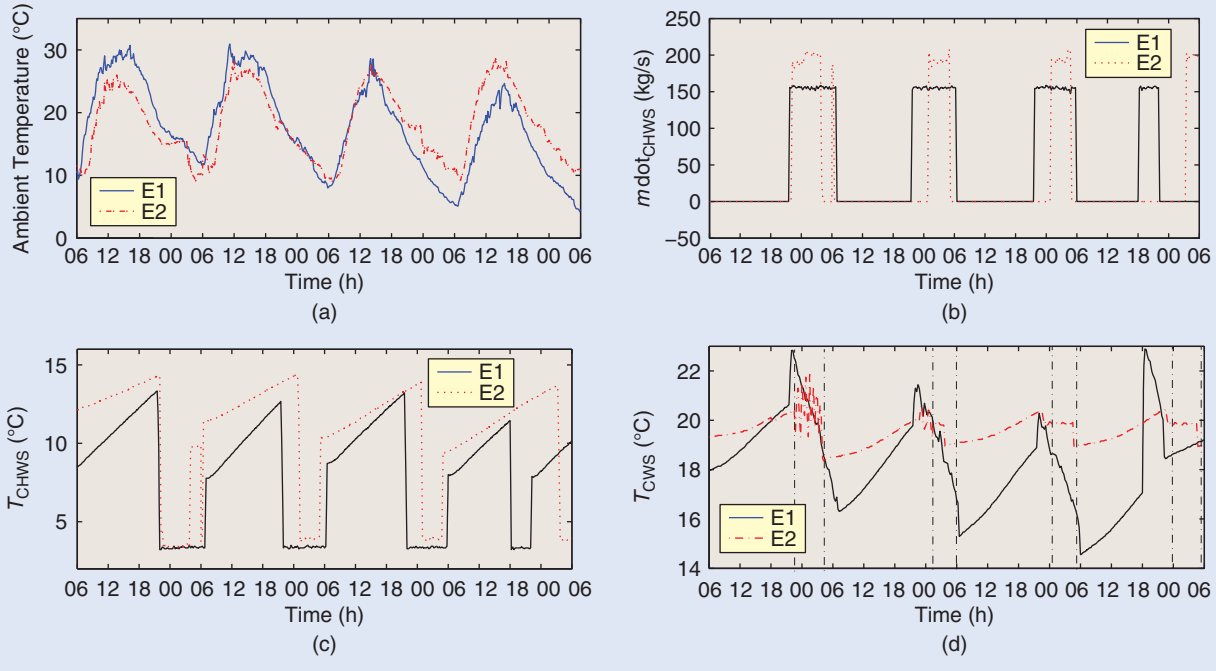

FIGURE 20 Experimental results of the central cooling plant with thermal storage at University of California, Merced, controlled by the high-level model predictive control (MPC) (E2). The plant coefficient of performance (COP) is improved by 19%, which corresponds to a total of US$1,280 weekly savings compared to the original baseline control logic (E1). We note that the MPC applies a higher mass flow rate of chilled water supply, and it schedules the chillers to operate during the period with the lowest ambient temperature. As a result, the combined chiller and cooling tower efficiency is increased. Also, MPC applies a slightly higher chilled water supply temperature, improving the COP of the chilling system according to its performance curve. (a) Ambient temperature during E1 (solid line) and E2 (dotted line), (b) mass flow rate of chilled water supply, (c) temperature of chilled water supply, and (d) temperature of condenser water supply.

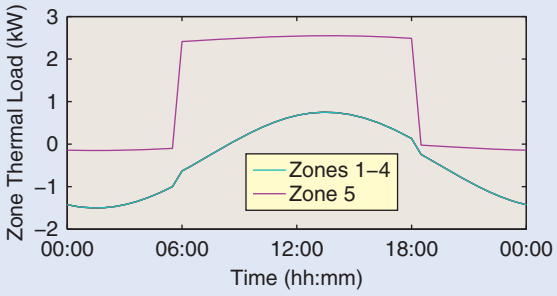

FIGURE 21 Zone thermal loads $\dot{Q}_i$. The first four zones have an equal load that requires heating, with the exception of a short period in the afternoon. Zone 5 has a high positive load that requires cooling during occupied hours, with a small negative load during unoccupied hours.

total of US$1,280 weekly savings compared to the original baseline control E1. The improvement can be explained by the control input profiles plotted in Figure 20. HMPC in E2 applies a higher mass flow rate of chilled water supply than E1 [Figure 20(b)], and uses a different schedule in order to operate the chilling system during the period with the lowest ambient temperature. As a result, the combined chiller and cooling tower efficiency is enhanced. HMPC in E2 applies a slightly higher chilled water supply temperature than E1, improving the COP of the chilling system according to the performance curve of the chilling system Table 1(1). Also, HMPC in E2 applies higher condenser water supply temperature than E1, shown in Figure 20(d). This adjustment enables the chilling system to better balance the loads on chillers and cooling towers. Further details can be found in [8] and [9].

MPC for Low-Level Control

The objective of LMPC is to minimize energy consumption in the form of cold water, hot water, and electricity while maintaining the thermal zones within the comfort range. The model for this level is described by Table 3(1) and (2) and (18) and (19). The control inputs optimized by the LMPC are AHU supply fan speed, recirculation damper position, cooling and heating coil water valve positions, and zone VAV damper positions. The cost function is the following combination of terms from Table 3(3)–(5).

$$J(x, u, k) = r_e(k)P_f(u) + r_e(k)P_c(u) + r_h(k)\sum_j P_h^j(u) \qquad (26)$$

where $r_e(k)$ and $r_h(k)$ refer to the utility rate in dollars per unit energy for electricity and heating fuel, respectively.

Simulation Results

To demonstrate LMPC results, we construct a simple five-zone building model with input thermal loads, as shown in Figure 21. The heat transfer between the five zones is neglected. The first four zones have equal and negative loads that require heating (except briefly in the afternoon). Zone 5 has a high positive load that requires cooling during occupied hours, with a small negative load in unoccupied hours.

The nominal LMPC results in Figure 22 show the tradeoff between supply temperature and mass flow rate. Between 6:30 and 10:00, we can see economizer and temperature reset-like operation where cooling of Zone 5 is performed using outside air, warmer supply temperatures, and high mass flow rates. For details on economizer and temperature reset operation see "Current Building Operation and Control Logic." This control scheme saves energy because the rest of the zones are in heating mode during this period, and any cooler supply temperature would require reheating to keep those zone temperatures above their lower bounds. Once all of the zones are in cooling mode, controlling the cooling coil to the minimum feasible supply temperature and using lower flow rates becomes a more efficient strategy. We see a brief supply temperature reset behavior again near the end of the occupied hours at 18:00. Anticipating less cooling demand for the unoccupied period, the LMPC controller starts increasing the supply temperature early.

To compare against the nominal case LMPC results in Figure 22(a) and Figure 22(b), we repeat the calculations with modified versions of the cost function. First, in Figure 23(a), we modify the electric utility rate $r_e(k)$ to have a higher value between 12:00 and 16:30. In Figure 23(b), the utility rates are constant throughout but we add an additional penalty term to the cost function to minimize the peak electric power over the entire day. The modified cost function is

$$J_{\text{mod}}(x, u, k) = r_e(k)P_f(u) + r_e(k)P_c(u) + r_h(k)\sum_j P_h^j(u) + \varphi \max_k (P_f(u) + P_c(u)), \qquad (27)$$

where φ is a penalty weighting factor in dollars per unit power. Both these modified cases in Figure 23(a) and 23(b) demonstrate precooling of Zone 5 and lengthened cooling of Zones 1–4, but with different timing and intent. In Figure 23(a), the peak electric power is not penalized in the cost function but the electric utility rate has a higher value between 12:00 and 16:30. As a consequence, precooling is only performed immediately before noon, with a corresponding spike in cooling power, so that less cooling energy is used between 12:00 and 16:30. In Figure 23(b), the peak electric power is included in the cost function, so Zone 5 is precooled earlier in the morning. As a result, cooling power is increased at a time when it would normally be low, shifting electric power use away from the times it would normally be at maximum.

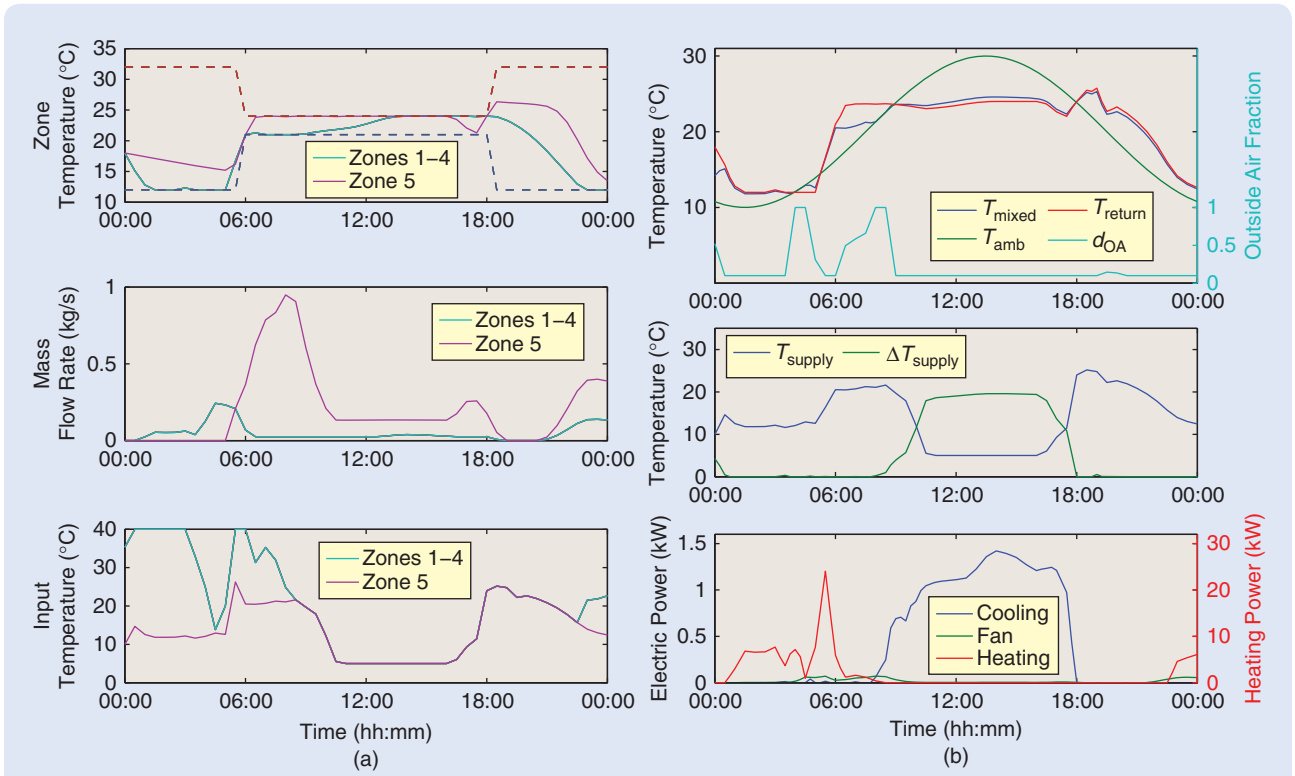

FIGURE 22 Simulation results of five thermal zones controlled by model predictive control (MPC). The results show a complicated tradeoff between supply temperature and mass flow rate. Between 6:30 and 10:00, we can see economizer and temperature reset-like operation where cooling of Zone 5 is performed using outside air, warmer supply temperatures, and high mass flow rates. This control scheme saves energy because the rest of the zones are in heating mode during this period, and any cooler supply temperature would require reheat to keep those zone temperatures above their lower bounds. Once all of the zones are in cooling mode, controlling the cooling coil to the minimum feasible supply temperature and using lower flow rates becomes a more efficient strategy. We see a brief supply temperature reset behavior again near the end of the occupied hours at 18:00. Anticipating less cooling demand for the unoccupied period, MPC starts increasing the supply temperature early. (a) Nominal case zone results and (b) nominal case air handling unit results.

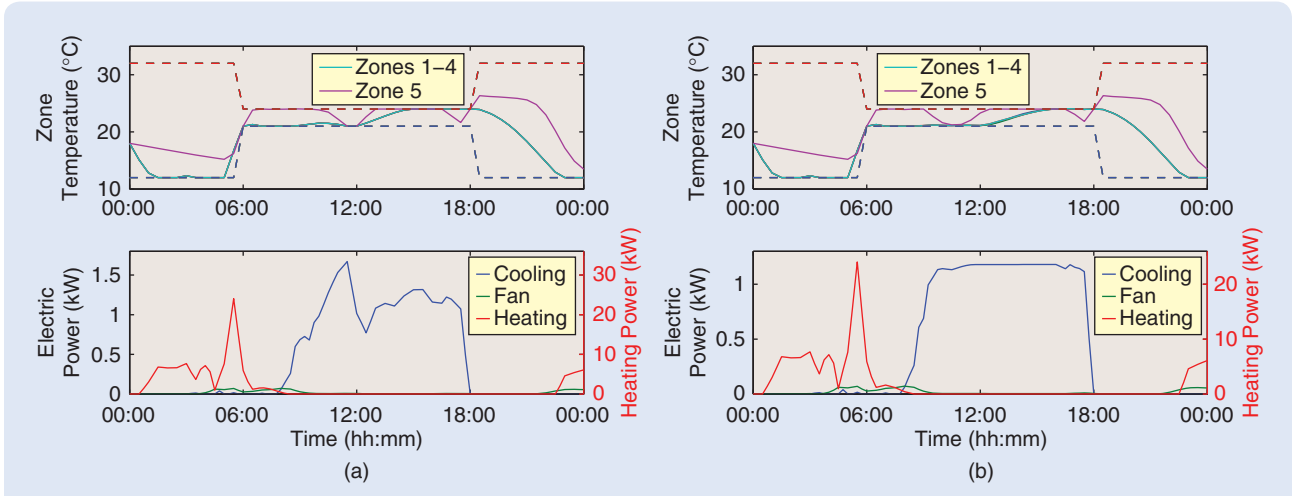

FIGURE 23 Simulation results of five thermal zones controlled by model predictive control. In (a), the electric utility rate has a higher value between 12:00 and 16:30. In (b), the utility rates are constant throughout, but the maximum electric power over the entire day is penalized in the cost function. Both figures demonstrate the precooling of Zone 5 and lengthened cooling of Zones 1–4, but with different timing and intent. In (a), precooling is only performed immediately before noon, with a corresponding spike in cooling power, so that less cooling energy is used between 12:00 and 16:30. In (b), the peak electric power is included in the cost function so Zone 5 is precooled beginning earlier in the morning. As a result, the cooling power is increased at a time when it would normally be low, shifting electric power use away from the times it would normally be at maximum. (a) Variable utility rate case and (b) peak power limiting case.

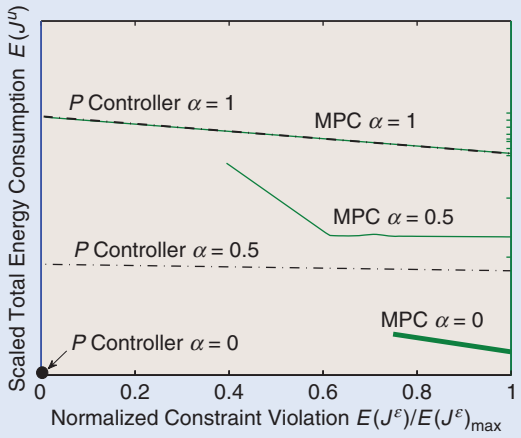

FIGURE 24 The model predictive control (MPC) is designed assuming a perfect prediction of occupancy profile. In reality, the probability of the occupancy profile being correct is α. The figure shows the expected control input cost $\mathbb{E}[J^u]$ and constraint violation $\mathbb{E}[J^\varepsilon]$ for MPC and for the proportional controller in a closed-loop. The closed-loop simulations use different occupancy load profiles depending on the chosen probability α. When the prediction is perfect ($\alpha = 1$), the performance of MPC is the same as the proportional controller in terms of total energy consumption and constraint violation. However, the MPC performance deteriorates as α decreases. In fact, MPC fails to keep the zone temperature within the comfort constraints due to the misleading predictions. MPC consumes more energy than the proportional controller for $\alpha = 0.5$ and $\alpha = 0$ because MPC is precooling even if occupants do not enter the space.

MPC MAIN ISSUES AND CURRENT RESEARCH

The appealing advantages of MPC shown though simulations and experiments in the previous sections do not come without a price. Several issues have to be considered while designing and implementing MPC for buildings. In the remainder of this section, we present some design and implementation considerations.

Design Considerations

Stability and Feasibility

The stability and feasibility of MPC are well-studied issues [3]. In particular, it has been shown that stability and feasibility are not ensured by the MPC law without terminal cost and constraints [3]. Typically the terminal constraint is a robust control invariant set so that the persistent feasibility of the MPC strategy is guaranteed. Persistent feasibility ensures that if the MPC (20)–(25) is feasible for a given initial state $x(0)$, then it is feasible for all $t \geq 0$. Definitions and properties of invariant sets can be found in [3] and [26]. In the specific context of the MPC considered in this article, the terminal set ensures that enough energy is actively stored in thermal storage elements to counteract a bounded unpredicted change in demand. A treatment of sufficient conditions guaranteeing a persistent feasibility of MPC problems goes beyond the scope of this work and can be found in the survey [3] and in [6] for the specific case of the University of California, Merced, study.

Prediction Uncertainty

The example in the first section of this article showed benefits of MPC under the assumption that MPC has perfect knowledge of predicted disturbances and system dynamics. This section tries to highlight potential issues associated with this assumption. We focus on total energy consumption using the simple MPC problem (3)–(6) with $\kappa = 0$. The control design assumes that weather prediction in Figure 1(a) is perfect and occupancy load prediction in Figure 1(b) is perfect. This time we assume that in reality the occupancy load differs from what was predicted. Two scenarios are considered. In scenario S1 the future occupancy load is exactly the same as predicted in Figure 1(b), with probability $P(S1)$ equal to α. In scenario S2, the occupancy load is zero over the entire day with probability $P(S2)$ equal to $1 - \alpha$. In short, the controller is designed based on S1 but the probability of S1 happening is α.

The expected value of the control input cost $\mathbb{E}[J^u]$ and constraint violation $\mathbb{E}[J^\varepsilon]$ for MPC C2 and proportional controller C1 are computed in a closed-loop. The closed-loop simulations use different occupancy load profiles depending on the chosen probability α.

Simulation results for various values of α and various tunings for MPC C2 and proportional controller C1 are summarized in Figure 24. When the prediction is perfect with $\alpha = 1$, the performance of MPC C2 is the same as the proportional controller C1 in terms of total energy consumption and constraint violation. However, the MPC performance deteriorates as α decreases. In fact, MPC fails to keep the zone temperature within the comfort constraints due to the misleading predictions. MPC consumes more energy than the proportional controller for $\alpha = 0.5$ and $\alpha = 0$ because MPC is performing precooling even if occupants do not enter the space.

Stochastic MPC [23], [27] might be a better approach to address this issue when probability distribution functions of the loads are available. In this case, we would minimize expected costs and satisfy constraints with a given probability. We are currently investigating this research direction and its real-time computational complexity.

Implementation Considerations

Convergence to Suboptimal Solutions

The product between air temperatures and mass flow rates in the thermodynamic energy balance (18) and (19) leads to a nonconvex MPC problem which might have distinct locally optimal solutions. Fast computational techniques for solving nonconvex optimization such as sequential

quadratic programming (SQP) can only provide certificates of local optimality. These locally optimal solutions might be less efficient than those obtained with a simpler control design. We are currently analyzing different types of local optima and their physical interpretation. The analysis can be used to derive branch and bound rules which allow an SQP solver to converge to globally optimal control sequences.

Computational Complexity of Model Predictive Control

As the complexity of the building model increases, centralized MPC might become computationally intractable due to the limited computational resources available on current building control platforms. This limitation is critical at the low level of the control architecture where distributed inexpensive computing platforms are common. The limitation might be overcome by efficient numerical solvers tailored to the specific hardware or with the use of distributed MPC [28], [29]. In distributed MPC, the centralized problem is decomposed into a set of smaller problems which can be associated with different subsystems such as VAV boxes and AHUs. Each subsystem solves local small MPC problems with information from local and neighboring subsystems. The local MPC modules communicate with each other to converge to an optimal solution [29].

An alternative approach to address the computational complexity of MPC is to precompute the control action for a set of initial states and external parameters. A lookup table can be generated by gridding the space of parameters and states and solving the optimization problem offline for each grid point. For linear and switched linear systems the gridding can be avoided using multiparametric optimization [30]. In the specific context of MPC for buildings, the authors of [31] present a rule extraction approach. Rather than running an online MPC in real time, many simulations of the MPC are executed offline. The simulation results are then used to generate simplified rule-based controllers as functions of operating conditions.

Equipment Retrofitting

MPC requires sensor data from a building in order to initialize simulations and make predictions. Additionally, there must be some way to communicate the computed optimal control inputs either to lower level controllers (for setpoint tracking) or directly to the control actuators. Modern digital building automation systems satisfy these requirements, but are only present in new buildings. In order to apply MPC to the existing stock of older buildings, HVAC equipment must be retrofitted for digital control and additional sensors need to be added or existing sensors replaced with digital versions. This can be prohibitively expensive, and must be offset by the operational energy cost savings of MPC versus the baseline control.

CONCLUSIONS

A simple thermal mass model has been used to show the basic mechanism of active thermal storage and how this mechanism naturally emerges in a predictive control scheme. The model is also used to demonstrate a fundamental tradeoff involving savings, losses, and uncertainty in load shifting. In the second part of this article, we have provided the main ingredients of a predictive control framework implementable in a building equipped with thermal storage. Both the energy conversion and the energy distribution problems have been solved using an MPC scheme. Simulations and experimental results have shown the effectiveness of the control scheme. In particular, the performance of MPC exhibits several aspects of heuristic HVAC control sequences in a coordinated manner. The delivery of the proposed predictive control technologies is contingent on gaining an understanding of several design and implementation aspects specific to the building industry. Some of these considerations have been reported in the sidebars and in the last part of the article.

ACKNOWLEDGMENTS

This research has been partially supported by the National Science Foundation under Grant No. 0844456. We want to thank the anonymous reviewers for their helpful comments on the original version of the manuscript. The results of the project at the University of California, Merced, presented in this article were carried out in collaboration with Brandon Hencey, Sorin Bengea, Brian Coffey, and Phillip Haves.

AUTHOR INFORMATION

Yudong Ma (myd07@berkeley.edu) received a B.S. in the department of mechanical engineering with a minor in mathematics from Tsinghua University, China, in 2007. He is currently a Ph.D. student in mechanical engineering at the University of California, Berkeley, conducting research on applications of MPC for energy efficiency. He can be contacted at 2169 Etcheverry Hall, 2521 Hearst Ave, Berkeley, CA 94709, USA.

Anthony Kelman received a B.S. in mechanical engineering from the California Institute of Technology in 2007. From 2007 to 2009, he worked on dynamic modeling and controls development of aircraft environmental control systems at Honeywell Aerospace in Torrance, California. He is currently a Ph.D. student in mechanical engineering at the University of California, Berkeley, conducting research on applications of MPC for energy efficiency.

Allan Daly is a registered mechanical engineer and a principal at Taylor Engineering, LLC, designing and implementing innovative and sustainable mechanical systems. He specializes in energy efficient and environmentally responsible HVAC system designs that maximize occupant health and comfort, and his current work is focused in institutional projects, commissioning work, energy modeling, and green building projects. They

include the campus of University of California, Merced, which is LEED Silver rated. He teaches and lectures widely to both professional and academic audiences on energy efficiency, integrated design, building simulation, and underfloor air distribution system design. He is a member of the U.S. Green Building Council Energy and Environment Technical Advisory Group, helping to guide the development of LEED 2012.

Francesco Borrelli received a Laurea degree in computer science engineering in 1998 from the University of Naples Federico II, Italy. In 2002 he received a Ph.D. from the Automatic Control Laboratory at ETH-Zurich, Switzerland. He is currently an associate professor in the department of mechanical engineering at the University of California, Berkeley, USA. He is the author of more than 60 publications in the field of predictive control. He is author of the book *Constrained Optimal Control of Linear and Hybrid Systems* (Springer Verlag), the recipient of the Innovation Prize 2004 from the ElectroSwiss Foundation, and a recipient of the 2009 NSF Career award. In 2008, he was appointed the chair of the IEEE Technical Committee on Automotive Control. His research interests include constrained optimal control, MPC and its application to advanced automotive control, and energy efficient building operation.

REFERENCES

[1] G. Henze, C. Felsmann, and G. Knabe, "Evaluation of optimal control for active and passive building thermal storage," *Int. J. Thermal Sci.*, vol. 43, no. 2, pp. 173–183, 2004.

[2] S. Liu and G. Henze, "Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory, Part 1: Theoretical foundation," *Energy Build.*, vol. 38, no. 2, pp. 142–147, 2006.

[3] D. Mayne, J. Rawlings, C. Rao, and P. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, pp. 789–814, June 2000.

[4] J. Braun, "Reducing energy costs and peak electrical demand through optimal control of building thermal storage," *ASHRAE Trans.*, vol. 96, no. 2, pp. 976–887, 1990.

[5] T. Nagai, "Optimization method for minimizing annual energy, peak energy demand, and annual energy cost through use of building thermal storage," *ASHRAE Trans.*, vol. 108, no. 1, pp. 976–887, 2002.

[6] Y. Ma, F. Borrelli, B. Hencey, A. Packard, and S. Bortoff, "Model predictive control of thermal energy storage in building cooling systems," in *Proc. 48th IEEE Conf. CDC/CCC*, Dec. 2009, pp. 392–397.

[7] P. Haves, B. Hencey, F. Borrellli, J. Elliott, Y. Ma, B. Coffey, S. Bengea, and M. Wetter. (2010). Model predictive control of HVAC systems: Implementation and testing. Univ. California, Merced, Final Rep. US DOE and CEC PIER. [Online]. Available: http://www.escholarship.org/uc/item/3pt2d32h.pdf

[8] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," *IEEE Trans. Contr. Syst. Technol.*, vol. PP, no. 99, pp. 1–8, 2011.

[9] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," in *Proc. American Control Conf.*, June 2010, pp. 5106–5111.

[10] M. Gwerder, B. Lehmann, J. Tödtli, V. Dorer, and F. Renggli, "Control of thermally-activated building systems (TABS)," *Appl. Energy*, vol. 85, no. 7, pp. 565–581, 2008.

[11] J. H. Lienhard, *A Heat Transfer Textbook*. Cambridge, MA: Phlogiston Press, 2002.

[12] K. Marik, Z. Schindler, and P. Stluka, "Decision support tools for advanced energy management," in *Proc. 17th Int. Congr. Chemical and Process Engineering*, June 2008, vol. 33, pp. 858–873.

[13] J. Braun and N. Chaturvedi, "Model for transient building load prediction," *HVAC&R Res.*, vol. 8, no. 1, pp. 73–99, 2002.

[14] J. Jang. (2008). System design and dynamic signature identification for intelligent energy management in residential buildings. Ph.D. dissertation, Univ. California, Berkeley. [Online]. Available: http://escholarship.org/uc/item/0v83w3kw.pdf

[15] C. Liao and P. Barooah, "An integrated approach to occupancy modeling and estimation in commercial buildings," in *Proc. American Control Conf.*, June 2010, pp. 3130–3135.

[16] V. Erickson, Y. Lin, A. Kamthe, B. Rohini, A. Surana, A. Cerpa, M. Sohn, and S. Narayanan, "Energy efficient building environment control strategies using real-time occupancy measurements," in *Proc. 1st ACM Workshop Embedded Sensing Systems for Energy Efficiency in Buildings*, 2009, pp. 19–24.

[17] S. Meyn, A. Surana, Y. Lin, S. Oggianu, S. Narayanan, and T. Frewen, "A sensor-utility-network method for estimation of occupancy in buildings," in *Proc. 48th IEEE Conf. CDC/CCC*, Dec. 2009, pp. 1494–1500.

[18] F. Romie, "Transient response of the counterflow heat exchanger," *J. Heat Transfer*, vol. 106, no. 3, pp. 620–626, 1984.

[19] R. Chillar and R. Liesen, "Improvement of the ASHRAE secondary HVAC toolkit simple cooling coil model for simulation," in *Proc. SimBuild Conf.*, Boulder, CO, Aug. 2004, pp. 1–7.

[20] Y. Wang, W. Cai, Y. Soh, S. Li, L. Lu, and L. Xie, "A simplified modeling of cooling coils for control and optimization of HVAC systems," *Energy Convers. Manage.*, vol. 45, no. 18–19, pp. 2915–2930, 2004.

[21] S. V. Raković, E. C. Kerrigan, D. Q. Mayne, and K. I. Kouramas, "Optimized robust control invariance for linear discrete-time systems: Theoretical foundations," *Automatica*, vol. 43, no. 5, pp. 831–841, 2007.

[22] A. Bemporad, F. Borrelli, and M. Morari, "Robust model predictive control: piecewise linear explicit solution," in *Proc. European Control Conf.*, Porto, Portugal, Sept. 2001, pp. 939–944.

[23] M. Cannon, B. Kouvaritakis, and X. Wu, "Probabilistic constrained MPC for multiplicative and additive stochastic uncertainty," *IEEE Trans. Automat. Contr.*, vol. 54, no. 7, pp. 1626–1632, 2009.

[24] D. Van Hessem, C. Scherer, and O. Bosgra, "LMI-based closed-loop economic optimization of stochastic process operation under state and input constraints," in *Proc. 40th IEEE Conf. Decision and Control*, Dec. 2001, vol. 5, pp. 4228–4233.

[25] P. Couchman, M. Cannon, and B. Kouvaritakis, "Stochastic MPC with inequality stability constraints," *Automatica*, vol. 42, no. 12, pp. 2169–2174, 2006.

[26] F. Blanchini, "Set invariance in control—A survey," *Automatica*, vol. 35, pp. 1747–1768, Nov. 1999.

[27] F. Oldewurtel, A. Parisio, C. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, "Energy efficient building climate control using stochastic model predictive control and weather predictions," in *Proc. American Control Conf.*, Baltimore, MD, June 2010, pp. 5100–5105.

[28] J. Rawlings and B. Stewart, "Coordinating multiple optimization-based controllers: New opportunities and challenges," *J. Process Control*, vol. 18, no. 9, pp. 839–845, 2008.

[29] Y. Ma and F. Borrelli, "A distributed predictive control approach to building temperature regulation," in *Proc. American Control Conf.*, June 2011, pp. 2089–2094.

[30] F. Borrelli, *Constrained Optimal Control of Linear and Hybrid Systems*, vol. 290. Berlin: Springer-Verlag, 2003.

[31] P. May-Ostendorp, G. Henze, C. Corbin, B. Rajagopalan, and C. Felsmann, "Model predictive control of mixed-mode buildings with rule extraction," *Build. Environ.*, vol. 46, no. 2, pp. 428–437, 2011.

[32] J. Grabe and S. Winter. (2007, Oct.). The correlation between PMV and dissatisfaction on the basis of the ASHRAE and the McIntyre scale—Towards an improved concept of dissatisfaction. Int. Soc. Built Environ., Tech. Rep. [Online]. Available: http://ibe.sagepub.com/content/17/2/103.full.pdf+html

[33] B. Olesen and G. Brager. (2004, Aug.). A better way to predict comfort: The new ASHRAE Standard 55-2004. ASHRAE J. [Online]. Available: http://escholarship.org/uc/item/2m34683k.pdf

[34] B. Coffey, *Using Building Simulation and Optimization to Calculate Lookup Tables for Control*. Berkeley, CA: University of California at Berkeley, 2003, ch. 7, pp. 94–117.