

# Experimental Investigation of a Capacity-Based Demand Response Mechanism for District-Scale Applications

Jacques A. de Chalendar  
Stanford University  
[jdechalendar@stanford.edu](mailto:jdechalendar@stanford.edu)

Peter W. Glynn  
Stanford University  
[glynn@stanford.edu](mailto:glynn@stanford.edu)

Sally M. Benson  
Stanford University  
[smbenson@stanford.edu](mailto:smbenson@stanford.edu)

## Abstract

*District heating and cooling systems incorporating heat recovery and large-scale thermal storage dramatically reduce energy waste and greenhouse gas emissions. Electrifying district energy systems also has the effect of introducing city-scale controllable loads at the level of the electrical substation. Here we explore the opportunity for these systems to provide energy services to the grid through capacity-based demand response mechanisms. We present both a planning approach to estimate available demand-side capacity and a control framework to guide real-time scheduling when the program is active. These tools are used to assess the technical feasibility and the economic viability of participating in capacity-based demand response in the context of a real-world, megawatt-scale pilot during the summer of 2018 on the Stanford University campus.*

## 1. Introduction

### 1.1 Motivation

As century-old power grids evolve to accommodate increasing penetrations of renewables, an increasing body of literature suggests the need for a greater role for responsive loads in the hierarchical control mechanism for the grid [1], [2]. Determining which energy services can be provided by which demand-side assets and at what price will be key to unlocking their potential. Concurrently, recent advances in the design and “smart control” of district energy networks [3], [4] raise the question of the role that they and growing urban environments can play in the power grid, in the context of calls for deeper integrations of our energy systems across energy pathways and scales [5].

Here we present a case study of the potential for district-scale electric heating and cooling, combined with large-scale thermal energy storage, to provide grid services. As the main supplier of heating and cooling to over 150 buildings on campus, the California-based Stanford Energy System Innovations (SESI) project provides an ideal case study and testbed. In a major

overhaul completed in 2015, the campus district energy system switched from a gas-fired co-generation-based system with steam distribution to the current electrified, integrated heating and cooling system with low-temperature and hot water distribution, adopting several of the latest innovations in district energy [3]. In the new design, shown in Figure 1, the bulk of thermal loads are met with waste-heat-driven electric heat pumps, complemented by conventional electric chillers in the summer and gas-fired boilers in the winter. The new design reduces costs, energy losses and carbon emissions and tightens the integration of the district network with the power system, as campus energy needs are now almost exclusively met by the power grid. Substantial additional grid, carbon and economic benefits can be obtained by optimizing the operations of this electrified district energy system [6], [7].

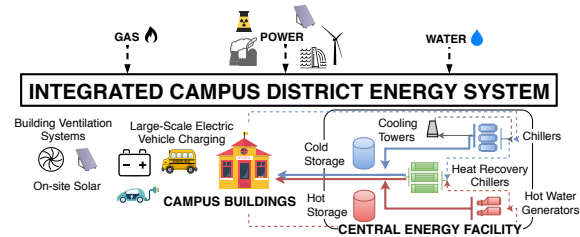


Figure 1. Energy system schematic.

In this paper, we explore opportunities for even greater benefits from the coupling between the thermal and electric systems and the buffer offered by thermal storage (here, insulated steel tanks), by designing optimal operation strategies for district energy systems enrolled in a capacity-based demand response (DR) program like Pacific Gas & Electric’s Capacity Bidding Program (PG&E CBP). Such programs are designed to shift a portion of the risk associated with prediction errors on variable generation and demand from the utility to the consumer. In compensation for providing this service, the facility operator is rewarded with both a demand-response capacity payment and an energy payment.

In the CBP, the enrolled customer bids a monthly load-shedding capacity. During the operating month, the

utility can then call a certain number of event days, during which the customer is expected to shed load for a pre-specified number of hours. Importantly, once the customer has made their capacity nomination, which is a power commitment, they are locked in for the month and can be called upon for arbitrary amounts of energy (within certain bounds). Accordingly, the bulk of the reward for customers comes from a capacity component, although there is an energy component as well. We seek to gain insights both on the questions of technical feasibility, *i.e.* what are the scale and characteristics of the service that can be provided by the campus, and economic viability, *i.e.* whether energy costs can be lowered by participating, given present-day pricing. The operation strategies we build are optimal from the point of view of customers enrolled in CBP, and therefore also allow us to assess their rational behavior and whether this achieves system benefits for the power grid.

## 1.2 Context and key contributions

District energy systems are changing from traditional waste-heat, steam-based networks to integrated heating-and-cooling multi-energy systems [3], [4], offering new opportunities for control and integrations with the power grid [7]–[10]. An example of recent interest in making district heating networks more responsive is in the North of China, where wind penetration is constrained by inflexible Combined Heat and Power plants that simultaneously produce heat and electricity in the winter [11]. Demand-side management strategies have been gaining traction since the 1980s. There has been a particular focus on Thermostatically Controlled Loads (TCLs, such as refrigerators and air conditioners), with the development of methods to coordinate large collections of devices [12], [13], including randomization techniques to avoid synchronization (*e.g.* [2]). Other small loads such as pool pumps or electric vehicles have been shown to have potential [14], [15], but also larger ones such as data centers [16] and industrial facilities [17]. In the context of urban environments, there have been evaluations of the potential for DR at the building level [18], including experimental demonstrations of second-scale response times [19]. Most of these examples focus on the provision of energy services at short time scales of days to seconds however, in contrast to the PG&E CBP, where power commitments are made on a monthly basis. Capacity-based mechanisms have received some, but much less modeling attention [20], [21], even though they represent the bulk of revenue for current DR programs (*e.g.* see Figure 1 of [22] for data on the PJM Interconnection). The PG&E CBP was studied at the Google MtV campus [23], but the ability to control

energy assets was limited, included only a fraction of the buildings, and the response was performed at a building rather than district level. Accordingly, efforts focused on the variability in the response of the loads to DR signals, and on a statistical characterization of drops in demand to inform risk-aware capacity nominations. In contrast, the work presented here introduces a general framework to study participation strategies in capacity-based DR programs that can be applied to controllable loads, as long as their operations scheduling can be modeled as in equation (2). Key contributions presented in this paper include:

- (a) A simple deterministic model for the rational behavior of customers enrolled in the CBP,
- (b) A two-stage stochastic optimization approach to planning capacity nominations,
- (c) An approach based on Receding Horizon Control (RHC) to scheduling energy operations when the program is active,
- (d) Results and insights from the application of this framework to a real-world pilot at the district level and on a megawatt-scale.

## 1.3. Preliminary: operations scheduling for district energy systems

Previous work by the same authors [7] introduced a framework for the operations scheduling of district energy systems and applied it to the SESI project. Such systems provide different energy streams to their consumers, *e.g.* cooling, heating, and electric power. Heating and cooling is generated using electricity and gas at a central on-site location, that we will generically refer to as the central energy facility (CEF), before being distributed through a network of pipes. The aggregate electric and gas load can be divided in (i) a fraction that is managed by a central controller, *e.g.* from the CEF or from electric vehicle charging, and (ii) a fraction that is not dispatchable and depends on the end-users.

We call the operations scheduling problem that of choosing the schedule for the controllable energy draws such as to minimize operating costs for the aggregate system. Typically, an hourly volumetric price is paid for electricity and gas, and a monthly flow rate price, also called a demand charge, is paid for the maximum monthly electricity consumption (peak demand). In the case of the SESI project, the controllable energy draws are determined by the production schedule of each of the machines at the central energy facility. With  $e_t, g_t$  the hourly aggregate electricity and gas consumption, and  $\pi_{e,t}, \pi_{g,t}$  the corresponding hourly electricity and gas prices,  $z_j$  monthly auxiliary variables to represent monthly electricity peak demand and  $\pi_{z,j}$  the monthly

demand charge, it is shown in [7] that the operations scheduling problem can be written as:

$$\operatorname{argmin} \sum_{t \in \mathcal{T}} \pi_{e,t} e_t + \pi_{g,t} g_t + \sum_{j \in \mathcal{M}} \pi_{z,j} z_j, \quad (1)$$

subject to various technology constraints for the different machines, and storage dynamics for the heat storages. Here  $\mathcal{M}$  represents the set of months and  $\mathcal{T}$  the set of hours in the year. Equation (1) represents the campus annual energy bill. The program we solve in [7] also includes quadratic penalties for auxiliary variables that represent unmet loads, these were omitted here for simplicity. The operations scheduling problem can be written more concisely by recognizing that it is a Quadratic Program (QP), of the form:

$$\begin{aligned} \operatorname{argmin} f(x) &= c^T x + x^T Q x, \\ \text{subject to: } Ax &= b, \\ x &\geq 0. \end{aligned} \quad (2)$$

In this more general formulation, the decision vector  $x$  encodes hourly operations schedules and the monthly peak demands, the vector  $c$  and matrix  $Q$  are used to encode the operating costs, matrix  $A$  is called the technology matrix and vector  $b$  includes the energy loads that must be met (as well as terms to represent operational constraints for the machines). As a reference, the problem of scheduling hourly energy operations for a year for the SESI project can be represented as a QP with 150k decision variables and 250k constraints and is solvable in seconds on a laptop. In the remainder of this paper, we will assume that the operations of the energy system we are modeling are captured by equation (2). Our goal will be to devise planning and control strategies to participate in CBP, given that energy operations can be modeled by equation (2).

## 2. Modeling capacity-based demand response

We now describe the PG&E Capacity Bidding Program (CBP) [24] and build a simple deterministic model for customer participation. As of 2018, PG&E is entering commitments from the CBP as Proxy Demand Resource (PDR) assets in the California Independent System Operator (CAISO) markets. The PDR mechanism provides a way for demand response to participate in energy markets that were traditionally reserved for supply-side assets. We model the *Prescribed* option of the CBP referenced in [24] where PG&E chooses the price at which CBP commitments are entered the market (likely to be between \$70-85 MWh<sup>-1</sup>). Two new options were introduced in 2018 (*Elect* and *Elect +*) to give customers more flexibility but are not considered here.

The program season is May through September. Twenty-five days before the start of each month,

participants in the program, also called aggregators, must submit a capacity bid (in MW) to PG&E. At any time during the month, aggregators can then be called upon by PG&E to deliver their capacity in the form of load shedding, within certain limits each month: (i) program hours are eleven a.m. to seven p.m., (ii) events last one to four hours, and (iii) there are caps at thirty hours, five days or three consecutive days, whichever is most constraining. Events can be expected to be triggered (i) when either the CAISO Locational Marginal Price (LMP), or the load that PG&E is expected to serve the following day, or forecasted temperatures, reach certain thresholds, (ii) when the CAISO dispatches PG&E for a PDR that is part of the CBP, and (iii) at PG&E's discretion. Notifications are made at four p.m. on the day before an event. Payments under the CBP are divided into two portions: a capacity payment and an energy payment. The energy payment is passed on directly from CAISO's PDR mechanism and corresponds to the difference between a day-ahead and a real-time price. Monthly, energy payments are expected to represent less than 10% of capacity payments and will depend on LMPs that are difficult to predict one month ahead. For the remainder of the paper, they will not be considered. The capacity payment can be calculated from the capacity price  $\pi_c^{CBP}$  (\$·kW<sup>-1</sup>), the capacity bid  $y$  (kW, both known before the fact), and the aggregator performance (% adjustment, measured after the fact). We call nominal capacity payment  $p_c^{CBP} = y\pi_c^{CBP}$ , the maximum reward that the aggregator achieves by delivering exactly as was planned by their bid, also called a nomination, at the beginning of the month. Illustrative numerical values are given in Table 1. If the campus were to participate at the 10 MW level, this would represent monthly savings of 1-12% for the six months that the program is active (3% annually).

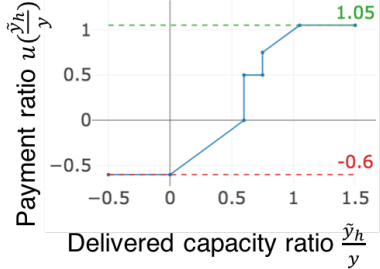
As in many DR programs, measuring hourly performance is key. We call  $\mathcal{H}$  the set of event hours. PG&E defines the delivered capacity  $\tilde{y}_h$  at event hour  $h$  as the observed drop in electric load  $e_h$  from the baseline  $e_h^b$ . In the CBP, the baseline for a given hour corresponds to the average of the ten previous days at the same hour of the day (excluding weekends, holidays, and event days). The baseline can also be adjusted on the day of an event (by +/-40%), but we omit that to be conservative in our estimates here. The monthly capacity payment  $g(y, \tilde{y})$  is computed from the hourly payment ratios  $u(\tilde{y}_h/y)$ , which are a function of the hourly ratios of delivered capacity to nominated capacity  $\tilde{y}_h/y$ :

$$g(y, \tilde{y}) = \sum_{h \in \mathcal{H}} \pi_c^{CBP} \frac{y}{|\mathcal{H}|} u\left(\frac{\tilde{y}_h}{y}\right), \quad (3)$$

$$= p_c^{CBP} \frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} u\left(\frac{\tilde{y}_h}{y}\right). \quad (4)$$

**Table 1. Value of nominal capacity payments from bidding in the PG&E CBP [24].**

	Jul	Aug	Sep	May-Sep
1 kW	\$16.3	\$22.6	\$13.9	\$62.1
5 MW	\$82k	\$113k	\$69.5k	\$311k
10 MW	\$163k	\$226k	\$139k	\$621k



**Figure 2. Risk-averse capacity rewards as a function of hourly delivered capacity.**

Shown in Figure 2, the function  $u(\tilde{y}_h/y)$  is piecewise linear, takes values between -60% and +105%, and is negative if delivered capacity drops below 60% of the capacity nomination. It is designed to make the enrolled customers risk-averse (but is not concave), so that they reliably deliver the capacity they promised during each event. Although performance is measured on an hourly basis, aggregator gains do not increase proportionally with the number of events (and therefore energy delivered). The maximum aggregator gain only depends on the capacity bid that was made at the beginning of the month. Consequently, the CBP rewards a service measured in units of power (or capacity) rather than energy.

As a back-of-the-envelope evaluation of the risk from participating in the CBP, we calculate a condition for a null payment for capacity, assuming the aggregator responds either fully to an event or not at all. Calling  $k$  the number of the event hours for which performance is perfect, we can write the null capacity payment in this case as:

$$0 = \frac{k}{|\mathcal{H}|} p_c^{CBP} - 0.6 \frac{|\mathcal{H}| - k}{|\mathcal{H}|} p_c^{CBP} \quad (5)$$

According to equation (5), aggregators can fail for up to 63.5% of the event hours without losing money by participating in the program. The payment ratios in equation (4) introduce non-linearities in the objective function. Standard methods exist to deal with piecewise-linear functions, fractions and products of decision variables [25] but for simplicity, here we instead choose to penalize deviations, and redefine  $g$  as:

$$\hat{g}(y, \tilde{y}) := y \pi_c^{CBP} - \lambda \sum_{h \in \mathcal{H}} (\tilde{y}_h - y)^2, \quad (6)$$

where  $\lambda$  is a tunable parameter and the choice of the 2-norm over the 1-norm is motivated by our risk-averse setting: using a 2-norm will produce solutions where

penalties are “spread” over multiple hours, whereas a 1-norm penalty will direct the program towards sparse solution vectors.

To summarize, the deterministic participation of a customer whose energy operations are well captured by equation (2) and is enrolled in the CBP can be modeled by the following program:

$$\begin{aligned} \min & f(x) - \hat{g}(y, \tilde{y}), \\ \text{subject to: } & Ax = b, \\ & v(x, \tilde{y}) = 0, \\ & x, y \geq 0. \end{aligned} \quad (7)$$

Here the function  $v$  is used to write the defining constraint for the auxiliary variables  $\tilde{y}_h$  that represent the delivered capacities at each hour  $h$ :

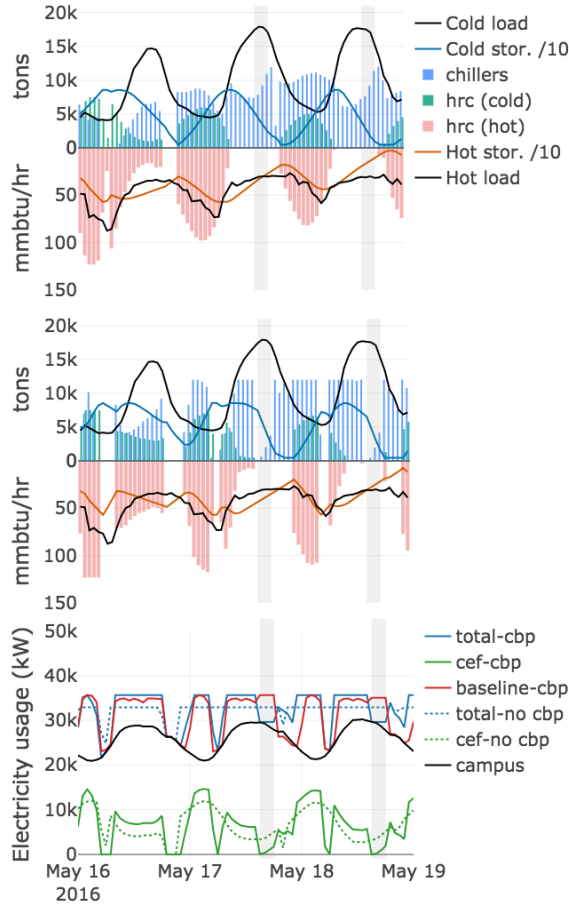
$$v(x, \tilde{y})_h = \tilde{y}_h - (e_{d(h)}^b - e_{d(h)}). \quad (8)$$

This can be more concisely written in matrix form:

$$v(x, \tilde{y}) = \tilde{y} - Dx. \quad (9)$$

In equations (8) and (9),  $d$  and  $D$  both represent the mapping from a given hour  $h$  in the set of event hours  $\mathcal{H}$  to the corresponding hour  $t$  in the set of hours in the month  $\mathcal{T}$ . Using our simplified form in equation (6) for the capacity payment results in a (convex) quadratic objective in problem (7), so that this now defines a QP. Consequently, our simplified deterministic model for CBP is not much more difficult than the one in equation (2). The delivered capacity  $\tilde{y}_h$  is defined as the difference between a baseline and actual load, so inflating the baseline by raising consumption on non-event days can be just as economically efficient as reducing load to meet a requested drop. This is not the intended behavior however, and plant operators will be reluctant to allow their demand charges to increase due to higher peak loads, so we include additional constraints on the monthly peak variables  $z_j, j \in \mathcal{M}$ :  $z_j \leq \bar{z}_j$ , where the  $\bar{z}_j$  are manually set (for instance from a previously computed solution to problem (2)). These constraints concern operations scheduling, and they will be encoded in matrix  $A$  and vector  $b$  in the remainder of this paper.

Figure 3 shows the thermal dispatch schedules and corresponding power draws that result from solving the deterministic versions of the operations scheduling problem (2) and the CBP operations scheduling problem (7) for a few days in May of 2016. In the thermal dispatch schedules, the top portion of the plot shows the cooling loads, the cooling output of the chillers and heat recovery chillers (HRCs), and the state of charge of the chilled storage. The bottom portion shows the heating output of the HRCs and gas heaters, as well as the state of charge of the hot storage. In the spring and summer, the extent to which the HRCs can be used to produce chilled water is constrained by the daily hot water consumption.



**Figure 3. Top/middle: thermal schedule without/with CBP. Bottom: Corresponding power draws without/with CBP.**

The thermal storage tanks are used as buffers and provide the main source of flexibility of the system, as described in [7]. The third plot in Figure 3 shows the corresponding power draws. The electricity flowing to the chillers and heat recovery chillers constitutes the bulk of the controllable Central Energy Facility (CEF; green) electric load. This is added to the non-dispatchable campus loads (black) to constitute the total electric load that Stanford is billed for (in blue). In this example, events are called on May 17<sup>th</sup> and 18<sup>th</sup> (shown in light gray), and the CEF provides 5 MW of capacity to PG&E. The CEF deviates from the optimal schedule for problem (2) to participate in the DR program, which causes an increase in peak load.

### 3. Planning approach

#### 3.1. Problem formulation

The “planning problem” consists of choosing the capacity nominations twenty-five days before the start

of each month. Our approach to the planning problem relies on solving a two-stage stochastic optimization program over the entire program season, where capacity nominations are decided at the first stage, and hourly operations are scheduled at the second one. We write our problem as:

$$\operatorname{argmin} \left\{ q(y) := \pi_C^{CBP T} y + \mathbb{E}Q(y, \xi) \right\}, \quad (10)$$

where  $y$  and  $\pi_C^{CBP}$  are now vectors with the monthly capacity nominations and prices, respectively, and  $Q(y, \xi)$  is the solution to the following second-stage problem:

$$\begin{aligned} Q(y, \xi) = \min & f(x) + \hat{g}(y, \tilde{y}) \\ \text{subject to: } & Ax = b, \\ & \tilde{y} = Dx, \\ & x \geq 0. \end{aligned} \quad (11)$$

Here  $\xi = (D, b, c)$ , where  $c$  represents the cost vector in problem (2), encodes the second-stage uncertainty: we assume that there is uncertainty in the energy loads and prices, as well as in when events are called. One realization of  $\xi$  corresponds to a given trajectory for loads, prices, and event times for the entire year. The program defined by equation (11) is slightly more restrictive than what is imposed by the CBP because it assumes that all nominations must be chosen at the beginning of the year and does not account for the fact that they can be changed every month. It does account for the coupling of the months through the baseline calculation however. Since the baseline can extend up to twenty days in the past, this coupling is quite strong and treating the months independently would be much less realistic.

If the support of the uncertainty is discrete, the planning problem can be represented by its certainty-equivalent counterpart, a deterministic optimization program that has the form:

$$\begin{aligned} \min & \sum_k p_k (f(x_k) + \hat{g}(y, \tilde{y}_k)), \\ \text{subject to: } & Ax_k = b_k, \quad k = 1 \dots K, \\ & \tilde{y}_k = D_k x_k, \quad k = 1 \dots K, \\ & x_k, y \geq 0, \quad k = 1 \dots K. \end{aligned} \quad (12)$$

In equation (12), there are  $K$  scenarios that each have weight  $p_k$ .

#### 3.2. Planning problem results

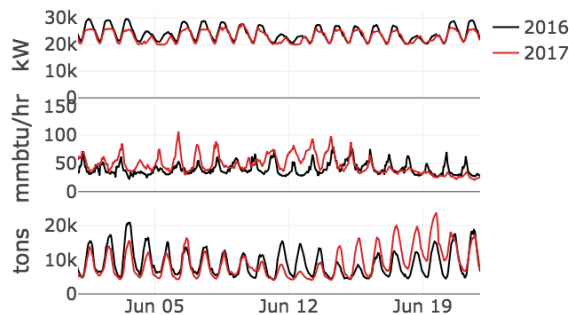
The planning problem formulation that was just described was used with real load data from 2016 and 2017 to inform the choice of capacity nominations for the CBP 2018 season at Stanford University. An extract of the input data is shown in Figure 4. It is likely that events will be called on days where temperatures are high, and the campus is itself under stress. To obtain a conservative estimate for available capacity, we assume that CBP events are called on the five days where

aggregate campus cooling load is highest each month. Since CBP events are only called on weekdays, we shift the timestamps of the 2017 data by 364 days instead of a full year and use the 2016 calendar. This ensures that weekday/weekend profiles for the two datasets correspond (as seen on the electricity plot in Figure 4), and that events can only be called on the (harder) days with weekday consumption profiles. There were several differences between the two years as evidenced by the data shown in Table 2, most notably: (i) the summer of 2016 was much milder than that of 2017, (ii) the summer of 2017 experienced two heat waves, where cooling loads rose to unprecedented levels, (iii) from the summer of 2016 to that of 2017 the on-site solar generation capacity on the Stanford campus was increased from 400 kW to 4.5 MW, the effect of which can be seen on the top plot for electricity usage in Figure 4.

We first compute the optimal CBP participation strategy by solving the program in equation (12), assuming that the two datasets should be given equal weight. Table 3 shows the expected value changes in different components of the campus energy bill from participating in the CBP, as well as the monthly capacity bids and the increase in peak load for July and August. The monthly energy bill is comprised of three main components: an energy charge (weighted integral of hourly electrical consumption timeseries), a demand charge (maximum of hourly electrical consumption timeseries), and a capacity payment if participating in CBP (a negative cost). July and August are the two most profitable months in the CBP [24].

**Table 2. Difference in daily consumption percentiles from 2016 to 2017.**

	P50	P75	P90	P95
Cooling	+9%	+13%	+23%	+33%
Heating	-6%	-2%	+8%	+12%
Electricity	-4%	-5%	-5%	-5%



**Figure 4. Campus aggregate energy load data for twenty days in June.**

Table 3 shows that the largest impact from participating in the CBP is to increase peak monthly loads (capped at 15% of the pre-CBP peaks here), thus raising demand charges, but that this increase is more than offset by the CBP capacity payments. The increase in peak load also gives the program more freedom to shift consumption to lower price periods, which accounts for the slight reduction in energy costs.

In Figure 5, we vary the weights of the different scenarios in formulation (12). We show the monthly capacity nominations as well as the expected value of the cost reduction from participating in the CBP as a function of the relative weight  $\alpha$  between 2016 and 2017. Low values for  $\alpha$  mean that experiencing a year like 2016 is more probable. We observe that participating in CBP is more challenging for the months of September and May in 2017 than 2016, and that the opposite is true for June. It is interesting and maybe counterintuitive to note that the June 2017 heatwave (June 18-19 in Figure 4) does not present a significant challenge either for demand charge mitigation or participation in the CBP, because the high cooling loads occurred on a weekend, when the campus electric load is lower. By contrast, the relatively lower cooling loads on June 3<sup>rd</sup> 2016, a Friday, are much more difficult to manage. The relative difficulty of meeting cooling loads during those two events is confirmed numerically from the Lagrange multiplier (shadow price) associated with the constraint on chilled storage capacity in both cases, which quantifies the objective function decrease from adding one unit of capacity at that hour. This example also illustrates the coupling between the different energy streams consumed on campus and motivates an integrated approach to managing them.

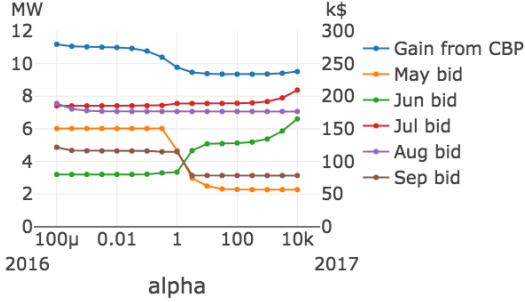
**Table 3. Expected value of bill changes for the here-and-now solution to the planning problem (12) using 2016 and 2017 data.**

	Jul	Aug	Total (May-Oct)
Energy (k\$)	-0.9	10.6	<b>-0.1</b>
Peak (MW)	4.8	4.0	NA
Demand (k\$)	35.1	28.9	<b>129.7</b>
CBP bid (MW)	7.6	7.1	NA
CBP (k\$)	-123	-159	<b>-374</b>
<b>Bill (k\$)</b>	<b>-90</b>	<b>-120</b>	<b>-245</b>

## 4. Control approach

### 4.1. Algorithm

In this section, we describe the general architecture of the control framework that was implemented for the 2018 CBP season at Stanford.



**Figure 5. Monthly capacity nominations and expected bill reduction for the CBP season as a function of the weight  $\alpha$  between 2016 and 2017 data.**

The control framework is used to make real-time control recommendations to the CEF operators. Informally described in Algorithm 1, the framework is based on Receding Horizon Control (RHC) [26] (also known as Model Predictive Control). At each time step  $t$ , we solve a deterministic CBP operations scheduling problem based on problem (7) that uses available historical data and an  $H$ -step ahead forecast. In our formulation, the state includes the energy in the thermal storage tanks, the peak electrical load to date and the information available for forecasts. The control corresponds to the electrical energy drawn by each CEF machine for the next  $H$  time steps.

In problem (13), the uncertain parameters are the campus energy draws for heating, cooling, and electricity, encoded in vector  $\hat{b}_t$ , and the dates of CBP events, encoded in matrix  $\hat{D}_t$ . In particular, the cost for operations  $f_t$  and for participating in the CBP  $g_t$ , as well as the technology matrix  $A$ , are assumed fully known. The uncertainty on the efficiency of the machines will result in flows to the energy storage tanks that are different from expected, but this will be absorbed by the uncertain parameter  $\hat{b}_t$ . In contrast with problem (7), the capacity nomination is not a decision variable but now set at value  $y_0$  (e.g. chosen when solving the planning problem (12)).

Real-time energy load data is input to the controller from the campus data historian and used to update the system state at every hour. The forecast has two components: energy draws and CBP event times. To forecast energy draws during the summer 2018 pilot, we use proprietary Johnson Controls software purchased by the university [27]. In the context of this paper however, the performance of Algorithm 1 is evaluated using the error model described in section 4.2. To deal with the uncertainty in CBP event dates and times, we use a simple strategy: (i) as soon as we are notified of an event by PG&E (4 p.m. on the day before), that event is included in the forecast; (ii) the forecast always includes an event day at the end of the forecast horizon. The goal

of this strategy is to ensure that the CEF will always be in one of two operating modes: if an event was called for the following day, then the CEF is preparing for the event, e.g. by making sure that the energy storage tanks are fully charged before the start of the event. If no event was called for the following day, then the CEF is preparing for an event at the end of its forecast horizon, e.g. by making sure its baseline is high (without raising peak load) so as to make responding to future events easier.

As discussed in section 4.3, the control algorithm is more sensitive to uncertainty in electric loads than in heating and cooling loads. One reason for this is the bill structure, and in particular the demand charge that is paid for the maximum monthly electrical load. Optimal solutions to problems (2) or (7) typically display a consumption pattern where aggregate electrical load is exactly at its monthly peak for the majority of the time steps, so that the corresponding load profile is kept as flat as possible (see e.g. the bottom plot in Figure 3 for an example of this behavior). Consequently, a direct implementation of Algorithm 1 generates control trajectories where the peak electrical load increases steadily over the course of the month, as shown in Figure 7. As the controller progresses in time, it tries to keep aggregate load as close as possible to the peak. If the non-dispatchable electrical load was higher than forecasted, then the value of the historical peak rises, and is subsequently used as the new target by the controller at the next time step. To correct this undesirable behavior, Algorithm 1 is modified so that a constant buffer term is now added to the electrical load forecast. Numerically, a buffer of  $+3\sigma_e$  kW was found to produce satisfactory results, where  $\sigma_e$  is an estimate for the standard deviation of the 1-hour ahead prediction error (assumed to be available from historical data).

---

**Algorithm 1:** RHC for CBP-aware CEF operations

---

Initialize state of charge of the storage tanks, choose capacity nomination  $y_0$  for the month;

At time step  $t = 1 \dots T$ :

1. Make forecasts for parameters  $\hat{b}_t$  and  $\hat{D}_t$ ,
2. Solve problem (12) at time  $t$ :
 
$$\min f_t(x) + g_t(y, \tilde{y}), \quad (13)$$
 subject to:  $Ax = \hat{b}_t$ ,  
 $\tilde{y} = \hat{D}_t x$ ,  
 $y = y_0$ ,  
 $x \geq 0$ ,

to obtain the control inputs to the CEF machines for time step  $t$ ,

3. Update state: observe real loads, update state of charge of the thermal storage tanks and historical peak electricity load for time step  $t$ .
-

## 4.2. Controller performance evaluation method

In order to evaluate the performance of our control algorithm, we use an approach similar to that in [28], where prediction errors are modeled using a martingale forecasting process to represent an unbiased prediction process that improves over time. We will use this error model to generate noisy forecasts from 2016 actuals for each of the energy streams that are consumed by the campus. Specifically, for a given quantity  $q$ , the prediction error at time  $t$  for time  $\tau$ ,  $q_t(\tau) - q(\tau)$  is modeled as a sum of normal random variables:

$$q_t(\tau) = q(\tau) + \sum_{s=t}^{\tau} n_s \sqrt{\frac{\sigma^2}{\tau - s + 1}}. \quad (14)$$

Here the random variables  $n_s(\tau) \sim \mathcal{N}(0,1)$  are independent and identically distributed. When evaluating our control algorithm, we will generate forecasts up to  $H$  steps ahead  $\{q_0(\tau), \tau = 1 \dots H\}$ . The root-mean-square error (RMSE)  $\tau$  steps ahead for this process is:

$$RMSE(\sigma, \tau) = \sigma \sqrt{\mathbb{E} \left( \sum_{s=1}^{\tau} \frac{n_s}{\sqrt{\tau - s + 1}} \right)^2} \quad (15)$$

We plot this quantity for  $\sigma = 1$  as a function of  $\tau$  in Figure 6 (evaluated numerically using 100k samples). The parameter  $\sigma$  will be used to control the level of uncertainty. The 1-hour and 7-day ahead RMSEs are approximately  $\sigma$  and  $2.4 \sigma$ . We will use the notation  $\sigma_h$ ,  $\sigma_c$ , and  $\sigma_e$  for the uncertainty-controlling parameters for heating, cooling and electricity, respectively, and express them as a percentage of a maximum value for each of the loads (25 mmbtu/hr, 1.6k tons and 850 kW).

## 4.3. Controller performance results

Figure 7 shows a comparison of the trajectory of a controller using Algorithm 1 with and without a buffer for the electrical load forecast to the optimal trajectory obtained with perfect information over the entire horizon for problem (2). This plot illustrates how a direct implementation of Algorithm 1 is derailed by uncertainty in the electrical load forecast. Adding a buffer makes the controller more conservative and avoids the steady increase in peak load. In Figure 8, we show the behavior of the controller in the CBP problem (7) under a 5 MW nomination, as well as the plan made by the controller at 3 p.m. on July 12<sup>th</sup>, just before it is notified of the July 13<sup>th</sup> event. Under Algorithm 1's policy, the baseline is slightly lower (500 kW) during events than in the perfect information case, so more load must be dropped to meet the nomination.

In Figure 9, we assess the performance of Algorithm 1 as a function of  $\sigma_c$  (CW) and  $\sigma_e$  (KWH) in the context of the CBP problem (7). Suboptimality manifests itself primarily through unmet cooling loads, so we use this as our performance criterion. The two other main components of the objective function (hourly energy cost and demand charge) do not vary significantly with increased levels of uncertainty (a constraint was imposed on peak load in problem (7)). In Figure 9, the x-axis is scaled with the maximum error we use in each case (1.6k tons and 850 kW). The y-axis is scaled by the mean hourly cooling load (9.3k tons), so that we report hours of lost cooling. We show the first, second and third quartiles for algorithm performance over 200 trials. The distributions are much tighter for prediction errors in electric load than in cooling load. This can be explained by (i) the much stronger variability in cooling loads than in electric load (see *e.g.* Figure 4), which means that although there are only a few days per month where greater than expected cooling loads are a problem, uncertainty on electrical load impacts the system similarly every day; and (ii) the demand charge, under which the system operates close to its peak load for prolonged periods of time. Hot water loads are lower in the summer, so corresponding prediction errors do not affect the quality of the solution and are not shown here.

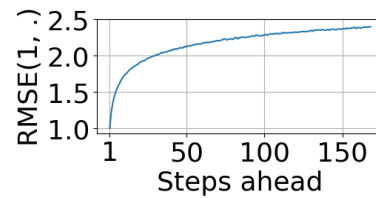


Figure 6. Normalized RMSE as a function of time for the prediction error model (14).

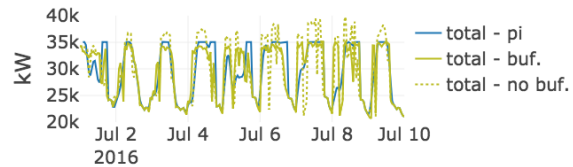


Figure 7. Problem (2): RHC with/without buffer term (buf.) vs. perfect information (pi).

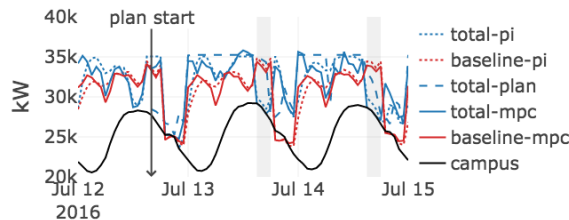
## 5. Discussion

Although we present its application to a specific case study, the framework introduced in this paper is quite general. The PG&E CBP is structured closely to the standard Demand Response Auction Mechanism (DRAM) contracts that were introduced by the California Independent System Operator to integrate

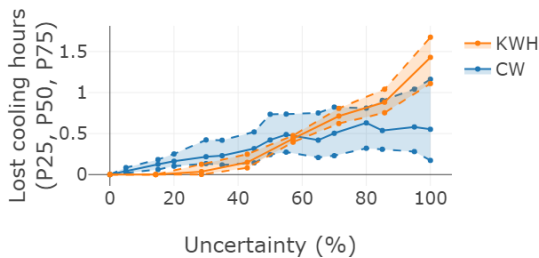


DR in their energy markets. Even though DRAM contracts typically do not include a capacity payment, they usually require the same type of monthly power commitments as the CBP. The problem formulation (2) is used here to describe a district energy system, but it can be generically applied to a range of energy assets, including electric vehicle charging stations, controllable industrial loads or electrochemical energy storage.

The goal of capacity-based DR is to provide value to the power system by identifying a certain number of controllable loads that can be called upon to reduce consumption when the grid is under stress. These stress periods will also tend to be the ones where the controllable loads are under stress themselves (the CBP typically targets the five hottest days of the month, for instance). By modeling the rational behavior of demand-side assets entered in such a program, we are in a position to assess whether such mechanisms can actually provide value, both to the utility and to the customer. Measuring performance during DR events will be key moving forward. Alternatives have been proposed to the current method for calculating baselines [36], under which today’s customers are incentivized to increase baselines such as to make responding to events easier. To a first order, increased baselines will make financial sense when the demand charge is lower than the capacity price (true for half of the CBP season at Stanford).



**Figure 8. Problem (7): RHC vs. perfect information (pi) trajectories and controller plan just before the July 13<sup>th</sup> event is announced.**



**Figure 9. Control algorithm performance as a function of prediction error for problem (7).**

Under the current rate structure, an 8 MW participation in the CBP reduces heating-and-cooling-related electricity costs by 28% in August and July (9%

annually). These costs represent only 25% of the aggregate campus electric bill however. Against the total bill, annual savings are 2%, which does not provide a strong business case for engaging in this type of DR. More precise assessments of DR's value to the grid (e.g. from distribution upgrade deferrals) could make that case stronger.

A natural extension of our current RHC-based control framework would be to solve a stochastic program at each time iteration. This would probably generate more conservative controls and remove the need for a buffer on electrical load. Other approaches than those presented in this paper also exist, in particular within the stochastic and dynamic programming communities [29]. In any case, participating in the CBP is complex and requires computational tools to predict available capacity before the start of the month and control energy operations. This is especially true in a district network where multiple energy streams are coupled. Many demand-side assets do not have such sophisticated controls available, yet they could probably provide significant value to the grid.

The Stanford CBP pilot highlights the inherent flexibility available in district energy systems with large thermal storage. The question of how to value this flexibility and how best to use it remains open. While current DR mechanisms are overwhelmingly based on load shedding, shifting to a demand dispatch paradigm could be a way of incentivizing the ability to increase load as well. The capacity bidding and control framework presented here could also be used for the more general problem of demand response.

## 6. Conclusion

This paper has focused on developing modeling tools to enable the real-life participation of electrified district energy systems in capacity-based demand response programs. The framework was applied and tested in the context of a pilot that is active during the summer of 2018 on the Stanford University campus to participate in PG&E's Capacity Bidding Program. The formulation that was derived can be used for arbitrary energy assets, as long as their operations scheduling problem can be described by equation (2). According to results from our planning approach, up to 8 MW of capacity could be provided by the 15 MW Central Energy Facility while maintaining the increase in monthly peak campus electrical loads below 15%. For the most lucrative months (July and August), net rewards from the program are expected to represent 5-7% of the monthly campus energy budget (or 28% of heating and cooling electricity costs). The RHC-based control approach we develop is most sensitive to uncertainty in the electrical loads, which are also the

easiest ones to forecast. This work provides a foundational framework for assessing the value of capacity-based DR for the power system and should inform future DR mechanism design.

## References

- [1] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419–4426, 2008.
- [2] Y. Chen et al., "Distributed Control Design for Balancing the Grid Using Flexible Loads," in IMA volume on the control of energy markets and grids, Springer., 2017.
- [3] H. Lund et al., "4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems.," *Energy*, vol. 68, pp. 1–11, 2014.
- [4] G. Chicco and P. Mancarella, "Distributed multi-generation: A comprehensive view," *Renew. Sustain. Energy Rev.*, vol. 13, no. 3, pp. 535–551, 2009.
- [5] P. Mancarella, "Unlocking Flexibility," *IEEE Power Energy Mag.*, no. February, pp. 43–52, 2017.
- [6] M. J. Wenzel, R. D. Turney, and K. H. Drees, "Model Predictive Control for Central Plant Optimization with Thermal Energy Storage," in 4th International High Performance Buildings Conference at Purdue, 2014.
- [7] J.A. de Chalendar, P.W. Glynn, and S.M. Benson, "City-scale experiments with integrated energy systems", *Manuscript submitted for publication*, 2018.
- [8] E. A. M. Cesena and P. Mancarella, "Energy Systems Integration in Smart Districts : Robust Optimisation of Multi-Energy Flows in Integrated Electricity , Heat and Gas Networks," *IEEE Trans. Smart Grid*, 2018.
- [9] B. Bach, J. Werling, T. Ommen, M. Münster, J. M. Morales, and B. Elmegaard, "Integration of large-scale heat pumps in the district heating systems of Greater Copenhagen," *Energy*, vol. 107, pp. 321–334, 2016.
- [10] M. G. Nielsen, J. M. Morales, M. Zugno, T. E. Pedersen, and H. Madsen, "Economic valuation of heat pumps and electric boilers in the Danish energy system," *Appl. Energy*, vol. 167, pp. 189–200, 2016.
- [11] X. Chen et al., "Increasing the Flexibility of Combined Heat and Power for Wind Power Integration in China: Modeling and Implications," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1848–1857, 2015.
- [12] R. Malhamé and C. Y. Chong, "Electric Load Model Synthesis by Diffusion Approximation of a High-Order Hybrid-State Stochastic System," *IEEE Trans. Automat. Contr.*, vol. 30, no. 9, pp. 854–860, 1985.
- [13] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Convers. Manag.*, vol. 50, no. 5, pp. 1389–1400, 2009.
- [14] S. P. Meyn, P. Barooah, A. Busic, Y. Chen, and J. Ehren, "Ancillary service to the grid using intelligent deferrable loads," *IEEE Trans. Automat. Contr.*, vol. 60, no. 11, pp. 2847–2862, 2015.
- [15] W. Kempton and J. Tomić, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *J. Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.
- [16] Z. Liu, M. Lin, A. Wierman, S. Low, and L. L. H. Andrew, "Greening geographical load balancing," *IEEE/ACM Trans. Netw.*, vol. 23, no. 2, pp. 657–671, 2015.
- [17] D. Todd, M. Caufield, B. Helms, M. Starke, B. Kirby, and J. Kueck, *Providing Reliability Services through Demand Response: A Preliminary Evaluation of the Demand Response Capabilities of Alcoa Inc.* 2009.
- [18] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, "Quantifying changes in building electricity use, with application to demand response," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 507–518, 2011.
- [19] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental Evaluation of Frequency Regulation From Commercial Building HVAC Systems," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 776–783, 2015.
- [20] P. Cappers, C. Goldman, and D. Kathan, "Demand response in U.S. electricity markets: Empirical evidence," *Energy*, vol. 35, no. 4, pp. 1526–1535, 2010.
- [21] H. A. Aalami, M. P. Moghaddam, and G. R. Yousefi, "Demand response modeling considering Interruptible/Curtailable loads and capacity market programs," *Appl. Energy*, vol. 87, no. 1, pp. 243–250, 2010.
- [22] PJM Interconnection, "Demand Response Strategy," 2017.
- [23] A. Radovanovic, W. D. Heavlin, and S. Kiliccote, "Optimized Risk-Aware Nomination Strategy in Demand Response Markets," *Proc. 3rd ACM Int. Conf. Syst. Energy-Efficient Built Environ. - BuildSys '16*, pp. 99–108, 2016.
- [24] Pacific Gas & Electric, "Electric Schedule E-CBP: Capacity Bidding Program," 2018. [Online]. Available: [https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC\\_SCHEDULES\\_E-CBP.pdf](https://www.pge.com/tariffs/assets/pdf/tariffbook/ELEC_SCHEDULES_E-CBP.pdf). [Accessed: 29-Apr-2018].
- [25] Johannes Bisschop, "AIMMS Modeling Guide - Optimization Modeling," 2012.
- [26] J. H. Lee, "Model predictive control: Review of the three decades of development," *Int. J. Control. Autom. Syst.*, vol. 9, no. 3, pp. 415–424, 2011.
- [27] M. N. ElBsat and M. J. Wenzel, "Load and Electricity Rates Prediction for Building Wide Optimization Applications," *4th Int. High Perform. Build. Conf. Purdue*, 2016.
- [28] L. Gan, A. Wierman, U. Topcu, N. Chen, and S. H. Low, "Real-time deferrable load control: handling the uncertainties of renewable generation," *Proc. Fourth Int. Conf. Futur. Energy Syst. (e-Energy '13)*, pp. 113–124, 2013.
- [29] W. B. Powell, "A unified framework for stochastic and dynamic programming," 2018. [Online]. Available: [http://castlelab.princeton.edu/html/Papers/Powell-UnifiedFrameworkStochasticOptimization\\_July222017.pdf](http://castlelab.princeton.edu/html/Papers/Powell-UnifiedFrameworkStochasticOptimization_July222017.pdf).