Model Predictive Control Approach to Online Computation of Demand-Side Flexibility of Commercial Buildings HVAC Systems for Supply Following

Mehdi Maasoumy†, Catherine Rosenberg$^{\diamond}$, Alberto Sangiovanni-Vincentelli*, Duncan S. Callaway$^{\triangle}$

Abstract— Commercial buildings have inherent flexibility in how their HVAC systems consume electricity. We investigate how to take advantage of this flexibility. We first propose a means to define and quantify the flexibility of a commercial building. We then propose a contractual framework that could be used by the building operator and the utility to declare flexibility on the one side and reward structure on the other side. We then design a control mechanism for the building to decide its flexibility for the next contractual period to maximize the reward, given the contractual framework. Finally, we perform at-scale experiments to demonstrate the feasibility of the proposed algorithm.

I. INTRODUCTION

Consumers of energy usually do not pay attention to *when* they use energy. Demand for electricity tends to rise specially at times when it seems natural to use it; So natural in fact that we all tend to use it at the same time and in similar ways. On days when demand for electricity is high, extra power plants are needed to meet the demand.

There are environmentally-friendly alternatives to adding more generation. Electricity customers can voluntarily trim their electricity use during peak periods or choose to use electricity at a less congested time, a strategy called "Supply-Following" (SF). With SF programs, utilities provide *incentives* to encourage consumers to reduce their demand during peak periods. Smart grids are enabling the transformation from a demand-following strategy to a supply-following one.

Smart buildings are essential elements of the smart grid and can play a significant role in its operation. A building has some inherent flexibility in the way its HVAC system consumes electricity while respecting the comfort of its occupants. This flexibility could be used to reduce costs if the electricity price is time-varying, or could be traded (i.e., sold to the utility) if a proper contractual framework exists. Typically the electricity consumption of a commercial building is controlled by a process that aims at reducing costs. Among the many possible ways to design such a process, Model Predictive Control (MPC) stands out. In order to use the potential of commercial buildings as providers of *flexibility* to the smart grid, we need to fundamentally redesign the way a building, and in particular its HVAC,

†Department of Mechanical Engineering, University of California, Berkeley, CA 94720, USA. Corresponding author. Email: mehdi@me.berkeley.edu
$^{\diamond}$Department of Electrical and Computer Engineering, University of Waterloo, Canada.
*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA.
$^{\triangle}$Energy and Resources Group, University of California, Berkeley, CA 94720, USA.

is controlled. In this paper, we identify and quantify "flexibility". Then, a contractual framework between the utility and the building operator is designed so that the building can "declare" its flexibility and be rewarded for it. Finally, a control algorithm is proposed that allows the operation of building under this framework and at-scale experiments are carried out to demonstrate the high potential of commercial buildings as a source of flexibility, and the feasibility of the developed algorithm.

II. BASELINE SYSTEM AND CONTRACT

We focus on commercial buildings due to the following reasons: 1) Commercial buildings account for more than 35% of electricity consumption in the US. 2) More than 30% of commercial buildings have adopted Building Energy Management System (BEMS) technology which facilitate the communication with the grid system operators for providing flexibility. The majority of these buildings are also equipped with variable frequency drives, which in coordination with BEMS, can modulate the heating, ventilation and air conditioning (HVAC) system power consumption frequently (in the order of seconds). 3) Compared to a typical residential building, commercial buildings typically have larger HVAC systems and therefore consume more electricity. About 15% of electricity consumption in commercial buildings is related to the fans of HVAC systems. Fans are the main drivers to move the conditioned air from the air handling units (AHU) to the rooms for climate control. For instance, the main supply fans that feed Sutardja Dai Hall on UC Berkeley campus can spin at variable speeds, with the corresponding power consumption which is proportional to the cube of fan speed, with the maximum rated power of 134 KW or about 14% of the maximum power consumed in that building. Moreover, we can directly control their power, upward or downward making it an ideal candidate for ancillary demand-response. See [1]–[5] for more information about the physics and control of HVAC systems.

A. The Baseline System

We consider a commercial building that has an HVAC system controlled by an MPC as our baseline. The MPC has the objective of minimizing the total energy cost (in dollar value). Let the time be slotted with τ as the length of each time slot, and let H^m be the prediction horizon (in number of time-slots) of the MPC. System dynamics is also discretized with a sampling time of τ. Typical values for τ and H^m range from 15 min to 1 hr for τ and a

few hours to a few days, e.g., 3 to 72 time slots (with the assumption of 1 hr for each time slot) for H^m. The choice of H^m depends on how far in the future the estimation of the predicted values have an acceptable accuracy, and also on how far in the future the required information (e.g. cost of energy) is available. In this paper we use $\tau = 1$ hr and $H^m = 24$ hrs. At each time t, the predictive controller solves the following optimal control problem to compute the optimal vector $\vec{u}_t = [u_t, \ldots, u_{t+H^m-1}]$, where u_{t+k} is the air mass flow to the thermal zones of the building. The inputs to the optimal control problem are the states (i.e. zone temperatures) x_t (as initial condition), the set of *electric* energy prices $\{\pi_t^e, \ldots, \pi_{t+H^m-1}^e\}$, the *non-electric* and non-fan energy prices, such as gas price for heating $\pi^{ne,h}$, and cooling $\pi^{ne,c}$ which are considered time-invariant, a set of constraints $\mathcal{X}_{t+k}$ on the system states of the type: "x_{t+k} should be in $\mathcal{X}_{t+k}$ for all times $t+k$ where $k \in \{1, \ldots, H^m\}$)", a set of constraints $\mathcal{U}_{t+k}$ on system inputs for all $k \in \{0, \ldots, H^m - 1\}$, and an estimate on unmodelled disturbances d_{t+k} (e.g., outside temperature or building occupancy [6]) for all $k \in \{0, \ldots, H^m - 1\}$ for the next H^m time slots. The optimization problem is as follows:

$$\min_{\vec{u}_t} \sum_{k=0}^{H^m-1} C_{\text{hvac}}(u_{t+k}, \pi_{t+k}^e, \pi^{ne,c}, \pi^{ne,h}, T_{t+k}^{\text{out}}) \quad (1)$$

$$\text{s.t.} \quad x_{t+k+1} = f(x_{t+k}, u_{t+k}, d_{t+k}), \quad k = 0, \ldots, H^m - 1$$

$$x_{t+k} \in \mathcal{X}_{t+k}, \quad\quad\quad\quad k = 1, \ldots, H^m$$

$$u_{t+k} \in \mathcal{U}_{t+k}, \quad\quad\quad\quad k = 0, \ldots, H^m - 1$$

where T^{out} is the outside air temperature, f captures the system dynamics, and the total HVAC power consumption cost $C_{\text{hvac}}(t)$ is the summation of fan power, cooling power and heating power, given by:

$$C_{\text{hvac}}(t) = \pi^e(t)P_f(t) + \pi^{ne,c}P_c(t) + \pi^{ne,h}P_h(t) \quad (2)$$

where with the assumption of no recirculation of air, and without loss of generality, these power consumptions for time slot $[t, t+1]$ (with the assumptions of constant air mass flow and price of energy over $[t, t+1]$) are calculated as follows:

$$P_f(u_t) = c_1.u_t^3 + c_2.u_t^2 + c_3.u_t + c_4 \quad (3)$$

$$P_h(u_t, T_t^{out}) = c_p.u_t.(T^s - T_t^{out})/COP_h \quad (4)$$

$$P_c(u_t, T_t^{out}) = c_p.u_t.(T_t^{out} - T^s)/COP_c \quad (5)$$

where c_1, c_2, c_3, and c_4 are constants of the fan as shown in Fig. 1, c_p is the specific heat of air, and COP_h, and COP_c are the coefficients of performance for the heating system and the cooling system, respectively, and the supply air temperature T^s is considered constant. To move the coolant fluid around, heating and cooling systems use pumps which consume electric power. However, we assume that electric power consumption of pumps is negligible compared to the non-electric heating and cooling powers of these systems. (see [3], [4] for more details).

Hence, the MPC solution to (1) is an optimal air mass flow trajectory (also called an air mass flow profile) $\vec{u}_t^* = [u_t^*, \ldots, u_{t+H^m-1}^*]$. Only the first entry of $\vec{u}_t^*$ is implemented at time t. At the next time step, $t+1$, the horizon of MPC
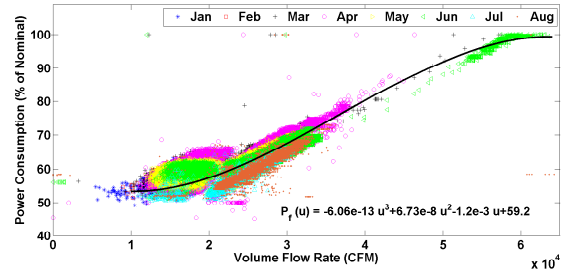

Fig. 1. Fan power consumption versus volume flow rate. Data is for January through August of 2013.

is receded by one step, a new MPC is set up, and solved to obtain $\vec{u}_{t+1}^* = [u_{t+1}^*, \ldots, u_{t+H^m}^*]$. Again, the first entry is implemented on the system, and the horizon is receded, and this process repeats until the whole time frame of interest is covered. At each time t, the optimal air mass flow vector for the next H time steps is given by $\vec{u}_t^* = [u_t^*, \ldots, u_{t+H^m-1}^*]$, with corresponding fan power flow profile of $P_f(\vec{u}^*) = [P_f(u_t^*), \ldots, P_f(u_{t+H^m-1}^*)]$. In the following, we will omit the reference to t in the profile and define a power profile as a vector $P_f(\vec{u}^*) = [P_f(u_0^*), \ldots, P_f(u_{H^m-1}^*)]$.

B. The Baseline Contract

The *baseline contract* corresponding to this baseline system is very simple. Namely, the building can consume what it wants at all times and it pays π_t^e per unit of consumed electricity at time t (i.e., if the time span is of duration Δ, then it pays $\sum_{k=t}^{t+\Delta} \pi^e(k)P_f(k)$ for electric power). Hence, the only information being exchanged between the utility and the BM is the real-time price vector $[\pi_0^e, \ldots, \pi_{H^m-1}^e]$ in (\$/kWh) sent by the utility to the building ($\pi^{ne,c}$ and $\pi^{ne,h}$ are considered constant and known). Then the building solves the MPC (1) to compute what its electricity consumption profile/trajectory would be in the next H^m time slots. As described, this MPC would minimize its total cost in the next H^m time slots while respecting the building and the occupants (comfort) constraints. Let $\vec{u}^*$ denote the trajectory that is the solution to this MPC and let $C(\vec{u}^*)$ be its cost.

III. FLEXIBILITY, CONTRACTUAL FRAMEWORK AND CONTROL ALGORITHM

A. Definition of Flexibility

We define flexibility in terms of two envelope air mass flow profiles $\vec{e^l} = [e_{t_{cs}}^l, \ldots, e_{t_{ce}}^l]$ a lower envelope, and $\vec{e^u} = [e_{t_{cs}}^u, \ldots, e_{t_{ce}}^u]$ an upper envelope. t_{cs} and t_{ce} are contract start and end times. Based on the requirement from the utility that flexibility of each building has to be determined ahead of time we pick t_{cs}, and consequently t_{cs} as follows:

$$t_{cs} \geq 1 \quad (6)$$

$$t_{cs} + 1 \leq t_{ce} \ll H^m - 1 \quad (7)$$

$$t_{ce} - t_{cs} = H^c \quad (8)$$

where H^c is the length of the contract. Typical values for H^c are much smaller than H^m, and can take values from one slot to a few time slots. The BEMS computes $\vec{e^l} =$

Parameter	Definition
τ	Sampling time for discretizing continuous system dynamics
H^m	Prediction horizon of MPC
H^c	Horizon (length) of the proposed contract
d_t	Disturbance to the system at time t which comprises outside temperature, occupancy, solar radiation, etc.
T_t^{out}	Outside air temperature at time t
T^s	Supply air temperature exiting air handling unit (AHU)
$\mathcal{X}_t$	Set of permissible states at time t
$\mathcal{U}_t$	Set of permissible inputs at time t
π_t^e	Per-unit price of electric energy at time t – ($/kWh$)
$\pi^{ne,c}$	Per-unit price of non-electric cooling energy – ($/kWh$)
$\pi^{ne,h}$	Per-unit price of non-electric heating energy – ($/kWh$)
COP_h	Coefficient of performance of heating system
COP_c	Coefficient of performance of cooling system
t_{cs}	Contract start time
t_{ce}	Contract end time
$\{\overline{\beta}_t, \underline{\beta}_t\}$	Reward paid from the utility to the building at time t for providing upward flexibility ($\overline{\beta}_t$),and down flexibility ($\underline{\beta}_t$) – ($/kW$)

Variable	Definition
x_t	State of the system at time t
u_t	Input to the system at time t
w_t	Uncertainty variable introduced to derive the worst-case (*robustified*) optimal control problem
e_k^u	Upper envelope for safe air mass flow
e_k^l	Lower envelope for safe air mass flow
$\{\overline{\varphi}_t, \underline{\varphi}_t\}$	Provided upward ($\overline{\varphi}$), and downward ($\underline{\varphi}$) flexibility by building at time t in unit of air flow
$\{\overline{\psi}_t, \underline{\psi}_t\}$	Provided upward ($\overline{\psi}$), and downward ($\underline{\psi}$) flexibility by building at time t in unit of power
$C_{\text{hvac}}(u_t, \pi_t^e)$	Total HVAC energy consumption cost at time t
P_f, P_h, P_c	Power consumption of fan, heating and cooling systems
$R(\Phi, \mathcal{B})$	Reward from utility to building for providing flexibility

$[e_0^l, \ldots, e_{H^m-1}^l]$ a lower envelope and $\vec{e^u} = [e_0^u, \ldots, e_{H^m-1}^u]$ an upper envelope (using the algorithm described later in the paper), so that any air mass flow profile $\vec{u} = [u_0, \ldots, u_{H^m-1}]$ such that for all $k \in \{0, \ldots, H^m - 1\}$, $e_k^l \leq u_k \leq e_k^u$ is feasible; i.e. no building constraints are violated at anytime. The corresponding fan power consumption envelopes are $P_f(\vec{e^l}) = [P_f(e_0^l), \ldots, \vec{P}_f(e_{H^m-1}^l)]$ and $P_f(\vec{e^u}) = [P_f(e_0^u), \ldots, P_f(e_{H^m-1}^u)]$. However the building only *declares* the first H^c values of the envelope as flexibility, namely, $\vec{e^l} = [e_{t_{cs}}^l, \ldots, e_{t_{cs}+H^c}^l]$ a lower envelope, and $\vec{e^u} = [e_{t_{cs}}^u, \ldots, e_{t_{cs}+H^c}^u]$ an upper envelope, at the beginning of each contract. The reason for declaring a subset of the obtained envelopes is that due to model mismatch, imperfect predictions of disturbance and so on, the later values in the H^m-step envelopes may not be accurate and need to be updated in the next time step. By declaring these two envelopes, the building has essentially declared its flexibility for the next H^c time slots. Note that there is NO objective function and energy cost here, we define flexibility with respect to feasibility criteria.

By declaring these two envelopes, the building manager is telling the utility: "I allow you to select any power trajectory $P_f(\vec{u}) = [P_f(u_{t_{cs}}), \ldots, P_f(u_{t_{ce}})]$ such that for all $k \in \{t_{cs}, \ldots, t_{ce}\}$, $P_f(e_k^l) \leq P_f(u_k) \leq P_f(e_k^u)$".

To quantify the flexibility provided at each time step k, we need to define a single *metric*. The metric that is natural to use is the difference between the upper and lower power envelope that can be consumed by the building without violating any constraints. Hence,

$$\text{Flexibility}(k) \triangleq P_f(e_k^u) - P_f(e_k^l) \qquad (9)$$

B. Contract Framework

In the proposed framework, the building operator can declare its flexibility contract to the utility. The flexibility declared by the building operator would only be significant if the reward is appropriate. It is important to understand that by allowing the utility to select any trajectory $P_f(\vec{u}) = [P_f(u_{t_{cs}}), \ldots, P_f(u_{t_{ce}})]$ or correspondingly $\vec{u} = [u_{t_{cs}}, \ldots, u_{t_{ce}}]$ such that for all $k \in \{t_{cs}, \ldots, t_{ce}\}$ $e_k^l \leq u_k \leq e_k^u$, the building might consume more or be in a worse state at the end of the H^c time slots. We define the "flexibility" contract as follows:

The utility charges the building operator for its baseline power consumption $P_f(\vec{u}^*)$ which is agreed upon at the time of the contract, irrespective of the deviations from the baseline power consumption due to flexibility signals from the utility, and the utility rewards the building operator for its declared flexibility.

Hence, this contract is deterministic in the sense that from the beginning of the contract, the utility and the building operator both know how much money each has to pay; the building pays for consuming $P_f(u_t^*)$ at rate π_t^e, and the utility pays the building for providing downward flexibility at rate $\underline{\beta}_t$ and for providing upward flexibility at rate $\overline{\beta}_t$.

In short, the contract implementation steps are as follows:

1) The building and the utility agree upon a contract length, H^c.
2) The utility declares $[\pi_0^e, \ldots, \pi_{H^m-1}^e]$, as per unit price of electric energy, $[\underline{\beta}_0, \ldots, \underline{\beta}_{H^m-1}]$ in ($/kW$), as reward for down flexibility, and $[\overline{\beta}_0, \ldots, \overline{\beta}_{H^m-1}]$ in ($/kW$), as reward for upward flexibility. If the utility is not willing to commit to the flexibility rates for the time span beyond the next immediate contract period, i.e. $[\underline{\beta}_{H^c+1}, \ldots, \underline{\beta}_{H^m-1}]$, and $[\overline{\beta}_{H^c+1}, \ldots, \overline{\beta}_{H^m-1}]$, the building operator can then obtain an estimate of these values from historical data.
3) The building operator computes the baseline air mass flow u_k^* and the two envelopes e_k^l and e_k^u, for the time frame $k = 0, 1, \ldots, H^m - 1$, with its overall cost if it uses the flexibility contract as follows:

$$C_f = \overbrace{\sum_{k=0}^{H^m-1} C_{\text{hvac}}(u_k, \pi_k^e)}^{\text{Building Operator Payment}} \qquad (10)$$
$$\underbrace{- \sum_{k=1}^{H^m-1} \underline{\beta}_k \underline{\psi}(u_k, e_k^l) - \sum_{k=1}^{H^m-1} \overline{\beta}_k \overline{\psi}(u_k, e_k^u)}_{\text{Reward for Providing Flexibility}}$$

where $\overline{\psi}$, and $\underline{\psi}$ are defined as:

$$\overline{\psi}(u_k, e_k^u) \triangleq P_f(e_k^u) - P_f(u_k) \tag{11a}$$

$$\underline{\psi}(u_k, e_k^l) \triangleq P_f(u_k) - P_f(e_k^l) \tag{11b}$$

4) The building operator then declares $P_f(\vec{e^l})$ and $P_f(\vec{e^u})$, the baseline profile $P_f(\vec{u}^*)$ that it will consume and the length of the contract H^c to the utility.

5) In the next H^c time slots, the utility will send signals (s_k)'s, such that $P_f(e_k^l) \leq s_k \leq P_f(e_k^u)$ and the building operator has to obey the signals, i.e., has to consume power in time slot k equal to s_k. Flexibility signal s_k may arrive as frequently as every few seconds, as mentioned earlier.

C. Proposed Control Mechanism

As discussed in Section III-B, if the utility provides a vector $\vec{\pi}^e = [\pi_0^e, \ldots, \pi_{H^m-1}^e]$ of per-unit of energy prices per time slot, then BEMS could try to minimize its total energy cost $\sum_{k=0}^{H^m-1} C_{\text{hvac}}(u_k, \pi_k, T_k^{out})$. In that case, BEMS uses building flexibility selfishly to minimize its cost. Hence, BEMS does value (and does use) its flexibility. If the utility values the flexibility that the building HVAC system can offer, it has to provide the right incentive and the right mechanism to declare this flexibility.

1) Formalizing Flexibility and Incentive in the MPC Framework: We say that the building can offer a flexibility $\Psi := \{\vec{\underline{\psi}}, \vec{\overline{\psi}}\}$ in fan power or equivalently a flexibility $\Phi := \{\vec{\underline{\varphi}}, \vec{\overline{\varphi}}\}$ in air mass flow which comprises down flexibility, $\vec{\underline{\varphi}}$, and upward flexibility $\vec{\overline{\varphi}}$, from the contract start time $t_{cs} = t + 1$ to the contract end time $t_{ce} := t + 1 + H^c$ for $1 \leq H^c \leq H^m$ time slots (starting from x_0) if there exist two trajectories $\vec{e^l} = \vec{u} + \vec{\underline{\varphi}}$ and $\vec{e^u} = \vec{u} + \vec{\overline{\varphi}}$, that satisfy:

$$\underline{\varphi}_k \leq 0, \quad \overline{\varphi}_k \geq 0 \qquad \forall k \in \{t_{cs}, \ldots, t_{ce}\} \tag{12}$$

$$f(x_k, u_k + \underline{\varphi}_k, d_k) \in \mathcal{X}_{k+1} \quad \forall k+1 \in \{t_{cs}, \ldots, t_{ce}\} \tag{13}$$

$$f(x_k, u_k + \overline{\varphi}_k, d_k) \in \mathcal{X}_{k+1} \quad \forall k+1 \in \{t_{cs}, \ldots, t_{ce}\} \tag{14}$$

$$u_k + \underline{\varphi}_k \in \mathcal{U}_k \qquad \forall k \in \{t_{cs}, \ldots, t_{ce}\} \tag{15}$$

$$u_k + \overline{\varphi}_k \in \mathcal{U}_k \qquad \forall k \in \{t_{cs}, \ldots, t_{ce}\} \tag{16}$$

What we defined here is a flexibility over multiple time slots. Assume that BEMS declares $\vec{\underline{\varphi}}$ and $\vec{\overline{\varphi}}$ (i.e., the utility might choose any fan power (and consequently air flow $\hat{u}_k$) in time step $t_{cs} \leq k \leq t_{ce}$ as long as $u_k^* + \underline{\varphi}_k \leq \hat{u}_k \leq u_k^* + \overline{\varphi}_k$), where u_k^* is the *baseline* air mass flow. Hence, we "center" the flexibility around u^*.

The flexibility-aware optimal control problem should minimize the cost function which is composed of the *cost* for the baseline HVAC power consumption, C_{hvac}, (i.e., the cost of problem (1)), minus the *reward* for the flexibility, R, which is computed as follows:

$$R(\vec{\Phi}, \vec{\mathcal{B}}) = \vec{\overline{\beta}}^T . \vec{\overline{\psi}}(\vec{u}, \vec{\overline{\varphi}}) + \vec{\underline{\beta}}^T . \vec{\underline{\psi}}(\vec{u}, \vec{\underline{\varphi}}) \tag{17}$$

where $\vec{\mathcal{B}} := \{\vec{\underline{\beta}}, \vec{\overline{\beta}}\}$, and $\underline{\psi}(.)$, and $\overline{\psi}(.)$ are given by (11), in which $\vec{e^l} = \vec{u} + \vec{\underline{\varphi}}$ and $\vec{e^u} = \vec{u} + \vec{\overline{\varphi}}$.

2) Putting It All Together: We assume at each time step t, that the current state of the building is known. Furthermore, the prediction of the outside temperature, inside heat generation, and constraints on the system states and inputs are known. The outputs of the algorithm is the nominal power consumption of the building, u_{t+k}^* for $k \in \{0, \ldots, H^m - 1\}$, the flexibility that the building can provide, Φ_{t+k+1}^* for $k \in \{0, \ldots, H^m - 1\}$, for future time steps, without violating constraints. Based on this set of information, the BEMS can decide how much flexibility to offer. It may declare very small flexibility and hence get a cost close to C_{hvac}.

The algorithm described above can be formulated as a *min-max* MPC problem. At time t we are interested in solving the following robust optimal control problem:

Min-max Problem

$$\min_{\vec{u}_t, \vec{\Phi}_{t+1}} \max_{\vec{w}_t} \sum_{k=0}^{H^m-1} C_{\text{hvac}}(u_{t+k}, \pi_{t+k}) - R(\Phi_{t+k+1}, \mathcal{B}_{t+k+1}) \tag{18a}$$

subject to:

$$x_{t+k+1} = f(x_{t+k}, u_{t+k} + w_{t+k}, d_k)$$
$$\forall \quad k = 0, \ldots, H^m - 1 \tag{18b}$$

$$\forall w_t \text{ s.t. : } \underline{\varphi}_t \leq w_t \leq \overline{\varphi}_t \tag{18c}$$

$$\forall w_{t+k} \text{ s.t. : } \underline{\varphi}_{t+k} \leq w_{t+k} \leq \overline{\varphi}_{t+k}$$
$$\forall k = 1, \ldots, H^m - 1 \tag{18d}$$

$$\overline{\varphi}_{t+k} \geq 0, \quad \forall k = 1, \ldots H^m - 1 \tag{18e}$$

$$\underline{\varphi}_{t+k} \leq 0, \quad \forall k = 1, \ldots H^m - 1 \tag{18f}$$

$$x_{t+k} \in \mathcal{X}_{t+k} \, \forall k = 1, \ldots, H^m \tag{18g}$$

$$u_{t+k} + w_{t+k} \in \mathcal{U}_{t+k} \, \forall k = 0, \ldots, H^m - 1 \tag{18h}$$

Note that $\underline{\varphi}_t$ and $\overline{\varphi}_t$ are computed in the previous time step and are constant values in this formulation, while $\underline{\varphi}_{t+k}$ and $\overline{\varphi}_{t+k}$ for $k \in \{1, \ldots, H^m-1\}$ are optimization variables and will be computed in the current time step by solving the optimal control problem (18).

The inner maximization problem robustifies the optimization problem and derives the worst-case scenario cost and constraints. The outer minimization problem solves for its arguments $(\vec{u}_t, \vec{\Phi}_{t+1})$ while it is guaranteed that the constraints are satisfied for all values of uncertainty w, as long as it is within the range $\underline{\varphi}_{t+k} \leq w_{t+k} \leq \overline{\varphi}_{t+k}$ for $k \in \{1, \ldots, H^m - 1\}$.

Theorem [7]:

Let $\mathbb{C}$ be a closed convex set and let $f : \mathbb{C} \to \mathbb{R}$ be a convex function. Then if f attains a maximum over $\mathbb{C}$, it attains a maximum at some extreme point of $\mathbb{C}$.

According to the theorem above, the *analytic* worst-case solution can be obtained. Closed convex sets can be characterized as the intersections of closed half-spaces (i.e. sets of points in space that lie on and to one side of a hyperplane). The feasible set for states, i.e. the temperature of the rooms in the building and inputs, i.e. air mass flow into the thermal
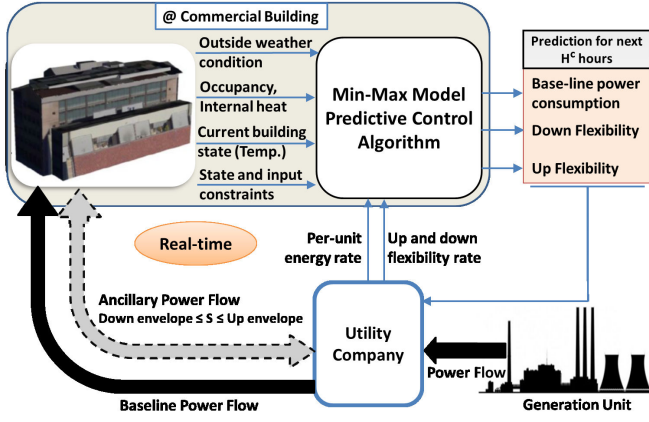
Fig. 2.   Schematic of the proposed architecture for contractual framework.

zones, are defined by:

$$\mathcal{X}_k := \{x \mid \underline{T}_k \le x \le \overline{T}_k\} \tag{19}$$

$$\mathcal{U}_k := \{u \mid \underline{U}_k \le u \le \overline{U}_k\} \tag{20}$$

where $\underline{T}_k$, and $\overline{T}_k$ are the upper and lower temperature limits and $\overline{U}_k$, and $\underline{U}_k$ are the upper and lower feasible air mass flow at time t. Therefore, the feasible set of (18) is closed and convex. The objective function is also convex on w, as the max is over variable w. In fact, w does not appear in the cost function. The objective function is a linear function on w and hence it is concave. We also consider a linearized state update equation as follows:

$$x_{t+1} = Ax_t + Bu_t + Ed_t \tag{21}$$

the nonlinear system dynamics has been linearized with the forward Euler integration formula with time-step $\tau = 1$ hr. Hence, the min-max problem (18) is equivalent to:

Minimization Problem

$$\min_{\vec{u}_t, \vec{\Phi}_{t+1}} \sum_{k=0}^{H^m-1} C_{\text{hvac}}(u_{t+k}, \pi_{t+k}) - R(\Phi_{t+k+1}, \mathcal{B}_{t+k+1}) \tag{22a}$$

s. t.:

$$\underline{x}_{t+k+1} = f(x_{t+k}, u_{t+k} + \underline{\varphi}_{t+k}, d_{t+k}) \tag{22b}$$
$$\forall k = 0, ..., H^m - 1$$

$$\overline{x}_{t+k+1} = f(x_{t+k}, u_{t+k} + \overline{\varphi}_{t+k}, d_{t+k}) \tag{22c}$$
$$\forall k = 0, ..., H^m - 1$$

$$\overline{\varphi}_{t+k} \ge 0, \quad \forall\, k = 1, ... H^m - 1 \tag{22d}$$

$$\underline{\varphi}_{t+k} \le 0, \quad \forall\, k = 1, ... H^m - 1 \tag{22e}$$

$$\underline{x}_{t+k} \in \mathcal{X}_{t+k} \quad \forall \quad k = 1, ..., H^m \tag{22f}$$

$$\overline{x}_{t+k} \in \mathcal{X}_{t+k} \quad \forall \quad k = 1, ..., H^m \tag{22g}$$

$$u_{t+k} + \underline{\varphi}_{t+k} \in \mathcal{U}_{t+k} \quad \forall\, k = 0, ..., H^m - 1 \tag{22h}$$

$$u_{t+k} + \overline{\varphi}_{t+k} \in \mathcal{U}_{t+k} \quad \forall\, k = 0, ..., H^m - 1 \tag{22i}$$

The *argmin* of the optimization problem (22) is the nominal power consumption, u^*_{t+k}, and the maximum available flexibility, $\Phi^*_{t+k}, \forall\, k = 0, ..., H^m - 1$. The building declares u^*_{t+k}, and Φ^*_{t+1+k} for the time slots: $\forall\, k = 0, ..., H^c - 1$ to the utility. After H^c time slots, the BEMS collects the updated parameters such as new measurements and disturbance predictions, and sets up the new MPC algorithm for the time step $k = H^c, H^c + 1, ..., H^c + H^m - 1$, and solves the new MPC for this time frame and uses only the first H^c values of baseline power consumption and flexibility, i.e. for $k = H^c, ..., 2H^c - 1$, and this process repeats.

Fig. 2 shows the schematic of the entire system architecture. The solid line power flow arrows correspond to the baseline system and contract. The ancillary power flow via the flexibility contract is shown with dashed line arrow. Real-time state of the system such as occupancy, internal heat, outside weather condition, building temperature, state and input constraints are passed as input to the algorithm. The utility also communicates information such as per-unit energy and upward and downward flexibility prices to the BEMS (or effectively to the algorithm). The output of the algorithm includes baseline power consumption, downward flexibility and upward flexibility values, and cost. Within the flexibility contract framework, this information is communicated to the utility. Consequently, a flexibility signal s_t is sent from the utility to the building to be tracked by the HVAC fan. Essentially, the utility has control of the building consumption for the next H_c time slots. The control strategy is actuated by sending flexibility signals (similar to frequency regulation signals). These signals can be sent as frequently as every few seconds. In fact, we show through experiments on a real building in Section V, that buildings with HVAC systems equipped with variable frequency drive (VFD) fans are capable of tracking such signals very fast (e.g. within a few seconds).

IV. COMPUTATIONAL RESULTS

The algorithm presented in Section III-C was implemented on a building model developed and validated against historical data in [4], [6]. We used Yalmip [8], an interface to optimization solvers available as a MATLAB toolbox, for rapid prototyping of optimal control problems. The nonlinear optimization problem solver Ipopt [9], was used to solve the resulting nonlinear optimization problem arising from the MPC approach. We use sampling time of $\tau = 1\ hr$. We assess the performance of the algorithm for the following scenarios.

Scenario I: Different reward rates have been considered for upward and downward flexibility at each time step. In particular, downward flexibility is rewarded more than upward flexibility ($\underline{\beta} > \overline{\beta}$), for most of the time. Since MPC maintains the comfort level using the least possible amount of energy, buildings that are operated under nominal MPC such as (1) have no down flexibility (for extended period of time, i.e. 1 hr or more). However we show that via proper incentives and by solving (22), it is possible to provide downward flexibility when most needed. The results of this scenario are shown in Fig. 3 for the case when ancillary signals are received from the utility every minute: no system constraint (e.g. temperature being within the comfort zone) will be violated when the fan speed enforcement is performed for arbitrary values of fan speed as long as fan power
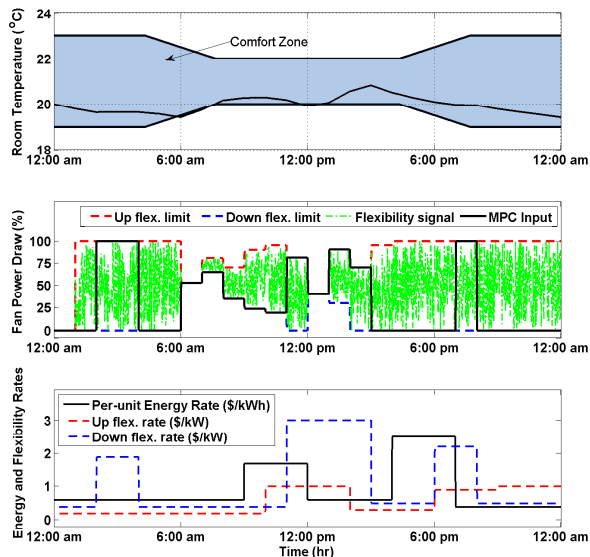
Fig. 3. **Scenario I:** The Per-unit energy rate, and upward and downward flexibility reward are shown in the lowest figure. The middle figure shows the resulting flexibility at each time, and the top figure shows the resulting room temperature. Flexibility signals are sent every minute from the utility to the building.

Fig. 4. **Scenario II:** In this case we consider $\overline{\beta} = \underline{\beta} = 0$. This results into a performance similar to the one of the nominal $\overline{\text{MPC}}$ (1).

consumption, and consequently fan speed is within the safe envelope calculated by the algorithm. It is shown in Fig. 3 that maximum flexibility (100%) is provided at times when the room temperature is far from the boundaries of the comfort zone. The flexibility decreases as the temperature of the room approaches the comfort zone boundary, and is minimum (about 0-15% in this case) when room temperature is fairly close to the boundaries of the comfort zone, and the reward for flexibility is not high enough.

Scenario II: In this scenario we test the algorithm when no *reward* for providing flexibility is provided (i.e. $\overline{\beta} = \underline{\beta} = 0$). Results are shown in Fig. 4. In this case the MPC "selfishly" uses the thermal capacitance of the building to minimize energy cost. This case yields to an MPC policy similar to nominal MPC (1). As expected, the resulting flexibility is very small throughout the day.

Scenario III: In this case, the amount of reward to the building for upward and downward flexibility is the same as Scenario I. Furthermore, the BM is encouraged to provide *equal* upward and downward flexibility as long as possible. This requirement has been enforced by adding an extra penalty term to the cost function as follows:

$$\text{cost} := \text{cost} + \gamma_t^T . |\overline{\varphi}_t - \underline{\varphi}_t| \qquad (23)$$

where $\gamma \gg 0$ is a constant, and $|.|$ returns the absolute value of its argument. Larger value of γ makes the equality of upward and downward flexibility to hold for a longer period. However, in this case, we lose the guaranteed maximization of the economic benefit of the building due to the newly added term in the cost function. Fig. 5 shows the result for this scenario. Maximum flexibility (100%) happens during
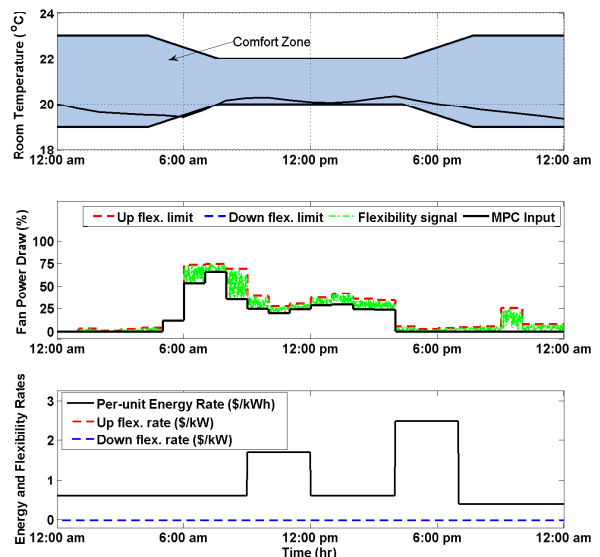
unoccupied hours and when the reward for flexibility is high. In this case minimum flexibility is 0% and it happens during 6 to 7 am, 9 to 10 am, and 12 to 1 pm. The room temperature is close to the comfort-zone boundary. The per-unit price of electricity is high, and the reward for upward and downward flexibility is very small. Hence, the algorithm provides no flexibility at these time periods. In Fig. 5 the frequency of flexibility signals is 1 hr. By picking a large value for the period of flexibility signal arrivals, we show that no system constraint will be violated when the fan speed is constrained even for long time periods, as long as the fan power consumption and consequently fan speed is within the safe envelope calculated by the algorithm.

V. EXPERIMENTAL RESULTS

A. Tracking Ancillary Signals by VAV-equipped Fans

We ran some experiments to demonstrate that it is indeed feasible to track the flexibility signals received from utility (s_k : where $e_k^l \leq s_k \leq e_k^u$) very fast. We run our experiments in a new building constructed four years ago (2009) on the UC Berkeley campus named Sutardja Dai Hall (SDH). It is a 141,000 square-foot modern building that houses several laboratories, including a nanofabrication lab, dozens of classrooms and collaborative work spaces. It contains a Siemens building management system called Apogee [10] which is connected to an sMAP server [11] co-located with the Apogee server. We run our experiments using a control platform built on sMAP, whereby control points are set programmatically upon approval from the building manager. The energy data of this building is stored in the cloud and is accessible to the public at [12].

The BEMS of the experiment set up is shown in Fig. 6. There are two sets of 9 fans namely AH2A and AH2B with a
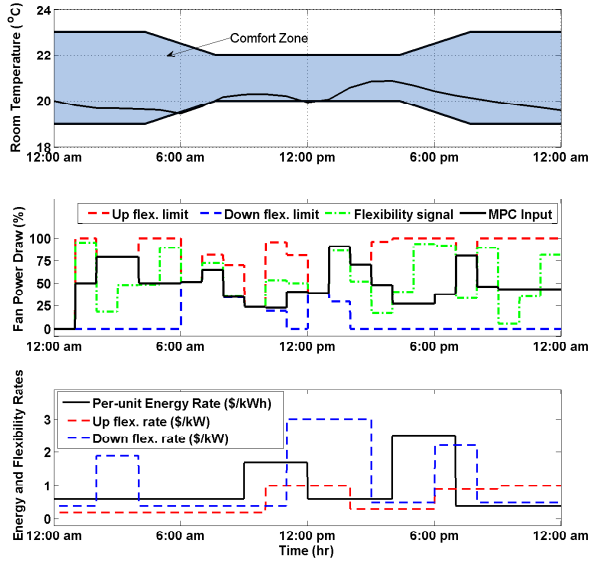
Fig. 5. **Scenario III:** In this case we use the cost function presented in (23) to push the algorithm to produce equal up and downward flexibility as long as possible. Frequency of flexibility signals is 1 hr.

Fig. 6. Building management system of the CITRIS hall, our experimental set up located on the campus of UC Berkeley.

total capacity of 180 hp (about 134 kW). The existing control mechanism is designed to maintain the pressure in the HVAC ducts at the set point value by setting the Supply Duct Static Pressure (SDSP) setpoint. If the pressure drops below the setpoint (e.g. due to the opening of more dampers at the room level, or to the increase of the SDSP setpoint) the two sets of fans will spin faster to compensate for the change in the system and increase the pressure in the duct work to reach the setpoint. And in the case that the pressure goes above the setpoint (e.g. due to the closing of more dampers or to the decrease of the SDSP setpoint), the fan speed will decrease to bring down the pressure to the setpoint. The air handling unit (AHU) maintains an almost constant supply temperature (T^s) in the supply air duct to be sent to the rooms. If one room needs more heating, the air is *reheated* locally at the Variable Air Volume (VAV) box designated for each room, to increase the supply air temperature for that specific room. The flow of air to each room ($\dot{m}_s$) is controlled by opening and closing of the dampers at the VAV level.

Several experiments were conducted on this test-bed to prove our hypothesis that it is indeed feasible to track the ancillary signal s_k received from the utility without affecting the comfort of the occupants. Here we report the two experiments performed on May 17 of 2013. Experiment 1 was conducted from 12:45 pm to 13:15 pm. and Experiment 2 was performed from 5:30 pm to 6:15 pm of the same day. Experiments were done by frequently changing the SDSP setpoint. SDSP is normally kept at 1.75 inch WC (inch Water Column). Due to safety reasons it is advised to always keep the duct pressure within 1.2 and 1.9 inch WC. Hence, we perform the experiments by changing the SDSP between these two values. In Experiment 1 the set
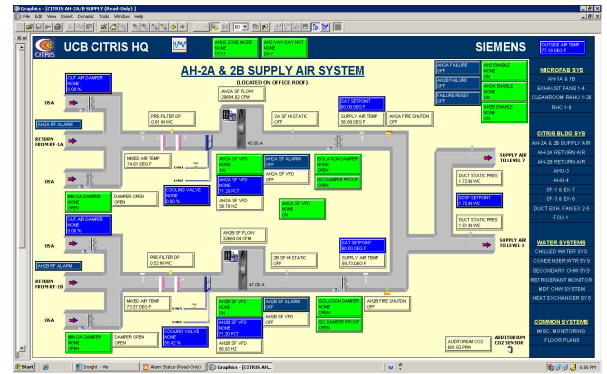
point was changed every minute to values randomly selected in the range $[1.2, 1.9]$ and held at that value for one minute. In Experiment 2, the same procedure was done but every three minutes, and the SDSP setpoint was kept at the new value for three minutes.

The results are shown in Fig. 7 and Fig. 8. In Fig. 7 we present the SDSP setpoint signal, as input of the experiment and the power consumption of the fan, as output of the experiment. It can be observed that the fan power consumption can vary up to 25% within a few seconds around its nominal power consumption at the time of the experiment (which itself is prescribed by the control algorithm running the HVAC system). Fig. 8 on the other hand shows the SDSP setpoint, the outside air temperature and temperature of 15 randomly selected rooms in the building. This figure is provided to show that the performed experiment created no significant and human-sensible change in the temperature of the building, and the building temperature was kept within the comfort zone at all times. Data of one day before the experiment (May 16) and the day of experiment (May 17) is provided for comparison.

Note that since this is a very large building with many occupants, we were not allowed to change the control algorithm of the building to test our algorithm for a long period of time. Nevertheless, the purpose of this experiment was to show that the high frequency component of the flexibility signals from utility, can indeed be tracked using the already existing BEMS software in commercial buildings.

Further experiments were performed on June 6 and 7, 2013 [13]. It was shown that a maximum of 20% flexibility in fan power consumption can be obtained within a few seconds of changing the SDSP setpoint. It is shown in [13] that such modulations of fan speed does not lead to any sensible change in the room temperature and also the percentage of the time in which the reheat valve is open does not change compared to similar times of day in previous days.

B. Total Flexibility of Commercial Buildings in the US

Using the algorithm developed in this paper it was shown that a certain percentage of fan power consumption flexibility can be obtained from commercial buildings and used as
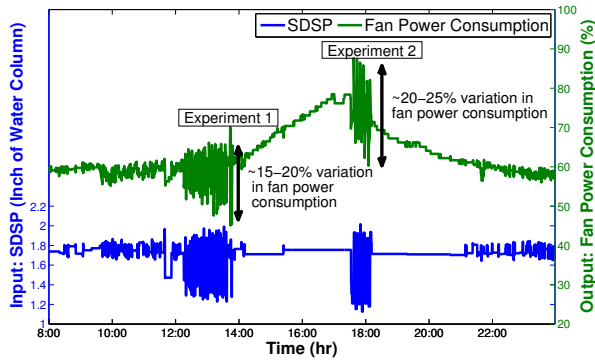
Fig. 7.   Fan power consumption can vary as quickly as in a few seconds by up to 25% by changing the SDSP setpoint.
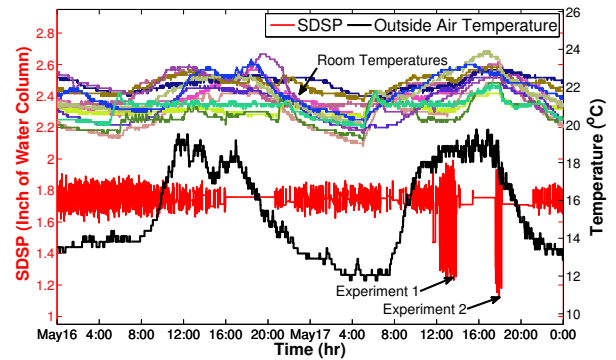

Fig. 8.   SDSP setpoint, outside air temperature, and temperature of 15 randomly selected rooms are shown for one day before (May 16, 2013) and the day of the experiment (May 17, 2013).

fast regulation reserve at each time of the day. Consider a case when 50% flexibility in power consumption is available. 50% flexibility is equivalent to 67 kW of power. According to the latest survey on energy consumption of commercial buildings, performed in 2003 [14], there are 4.9 million commercial buildings in the US which cover a total area of about 72 billion square foot. Almost 30% of such buildings are equipped with variable frequency drive fans. Assuming the same fan power consumption flexibility per square foot to that of the SDH for all commercial buildings, we estimate that at least 11.4 GW of fast ancillary service is readily available in the US at almost no cost, based on the 2003 data. Commercial building floor space is expected to reach 103 billion sq. ft. in 2035 [15]. With the same assumption of the above calculations, about 16.3 GW of regulation reserve will be available in 2035. Note that these numbers correspond to 50% flexibility. In the case of less or more percentage of available flexibility, these numbers will change proportionally.

VI. Conclusion and Future Work

We proposed a method to define and quantify the flexibility of a commercial building HVAC system. We also proposed a contractual framework that could be used by the building operator and the utility to declare the flexibility on the one side and the reward structure on the other side. We then designed a control mechanism for the building to decide how to declare its flexibility for the next contractual period to maximize its reward given the contractual framework. Finally, we performed at-scale experiments to demonstrate the high potential of commercial buildings as a source of flexibility, and the feasibility of the proposed algorithm.

In this paper we assumed that the building operator communicates directly with the utility. However a more realistic scenario would be when an electricity broker, an aggregator, seeks rate offers from suppliers for "bundled" groups of customers and acts on their behalf. In the future work we plan to study this scenario in details to generalize our approach appropriately.

VII. Acknowledgement

References

[1] 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 *American Control Conference (ACC), 2010.*   IEEE, 2010, pp. 5100–5105.

[2] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," in *American Control Conference (ACC), 2010.*   IEEE, 2010, pp. 5106–5111.

[3] M. Maasoumy, "Modeling and optimal control algorithm design for hvac systems in energy efficient buildings," Master's thesis, University of California, Berkeley, 2011. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-12.html

[4] M. Maasoumy, A. Pinto, and A. Sangiovanni-Vincentelli, "Model-based hierarchical optimal control design for HVAC systems," in *Dynamic System Control Conference (DSCC)*, 2011.

[5] H. Hao, A. Kowli, Y. Lin, P. Barooah, and S. Meyn, "Ancillary service for the grid via control of commercial building hvac systems," in *American Control Conference (ACC)*, 2013.

[6] M. Maasoumy and A. Sangiovanni-Vincentelli, "Total and peak energy consumption minimization of building hvac systems using model predictive control," *IEEE Design and Test of Computers*, Jul-Aug 2012.

[7] D. P. Bertsekas, "Nonlinear programming," 1999.

[8] J. Lofberg, " YALMIP : A Toolbox for Modeling and Optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[9] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, vol. 106, no. 1, pp. 25–57, 2006.

[10] Apogee Building Automation. [Online]. Available: https://www.hqs.sbt.siemens.com/gip/general/dlc/data/assets/hq/APOGEE–Building-Automation-A6V10301530-hq-en.pdf

[11] S. Dawson-Haggerty, X. Jiang, G. Tolle, J. Ortiz, and D. Culler, "smap: a simple measurement and actuation profile for physical information," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys'10. New York, USA: ACM, 2010.

[12] (2013, Sep) Openbms, a Berkeley campus energy portal. [Online]. Available: http://berkeley.openbms.org/

[13] M. Maasoumy, J. Ortiz, D. Culler, and A. Sangiovanni-Vincentelli, "Flexibility of commercial building HVAC fan as ancillary service for smart grid," in *IEEE Green Energy and Systems Conference (IGESC 2013)*, Long Beach, USA, Nov. 2013.

[14] Commercial Buildings Energy Consumption survey (CBECS) . [Online]. Available: http://www.eia.gov/consumption/commercial/data/2003/index.cfm?view=consumption

[15] EIA (2012) Annual energy Outlook 2012. . [Online]. Available: www.eia.gov/forecasts/aeo/pdf/0383(2012).pdf?