Contract Design for Frequency Regulation by Aggregations of Commercial Buildings

Maximilian Balandat, Frauke Oldewurtel, Mo Chen and Claire Tomlin

Abstract— **We investigate the contract design problem that an energy aggregator who participates in the wholesale market for Ancillary Services faces. Specifically, we consider a situation in which commercial buildings agree to adjust their heating, ventilation and air conditioning power consumption with respect to a nominal schedule, according to a signal sent by the aggregator. This signal may vary arbitrarily within a certain band, whose (time-varying) width is part of the agreed-upon contract between aggregator and building. In return, the aggregator offers monetary rewards to incentivize the individual buildings. This allows the aggregator to bundle the resulting capacities from different buildings and sell the total capacity in the spot market for frequency regulation. The aggregator's problem is to jointly determine nominal schedules, regulation capacities and monetary rewards to maximize its profit, while ensuring that the individual buildings have an incentive to participate. Assuming that there is no private information, we cast the contract design problem as a bilevel optimization problem, which we in turn reformulate as a mixed-integer program. We further show that if the building does not impose negative externalities on the aggregator, the problem reduces to a Linear Program.**

I. Introduction

The essence of reliable electricity grid operation is matching supply and demand at all times. Today, this is commonly achieved by having generators and load-serving entities (LSE) participate in wholesale electricity markets. These markets run on different time-scales to benefit from better short-term demand and supply forecasts. At some point, however, running an *energy* spot market becomes impractical. Instead, the grid operator procures generation *capacities*, so-called Ancillary Services (AS), that can be dispatched on fast time-scales. The highest quality AS is frequency regulation (up and down), over which the operator has near real-time control. Traditionally, these frequency control reserves have been provided by conventional generators. Due to the increasing penetration of renewable energy sources, more and more generation is uncertain and intermittent, which results in an increasing need for frequency regulation reserves [1], [2]. Instead of providing these capacity reserves solely using conventional generators, the mitigation of frequency deviations can also be supported by activating reserves on the demand side. This idea is not new; large industrial plants such as aluminum producers [3] have been offering grid services for a long time. More recently, there have also been pilot programs for demand response (DR), in which buildings are asked to shed load when particular grid events occur. However, load shedding is generally considered undesirable, as it negatively impacts user utility and comfort.

This paper addresses the problem of grouping several mid-to-large size commercial buildings together by an entity called *aggregator*, which uses the buildings' inter-temporal consumption flexibility in order to provide frequency regulation capacity in the AS market. Buildings have flexibility in parts of their power consumption due to their inherent thermal storage capacity. Intelligent control of heating, ventilation and air conditioning (HVAC) systems facilitates flexibility in power consumption without compromising occupants' thermal and air quality comfort. Commercial buildings seem particularly well suited for this task due to their comparatively large power consumption, advanced building management systems, and their HVAC systems' potential to follow a high-frequency regulation signal [4].

In order to avoid load shedding or violating user comfort it is desirable to plan ahead, i.e., to back off from comfort and actuator constraints to increase flexibility in power consumption. This can for example be achieved by using Model Predictive Control (MPC), see [5], [6], [7], [8] and references therein. The idea of using MPC to have commercial buildings provide frequency regulation has been proposed in [9], [10]. Like most of the initial work in this domain, [9] assumes that the objectives of aggregator and buildings are perfectly aligned. However, unless aggregator and buildings are under the same ownership, this will usually not be the case. Indeed, having buildings provide regulation capacity typically requires less aggressive control strategies and hence results in a more costly operation. An aggregator therefore will need to compensate the building for its service. So while there may be a large potential in using buildings to provide frequency regulation, a large part of this potential can only be tapped by providing the correct financial incentives. While [10] investigates this problem, it does so in a rather ad-hoc way, and does not provide a structured approach to determining "the best" incentives.

In this paper, limiting ourselves to the case without information asymmetries between buildings and aggregator, we formulate the optimal contract design problem an aggregator faces. We propose a contract structure in which a building agrees to adjust its power consumption according to a regulation signal within time-varying capacity bands around

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA. [balandat,mochen72,tomlin]@eecs.berkeley.edu oldewurtel@berkeley.edu

This work has been supported in part by NSF under CPS:ActionWebs (CNS-931843) and CPS:FORCES (CNS-1239166). The research of F. Oldewurtel has received funding from the European Union Seventh Framework Programme FP7-PEOPLE-2011-IOF under grant agreement number 302255, Marie Curie project 'Stochastic Model Predictive Control, Energy Efficient Building Control, Smart Grid'.

a nominal power schedule. In exchange, the building receives a monetary reward from the aggregator. The financial risk for the building due to following the regulation signal can be absorbed by the aggregator by charging the building for its energy consumption according to its nominal schedule. The resulting problem is a bilevel optimization problem in which the aggregator jointly optimizes over nominal schedule, capacities and rewards in order to maximize its expected profit, subject to the buildings' individual rationality constraints. Using techniques from integer programming we cast this bilevel problem as an equivalent mixed-integer program. Moreover, we show that if buildings do not impose externalities on the aggregator, determining the resulting "first-best contract" [11] reduces to solving a Linear Program.

Notation: For a matrix M let $(M)_{i:}$ and $(M)_{:j}$ denote the i-th row and j-th column of M, respectively. Further, define $M^+ := \max(M, 0)$ and $M^- := \max(-M, 0)$ (element-wise). By $\times$ we denote element-wise vector multiplication, and by $\pi_A(x)$ the Euclidean projection of x onto A.

II. Preliminaries

A. Wholesale Markets for Ancillary Services

We consider AS markets in deregulated wholesale electricity spot markets. Specifically, we focus on the frequency regulation capacity market run by CAISO, the California Independent System Operator[1].

Like energy markets, frequency regulation capacity markets run on different timescales, namely the Day-Ahead Market (DAM) and the Real-Time Market (RTM). Each market operates as follows: After submitting its bids (consisting of maximum capacities and prices for both regulation up and regulation down), a resource gets awarded a regulation capacity schedule based on a uniform price auction conducted by the ISO. The market clearing process is a large-scale optimization problem, in which capacity prices are determined from the dual variables (commonly referred to as "shadow prices"). At run-time, the resource then receives a Load Frequency Control (LFC) signal ω from the ISO, specifying how much it should deviate from its nominal power schedule (which is determined in the energy market). The LFC signal is constrained to the awarded regulation capacities. The ISO has direct access to measurements of the power output of the resource, and hence can verify whether it has fulfilled its obligations. If so, the resource receives a capacity payment according to the market clearing price[2].

B. The Role of the Aggregator

We are interested in using the temporal flexibility that buildings have in terms of their HVAC power consumption in order to offer regulation capacity to the grid. The role of the aggregator is to provide an interface between the spot market on one side and the individual buildings on the other, as illustrated in Figure 1. There are many reasons why

[1]The basic market setup is quite similar among different ISOs in the US.

[2]Recently, ISOs have been ordered by the federal regulator to implement "pay for performance" compensation [12], which rewards resources for performance (e.g. mileage payments or signal tracking accuracy) in addition to capacity. For simplicity we will restrict our attention to capacity payments.

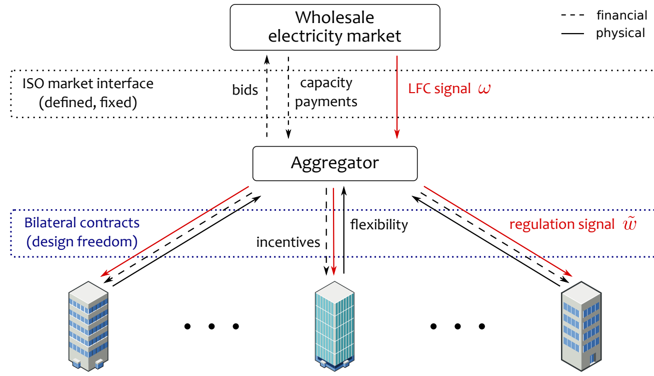

Fig. 1: The aggregator as an interface between market and buildings

individual buildings cannot easily participate in the wholesale market directly. Some markets have minimum bid size requirements [13] that often exceed a single building's power consumption. Moreover, participating in the market creates additional overhead – not only does it involve acquiring the necessary certifications from the ISO [14], but it also requires understanding market processes and bidding strategies.

An aggregator with expertise in these areas, on the other hand, can aggregate different buildings and bundle their individual flexibilities in such a way that the resulting product is suitable for sale in the spot market. In order for the individual buildings to be willing to participate in such a scheme, the aggregator must provide them with financial incentives. While the market interface is well defined and clearly specified, the aggregator has large freedom in choosing this incentive structure. In this paper we take a contract-theoretic point of view and consider the problem of maximizing the aggregator's profit over all contracts within a certain class.

C. The Basic Contract Structure

The basic structure of the bilateral contracts we propose is the following: For a given scheduling horizon, a building agrees to adjust its HVAC power consumption w.r.t. a nominal power schedule, based on a regulation signal w it receives from the aggregator that is guaranteed to lie within certain bounds. The building is charged for its energy consumption according to this nominal schedule, and the aggregator pays the building a one-time monetary reward R^b. In addition, in exchange for paying a penalty to the aggregator, the building may specify a certain amount of slack from the maximum up and down regulation signal, which it is permitted to violate. Nominal schedule, capacity bounds, slack and reward are all part of the contract specification.

Importantly, such a contract need not directly involve the building's power provider, e.g., the local utility. In fact, the provider simply bills the building for its actual consumption and need not even know of the contract (here we assume that the provider is oblivious to the "double payment" problem [15]). The difference between this amount and the hypothetical cost under the nominal schedule can then be settled ex post between building and aggregator. Charging the building only for its nominal consumption is an important feature of our contract, as it does not expose the building to any financial risk associated with the regulation signal.

D. Dealing With the Different Timescales of the Problem

An interesting characteristic of the multi-level problem we consider in this paper is that it takes place on three different timescales, as illustrated in Figure 2. On the market

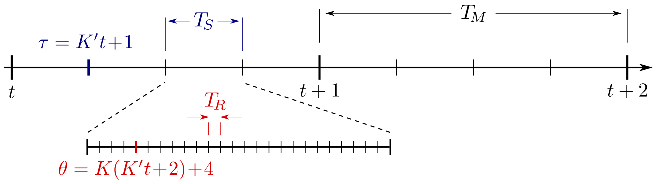

Fig. 2: Different time scales of the aggregation problem

level, bids, prices and awarded AS capacities are determined for market intervals of length T_M (typically $T_M = 1\,\mathrm{h}$ for DAM and $T_M = 15\,\mathrm{min}$ for RTM). On the other extreme, at the level of frequency control, the ISO sends a LFC signal ω that is piece-wise constant over intervals of length $T_R \ll T_M$ ($T_R = 4\,\mathrm{s}$ in CAISO AS regions). The control schedule for a building is determined for an intermediate interval length T_S, which depends on the time-constants of the building, the frequency with which set points can be changed, and possible computational challenges with the resulting problem. Commonly, T_S will lie in the range of $5-15\,\mathrm{min}$. Here we make the following Assumption:

Assumption 1: There exist constants $K, K' \in \mathbb{N}$ such that $T_S = KT_R$ and $T_M = K'T_S$, respectively.

We consider a contract design problem over a scheduling horizon of N_M market periods (e.g., $N_M = 24$ for the Day-Ahead-Market). Let $N_S = K'N_M$, $N_R = KN_S$ and

$$\mathbb{T}_S(t) = \left\{\tau \in N_S : t\,T_M \le \tau T_S < (t+1)T_M\right\}$$
$$\mathbb{T}_R(\tau) = \left\{\theta \in N_R : \tau\,T_S \le \theta T_R < (\tau+1)T_S\right\}$$

We also write $t_\tau = \mathbb{T}_S^{-1}(\tau)$ and $\tau_\theta = \mathbb{T}_R^{-1}(\theta)$.

III. A Building's Scheduling Problem

A. Building Model

Consider a building b whose HVAC dynamics for sampling time T_S are described by the discrete-time LTI system[3]

$$x^{\mathrm{b}}_{\tau+1} = A^{\mathrm{b}} x^{\mathrm{b}}_\tau + B^{\mathrm{b}}(u^{\mathrm{b}}_\tau + \tilde{w}^{\mathrm{b}}_\tau) + E^{\mathrm{b}} v^{\mathrm{b}}_\tau \tag{1a}$$
$$y^{\mathrm{b}}_\tau = C^{\mathrm{b}} x^{\mathrm{b}}_\tau + D^{\mathrm{b}}(u^{\mathrm{b}}_\tau + \tilde{w}^{\mathrm{b}}_\tau) + F^{\mathrm{b}} v^{\mathrm{b}}_\tau \tag{1b}$$

with known initial condition $x^{\mathrm{b}}_0 \in \mathbb{R}^{n^{\mathrm{b}}_x}$. Here $u^{\mathrm{b}}_\tau \in \mathbb{R}^{n^{\mathrm{b}}_u}$ represents the control inputs to the HVAC system, $y^{\mathrm{b}}_\tau \in \mathbb{R}^{n^{\mathrm{b}}_y}$ is the system output and $v^{\mathrm{b}}_\tau \in \mathbb{R}^{n^{\mathrm{b}}_v}$ is the effect of weather and occupancy in time interval t. The signal $\tilde{w}^{\mathrm{b}}_\tau \in \mathbb{R}^{n^{\mathrm{b}}_u}$ is a proxy for the regulation signal w^{b}_θ received by the aggregator[4]. We write $\mathbf{u}^{\mathrm{b}} = (u^{\mathrm{b}}_0, \ldots, u^{\mathrm{b}}_{N_S})$ and will use analogous notation for other variables throughout the paper.

We make the following assumption:

Assumption 2: For each building, the power consumption of each actuator is linear in the associated control input.

Under Assumption 2, building b's overall HVAC energy consumption Q^{b}_t in interval τ can be expressed as $Q^{\mathrm{b}}_\tau = (q^{\mathrm{b}})^\top u^{\mathrm{b}}_\tau = \sum_{i=1}^{n_y} q^{\mathrm{b}}_i\, u^{\mathrm{b}}_{\tau,i}$, where $q^{\mathrm{b}}_i \ge 0$ is the per-unit energy consumption of actuator i over an interval of length T_S.

As weather and occupancy effects are usually uncertain, we need to account for this uncertainty in the scheduling problem. There are different ways of accomplishing this; here we choose to formulate chance constraints that require the building's comfort constraints to be satisfied with given levels of probability. The weather and occupancy effect v^{b}_τ can be split into a known forecast $\hat{v}^{\mathrm{b}}_\tau$ and the forecast error $\tilde{v}^{\mathrm{b}}_\tau := v^{\mathrm{b}}_\tau - \hat{v}^{\mathrm{b}}_\tau$. We assume that – based on the analysis of historical data – a linear uncertainty model of the form

$$\tilde{v}^{\mathrm{b}}_{\tau+1} = G^{\mathrm{b}} \tilde{v}^{\mathrm{b}}_\tau + H^{\mathrm{b}} e^{\mathrm{b}}_\tau \tag{2}$$

has been identified, where $e^{\mathrm{b}}_\tau \overset{d}{=} \mathcal{N}(0, I)$ are i.i.d. Gaussian and $\tilde{v}^{\mathrm{b}}_0 \overset{d}{=} \mathcal{N}(0, \Sigma^{\mathrm{b}}_0)$ with e^{b}_τ and $\tilde{v}^{\mathrm{b}}_0$ independent for all τ.

Some algebra allows us to write the system output y^{b}_τ as

$$\begin{aligned} y^{\mathrm{b}}_\tau = \hat{y}^{\mathrm{b}}_\tau(\mathbf{u}^{\mathrm{b}}) &+ \textstyle\sum_{\iota<\tau} W^{\mathrm{b}}_{\tau-1-\iota} \tilde{w}^{\mathrm{b}}_\iota + D^{\mathrm{b}}_\tau \tau^{\mathrm{b}}_\tau \\ &+ \mathcal{G}^{\mathrm{b}}_\tau \tilde{v}^{\mathrm{b}}_0 + \textstyle\sum_{\iota<\tau} \mathcal{H}^{\mathrm{b}}_{\tau,\iota} e^{\mathrm{b}}_\iota \end{aligned} \tag{3}$$

where

$$\begin{aligned} \hat{y}^{\mathrm{b}}_\tau(\mathbf{u}^{\mathrm{b}}) = C^{\mathrm{b}} D^{\mathrm{b}} u^{\mathrm{b}}_\tau &+ C^{\mathrm{b}} F^{\mathrm{b}} \hat{v}^{\mathrm{b}}_\tau + C^{\mathrm{b}} (A^{\mathrm{b}})^\tau x^{\mathrm{b}}_0 \\ &+ \textstyle\sum_{\iota<\tau} C^{\mathrm{b}} (A^{\mathrm{b}})^{\tau-1-\iota} \big(B^{\mathrm{b}} u^{\mathrm{b}}_\iota + E^{\mathrm{b}} \hat{v}^{\mathrm{b}}_\iota\big) \end{aligned} \tag{4}$$

is the nominal system output, $W^{\mathrm{b}}_\tau = C^{\mathrm{b}}(A^{\mathrm{b}})^\tau B^{\mathrm{b}}$, and

$$\mathcal{G}^{\mathrm{b}}_\tau = \textstyle\sum_{\iota<\tau} C^{\mathrm{b}} (A^{\mathrm{b}})^{\tau-1-\iota} E^{\mathrm{b}} (G^{\mathrm{b}})^\iota + F^{\mathrm{b}} (G^{\mathrm{b}})^\tau \tag{5}$$

$$\begin{aligned} \mathcal{H}^{\mathrm{b}}_{\tau,\iota} = &\textstyle\sum_{\iota<\theta<\tau} C^{\mathrm{b}} (A^{\mathrm{b}})^{\tau-1-\theta} E^{\mathrm{b}} (G^{\mathrm{b}})^{\theta-1-\iota} H^{\mathrm{b}} \\ &+ F^{\mathrm{b}} (G^{\mathrm{b}})^{\tau-1-\iota} H^{\mathrm{b}} \end{aligned} \tag{6}$$

B. The Robust Optimal Scheduling Problem

Denote by $\mathbf{r}^{\mathrm{b}\uparrow} = (r^{\mathrm{b}\uparrow}_0, \ldots, r^{\mathrm{b}\uparrow}_{N_S})$ and $\mathbf{r}^{\mathrm{b}\downarrow} = (r^{\mathrm{b}\downarrow}_0, \ldots, r^{\mathrm{b}\downarrow}_{N_S})$ the sequences of non-negative vectors of "regulation up" and "regulation down" capacities for building b, respectively[5]. Suppose that at each time τ, the building receives a regulation signal $\tilde{w}^{\mathrm{b}}_\tau \in [-r^{\mathrm{b}\downarrow}_\tau, r^{\mathrm{b}\uparrow}_\tau]$. For the purpose of this section suppose that the capacity vectors $\mathbf{r}^{\mathrm{b}\uparrow}$ and $\mathbf{r}^{\mathrm{b}\downarrow}$ are fixed and known. It may be infeasible for a building to tolerate all possible sequences of the regulation signal within $[-\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}]$ without jeopardizing occupant comfort. Hence we allow the building to shrink the interval within which it is required to robustly follow the regulation signal to $[-r^{\mathrm{b}\uparrow}_\tau + s^{\mathrm{b}\uparrow}_\tau, r^{\mathrm{b}\downarrow}_\tau - s^{\mathrm{b}\downarrow}_\tau]$, where $s^{\mathrm{b}\uparrow}_\tau$ and $s^{\mathrm{b}\downarrow}_\tau$ with

$$0 \le s^{\mathrm{b}\downarrow}_\tau \le r^{\mathrm{b}\downarrow}_\tau, \quad 0 \le s^{\mathrm{b}\uparrow}_\tau \le r^{\mathrm{b}\uparrow}_\tau \qquad \forall\, \tau \le N_S \tag{7}$$

are the *slacks*. We define the *effective regulation signal*

$$\check{w}_\tau = \pi_{[-r^{\mathrm{b}\uparrow}_\tau + s^{\mathrm{b}\uparrow}_\tau,\, r^{\mathrm{b}\downarrow}_\tau - s^{\mathrm{b}\downarrow}_\tau]}(\tilde{w}^{\mathrm{b}}_\tau) \tag{8}$$

In exchange for the reduced uncertainty resulting from this, the building pays a penalty $\mathcal{P}^{\mathrm{b}}(\mathbf{s}^{\mathrm{b}})$ to the aggregator, where

[3]The assumption of time-invariance is only made for notational convenience; the extension to linear time-varying systems is immediate.

[4]Under some mild conditions, which are satisfied by most building models, one can show that robustness of (1) w.r.t. $\tilde{w}^{\mathrm{b}}_\tau \in [-r^{\mathrm{b}\downarrow}_\tau, r^{\mathrm{b}\uparrow}_\tau]$ implies robustness of the "true" fast-sampled system w.r.t. $w^{\mathrm{b}}_\theta \in [-r^{\mathrm{b}\downarrow}_{\tau_\theta}, r^{\mathrm{b}\uparrow}_{\tau_\theta}]$.

[5]Our convention is to use the classic point of view of generation markets in which "regulation up" means that a generator increases its power output and "regulation down" means that he decreases it. Thus, from the point of view of a load, "regulation up" means a *decrease* while "regulation down" means an *increase* in power consumption w.r.t. the nominal schedule.

$\mathbf{s}^{\mathrm{b}} = (\mathbf{s}^{\mathrm{b}\uparrow}, \mathbf{s}^{\mathrm{b}\downarrow})$. Given $\mathbf{r}^{\mathrm{b}}$ and $\mathbf{s}^{\mathrm{b}}$, the building seeks the nominal schedule of minimum expected cost that ensures constraint satisfaction for all possible sequences of $\check{\mathbf{w}}^{\mathrm{b}}$.

1) Input Constraints: The actual control inputs $u_\tau^{\mathrm{b}} + \check{w}_\tau^{\mathrm{b}}$ at time τ are restricted by physical limits on the actuators, which translate into the following constraints:

$$\underline{u}_\tau^{\mathrm{b}} \leq u_\tau^{\mathrm{b}} + \check{w}_\tau^{\mathrm{b}} \leq \overline{u}_\tau^{\mathrm{b}} \qquad \forall\, \tau \leq N_S \tag{9}$$

Since the regulation signal is unknown at scheduling time, the above constraints have to hold for any regulation signal $\check{w}_\tau^{\mathrm{b}} \in [-r_\tau^{\mathrm{b}\uparrow} + s_\tau^{\mathrm{b}\uparrow}, r_\tau^{\mathrm{b}\downarrow} - s_\tau^{\mathrm{b}\downarrow}]$. This can be achieved by imposing the following robustified version of (9):

$$\underline{u}_\tau^{\mathrm{b}} + r_\tau^{\mathrm{b}\uparrow} - s_\tau^{\mathrm{b}\uparrow} \leq u_\tau^{\mathrm{b}} \leq \overline{u}_\tau^{\mathrm{b}} - r_\tau^{\mathrm{b}\downarrow} + s_\tau^{\mathrm{b}\downarrow} \qquad \forall\, \tau \leq N_S \tag{10}$$

Importantly, (10) deals with the uncertainty introduced by $\check{\mathbf{w}}^{\mathrm{b}}$ in a robust fashion, i.e., it accounts for the worst case, so the building need not make any assumptions on the statistical properties of $\check{\mathbf{w}}^{\mathrm{b}}$.

2) Output Constraints: The building aims to ensure that, for all possible sequences of the signal $\check{\mathbf{w}}^{\mathrm{b}}$, the output $\mathbf{y}^{\mathrm{b}}$ satisfies the building's comfort constraints. Since small violations of comfort constraints are usually non-critical, we assume the building formulates chance constraints, which ensure that the constraints at any point in time are satisfied with a given level of probability. Specifically, suppose that, for all possible sequences of $\check{\mathbf{w}}^{\mathrm{b}}$, the comfort constraints $y_{\tau,i}^{\mathrm{b}} \geq \underline{y}_{\tau,i}^{\mathrm{b}}$ and $y_{\tau,i}^{\mathrm{b}} \leq \overline{y}_{\tau,i}^{\mathrm{b}}$ are to be satisfied with a probability of at least $1-\underline{\alpha}_{\tau,i}^{\mathrm{b}}$ and $1-\overline{\alpha}_{\tau,i}^{\mathrm{b}}$, respectively (here $\underline{\alpha}_{\tau,i}^{\mathrm{b}}, \overline{\alpha}_{\tau,i}^{\mathrm{b}} < 0.5$ are given, and typically $\underline{\alpha}_{\tau,i}^{\mathrm{b}}, \overline{\alpha}_{\tau,i}^{\mathrm{b}} \ll 1$).

Let $\mathcal{W}_\tau^{\mathrm{b}} = \big\{(\tilde{w}_0, \ldots, \tilde{w}_\tau) \colon -r_\iota^{\mathrm{b}\uparrow} + s_\iota^{\mathrm{b}\uparrow} \leq \tilde{w}_\iota \leq r_\iota^{\mathrm{b}\downarrow} - s_\iota^{\mathrm{b}\uparrow},\, \forall\, \iota \leq \tau\big\}$. The constraints on the output then read

$$\forall\, \tau, \forall\, i, \forall\, \tilde{w} \in \mathcal{W}_\tau^{\mathrm{b}}, \quad \begin{cases} \mathbb{P}\big(y_{\tau,i}^{\mathrm{b}} \geq \underline{y}_{\tau,i}^{\mathrm{b}}\big) \geq 1 - \underline{\alpha}_{\tau,i}^{\mathrm{b}} \\ \mathbb{P}\big(y_{\tau,i}^{\mathrm{b}} \leq \overline{y}_{\tau,i}^{\mathrm{b}}\big) \geq 1 - \overline{\alpha}_{\tau,i}^{\mathrm{b}} \end{cases} \tag{11}$$

Using techniques from robust optimization and chance-constrained programming, the above chance constraints for $\tau = 0, \ldots, N_S$ and $1 \leq i \leq n_y^{\mathrm{b}}$ can be expressed as

$$\begin{aligned} \hat{y}_{\tau,i}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) \geq \underline{y}_{\tau,i}^{\mathrm{b}} &+ \textstyle\sum_{\iota<\tau}\Big[\big(W_{\tau-1-\iota}^{\mathrm{b}}\big)_{i:}^{+}(r_\iota^{\mathrm{b}\uparrow} - s_\iota^{\mathrm{b}\uparrow}) \\ &+ \big(W_{\tau-1-\iota}^{\mathrm{b}}\big)_{i:}^{-}(r_\iota^{\mathrm{b}\downarrow} - s_\iota^{\mathrm{b}\downarrow})\Big] + (D^{\mathrm{b}})_{i:}^{+}(r_\tau^{\mathrm{b}\uparrow} - s_\tau^{\mathrm{b}\uparrow}) \\ &+ (D^{\mathrm{b}})_{i:}^{-}(r_\tau^{\mathrm{b}\downarrow} - s_\tau^{\mathrm{b}\downarrow}) + \sigma_{\tau,i}^{\mathrm{b}}\, \Phi^{-1}(1 - \underline{\alpha}_{\tau,i}^{\mathrm{b}}) \end{aligned} \tag{12a}$$

$$\begin{aligned} \hat{y}_{\tau,i}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) \leq \overline{y}_{\tau,i}^{\mathrm{b}} &- \textstyle\sum_{\iota<\tau}\Big[\big(W_{\tau-1-\iota}^{\mathrm{b}}\big)_{i:}^{+}(r_\iota^{\mathrm{b}\downarrow} - s_\iota^{\mathrm{b}\downarrow}) \\ &+ \big(W_{\tau-1-\iota}^{\mathrm{b}}\big)_{i:}^{-}(r_\iota^{\mathrm{b}\uparrow} - s_\iota^{\mathrm{b}\uparrow})\Big] - (D^{\mathrm{b}})_{i:}^{+}(r_\tau^{\mathrm{b}\downarrow} - s_\tau^{\mathrm{b}\downarrow}) \\ &- (D^{\mathrm{b}})_{i:}^{-}(r_\tau^{\mathrm{b}\uparrow} - s_\tau^{\mathrm{b}\uparrow}) - \sigma_{t,i}^{\mathrm{b}}\, \Phi^{-1}(1 - \overline{\alpha}_{\tau,i}^{\mathrm{b}}) \end{aligned} \tag{12b}$$

where $(\sigma_{\tau,i}^{\mathrm{b}})^2 = (\mathcal{G}_\tau^{\mathrm{b}})_{i:}\Sigma_0^{\mathrm{b}}(\mathcal{G}_\tau^{\mathrm{b}})_{i:}^\top + \sum_{\iota<\tau}(\mathcal{H}_\tau^{\mathrm{b}})_{i:}(\mathcal{H}_\iota^{\mathrm{b}})_{i:}^\top$ and Φ is the cdf of a standard normal random variable.

Note that, importantly, both (12a) and (12b) are linear in the control, the capacities and the slack variables.

3) The Robust Scheduling Problem for Fixed Capacities: Let $\mathbf{c}^{\mathrm{b}} = (c_0^{\mathrm{b}}, \ldots, c_{N_S}^{\mathrm{b}})$ be the vector of electricity prices for the building over the scheduling horizon[6]. As discussed in section II-C, in the contract we propose a building pays only for its power consumption according to nominal schedule. That is, the energy cost (excluding $\mathcal{P}^{\mathrm{b}}$) of building b is

$$\varphi_{\text{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) = \textstyle\sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} (q^{\mathrm{b}})^\top u_\tau^{\mathrm{b}}$$

In practice the building will still pay the utility

$$\varphi_{\text{act}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}, \check{\mathbf{w}}^{\mathrm{b}}) = \textstyle\sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} (q^{\mathrm{b}})^\top u_\tau^{\mathrm{b}} + \sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} \frac{(q^{\mathrm{b}})^\top}{K} \sum_{\theta \in \mathbb{T}_R(\tau)} \check{w}_\theta^{\mathrm{b}}$$

for its actual consumption, while the difference

$$\Delta\varphi^{\mathrm{b}}(\check{\mathbf{w}}^{\mathrm{b}}) = \textstyle\sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} \frac{(q^{\mathrm{b}})^\top}{K} \sum_{\theta \in \mathbb{T}_R(\tau)} \check{w}_\theta^{\mathrm{b}}$$

is settled ex post with the aggregator. Here the signal $\check{\mathbf{w}}^{\mathrm{b}}$ is the effective regulation signal received with frequency $1/T_R$.

The building's overall robust scheduling problem for given regulation capacity vectors $\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}$ is therefore

$$\begin{aligned} (\mathbf{u}^{\mathrm{b}*}, \mathbf{s}^{\mathrm{b}*})(\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}) = \underset{\mathbf{u}^{\mathrm{b}}, \mathbf{s}^{\mathrm{b}}}{\arg\min}\ & \varphi_{\text{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) + \mathcal{P}^{\mathrm{b}}(\mathbf{s}^{\mathrm{b}}) \\ \text{s.t.}\quad & (7),\ (10),\ (12) \end{aligned} \tag{13}$$

We make the following assumption:

Assumption 3: The building penalties $\mathcal{P}^{\mathrm{b}}$ are of the form

$$\mathcal{P}^{\mathrm{b}}(\mathbf{s}^{\mathrm{b}}) = \textstyle\sum_{\mathrm{b}} \sum_{\tau=1}^{N_S} (p_l^{\uparrow} s_\tau^{\uparrow} + p_q^{\uparrow}(s_\tau^{\uparrow})^2 + p_l^{\downarrow} s_\tau^{\downarrow} + p_q^{\downarrow}(s_\tau^{\downarrow})^2)$$

where $p_l^{\uparrow}, p_l^{\downarrow} \geq 0$ and $p_q^{\uparrow}, p_q^{\downarrow} \geq 0$ are fixed coefficients. Under Assumption 3, (13) is a Quadratic Program (QP).

IV. The Optimal Contract Design Problem

In this section we formulate the aggregator's optimal contract design problem for a given collection $\mathcal{B}$ of buildings. To this end we require some additional assumptions.

Assumption 4: The aggregator is risk-neutral.

Assumption 5: The aggregator has no market power in the wholesale market for frequency regulation capacity.

Assumption 6: Buildings do not possess any private information. That is, for each $\mathrm{b} \in \mathcal{B}$ the aggregator has full knowledge of the system model (1), constraints, energy consumption rates q^{b} and tariff $\mathbf{c}^{\mathrm{b}}$. Moreover, it has access to measurements of the initial state x_0^{b} and the HVAC power consumption on the frequency $1/T_R$ of the LFC signal ω.

Assumption 4 is standard and made primarily for mathematical convenience. An extension to a risk-averse aggregator, while desirable, appears quite challenging at this point.

Assumption 5 holds if the total capacity that can be provided by all buildings is small enough that the aggregator's bid does not significantly affect the overall supply curve. Then the aggregator acts as a price-taker and places its bid at true cost. Due to Assumptions 2 and 4 it suffices that the aggregator has access to an unbiased estimate of the expected capacity prices over the scheduling horizon.

Assumption 6 is more restrictive. In practice, an aggregator will not know the exact building models, which causes information asymmetries. It would be interesting to analyze how much of a so-called information rent a building could earn in this case, but this is outside the scope of this paper.

[6]We assume c^{b} is known over the whole scheduling horizon, but may otherwise be arbitrary (i.e. c^{b} can describe both flat and Time-of-Use tariffs).

A. Objective Function

Denote by $r_t^{M\uparrow}$ and $r_t^{M\downarrow}$ the regulation up and down capacities offered in the market in period t, respectively. Let $\mathbf{r}^M = (r^{M\uparrow}, r^{M\downarrow})$ and, similarly, $R = \{R^{\mathrm{b}} : \mathrm{b} \in \mathcal{B}\}$, $\mathbf{r} = \{(\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}) : \mathrm{b} \in \mathcal{B}\}$ and $\mathbf{s} = \{(\mathbf{s}^{\mathrm{b}\downarrow}, \mathbf{s}^{\mathrm{b}\uparrow}) : \mathrm{b} \in \mathcal{B}\}$.

The deviations $\epsilon_\theta^{\mathrm{b}} := \check{w}_\theta^{\mathrm{b}} - w_\theta^{\mathrm{b}}$ of building b's control from the dispatch signal $u_{\tau_\theta} + w_\theta^{\mathrm{b}}$ are given by

$$\epsilon_\theta^{\mathrm{b}} = (w_\theta^{\mathrm{b}} + r_{\tau_\theta}^{\mathrm{b}\uparrow} - s_{\tau_\theta}^{\mathrm{b}\uparrow})^- - (w_\theta^{\mathrm{b}} - r_{\tau_\theta}^{\mathrm{b}\downarrow} + s_{\tau_\theta}^{\mathrm{b}\downarrow})^+$$

Observe that $\epsilon_\theta^{\mathrm{b}} = 0$ if the slacks $s_{\tau_\theta}^{\mathrm{b}\uparrow}, s_{\tau_\theta}^{\mathrm{b}\downarrow}$ are zero. With

$$w_\theta^M = \begin{cases} -\omega_\theta \, r_{t_{\tau_\theta}}^{M\uparrow} & \text{if } \omega_\theta \geq 0 \\ -\omega_\theta \, r_{t_{\tau_\theta}}^{M\downarrow} & \text{if } \omega_\theta < 0 \end{cases} \tag{14}$$

denoting the aggregator's regulation signal derived from the LFC signal $\omega_\theta \in [-1, 1]$, the deviation ϵ_θ^M of the buildings' aggregate consumption from the ISO dispatch signal is

$$\epsilon_\theta^M = \textstyle\sum_{\mathrm{b}} (q^{\mathrm{b}})^\top w_\theta^{\mathrm{b}} - w_\theta^M + \sum_{\mathrm{b}} (q^{\mathrm{b}})^\top \epsilon_\theta^{\mathrm{b}}$$

Using the above we can write

$$\mathbb{E}\big[\textstyle\sum_{\mathrm{b}} \Delta\varphi^{\mathrm{b}}(\check{\mathbf{w}}^{\mathrm{b}})\big] = \sum_{\mathrm{b}} \sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} \sum_{\theta \in \mathbb{T}_R(\tau)} \frac{(q^{\mathrm{b}})^\top}{K} \mathbb{E}\big[w_\theta^{\mathrm{b}} + \epsilon_\theta^{\mathrm{b}}\big]$$

Under Assumption 4 the aggregator aims to maximize its expected profit

$$\begin{aligned} \psi(\mathbf{r}^M, \mathbf{r}, \mathbf{s}, R) = \mathbb{E}\Big[&\textstyle\sum_{t=1}^{N_M} \big(\rho_t^\uparrow \, r_t^{M\uparrow} + \rho_t^\downarrow \, r_t^{M\downarrow}\big) - \mathcal{P}^M(\boldsymbol{\epsilon}^M) \\ &- \textstyle\sum_{\mathrm{b}} \big(R^{\mathrm{b}} + \Delta\varphi^{\mathrm{b}}(\check{\mathbf{w}}^{\mathrm{b}}) - \mathcal{P}^b(\mathbf{s}^{\mathrm{b}})\big)\Big] \end{aligned} \tag{15}$$

which consists of the expected revenue generated in the spot market, a penalty $\mathcal{P}(\boldsymbol{\epsilon}^M)$ incurred for deviations from the market dispatch signal and, for each building, the reward R^{b}, the cost difference $\Delta\varphi^{\mathrm{b}}$ and the building's penalty $\mathcal{P}^{\mathrm{b}}$. A typical market penalty would for example be

$$\mathcal{P}(\boldsymbol{\epsilon}^M) = \textstyle\sum_{\theta=1}^{N_R} p_h (\epsilon_\theta^M)^+ + p_l (\epsilon_\theta^M)^- \tag{16}$$

where $p_h, p_l \geq 0$ are the per-unit penalty factors[7].

So far we have not discussed how the aggregator would determine the signals w_θ^{b} from the ISO's LFC signal ω_θ. While very interesting[8], this problem is beyond the scope of this paper. We make the following simplifying assumption:

Assumption 7: The contributions of the different buildings to the provided overall modulation of the demand are determined in a "pro rata" fashion, i.e., proportional to their respective contracted capacities. Specifically, we assume that

$$w_\theta^{\mathrm{b}} = \begin{cases} -\omega_\theta \, r_{\tau_\theta}^{\mathrm{b}\uparrow} & \text{if } \omega_\theta \geq 0 \\ -\omega_\theta \, r_{\tau_\theta}^{\mathrm{b}\downarrow} & \text{if } \omega_\theta < 0 \end{cases} \tag{17}$$

for all $b \in \mathcal{B}$ and $\theta = 0, \ldots, N_R$.

Not surprisingly, the aggregator's payoff will depend on the statistical properties of the LFC signal ω. Since this signal depends directly on the mismatch of supply and demand in the grid, which the ISO consistently tries to minimize, it is quite natural to make the following assumption:

[7] $-\mathcal{P}^M$ can also be interpreted as a "pay for performance" payment [12].

[8] The overall demand modulation must be split not only between the different buildings but also between different actuators within the buildings.

Assumption 8: For all $\theta = 0, \ldots, N_R$ the LFC signal ω_θ is a random variable supported on $[-1, 1]$ with a continuous distribution that is symmetric around zero.

Using Assumptions 7 and 8 we can compute

$$\mathbb{E}\big[w_\theta^{\mathrm{b}}\big] = \zeta_\theta\big(r_{\tau_\theta}^{\mathrm{b}\downarrow} - r_{\tau_\theta}^{\mathrm{b}\uparrow}\big) \tag{18a}$$

$$\begin{aligned} \mathbb{E}\big[\epsilon_\theta^{\mathrm{b}}\big] = \;&\mathbb{E}\big[\big((1 - \omega_\theta) r_{\tau_\theta}^{\mathrm{b}\uparrow} - s_{\tau_\theta}^{\mathrm{b}\uparrow}\big)^- \, \mathbf{1}_{\{\omega_\theta \geq 0\}}\big] \\ &- \mathbb{E}\big[\big(-(1 + \omega_\theta) r_{\tau_\theta}^{\mathrm{b}\downarrow} + s_{\tau_\theta}^{\mathrm{b}\downarrow}\big)^+ \, \mathbf{1}_{\{\omega_\theta < 0\}}\big] \end{aligned} \tag{18b}$$

where $\zeta_\theta := 0.5 \, \mathbb{E}[\omega_\theta \,|\, \omega_\theta \geq 0]$. Hence $\mathbb{E}\big[w_\theta^{\mathrm{b}}\big]$ is linear in the capacities, and this is independent of the distribution of ω_θ. This is not the case for $\mathbb{E}\big[\epsilon_\theta^{\mathrm{b}}\big]$, i.e., the effect of $r_{\tau_\theta}^{\mathrm{b}}$ and $s_{\tau_\theta}^{\mathrm{b}}$ on $\mathbb{E}\big[\epsilon_\theta^{\mathrm{b}}\big]$ on is distribution-dependent. E.g., if $\omega_\theta \sim U[-1, 1]$ one can find that $\mathbb{E}[\epsilon_\theta^{\mathrm{b}}] = \frac{1}{4}\big[(s_{\tau_\theta}^{\mathrm{b}\uparrow})^2 / r_{\tau_\theta}^{\mathrm{b}\uparrow} - (s_{\tau_\theta}^{\mathrm{b}\downarrow})^2 / r_{\tau_\theta}^{\mathrm{b}\downarrow}\big]$.

From now on we will assume for simplicity that $\zeta_\theta = \zeta$ for all θ, which is for example the case if the ω_θ are identically distributed. From (18) and Assumption 8 it follows that $\mathbb{E}\big[w_\theta^{\mathrm{b}} + \epsilon_\theta^{\mathrm{b}}\big] = 0$ if capacities and slacks are symmetric.

B. Individual Rationality Constraints

Observe that for all buildings $b \in \mathcal{B}$ we have that

$$\varphi_{\mathrm{oo}}^{\mathrm{b}*} := \varphi_{\mathrm{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}*}(0\,, 0)) \leq \varphi_{\mathrm{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}*}(\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}))$$

for any capacity vectors $\mathbf{r}^{\mathrm{b}\downarrow}, \mathbf{r}^{\mathrm{b}\uparrow}$. We refer to $\varphi_{\mathrm{oo}}^{\mathrm{b}*}$ as the value of the *outside option* of building b, since it represents the minimum (nominal) energy cost without contract. The aggregator must then ensure that the overall cost incurred by building b (including the monetary reward R^{b}) when participating in the contract does not exceed $\varphi_{\mathrm{oo}}^{\mathrm{b}*}$.

C. The Aggregator's Optimization Problem

The aggregator's optimal contract design problem can now be stated as the following bilevel optimization problem [16]:

$$\max_{\mathbf{r}^M, \mathbf{r}, \mathbf{u}, \mathbf{s}, R} \; \psi(\mathbf{r}^M, \mathbf{r}, \mathbf{s}, R) \tag{19a}$$

subject to

$$r_t^{M\uparrow} \geq 0, \; r_t^{M\downarrow} \geq 0 \qquad \forall\, t \leq N_M \tag{19b}$$

$$r_\tau^{\mathrm{b}\uparrow} \geq 0, \; r_\tau^{\mathrm{b}\downarrow} \geq 0 \qquad \forall\, \mathrm{b} \in \mathcal{B}, \; \forall\, \tau \leq N_S \tag{19c}$$

$$(\mathbf{u}^{\mathrm{b}}, \mathbf{s}^{\mathrm{b}}) \text{ solves (13)} \qquad \forall\, \mathrm{b} \in \mathcal{B} \tag{19d}$$

$$\varphi^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}, \mathbf{s}^{\mathrm{b}}) - R^{\mathrm{b}} \leq \varphi_{\mathrm{oo}}^{\mathrm{b}*} \qquad \forall\, \mathrm{b} \in \mathcal{B} \tag{19e}$$

In the above problem, (19d) requires each building to behave optimally w.r.t. its capacity vectors $\mathbf{r}^{\mathrm{b}\downarrow}$ and $\mathbf{r}^{\mathrm{b}\uparrow}$. Participation of each building is ensured by the individual rationality constraint (19e).

Observing that R^{b} does not affect (19d) and that (15) is strictly decreasing in R^{b}, it is clear that the individual rationality constraint (19e) will be tight at the optimum. Substituting the resulting equality into the objective we get

$$\begin{aligned} \tilde{\psi}(\mathbf{r}^M, \mathbf{r}, \mathbf{s}) = \mathbb{E}\Big[&\textstyle\sum_{t=1}^{N_M} \big(\rho_t^\uparrow \, r_t^{M\uparrow} + \rho_t^\downarrow \, r_t^{M\downarrow}\big) - \mathcal{P}^M(\boldsymbol{\epsilon}^M) \\ &- \textstyle\sum_{\mathrm{b}} \varphi_{\mathrm{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) - \varphi_{\mathrm{oo}}^{\mathrm{b}*} + \Delta\varphi^{\mathrm{b}}(\check{\mathbf{w}}^{\mathrm{b}})\big)\Big] \end{aligned} \tag{20}$$

and with that (19) is equivalent to

$$\max_{\mathbf{r}^M, \mathbf{r}, \mathbf{u}, \mathbf{s}} \; \tilde{\psi}(\mathbf{r}^M, \mathbf{r}, \mathbf{s}) \quad \text{s.t.} \quad \text{(19b), (19c), (19d)} \tag{21}$$

If the nominal scheduling problems for all buildings are feasible, then (19) will always have a solution: if $\mathbf{r}^M = \mathbf{r} = \mathbf{s} = 0$ and $\mathbf{u}^{\mathrm{b}} = \mathbf{u}^{\mathrm{b}*}(0\,,0)$ for all b then (19b) and (19c) hold with equality, and (19d) holds by definition of $\mathbf{u}^{\mathrm{b}*}$.

V. Solution Methodology

The aggregator's optimization problem (19a) is a bilevel optimization problem, for which different solution approaches exist. These include vertex enumeration techniques, branch-and-bound (BNB) algorithms, penalty function methods, descent algorithms and trust-region methods [16]. We focus on BNB techniques since they can provide certificates for global optimality of the solution based on duality theory.

A. The General Case

The basic idea of BNB methods for bilevel optimization problems is to replace the lower level problem by a set of suitable optimality conditions. In our case the lower level problem is a QP, so the corresponding Karush-Kuhn-Tucker (KKT) optimality conditions are necessary and sufficient. Hence the bilevel problem can be written equivalently as a standard optimization problem in which the KKT conditions of the lower level problem appear as constraints[9].

Since for QPs the primal feasibility, dual feasibility and Lagrangian stationarity conditions are all affine constraints, the only challenge are the complementary slackness (CS) conditions. For example, the corresponding CS conditions for the (upper) robustified input constraints (10) are given by

$$\underline{\lambda}_\tau^{\mathrm{b}} \times \left(-u_\tau^{\mathrm{b}} + \underline{u}_\tau^{\mathrm{b}} + r_\tau^{\mathrm{b}\uparrow} - s_\tau^{\mathrm{b}\uparrow}\right) = 0, \qquad \forall\, \tau = 0, \dots, N_S$$

where $0 \leq \underline{\lambda}_\tau^{\mathrm{b}} \in \mathbb{R}^{n_u^{\mathrm{b}}}$ are the dual variables.

There are two main ways of dealing with the CS conditions. One is to directly apply a specialized BNB algorithm, the other one is to reformulate the problem by introducing auxiliary binary variables and bounding primal constraints and dual variables[10]. Taking the latter approach, we introduce variables $\underline{z}_\tau^{\mathrm{b}} \in \{0,1\}^{n_u^{\mathrm{b}}}$, and express primal feasibility, dual feasibility and CS conditions jointly by

$$-M_{\underline{u}_\tau^{\mathrm{b}}} \times \underline{z}_\tau^{\mathrm{b}} \leq -u_\tau^{\mathrm{b}} + \underline{u}_\tau^{\mathrm{b}} + r_\tau^{\mathrm{b}\uparrow} - s_\tau^{\mathrm{b}\uparrow} \leq 0 \qquad \text{(22a)}$$

$$0 \leq \underline{\lambda}_\tau^{\mathrm{b}} \leq (1 - \underline{z}_\tau^{\mathrm{b}}) \times M_{\underline{\lambda}_{u,\tau}^{\mathrm{b}}} \qquad \text{(22b)}$$

where $M_{\underline{u}_\tau^{\mathrm{b}}}, M_{\underline{\lambda}_\tau^{\mathrm{b}}} \in \mathbb{R}_+^{n_u^{\mathrm{b}}}$ are sufficiently large constants. All other CS conditions can be handled analogously.

Thus, the bilevel problem (21) is equivalent to

$$\max_{\mathbf{r}^M, \mathbf{r}, \mathbf{u}, \mathbf{s}, \boldsymbol{\lambda}, \mathbf{z}} \tilde{\psi}(\mathbf{r}^M, \mathbf{r}, \mathbf{s}) \quad \text{s.t.} \quad \text{(19b), (19c), } \mathbf{(CS)}, \mathbf{(LS)} \qquad (23)$$

where **(CS)** denotes a set of constraints of the form (22), and **(LS)** represents the Lagragian stationarity constraint, which is an affine equality constraint.

In general, (23) is a Mixed-Integer optimization problem with linear constraints and nonlinear objective, the specific form of which depends on the kind of market penalty and the shape of the distribution of the LFC signal ω_θ. Note that (23) is highly structured. In particular, the constraints are completely decoupled. This suggests that the problem should be amenable to decomposition techniques that aim at separating the coupling term $\mathcal{P}^M$ in the objective.

An important feature of this formulation is that, since the KKT conditions of the buildings' problems appear explicitly as constraints, any feasible solution of (23) describes a situation in which all buildings behave optimally w.r.t. their regulation capacities. Hence it is not necessary to find the global optimum in order to ensure individual rationality.

B. The First-Best Contract

In general, the buildings' decisions impose externalities on the aggregator via their effect on the distributions of $\boldsymbol{\epsilon}^M$ and $\check{\mathbf{w}}^{\mathrm{b}}$. If the aggregator wants to avoid being penalized for deviations from schedule in the market and therefore insists that $\mathbf{s} = 0$, i.e., that all buildings are robustly providing their full capacities $\mathbf{r}^{\mathrm{b}}$, then the problem can be transformed into a single Linear Program.

Indeed, if $\mathbf{s}^{\mathrm{b}} = 0$ for all b, then $\mathcal{P}^{\mathrm{b}}(\mathbf{s}^{\mathrm{b}}) = 0$, $\check{\mathbf{w}}^{\mathrm{b}} = \mathbf{w}^{\mathrm{b}}$ for all b, and $\boldsymbol{\epsilon}^M = 0$. Hence we see from (20) that the only term in the objective that depends on building b's decision variables is $\varphi_{\text{nom}}^{\mathrm{b}}$. Moreover, it is clearly optimal for the aggregator to commit all available capacity in the market:

$$\left.\begin{aligned} r_t^{M\uparrow} &= \textstyle\sum_{\mathrm{b}} (q^{\mathrm{b}})^\top r_\tau^{\mathrm{b}\uparrow} \\ r_t^{M\downarrow} &= \textstyle\sum_{\mathrm{b}} (q^{\mathrm{b}})^\top r_\tau^{\mathrm{b}\downarrow} \end{aligned}\right\}, \quad \forall\, t \leq N_M,\ \tau \in \mathbb{T}_S(t) \qquad (24)$$

Writing $\mathcal{U}^{\mathrm{b}}(\mathbf{r}^{\mathrm{b}}) = \{\mathbf{u}^{\mathrm{b}} : (10), (12) \text{ hold with } \mathbf{s}^{\mathrm{b}} = 0\}$ and $\mathcal{R} = \{(\mathbf{r}^M, \mathbf{r}) : \mathbf{r}^M, \mathbf{r} \geq 0, (24) \text{ holds}\}$ we have

$$\begin{aligned} &\max_{(\mathbf{r}^M, \mathbf{r}) \in \mathcal{R}} \mathbb{E}\Big[\textstyle\sum_{t=1}^{N_M} \big(\rho_t^{\uparrow} r_t^{M\uparrow} + \rho_t^{\downarrow} r_t^{M\downarrow}\big) + \sum_{\mathrm{b}} \varphi_{\text{oo}}^{\mathrm{b}*} \\ &\qquad - \textstyle\sum_{\mathrm{b}} \big(\min_{\mathbf{u}^{\mathrm{b}} \in \mathcal{U}^{\mathrm{b}}(\mathbf{r}^{\mathrm{b}})} \varphi_{\text{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) + \Delta\varphi^{\mathrm{b}}(\mathbf{w}^{\mathrm{b}})\big)\Big] \\ &= \max_{\substack{(\mathbf{r}^M, \mathbf{r}) \in \mathcal{R} \\ \mathbf{u}^{\mathrm{b}} \in \mathcal{U}^{\mathrm{b}}(\mathbf{r}^{\mathrm{b}}), \forall \mathrm{b}}} \textstyle\sum_{t=1}^{N_M} \big(\mathbb{E}[\rho_t^{\uparrow}]\, r_t^{M\uparrow} + \mathbb{E}[\rho_t^{\downarrow}]\, r_t^{M\downarrow}\big) + \sum_{\mathrm{b}} \varphi_{\text{oo}}^{\mathrm{b}*} \\ &\qquad - \textstyle\sum_{\mathrm{b}} \big(\varphi_{\text{nom}}^{\mathrm{b}}(\mathbf{u}^{\mathrm{b}}) + \mathbb{E}\big[\Delta\varphi^{\mathrm{b}}(\mathbf{w}^{\mathrm{b}})\big]\big) \end{aligned} \qquad (25)$$

where

$$\mathbb{E}\big[\textstyle\sum_{\mathrm{b}} \Delta\varphi^{\mathrm{b}}(\mathbf{w}^{\mathrm{b}})\big] = \sum_{\mathrm{b}} \sum_{\tau=1}^{N_S} c_\tau^{\mathrm{b}} \sum_{\theta \in \mathbb{T}_R(\tau)} \frac{(q^{\mathrm{b}})^\top}{K} \zeta_\theta \big(r_{\tau\theta}^{\mathrm{b}\downarrow} - r_{\tau\theta}^{\mathrm{b}\uparrow}\big)$$

is linear in the decision variables. Finally $\mathcal{R}$ and $\mathcal{U}^{\mathrm{b}}$ are described by a set of linear constraints, so (25) is an LP. Economically speaking, there is no Moral Hazard, so the aggregator is able to implement what is known as the first-best contract [11] in contract theory (note that there is also no Adverse selection as there are no information asymmetries). Note however that the first-best contract in this setting need not be optimal for the aggregator. This is the case if the benefit from the buildings' reduced costs for adjusting their schedule under less strict robustness requirements outweighs the expected deviation penalties incurred in the market.

VI. Numerical Example

In this section we apply our methodology for the first-best contract to a simple numerical example with four buildings. For each building, we use a scaled version of a

[9]With the dual variables of the lower level problem now appearing as primal variables of the overall problem.

[10]This approach is often referred to as a "big-M" formulation [17].

simple one-dimensional model identified by applying semi-parametric regression techniques on a testbed on the UC Berkeley campus [18]. We emphasize that our simulations are primarily for illustration and for showing some important qualitative properties of the optimal contract. In order to get meaningful quantitative results on the economic feasibility of such a contract scheme, a much more thorough study is necessary. For our example we consider a scheduling horizon of $T_M = 24$ and scheduling intervals of length $T_S = 15$ min, so that $N_S = 96$ and $K' = 4$.

A. Model Parameters

1) Building Models: We generate four slightly perturbed versions of the model identified in [18], where the disturbance v is given by a weighted sum of outside temperature and heating load from occupants, equipment, and solar heating. For each model we assume a maximum power consumption typical for mid-to-large size commercial buildings, from which we determine the consumption factors q. For all models we have $C = E = 1$ and $D = F = 0$, with the remaining parameters given in Table I. Comfort constraints are chosen so as to have a tight temperature band during common working hours, and looser band outside common working hours. We assume that $\underline{\alpha}_{t,i} = \overline{\alpha}_{t,i} = 5\%$, $\Sigma_0 = 0.5$ and $\underline{u} = 0$ for all buildings.

b	A	B	x_0 [°C]	q [kWh]	$\overline{u}$	G	H	tariff
1	0.64	−2.64	23	43.75	0.5	0.2	0.15	A10_TOU
2	0.68	−2.55	22	62.5	0.75	0.25	0.2	E19_TOU
3	0.635	−2.7	24	31.25	0.65	0.15	0.1	A10_TOU
4	0.62	−2.65	22.5	50.0	0.5	0.1	0.15	A10

TABLE I: Building model parameters

2) Electricity Tariffs: Electricity costs $\mathbf{c}^{\mathrm{b}}$ were taken as the energy charges[11] of various PG&E commercial electricity tariffs (summer period) offered in California [19].

3) Capacity Prices and LFC signal: Figure 3 shows CAISO DAM regulation up and down prices for August 2012, with the respective averages (bold) for each hour across the whole month. We found that with these capacity prices

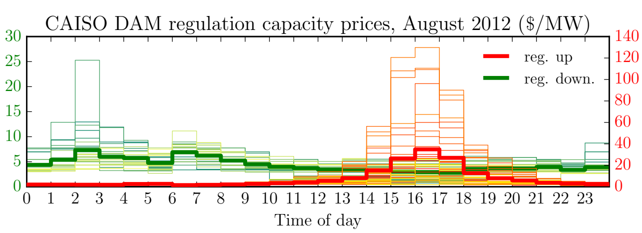

Fig. 3: CAISO DAM regulation capacity prices for August 2012

the optimal strategy for the aggregator is to not bid any capacity in the market. This is primarily because of the small thermal capacity and inefficient HVAC system of our simple model. As a result the increased cost for deviating from the cost-optimal power schedule outweighs the potential revenue in the capacity market. In order to still be able to illustrate some qualitative properties of the optimal contract, we have scaled the expected capacity prices $\rho^{\uparrow}$ and $\rho^{\downarrow}$ by a factor 20 and 15, respectively, for our simulations. Finally, we assume that the ISO's LFC signals ω_θ are i.i.d. uniformly on $[-1, 1]$, which implies that $\zeta = 0.25$.

B. Simulation Results

The optimal contract design problem for the DAM for the above parameters was solved in 0.09 seconds on a Laptop computer using Gurobi [20]. The aggregator's expected profit is \$20.7, the rewards R^{b} are \$28.4, \$24.1, \$9.0 and \$17.5 for buildings $1-4$. Hence for the simple building model used in this simulation, the aggregator must spend a large share of its revenue on incentivizing the buildings.

Figure 4 shows how the individual buildings contribute to the overall regulation capacities offered in the market. Observe how the contributions vary on the faster scheduling frequency on the building level to provide capacity on the slower market time scale in the most cost-effective way.

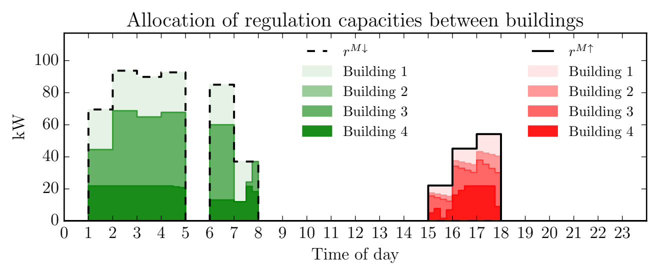

Fig. 4: Shares of the overall regulation capacities between buildings

Temperature evolutions and control inputs for Building 1 are shown in Figure 5, where u_{oo} and u^* are the optimal schedules of outside option and optimal contract, respectively. The regulation signal w has been sampled uniformly from within the contracted bounds. In order to be able to clearly observe the behavior of the optimal control, we have set $\tilde{v} = 0$ in Figure 5. The shaded areas in the lower

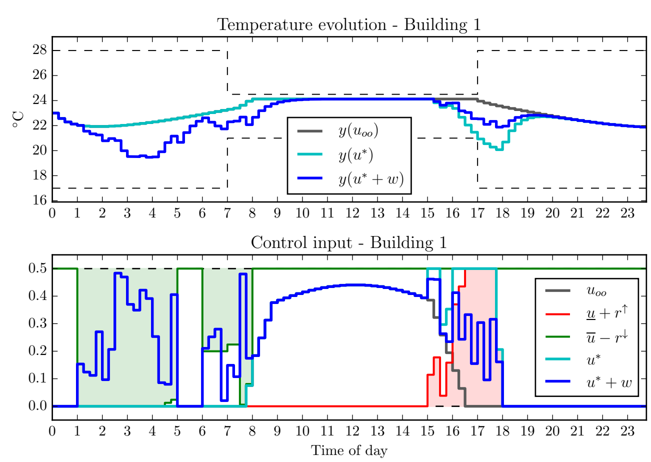

Fig. 5: Temperature and control for Building 1

plot represent the amounts by which the actuator constraints have been tightened in order to be able to ensure that the resulting nominal control schedule is robustly feasible

[11] In addition to the conventional energy charge, tariffs A10, A10_TOU and E19_TOU also have a monthly peak demand charge. For simplicity we assume here that this peak is not affected by the modified schedules. Prices given here are for the secondary distribution voltage level.

(observe that the control $u^* + w$ indeed never violates the actuator constraints).

Figure 6 shows the temperatures for Building 1 for 30 randomly sampled errors sequences $\tilde{v}$, again assuming a uniformly distributed LFC signal ω. The individual comfort constraints are violated in 0.57% of all time intervals.

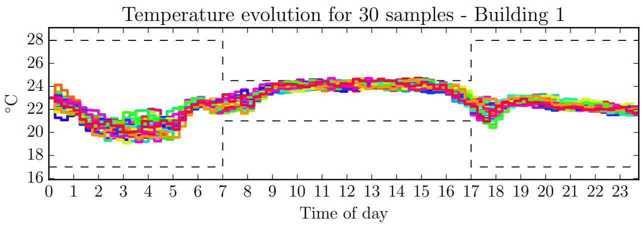

Fig. 6: Sample temperature trajectories for Building 1

Finally, Figure 7 shows the strong dependency of the total amount of contracted capacity over the scheduling horizon and the aggregator's expected profit on the parameter ζ, and hence on the shape of the distribution of ω. Noting that higher values of ζ correspond to a high variance of ω, this is not surprising. Indeed, when contracting regulation down capacity this would result in the cost difference $\Delta\varphi(w)$ likely being large, so in expectation the aggregator would need to pay a large amount to ensure that buildings are only charged according to nominal schedules. For small ζ the distribution of ω is concentrated around zero and this issue is less pronounced. The situation is mirrored for regulation up capacity; in this case large values of ζ are beneficial for the aggregator. Figure 7 also shows that the overall amount of

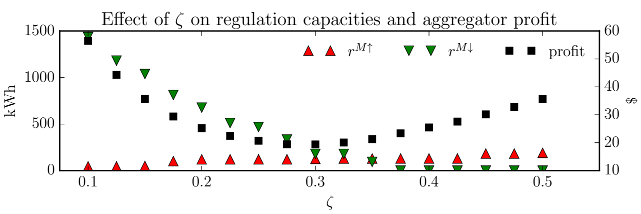

Fig. 7: Effect of the parameter ζ

capacity offered is much higher for regulation down. This is due to the fact that in many cases regulation down capacity can be offered without having to deviate from the optimal schedule of the outside option.

VII. Conclusion and Future Work

In this paper we formulated the optimal contract design problem an aggregator faces when using the inter-temporal flexibility of buildings' HVAC consumption in order to participate in the regulation capacity spot market. We showed how to formulate the resulting bilevel optimization problem as a mixed-integer problem, and how in the absence of Moral Hazard the first-best contract can be found by solving a Linear Program. Our methodology is quite general and can be used for a wide class of related problems characterized by the aggregation of agents with competing objectives. We illustrated the practicality of our approach by applying it to a simple numerical example.

We believe that the use of commercial buildings to provide frequency regulation capacity can be a viable option in certain market environments. With this work we provide a framework that can be used to analyze the feasibility of the business model of an aggregator. While our simulation results do provide some interesting qualitative insights into the contract, a more thorough quantitative study is necessary in order to assess the economic feasibility of this scheme in today's market structure. In the future, we hope to extend our work to more general building models, incorporate information asymmetries, and take into account new market instruments such as "Regulation Energy Management" [14].

Acknowledgement

We thank Clay Campaigne for interesting discussions.

References

[1] D. Halamay, T. Brekken, A. Simmons, and S. McArthur, "Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation," *Sustainable Energy, IEEE Transactions on*, vol. 2, no. 3, pp. 321–328, July 2011.

[2] Y. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational impacts of wind generation on california power systems," *Power Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 1039–1050, May 2009.

[3] D. Todd, M. Caufield, B. Helms, M. R. Starke, B. J. Kirby, and J. D. Kueck, "Providing Reliability Services through Demand Response: A Preliminary Evaluation of the Demand Response Capabilities of Alcoa Inc." 2009.

[4] 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*, June 2013, pp. 467–472.

[5] G. P. Henze, D. E. Kalz, S. Liu, and C. Felsmann, "Experimental analysis of model-based predictive optimal control for active and passive building thermal storage inventory," *HVAC&R Research*, vol. 11, no. 2, pp. 189–213, 2005.

[6] F. Oldewurtel, "Stochastic Model Predictive Control for Energy Efficient Building Climate Control," Ph.D. dissertation, ETH Zurich, 2011.

[7] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," *Control Systems Technology, IEEE Transactions on*, vol. 20, no. 3, pp. 796–803, May 2012.

[8] A. Aswani, N. Master, J. Taneja, A. Krioukov, D. Culler, and C. J. Tomlin, "Energy-Efficient Building HVAC Control Using Hybrid System LBMPC," in *Proceedings of the 4th IFAC Nonlinear Model Predictive Control Conference*, Leeuwenhorst, NL, 2012, pp. 496–501.

[9] E. Vrettos, F. Oldewurtel, F. Zhu, and G. Andersson, "Robust Provision of Frequency Reserves by Office Building Aggregations," in *Proceedings of the 19th IFAC World Congress*, August 2014.

[10] M. Maasoumy, C. Rosenberg, A. Sangiovanni-Vincentelli, and D. S. Callaway, "Model Predictive Control Approach to Online Computation of Demand-Side Flexibility of Commercial Buildings HVAC Systems for Supply Following," in *American Control Conference (ACC)*, Portland, OR, USA, 2014.

[11] J.-J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*. Princenton University Press, 2002.

[12] FERC, "Order no. 755: Frequency regulation compensation in the organized wholesale power markets," Oct 2011.

[13] E. Cutter, L. Alagappan, and S. Price, "Impact of market rules on energy storage economics," in *32nd IAEE Int. Conf.*, vol. 24, Jun 2009.

[14] California Independent System Operator Corporation, "Fifth Replacement FERC Electric Tariff," May 2014.

[15] B. S. Black and R. J. Pierce, Jr., "The Choice between Markets and Central Planning in Regulating the U. S. Electricity Industry," *Columbia Law Review*, vol. 93, no. 6, pp. 1339–1441, 1993.

[16] B. Colson, P. Marcotte, and G. Savard, "An overview of bilevel optimization," vol. 153, no. 1, pp. 235–256, 2007.

[17] H. P. Williams, *Logic and Integer Programming*, ser. Operations Research & Management Science. Springer, 2009.

[18] A. Aswani, N. Master, J. Taneja, V. Smith, A. Krioukov, D. Culler, and C. Tomlin, "Identifying models of hvac systems using semiparametric regression," in *American Control Conference (ACC)*, 2012.

[19] PG&E, "Commercial electricity tariffs," May 2014.

[20] Gurobi Optimization, "Gurobi Optimizer Reference Manual," 2014.