A FRAMEWORK FOR ENABLING THE UTILIZATION OF FLEXIBLE LOADS TO PROVIDE FREQUENCY REGULATION

BY

JUSTIN TODD HUGHES

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Associate Professor Alejandro D. Domínguez-García, Chair
Professor Thomas J. Overbye
Professor Kameshwar Poolla, University of California, Berkeley
Professor Peter W. Sauer

ABSTRACT

Frequency regulation is becoming increasingly important with deeper penetration of variable generation resources. This dissertation is about exploiting the flexibility of distributed energy resources (DERs) to provide frequency regulation though the framework of an aggregator, which groups DERs into simple, yet accurate, models; offers capability based on these models; and coordinates the DERs to provide the service.

Flexible loads have been proposed as a low-cost provider of frequency regulation. For example, the flexibility of loads with inherent thermal energy storage resides in their ability to vary their electricity consumption without compromising their end function. The aggregate flexibility of a collection of diverse residential air-conditioning loads has previously been shown to be well modeled as a virtual battery using first principles load models. In this dissertation, through developing control and parameter identification schemes, we show that the virtual battery can also model more complex loads such as buildings with large, multi-zone air conditioning systems.

The small power ratings and capacity constraints of individual flexible loads is an obstacle to their integration. Thus, we additionally propose a framework wherein an aggregator coordinates the response of many flexible loads and other types of distributed energy resources (DERs) (e.g., plug-in electric vehicles and microturbines) connected to electric power distribution networks to provide frequency regulation services. In this framework, the aggregator participates in the day-ahead and real-time ancillary services markets by submitting an offer to provide frequency regulation. If the offer is

accepted, the aggregator must coordinate the response of the DERs in order to provide the service. The DERs are compensated through bilateral contracts, the terms of which are negotiated in advance.

The DER coordination problem the aggregator is faced with is cast as an optimal control problem, and we propose a bilayer framework to obtain a sub-optimal solution. In the first layer, we utilize model-predictive control techniques driven by regulation signal forecasts and parameter estimates to obtain a reference control action for the DERs. A second control layer provides closed-loop regulation around the reference computed by the top layer, which minimizes the error that arises due to forecast error, plant-model mismatch, and the slower speed of the optimal control.

To Nina

ACKNOWLEDGMENTS

This dissertation would not have been possible without the assistance, guidance, and support provided by many people over the years.

First and foremost, my adviser and doctoral committee chair Alejandro D. Domínguez-García has provided abundant discussion, feedback, and mentorship for many years. The other members of the committee, Thomas J. Overbye, Kameshwar Poolla, and Peter W. Sauer, have also been instrumental in my education. As I have grown, they have been my teachers, advisers, co-authors, and colleagues. I offer them my sincere gratitude.

Finally, I recognize my friends and family. The confidence and enthusiasm of my peers and the encouragement of my family have been essential motivation while completing this program. A special thanks to my wife, who has shown patience and unwavering love at every step of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE	Area Control Error
AGC	Automatic Generation Control
DER	Distributed Energy Resource
DES	Distributed Energy Storage
HVAC	Heating, Ventilation, and Air Conditioning
MPC	Model Predictive Control
PEV	Plug-in Electric Vehicle
PI	Proportional–Integral
PV	Solar Photovoltaic
RTO	Regional Transmission Operator
TCL	Thermostatically Controlled Load
VAV	Variable Air Volume

CHAPTER 1

INTRODUCTION

In this chapter, we provide context for the remainder of the dissertation. We begin by discussing the background of the present changes underway in electric power systems and the benefits that can be gained from utilizing flexible loads. Related work will be discussed, and the main contributions of this thesis will be explained. Finally, we will summarize the contents of the following chapters.

1.1 Background

Electric power systems are undergoing dramatic transformations in structure and functionality in response to the US DoE *Smart Grid* vision [1], and its European counterpart *Electricity Networks of the Future* [2]. These transformations are enabled by (i) the integration of new renewable generation resources (e.g., solar photovoltaics (PV) installations) and energy-storage capable loads (e.g., plug-in electric vehicles (PEVs)), and (ii) the increased reliance on advanced communications, which enables the active control of other types of energy-storage capable loads such as thermostatically controlled loads (TCLs) (e.g., air conditioners, heat pumps, water heaters, and refrigerators).

These generation and controllable/storage-capable resources are commonly referred to as distributed energy resources (DERs). If properly coordinated, DERs provide new opportunities and added flexibility in the procurement of ancillary services such as frequency regulation and load following. For

instance, PEVs and TCLs can be utilized to provide active power for up and down regulation services, e.g., energy peak-shaving during peak hours and load-leveling at night [3–5].

Frequency regulation is a necessary ancillary service because instantaneous imbalances between power supply and demand cause the frequency of a power system to deviate from its nominal value, e.g., 60 Hz in the USA. Many devices rely on a fixed frequency supply for proper operation, so this deviation must be minimized. This imbalance is used alongside scheduled and actual interchange flows in the calculation of the so-called area control error (ACE). This value is used in the automatic generation control (AGC) system to automatically coordinate the response of generators so as to minimize the ACE.

It has been hypothesized for decades that demand-side resources could supplement or even replace the regulation service provided by conventional generators [6]. The massive integration of renewable resources and their often unpredictable and variable generation patterns are creating an increasing need for frequency regulation [7]; this need has renewed interest in demand-side resources. The increasing prevalence of technologies such as advanced controls, real-time metering, and continuously variable power electronic devices are enabling this radical shift in the way frequency regulation is performed in the bulk power system. Also, advances in power electronics allow more precise control over how much power loads consume, e.g., through cheaper and more reliable variable frequency drives. Many loads can modify their consumption more quickly than conventional generators, which is an advantage when providing this service. Together, these transformative changes are creating a perfect environment for the participation of load side resources in the procurement of ancillary services. In order to enable this added functionality that these new technologies may provide, it is necessary to develop an appropriate structure and control mechanism.

We are interested in the ability of flexible loads and other types of dis-

tributed energy resources (DERs) to provide frequency regulation. These resources could include microturbines, renewable generation resources (e.g., rooftop solar photovoltaics (PV) installations), chemical energy-storage capable loads (e.g., plug-in electric vehicles (PEVs)), and thermal energy-storage capable loads such as thermostatically-controlled loads (TCLs) (e.g., air conditioners, heat pumps, water heaters, and refrigerators).

It is important to note the essential role of an aggregator in enabling the utilization of DERs for frequency regulation. Most of these DERs will not be able to participate directly in a frequency regulation market because they may not meet minimum regulation capability and performance standards because they are smaller and less concentrated than traditional resources. For example, in the PJM interconnection, participation in such a market requires a minimum of 0.1 MW of regulation [8]; this exceeds the capability of most DERs. By aggregating the frequency regulation capacity that these DERs can collectively provide, the aggregator will be able meet the requirements of the frequency regulation market. If an aggregator is subsequently called upon to provide frequency regulation, the regional transmission operator (RTO) will issue a frequency regulation signal—updated every 2 s—and the aggregator needs to be able to track it by coordinating the collective response of the DERs.

The cost incurred by the aggregator includes the payments to the DERs, and the penalty that the aggregator needs to pay if it is not able to follow the frequency regulation signal set by the RTO. The profit of the aggregator is the difference between: the revenue obtained from selling the frequency regulation service in the real-time market, and the costs incurred by the payments to the DERs and the penalties for not being able to follow the frequency regulation signal. Thus, since the revenue is determined by the market clearing price, which is fixed before service delivery, the aggregator maximizes its profit by minimizing the total payments to the DERs and the penalties incurred for not being able to follow the frequency regulation signal.

1.2 Contributions of Dissertation

1.2.1 Load Flexibility Modeling and Control

In order for a load to provide frequency regulation, it must have a controller which can utilize its flexibility to follow a regulation signal. To this end, we propose one such controller, the sole function of which is to provide maximum regulation capability while respecting constraints, e.g., occupant comfort and equipment ratings. We note that there may be additional objectives when designing the control system, e.g., minimization of total energy use [9]. A natural extension of our proposed controller would balance minimizing energy costs while maximizing income from the regulation market so as to minimize total costs.

Beyond designing the controller described above, we are interested in quantifying the flexibility of loads. Flexibility is defined as the ability to tolerate perturbations from the baseline power, which is the power the building would have consumed were it not providing the regulation service. The details of flexibility and baseline power are calculated based on the rules of the relevant electricity markets. Our tool for quantifying flexibility of loads is the virtual battery, which is a first-order linear time-varying (LTV) model with constraints on the value that the state and the input takes. This model can easily be used by system operators procuring capacity in the ancillary services market because it is agnostic to the details of the underlying resource. The authors of [10] demonstrated the power of this model to simply and succinctly describe the aggregate flexibility of a large number of possibly heterogeneous TCLs. In this work, we show that the virtual battery model has the power to capture the flexibility of more complex loads, and we provide a method to identify its parameters.

Our proposed method for identifying the parameters of the virtual battery model requires a detailed model of the load and its control system. We use

this model to perform software-based tests to determine equivalent battery parameters including charge/discharge rate limits, capacity, dissipation, and initial charge. These tests stress the system by issuing carefully selected commands to the controller. By noting which commands cause the controller to fail to converge at what time, our method can deduce how the real system will react to certain inputs; this is useful to determine how much frequency regulation can be offered in the market.

1.2.2 Coordination of Distributed Energy Resources

Once the capability of a flexible load has been quantified, it may be necessary to coordinate it with other DERs in order to realize the full potential of these resources to provide frequency regulation; thus, in this dissertation, we propose a framework that enables such coordination.

In this framework, the aggregator participates in the day-ahead or real-time ancillary services market by submitting an offer to provide frequency regulation services. The aggregator does not own generation or storage assets. The DERs are compensated for participation through negotiated bilateral contracts. This compensation is agreed to ex-ante and may vary among the participating DERs. For example, the aggregator is likely to take into account the fuel cost of a microturbine when setting its monetary compensation; whereas for a TCL, the aggregator might consider and inconvenience costs, e.g., PEV less than fully charged or water heater being too cold. If the aggregator's offer is accepted, it must coordinate the response of a set of DERs. We propose a bilayer control method.

In the top control layer, the aggregator uses model-predictive control (MPC) techniques to minimize the costs incurred when providing regulation during a fixed service interval at time-scales consistent with existing real-time regulation markets.[1] The constraints in the optimal control problem include the

[1]In a real-time market, the duration of the period over which the aggregator offers to

inherent dynamics associated with the DERs power delivery process, their upper and lower power output limits, their upper and lower energy limits, as well as constraints on their ramping rates, i.e., the rate at which they can change their power output.

In the bottom layer, a closed-loop control, similar to that implemented in traditional automatic generation control (AGC) systems, regulates around the MPC solution calculated by the top layer. Separating fast and slow time scales when designing a controller is a well-established method for solving problems involving dynamics of differing speeds [13]. The authors of [14] have proposed replacing traditional AGC with MPC, but, to our knowledge, this dissertation is the first work to propose the aforementioned bilayered approach. This approach benefits from both the speed of traditional AGC-like control and the foresight of MPC, giving results with impressive accuracy, while the amounts of computational power necessary to obtain such accurate solutions are limited.

1.3 Related Work

1.3.1 Load Flexibility Modeling and Control

Previously, it has been shown that residential HVAC systems can be aggregated and used to provide frequency regulation by utilizing their thermal energy capacity and flexible energy consumption. The virtual battery model—a first-order linear dynamical model—was analytically shown to be an accurate and simple model to capture the flexibility of residential HVAC systems.

In this context, many recent papers have focused on the use of small residential thermostatically controlled loads (TCLs) with the ability to store

provide the regulation service is typically five to ten minutes, and the offer needs to be submitted in advance, e.g., two periods before the actual service is to be provided if the offer is accepted [11, 12].

thermal energy [10, 15–18]. Commercial buildings are another load with the ability to store thermal energy and with the flexibility to provide frequency regulation. The large power consumption and thermal mass of these buildings, combined with the ability to continuously vary power consumption, could eliminate the need to aggregate them with other similar loads; this simplifies the communication and control with remote devices. The authors of [19–25] focus on using commercial buildings to provide regulation, possibly in select frequency ranges of the regulation signal, or other ancillary services, but do not adopt the virtual battery model.

There is a small amount of literature in which techniques related to ours have been suggested. For example, in [26], charge rate limits and capacity parameters are identified for a collection of TCLs; our proposed method improves upon this technique in two ways. First, we identify the parameters in terms of a more accurate model that includes dissipation. Second, we do not rely on the ability to command a load to consume a maximum or minimum possible power. Although this may be simple for a collection of TCLs, it is not clear that this approach can easily be utilized in more complicated systems.

1.3.2 Coordination of Distributed Energy Resources

There have been many recent papers that exploit the salient features of DERs in different ways, e.g., [27] uses receding horizon model predictive control of flexible loads for energy arbitrage. The framework proposed in [28] utilizes stochastic dynamic programming to arrive at a solution to a Markov process in which flexible loads respond to broadcasted prices. A number of papers have looked at using air conditioners for spinning reserves or demand response [29–31]. We will focus on papers that directly control DERs to provide frequency regulation services. For example, in [3], the authors use mean field games to control a very large population of homogeneous PEVs.

In [20] and [18], the authors propose to use Markovian Decision Processes to control heating, ventilation and air conditioning (HVAC) systems in commercial buildings and deferrable loads such as pool pumps, respectively. Finally, in [5], the authors propose an allocation strategy for TCLs based on priority stacks. Many of these earlier works focus on a specific class of DERs, while we are interested in exploiting the distinct capabilities of diverse classes of DERs.

The formulation of the problem to be solved in the top layer of our architecture is similar to [32], in which the authors use MPC to coordinate frequency response of diverse types of resources. Here, we build on the framework introduced in that paper and make valuable contributions in several directions. First, we generalize the model of energy-limited DERs and introduce a cost function that explicitly takes into account the aggregator payments to the DERs, which allows us to introduce several related economic problems. Second, we formalize the decision-making that the aggregator is faced with when coordinating the DER responses as a stochastic optimal control problem, and we show that in the perfect information case (i.e., the regulation signal is known a priori), the optimal control problem reduces to a linear program. Additionally, we go beyond the assumption of a perfect forecast and a two step prediction horizon in [32], and investigate longer time-horizon schemes that incorporate a forecasting technique the aggregator can use to handle imperfect information. Our architecture adds a second layer which improves tracking of fast moving regulation signals and decreases the computational complexity involved in finding a solution. Finally, we provide empirical evidence via synthetic simulations that use a mix of real and simulated data to show the effectiveness of these schemes and conduct parameter sensitivity studies. An aggregator can use similar studies to optimize the control and determine appropriate DER portfolios to profitably offer regulation services.

The authors of [33] and [34] propose an ambitious multi-level control framework for coordinating aggregations of commercial buildings. The nicely inte-

grated result considers the interaction between the aggregator and the system operator/reserves market as well as the internal control of a single type of DER. In this framework, the reserve allocations for individual loads are calculated by the aggregator daily, while distributed controllers handle changes on the order of minutes and seconds. In contrast, we propose a framework in which the aggregator runs a fast, centralized control to coordinate the response. In our framework, the distributed controllers are not required to use predictive techniques for regulation. We believe this method is advantageous as the centralized controller is able to fully consider the entire pool of capability in real time, utilizing the strength of each heterogeneous DER as appropriate.

The forward-looking nature of the centralized top layer formulation is especially valuable when utilizing energy-constrained resources. A classic controller may push the resource to its limits for short term performance, whereas predictive control has the option of saving some capacity for an even more critical time period in the future. Because of this benefit, other authors have also recently taken interest in using MPC for frequency regulation [35], [36], [37]. Our proposal differentiates itself by being a two-level control strategy, which utilizes AGC-like control for fast time scales and MPC for slow time scales. Together, the two levels realize benefits from each of the control techniques.

1.4 Dissertation Organization

In Chapter 2, we discuss virtual battery models and how they can be created for various types of flexible loads. The initial research and the synthetic case study on input types from this chapter have been published in [38]. Another paper with improvements to the formulation and procedure, as well as a case study of a realistic building, has been published in [39].

In Chapter 3, we propose an aggregator framework that shows how virtual

battery models can be utilized alongside models of other DERs to provide frequency regulation.

In Chapter 4, we propose an aggregator framework that shows how virtual battery models can be utilized alongside models of other DERs to provide frequency regulation. Content from Chapters 3 and 4 has been prepared for upcoming publication.

Chapter 5 summarizes the main results of the dissertation and suggests directions for future work.

CHAPTER 2

LOAD FLEXIBILITY MODELING AND CONTROL

In this chapter, we introduce a general model of a flexible load and formulate our proposed controller using a first-order model as a motivating example. For clarity, we will refer to the load as a commercial building with an associated HVAC system and controller; however, application to other systems is straightforward (e.g., by replacing the state vector with tank levels in a pumping application or battery charge in an PEV application). We formulate the problem of identifying the parameters of the virtual battery model describing the ability of a load to provide frequency regulation and introduce the proposed algorithm for identifying the parameters of the virtual battery model. The algorithms are used to study effects of different test inputs using an artificial tests system. We also develop a model of an airport terminal building and present the results of the procedure used on this model.

2.1 Building Thermal Dynamics and Control

In this section, we introduce a generalized flexible commercial building model and a controller that enables it to provide frequency regulation. Throughout this section, we will use a first-order model to illustrate the ideas presented.

2.1.1 Flexible Load Model

We first introduce a model to describe the thermal dynamics of a building. Let T denote the vector of temperatures in different zones of the building, let

s denote the vector of control inputs (e.g., HVAC fan speed, air flow control damper positions, on/off control for vestibule heaters, etc.), and let $w(t)$ denote a vector of exogenous variables (e.g., outside ambient temperature, solar thermal load, internal thermal load from occupants and equipment). Then, the dynamics of the system can be described by a nonlinear state space model of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t) = h_1(T(t), s(t), w(t)), \tag{2.1}$$

where $h_1(\cdot, \cdot, \cdot)$ describes the change in zone temperature as a function of the current state and control variables.

Additional variables relate the dynamics in (2.1) to the electric power consumed by the building HVAC system, which we denote $P(t)$; specifically, we can write

$$P(t) = h_2(T(t), s(t), w(t)). \tag{2.2}$$

Finally, we need to consider the constraints which arise from acceptable occupant comfort and those that arise from the ratings of the equipment:

$$h_3(T(t), s(t)) \leq 0. \tag{2.3}$$

Next, we illustrate the above concepts using a first-order model as an example.

Example 1 (First-Order System): Consider a single room building with heat transfer from outside ambient as well as interior thermal loads. Assume that the building is cooled by a variable speed air conditioner which recycles some fraction of the interior air. Let $T(t)$ be the temperature of the room and the control input $s(t)$ be equal to the mass flow rate of the conditioned air $\dot{m}$; then, the expression for (2.1) in this case is

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t) = \frac{1}{m}\left(\dot{q} - rT(t) + c_p\dot{m}(t)(T_c - T(t))\right), \tag{2.4}$$

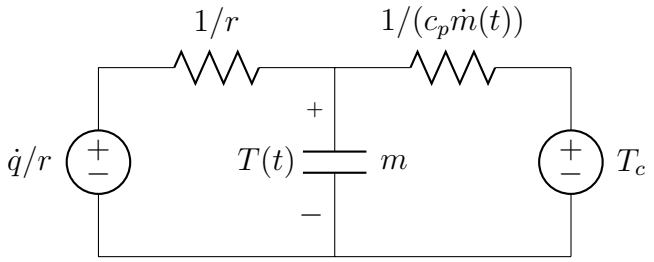

Figure 2.1: Equivalent first-order model circuit diagram.

where m is the thermal mass of the room, r is the thermal conductance, c_p is the specific heat capacity of air, and T_c is the temperature of the conditioned air. The external variable w equals $\dot{q}$, which is the thermal load that includes effects from the ambient temperature as well as objects inside the room. An equivalent circuit diagram of the thermal dynamics of the building is shown in Fig. 2.1.

The total power consumed, $P(t)$, is the sum of the fan power, which is assumed to be quadratic in mass flow rate, and the cooling power, which is the power required to maintain the cooling coil temperature as it cools the passing air; thus the expression for (2.2) in this case is

$$P(t) = \kappa_f \dot{m}(t)^2 + \frac{c_p}{\eta_c} \dot{m}(t) \left((1 - d_r) T_{oa} + d_r T - T_c \right), \tag{2.5}$$

where κ_f is a function of the properties of the fan, η_c is the cooling system coefficient of performance, d_r is the fraction of return air that is recycled, and T_{oa} is the outside ambient temperature.

We enforce limits on interior temperature to ensure the comfort of occupants and limits on mass flow rate to keep the AC system functioning correctly; in this case, the constraints in (2.3) are as follows: $\underline{T} \leq T(t) \leq \overline{T}$, $\underline{\dot{m}} \leq \dot{m}(t) \leq \overline{\dot{m}}$. This completes the example.

2.1.2 Baseline Power

We define the regulation power at time t as the difference between the actual power consumed by the load, $P(t)$, and some baseline power, denoted by P^0, which is the total electric power consumed by the system were it not providing the regulation services. In practice, it may be challenging to calculate the baseline power in a way that is fair and auditable.

In this work, we consider P^0 to be the value obtained from the steady state solution of (2.1), with the zone temperatures set to some nominal value, T^{m}. In subsequent developments, we will assume this solution satisfies (2.3); thus, by setting the left hand side of (2.1) to zero, we can implicitly solve for the baseline control input, s^0. If the problem is underdetermined, some secondary selection criterion would need to be utilized, e.g., energy efficiency. Using s^0, we can calculate the baseline power from (2.2), which results in the baseline power P^0.

Example 2 (First-Order System): Consider the same system as in Example 1; then, solving for the steady state of (2.4) gives us the baseline mass flow rate:

$$\dot{m}^0 = \frac{rT^m - \dot{q}}{c_p(T_c - T^m)},$$

from which we can calculate the baseline power using (2.5); which results in

$$P^0 = \kappa_f(\dot{m}^0)^2 + \frac{c_p}{\eta_c}\dot{m}^0\left((1-d_r)\,T_{oa} + d_rT^m - T_c\right).$$

This completes the example.

2.1.3 Controller Design

The controller must enforce a constraint so the load consumes the commanded power output, $P^*(t)$, which is equal to the desired regulation plus the baseline power. The output of the controller is an optimal control, $s^*(t)$,

which causes the HVAC system to consume the requested amount of power while also respecting the limits in (2.3). In the most general form, we can write

$$\begin{aligned} s^*(t) = \arg\min_{s(t)} \quad & h_4(T(t), s(t)) \\ \text{subject to} \quad & h_3(T(t), s(t)) \leq 0 \\ & |P^*(t) - P(t)| \leq \delta, \end{aligned} \tag{2.6}$$

where h_4 is an objective function that is used to weigh different possibilities if there are multiple solutions that satisfy the hard constraints and δ is some small value considered acceptable by the system operator.

The controller must be designed with the structure of the system and priorities of the owner in mind. For example, on a hot day the building HVAC system may be able to increase the amount of outside air being brought in (decreasing the recirculation fraction).[1] This would lower efficiency, causing extra power to be used without decreasing the temperature. This inefficient use of energy would result in a battery model with large capacity, but this would have to be balanced against wasting cool air on a hot day. An ideal objective function will have a unique solution that balances the capacity of the battery model, temperature variations, equipment cycling, and losses. The controller should also have defined behavior if it is unable to meet the commanded power.

Example 3 (First-Order System): For the first-order system in Examples 1 and 2, there is only one solution that satisfies the constraints in (2.6) if $\delta = 0$, so h_4 in (2.6) can be chosen arbitrarily. In fact, we can find the solution analytically, so an optimization procedure is not required. To find this solution, note that at any given time we can find the mass flow rate that

[1] This is known as economizing mode and is desirable under certain combinations of ambient and zone temperatures.

will cause a power consumption P^* by using the quadratic equation

$$P^*(t) = \kappa_f(\dot{m}^*(t))^2 + \frac{c_p}{\eta_c}\dot{m}^*(t)\left((1-d_r)\,T_{oa} + d_r T(t) - T_c\right)$$

and solving for $\dot{m}^*$. There will be only one meaningful solution assuming realistic parameters, which is given by

$$\begin{aligned}\dot{m}^*(t) = &\frac{c_p(-d_r T(t) + (d_r-1)T_{oa} + T_c)}{2\eta_c\kappa_f} \\ &+ \frac{\sqrt{c_p^2(-d_r T(t) + (d_r-1)T_{oa} + T_c)^2 + 4\eta_c^2\kappa_f P^*(t)}}{2\eta_c\kappa_f}.\end{aligned} \tag{2.7}$$

Plugging (2.7) into (2.4) gives us a an initial value problem which can only be solved numerically unless $d_r = 0$; this concludes the example.

2.2 Virtual Battery Parameter Estimation

We first define a procedure which incorporates software-based stress tests to determine which regulation signals the building/HVAC system is capable of following. Then, we introduce a reduced-order model—the virtual battery model—that we will use to compactly represent the flexibility of the building/HVAC system outfitted with the controller proposed in Section 2.1.3. Using these, we formulate a criterion for the quality of the virtual battery model for describing the behavior of the full nonlinear system model. The problem is then to find the parameters that optimize this criterion.

2.2.1 Violation Time Function

We define a scalar input $u_i(t) = P_i^*(t) - P^0$ which is the desired deviation from the baseline power consumption profile. Assume we are free to choose $u_i(t)$, but have no knowledge of the structure or parameters of the underlying system in (2.1)–(2.3), and (2.6), and cannot make measurements beyond

checking whether or not constraints have been violated. Then, for some input $u_i(t)$, define a function $f(u_i(t), \overline{\tau})$ such that if there is a constraint violation at or before $t = \overline{\tau}$, it takes the value of the time τ_i at which a constraint was violated, otherwise it takes the value ∞; in other words:

$$f(u_i(t), \overline{\tau}) = \begin{cases} \infty & \text{if } \exists \text{ solution to (2.6) } \forall\ t \leq \overline{\tau} \\ \tau_i & \text{otherwise,} \end{cases} \tag{2.8}$$

where $\tau_i = \min t$ such that there is no solution to (2.6).

2.2.2 Virtual Battery Model

The virtual battery been shown to accurately model the flexibility of certain buildings [10], [38]. Even in the case of a nonlinear, high-order building model, heat transfer is governed by Fourier's law. With the appropriate control, many buildings can act as a battery-type first-order model, the dynamics of which is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = -ax(t) - u_i(t), \tag{2.9}$$

where $x(t) \in \mathbb{R}$, $u_i(t) \in \mathbb{R}$, $a > 0$ is a constant, and $x(0) = x_0$. There are upper and lower bounds constraining $x(t)$ and $u_i(t)$, i.e.,

$$-C \leq x(t) \leq C, \quad -\underline{n} \leq u_i(t) \leq \overline{n}, \tag{2.10}$$

where $C > 0$, $\underline{n} > 0$, $\overline{n} > 0$ are constant. If a constraint is violated, the behavior is undefined. We group the parameters into a vector $\boldsymbol{\phi} = [a, C, \overline{n}, \underline{n}, x_0]^T$ to make the notation more compact.

For some input $u_i(t)$, define a function $b(u_i(t), \boldsymbol{\phi}, \overline{\tau})$ such that if a constraint in (2.10) is violated by (2.9) before time $\overline{\tau}$, it takes the value of the time τ_i at which a constraint was violated; otherwise it takes the value ∞. Thus,

similar to (2.8), we have that

$$b(u_i(t), \boldsymbol{\phi}, \overline{\tau}) = \begin{cases} \infty & \text{if (2.9) – (2.10) hold } \forall\ t \leq \overline{\tau} \\ \tau_i & \text{otherwise,} \end{cases}$$

where $\tau_i = \min t$ such that (2.9) and (2.10) are not satisfied.

2.2.3 Problem Statement

We want to find the values of the virtual battery model parameters in (2.9) and (2.10) which will allow us to predict the behavior of the dynamic model in (2.1)–(2.3) and (2.6). The quality of the fit is inversely related to the difference between the violation times predicted by the nonlinear system and those predicted by the virtual battery model. If the fit is not exact, we wish to err on the side of caution by constraining the battery model to predict a violation time smaller than that incurred by the nonlinear model. This ensures that if an input does not cause a violation on the identified battery model, it will not cause a violation in the nonlinear model. Mathematically, the problem can be formulated as finding a set of parameters $\boldsymbol{\phi}^*$ such that

$$\begin{aligned} \boldsymbol{\phi}^* = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \max_{u_i(t)} \quad & |b(u_i(t), \boldsymbol{\phi}, \overline{\tau}) - f(u_i(t), \overline{\tau})| \\ \text{subject to} \quad & b(u_i(t), \boldsymbol{\phi}, \overline{\tau}) \leq f(u_i(t), \overline{\tau}). \end{aligned} \tag{2.11}$$

2.3 Estimation Algorithms

In this section, we propose algorithms for identifying the parameters of the virtual battery model capturing the flexibility of the HVAC system of a commercial building as described by the dynamic model in (2.1)–(2.3) and (2.6). The basic structure of the proposed identification setup is shown in Fig. 2.2, where P^* is the commanded power, and s is a vector of control signals.

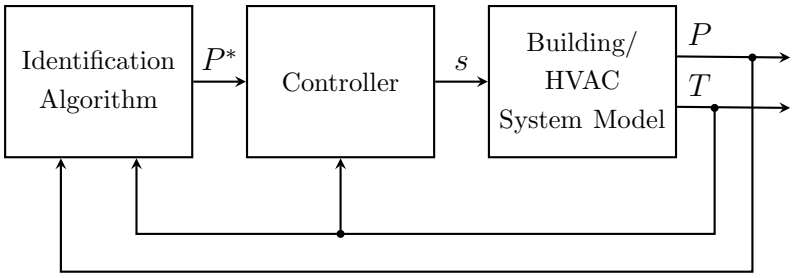

Figure 2.2: System identification setup.

Feedback includes the state vector T and the actual power P consumed by the building HVAC system.

2.3.1 Estimation of Rate Limits

The first step of the proposed procedure is to identify the rate limits $\overline{n}$ and $\underline{n}$. If the initial state is within its temperature bounds and we apply an input which causes a constraint to be immediately violated (i.e., $f(u_i(0), 0) = \infty$), we know it was due to the input constraints; this is because some finite time is required for an input to affect the value of the state.

To begin, we will assume w and $\boldsymbol{\phi}$ are constant. To be conservative, we must calculate $\underline{n}$ under the worst case scenario, which is when the state values are at their upper limit $\overline{T}$; likewise, $\overline{n}$ would need to use $\underline{T}$. The drawback of this conservative method is that the actual limits will be underestimated during normal operating conditions.

We know that $\overline{n} > 0$ ($\underline{n} > 0$), but we do not know an upper bound on these values. We therefore perform a one-sided binary search to find such an upper bound. Once we have an upper bound, we use it together with the greatest known lower bound in a binary search procedure to find $\overline{n}$ ($\underline{n}$) to arbitrary precision ϵ. The details of this procedure for estimating $\overline{n}$ and $\underline{n}$ are laid out in Algorithm 1. The procedure for $\underline{n}$ is similar, but with $f(\cdot, 0)$ replaced with $g(\cdot, 0) = f(-\cdot, 0)$.

Algorithm 1 Rate limit search algorithm

```
1: procedure SEARCH(ε)                                  ▷ ε > 0
2:     α ← 0                                            ▷ Lower bound
3:     β ← 1                                            ▷ Upper bound
4:     while f(β, 0) = ∞ do                             ▷ No instant violation
5:         α ← β
6:         β ← 2 · β
7:     end while
8:     while (β − α) > ε do
9:         γ ← (α + β)/2                                ▷ New bound to be tested
10:        if f(γ, 0) = ∞ then
11:            α ← γ
12:        else
13:            β ← γ
14:        end if
15:    end while
16:    return α                                         ▷ Less than ε below true value
17: end procedure
```

2.3.2 Estimation of Capacity and Dissipation Constant

If we respect the identified rate limits, we can guarantee that any constraint violation error is due to the capacity limit. In general, dissipation cannot be neglected when solving for the capacity limit, such that the two must be solved for simultaneously.

Because we are trying to fit the behavior of a linear model to that of a nonlinear one, we must look for a *sufficient* solution instead of an *exact* one. We say a solution is sufficient in the sense that verifying that an input does not cause any violations in the virtual battery model is sufficient to guarantee that the same input will not cause violations in the full nonlinear model. We are unable to mathematically prove that a solution is sufficient, i.e., the constraint in (2.11) will hold for all $u_i(t)$, because of the nonlinearity in (2.2); instead, we propose a heuristic. In Section 2.5 we will provide empirical evidence for the effectiveness of this heuristic procedure.

We start the procedure by picking a large value of $\overline{\tau}$ and generating $u_1(t), \ldots, u_n(t)$ such that violation times will be finite. Constant functions

with different (log-distributed) magnitudes are a natural first choice, but others can be used. Then, we can construct a vector of violation times for the nonlinear model: $F = [f(u_1(t), \overline{\tau}), \ldots, f(u_n(t), \overline{\tau})]^T$. We also construct a vector of violation times for the linear model: $B(a, C, x_0) = [b(u_1(t), \boldsymbol{\phi}, \overline{\tau}), \ldots, b(u_n(t), \boldsymbol{\phi}, \overline{\tau})]^T$. The values of the components of the difference vector $B - F$ can vary by many orders of magnitude; in order to prevent the larger values from dominating the optimization, we define a new vector $G(a, C, x_0) = [\log|b(u_1(t), \boldsymbol{\phi}, \overline{\tau}) - f(u_1(t), \overline{\tau})|, \ldots, \log|b(u_n(t), \boldsymbol{\phi}, \overline{\tau}) - f(u_n(t), \overline{\tau})|]^T$, which is the natural logarithm of each component of $B - F$.

We wish to find the values of a, C, and x_0 so the two models have similar violation times. If the difference in times cannot be reduced to zero, we require that the violation time predicted by the virtual battery model be less than that of the nonlinear model. We can cast this problem as a constrained nonlinear least squares problem of the form

$$\begin{aligned} \underset{a,C,x_0}{\text{minimize}} \quad & \|G(a, C, x_0)\|_2 \\ \text{subject to} \quad & B(a, C, x_0) \leq F. \end{aligned} \tag{2.12}$$

If $u_i(t) = k$ is constant and $x(0) = x_0$, we can write an analytic expression for b. The solution to (2.9) is given by

$$x(t) = \left(x_0 + \frac{k}{a}\right) e^{-at} - \frac{k}{a}.$$

Then, by setting $x(\tau_i) = -C$ and solving for τ_i, we obtain

$$\tau_i = b(k, \boldsymbol{\phi}, \overline{\tau}) = \frac{1}{a} \log\left(\frac{-ax_0 - k}{aC - k}\right),$$

if $k > aC$. However, in general we cannot analytically solve for the first τ_i at which the equality given by

$$x(\tau_i) = x_0 + \int_0^{\tau_i} e^{-a(\tau_i - \tau)} u_i(\tau)\, d\tau = \pm C$$

holds. If this is the case, we can use a numerical method to evaluate the function b, just like we must use for f. Given x_0 and $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = -ax(t) - u_i(t)$, we can approximate the system trajectory, $x(t)$, $t > 0$, until $|x(\tau_i)| = C$, and record τ_i. This change causes the computation of B to take orders of magnitude longer, but we found that solving (2.12) is still computationally tractable. Because F needs to be calculated only once, we expect the procedure to scale reasonably well to large systems.

2.4 Case Study: Artificial Test System

We begin this section by examining the behavior of the controller using an artificial test system. We will then test the performance of the identification procedure described in Section 2.3 on the aforementioned system.

2.4.1 Building Model

For this study we adopt a variable air volume (VAV) building HVAC model. Figure 2.3 illustrates such a system. The building comprises different zones, which are assumed to have first order thermal dynamics. The air handling unit (AHU) includes the supply fan and cooling coil. The supply fan is able to adjust its speed (and thus airflow), while the cooling coils are regulated to a setpoint temperature by chilled water. A VAV box near the duct terminals contains heating elements and dampers to control air flow. Return air from the zones is partially exhausted and partially recycled alongside outside air to create the supply air.

Let T denote the vector of building zone temperatures, and let $\dot{m}$ denote the vector of mass flow rates of cooled air into each zone. Also, let d_r denote the fraction of return air that is recycled into the system, and let T_c be the cooling coil outlet air temperature. Additionally let $\dot{Q}_{\text{offset}}$ be the vector of thermal loads independent of zone temperatures. Then, the dynamics of the

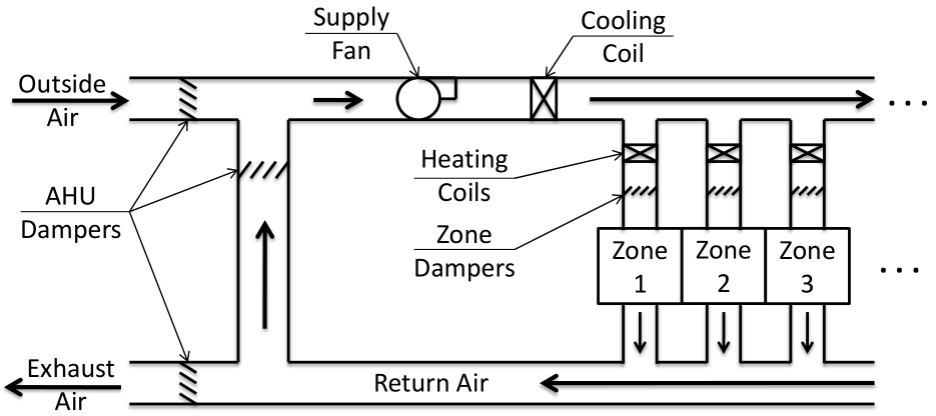

Figure 2.3: Illustration of typical variable air volume HVAC system [9].

system can be described by

$$M\frac{\mathrm{d}}{\mathrm{d}t}T(t) = RT(t) + \dot{Q}_{\text{offset}}(t) + c_p\dot{m}(t).^*(\mathbb{1}T_c(t) - T(t)), \tag{2.13}$$

where $\mathbb{1} = [1, \ldots, 1]^T$, $.^*$ denotes an elementwise product, M is a diagonal matrix of thermal capacitances associated with each building zone, R is a matrix of thermal resistances associated with each building zone, and c_p is the specific heat capacity of air.

Additional variables relate the dynamics in (2.13) to the electric power consumed by the HVAC system. Specifically, the electric power consumed by the supply fan, P_f, is given by

$$P_f(t) = \kappa_f \left(\mathbb{1}^T\dot{m}(t)\right)^2, \tag{2.14}$$

and the electric power consumed by the cooling coils, P_c, is given by

$$P_c(t) = \frac{c_p}{\eta_h}\mathbb{1}^T\dot{m}(t)(T_m(t) - T_c(t)), \tag{2.15}$$

where T_m is the cooling coil inlet air temperature which is given by

$$T_m(t) = (1 - d_r(t))T_{oa}(t) + d_r(t)T_r(t),$$

where T_r is the average return air temperature, which can be obtained as follows:

$$T_r(t) = \frac{\dot{m}(t)^T T(t)}{\mathbb{1}^T \dot{m}(t)}.$$

Finally, we need to consider the constraints which arise from acceptable occupant comfort:

$$\underline{T} \leq T(t) \leq \overline{T}, \quad 0 \leq d_r(t) \leq \overline{d}_r, \tag{2.16}$$

as well as those that arise from the ratings of the equipment:

$$\underline{\dot{m}} \leq \dot{m}(t) \leq \overline{\dot{m}}, \quad \underline{T}_c \leq T_c(t) \leq T_m. \tag{2.17}$$

In the remainder, we will assume that $d_r = 1$ and $T_c < \min(\underline{T})$. These assumptions are not a requirement for any future development, but they result in a cleaner formulation which is better for illustration purposes. Also, we will define the control input, $s(t)$, as a function of flow rates as follows: $s(t) = c_p \dot{m}(t).^*(T_c \mathbb{1} - T(t))$. Finally, we will neglect inter-zonal energy transfer. The matrix R becomes a diagonal matrix proportional to the difference between ambient and zonal temperatures, and $\dot{Q}_{\text{offset}}$ is a vector of thermal loads that are independent of both ambient and zone temperatures. With these simplifications, the dynamic model in (2.13) becomes

$$M \frac{\mathrm{d}}{\mathrm{d}t} T(t) = R(T_{oa}\mathbb{1} - T(t)) + \dot{Q}_{\text{offset}} + s(t). \tag{2.18}$$

The expression for the fan power in (2.14) becomes

$$P_f(t) = \frac{\kappa_f}{c_p} \left(\mathbb{1}^T \left(s(t)./(T_c \mathbb{1} - T(t)) \right) \right)^2, \tag{2.19}$$

where ./ denotes an elementwise division, and the expression for the cooling

coil power in (2.15) becomes

$$P_c(t) = -\mathbb{1}^T \frac{s(t)}{\eta_c}. \tag{2.20}$$

The constraints in (2.16) – (2.17) result in

$$\underline{T} \leq T(t) \leq \overline{T}, \tag{2.21}$$

$$\underline{s}(t) \leq s(t) \leq \overline{s}(t). \tag{2.22}$$

Numerical values of system parameters used in this section are given in Table 2.1.

Table 2.1: Parameters used in numerical study

Parameter	Value	Unit
n	5	zones
Δt	varied	s
c_p	1	kJ/(kg K)
mc_i	1000	kJ/K
R	0.1	kW/K
η_h	0.9	dimensionless
κ_f	0.065	kWs^2/kg^2
$\underline{T}_{zi}$	21	°C
$\overline{T}_{zi}$	24	°C
$\overline{T}_{oa}$	30	°C
$\underline{m}_{zi}$	0.025	kg/s
$\overline{m}_{zi}$	1.5	kg/s
$\dot{Q}_{\text{offset}}$	0	kW

2.4.2 Baseline Power

We define the regulation power at time t as the difference between the actual power consumed by the fan and cooling coils, i.e., $P_f(t) + P_c(t)$, and some baseline power, denoted by P^0, which is the total electric power consumed by the system were it not providing the regulation services. In this chapter, we

consider P^0 to be the value obtained from the steady state solution of (2.18), with the zone temperatures set to their midpoint values $T^{\mathrm{m}} = \frac{1}{2}(\underline{T} + \overline{T})$. In subsequent developments, we will assume this solution satisfies (2.21) and (2.22). Thus, by setting the left hand side of (2.18) to zero, it immediately follows that

$$s^0 = -(RT^{\mathrm{m}} + \dot{Q}_{\mathrm{offset}}). \tag{2.23}$$

From (2.23), we can calculate the baseline power using (2.19) and (2.20), which results in

$$P^0 = -\mathbb{1}^T \frac{s^0}{\eta_c} + \frac{\kappa_f}{c_p} \left(\mathbb{1}^T \left(s^0./(T_c\mathbb{1} - T^{\mathrm{m}})\right)\right)^2.$$

2.4.3 Controller Design

Previous work on TCLs has proposed various controllers including a priority stack scheme [40]. Such a design is not applicable to this system because we have continuous control inputs rather than a number of binary ones, thus we propose a new controller which is appropriate for more general systems.

The controller's input is a commanded power output, P^*, which is equal to the desired regulation plus the baseline power. The output is a control, $s(t),\ t > 0$, which causes the HVAC system to consume the requested amount of power while also respecting the limits in (2.21) and (2.22). First, we check for feasibility, and if there exists an input, $s(t),\ t > 0$, such that all constraints are satisfied, we choose to optimize $s(t),\ t > 0$, so temperatures are driven towards their midpoints. We can pose this problem as a nonlinear

least square error estimation problem:

$$\begin{aligned} s^*(t) = \underset{s(t)}{\arg\min} \quad & \|T(t+\Delta t) - T^{\mathrm{m}}\|_2 \\ \text{subject to} \quad & \underline{T} \le T(t+\Delta t) \le \overline{T} \\ & \underline{s}(t) \le s(t) \le \overline{s}(t) \\ & P^*(t) - P_f(t) - P_c(t) = 0. \end{aligned} \tag{2.24}$$

2.4.4 Controller Performance Verification

In the first study, we examine the behavior of the controller proposed in Section 2.1 and verify its functionality. The baseline power is calculated to be 4.17 kW. Figure 2.4 shows that initial zonal temperatures are evenly distributed through the acceptable range indicated by dashed lines. We see that the controller initially drives temperatures toward the midpoints, as desired. Minimum flow rate constraints are initially binding for zones 1 through 4. The commanded power is then stepped from 0 kW to 1 kW. The controller initially issues commands that perfectly meet the request. Temperatures decrease until temperature constraints become binding. At this point the controller is unable to meet the requested power, so there is a positive error.

2.4.5 Charge Rate Limit Estimation

We next use the estimation procedure outlined in Section 2.3 to identify the positive charge rate limits $\overline{n}$ and $\underline{n}$. We obtain that $\overline{n} = 103.7\,\mathrm{kW}$ and $\underline{n} = 2.44\,\mathrm{kW}$; the asymmetry is quite large because, for the chosen parameters, the air conditioning system is capable of blowing much more cold air than required to maintain a steady temperature. On the other hand, the baseline air flow is quite close to the lower limit, so it cannot consume much less than the baseline power value. For this reason, this system would

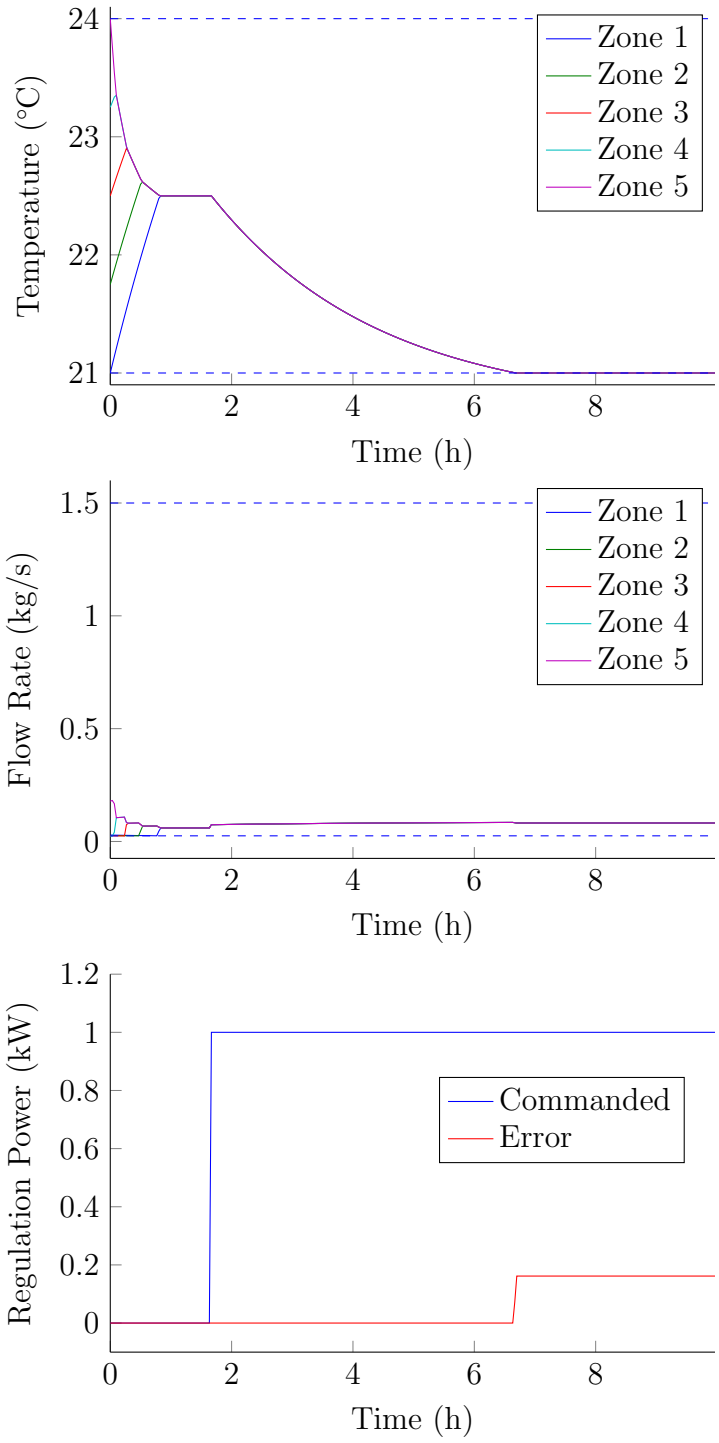

Figure 2.4: Response to step regulation signal.

be much better utilized in a market that treats up and down regulation as two distinct services.

Table 2.2: Estimated parameters

		Parameter	
Input	b Solver	a (s^{-1})	C (kWh)
Step	Analytic	1.003×10^{-4}	2.321
Step	Numerical	1.003×10^{-4}	2.321
Ramp	Analytic	1.002×10^{-4}	2.324
Ramp	Numerical	1.002×10^{-4}	2.324
RC Step	Numerical	1.003×10^{-4}	2.321
Monomial	Numerical	1.003×10^{-4}	2.322
RegD	Numerical	9.966×10^{-5}	2.334

2.4.6 Capacity and Dissipation Estimation

The next step is to identify capacity and dissipation parameters. For this task, a set of test inputs needs to be chosen. We investigate a number of different families of inputs and compare their performance. The parameters identified using the different techniques are remarkably consistent; the results are summarized in Table 2.2.

Step Input

We first test the use of inputs of the form $u(t) = k,\ t \geq 0$, with $k \in \mathbb{R}$. Fifty values of k were chosen logarithmically distributed between a value just above aC (which can be found using a search procedure similar to the one outlined in Algorithm 1) up to $\overline{n}$. Figure 2.5 provides a plot of violation time versus input magnitude. If $a = 0$, we would expect a straight line with slope -1. The line curves upward for small inputs because there is more time for the effects of the dissipation to manifest themselves.

Step inputs are simple enough that we can find an analytic solution to (2.9); thus we have a choice of calculating b using an analytic expression or a numerical solver. In this regard, the optimization procedure in (2.12) runs

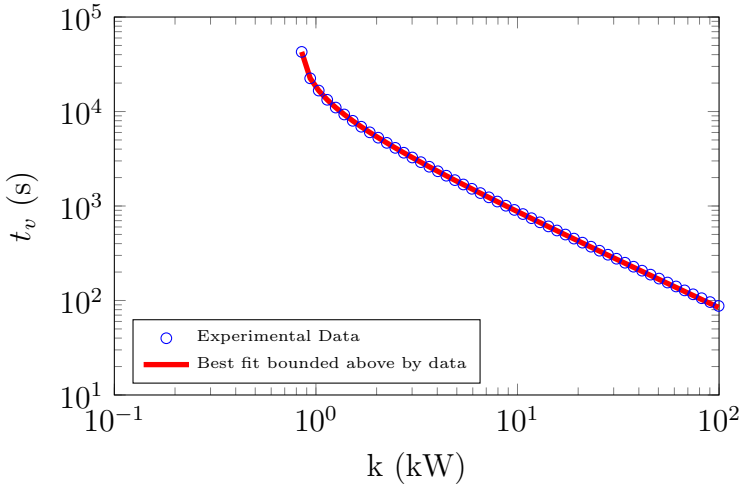

Figure 2.5: Best battery model bounded above by constant input experimental data.

orders of magnitude faster with the analytic expression, and the identified parameters were confirmed to be nearly identical in either case.

Ramp Input

Figure 2.6 shows the result for inputs of the form $u(t) = kt,\ t \geq 0$, with $k \in \mathbb{R}$. Again, we calculate the parameters using both an analytic expression, which is given by

$$t_v = \frac{1}{a}\left(1 + \frac{a^2 C}{k} + W(-e^{-1-\frac{a^2 C}{k}})\right),$$

where W is the Lambert W function [41], and a numerical solver; both approaches yield identical results. The identified parameters also agree with those identified using the step inputs.

RC Step Input

Figure 2.7 shows the result for inputs of the form $u(t) = k(1 - e^{-\lambda t}),\ t \geq 0$, with $k \in \mathbb{R}$, and $\lambda = 5 \times 10^{-5}\,\mathrm{s}^{-1}$. For a given k, violation times with this input are larger than the instantaneous step input because $u(0) = 0$ and $u(t)$

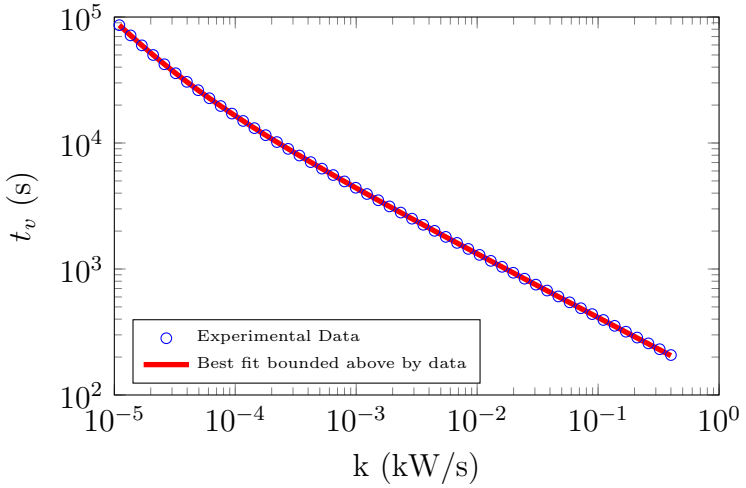

Figure 2.6: Best battery model bounded above by ramp input experimental data.

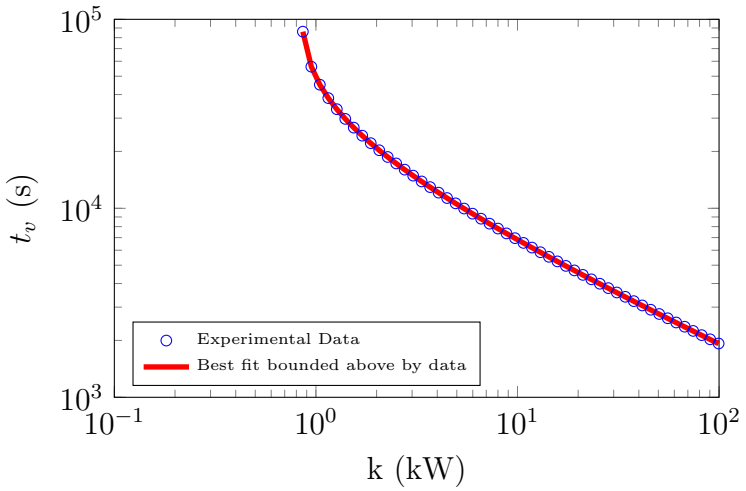

Figure 2.7: Best battery model bounded above by RC charging step input.

approaches k asymptotically. For this type of input and the following types, there is not an analytical expression for the violation time, so we only test the numerical methods. In the end, the identified parameters agree with the previous values identified using step functions.

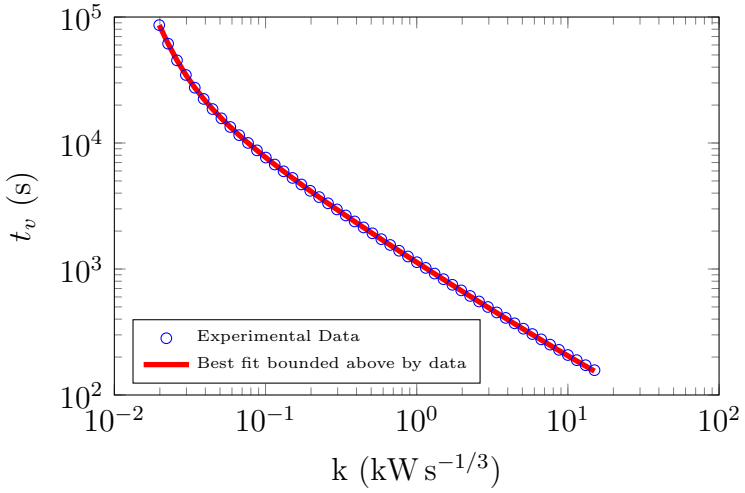

Figure 2.8: Best battery model bounded above by monomial input.

Monomial Input

Figure 2.8 shows the result for inputs of the form $u(t) = kt^{\lambda},\ t \geq 0$, with $k \in \mathbb{R}$, and $\lambda = \frac{1}{3}$. The parameters obtained using this input provide further evidence for the consistency of the results among different input types.

Regulation Signal Step Input

Figure 2.9 shows violation time vs. input constant using a modified regulation signal $u(t) = r(t)(k(r(t) > 0) + \underline{n}(r(t) < 0))$, where $r(t)$ is the PJM dynamic regulation signal from [42]. A representative segment of this signal is shown in Fig. 2.10. This asymmetric signal was selected because of the asymmetric nature of the charge rate constraints. A symmetric signal that respects $\underline{n}$ would never violate a capacity limit and would not provide limited economic benefit. Providing asymmetric regulation is possible in markets such as CAISO, where up regulation and down regulation are treated as different services.

Of all the tests, this had the most issues with convergence, step sizes, and tolerances. This is likely due to the input function not being monotonic. With a monotonic function, a small integration or interpolation error will

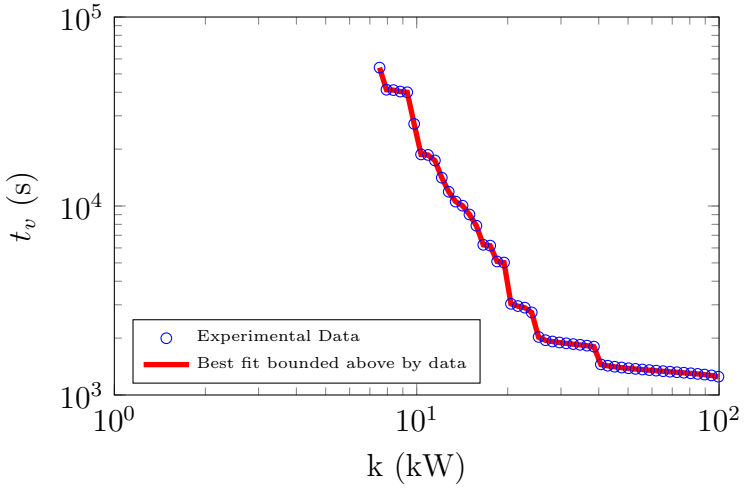

Figure 2.9: Best battery model identified by regulation signal input.

lead to a small change in violation time. This is not the case with this input signal. A small difference (for example, the nonlinear data uses Euler's method, but the battery model uses Runge-Kutta) can lead to a much bigger difference in violation time. Even with this difficulty, the estimated parameters using this input match those obtained with the other aforementioned approaches (see Table 2.2). Conversely, the parameters identified using the other inputs performed practically identically in predicting violation times from the regulation signal. Overall this is excellent empirical evidence to support our proposed stress-based estimation procedure for this system.

2.5 Case Study: Real Commercial Building

We begin this section by formulating a specific building/HVAC system model and a controller of the very general form presented in Section 2.1. We will then test the performance of the identification procedure described in Section 2.3 on the aforementioned system.

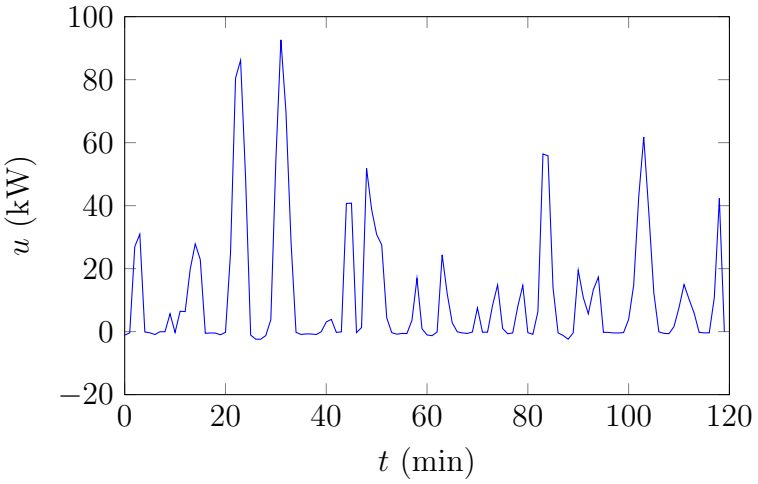

Figure 2.10: Portion of modified dynamic regulation input signal, $k = 99.5$.

2.5.1 Commercial Building/HVAC System Model

Our building/HVAC system dynamic model and controller formulation will now be specifically tailored to describe the thermal dynamics of the University of Illinois Willard Airport terminal building. The formulation is a generalization of the commercial building model used in Section 2.4, which has a structure that is insufficient to accurately model the terminal. Key changes include allowing for multiple air handling units, generalizing the fan power consumption formula, and accommodating unconditioned zones.

Let T denote the vector of building zone temperatures; M denote the (diagonal) matrix of zone thermal inertias; T_{oa} denote the outside ambient temperature; c_p denote the specific heat capacity of air (assumed constant); $\dot{m}$ denote the vector of conditioned air mass flow rates to each zone; T_c denote the air temperature at the output of the cooling coils; $\dot{Q}_{\text{people}}$, $\dot{Q}_{\text{lighting}}$, $\dot{Q}_{\text{equipment}}$, $\dot{Q}_{\text{solar}}$ denote vectors of thermal loads due to occupants, electric lighting, miscellaneous machinery and office equipment, and solar radiation, respectively. Then, the dynamics of the building/HVAC system are described

by

$$M\frac{\mathrm{d}}{\mathrm{d}t}T(t) = R_1T(t) + R_2(T_{oa} - T(t)) + \dot{Q}_{\text{people}} + \dot{Q}_{\text{lighting}} + \dot{Q}_{\text{equipment}} + \dot{Q}_{\text{solar}} + c_p\dot{m}.^*(T_c\mathbb{1} - T(t)), \tag{2.25}$$

where $.^*$ indicates element-wise multiplication; R_1 is a sparse, symmetric matrix with off-diagonal elements (i, j) that are the thermal conductance between zones i and j, and diagonal elements (i, i) equal to the negative sum of the off-diagonal elements in row i; R_2 is a diagonal matrix of external conductance values.

By defining $R = R_1 - R_2$ and $\dot{Q} = R_2T_{oa} + \dot{Q}_{\text{people}} + \dot{Q}_{\text{lighting}} + \dot{Q}_{\text{equipment}} + \dot{Q}_{\text{solar}}$, we arrive at a more compact version of (2.25):

$$M\frac{\mathrm{d}}{\mathrm{d}t}T(t) = RT(t) + \dot{Q} + c_p\dot{m}(t).^*(T_c\mathbb{1} - T(t)). \tag{2.26}$$

The Willard Airport terminal building has five air handling units (AHUs). Let $\dot{M}$ be the vector of mass flow rates through each of the air handler units. We assume ideal ducts, so each element can be obtained by summing the elements of $\dot{m}$ corresponding to the appropriate air handler. Then, the expression for power consumed by the air handling fans is a second-order function of mass flow rate:

$$P_f(t) = \kappa_{f2}\dot{M}(t)^T\dot{M}(t) + \kappa_{f1}\mathbb{1}^T\dot{M}(t) + \kappa_{f0}.$$

If we assume the return air is a weighted average of the conditioned zones, the expression for the cooling coil power is given by

$$P_c(t) = \frac{c_p}{\eta_c}\mathbb{1}^T\dot{m}(t)\left(T_m - T_c\right), \tag{2.27}$$

where the temperature of the input to the AHU is

$$T_m(t) = (1 - d_r)T_{oa} + d_r T_r,$$

and the return air temperature is

$$T_r(t) = \frac{\dot{m}(t)^T T(t)}{\mathbb{1}^T \dot{m}(t)}.$$

Fast chiller dynamics are assumed in (2.27), and fixed supply air temperature is assumed in (2.26) and (2.27). While we note there is other literature that assumes fast chiller dynamics [9], some papers use a first order time delay [19] or utilize only the fan power [35] to provide regulation. Extensive chiller modeling is outside the scope of this dissertation, but future work could classify typical equipment into classes where 1) our assumption effectively holds, 2) there is a noticeable, but acceptable lag in the chiller response, or 3) the response is delayed so much that is no longer useful for frequency regulation.

The model is composed of zones which may be conditioned or unconditioned. Let T_1 be the vector of conditioned zone temperatures, and $\dot{m}_1$ be the vector of conditioned zone mass flow rates. Similarly, let T_2 be the vector of unconditioned zone temperatures, and $\dot{m}_2 = 0 \cdot \mathbb{1}$ be the vector of unconditioned zone mass flow rates.

The constraints will then be $\underline{T} \leq T_1(t) \leq \overline{T}$ and $\underline{\dot{m}} \leq \dot{m}_1(t) \leq \overline{\dot{m}}$. Additional constraints, $\underline{\dot{M}} \leq \dot{M}(t) \leq \overline{\dot{M}}$, limit the total flow through each AHU.

The optimization problem in (2.6) is chosen to be

$$\begin{aligned} s^*(t) = \arg\min_{s(t)} \quad & \left\| (T_1(t+\Delta t) - T_1^{\mathrm{m}}) \,./\, \left(\overline{T} - \underline{T}\right) \right\|_\infty \\ \text{subject to} \quad & \underline{T}_1 \leq T_1(t+\Delta t) \leq \overline{T}_1 \\ & \underline{s}(t) \leq s(t) \leq \overline{s}(t) \\ & |P^*(t) - P_f(t) - P_c(t)| \leq \delta, \end{aligned} \tag{2.28}$$

where ./ indicates element-wise division and $s = \dot{m}_1$. The objective function is chosen so that the controller will attempt to keep all conditioned zones near their midpoint temperature. If the regulation signal pushes the system to its limits, the controller will bring all zones to their temperature limit at the same time, at which point there will be a constraint violation.

The problem in (2.28) is solved using sequential quadratic programming using the baseline control input as an initial condition. As overly restrictive constraints can cause problems with the numerical solver, if no feasible solution to (2.28) can be found with $\delta = 0.1\,\mathrm{W}$, δ is increased until a solution is found. Then, δ is decreased in gradual steps, using the previous solution as an initial condition for solving each new optimization problem. Then, if the optimization finds a solution with $\delta = 0.1\,\mathrm{W}$, the simulation continues. If there is truly no feasible solution, the controller considers this a violation and returns the solution that converged with the smallest δ. This alternative solution minimizes the error without violating state or input constraints. This graceful degradation is not required for the parameter identification algorithm proposed in Section 2.3, but it would be desirable when implementing the controller in a real system.

To find the baseline power, (2.26) can be broken into

$$\begin{bmatrix} M_c & 0 \\ 0 & M_u \end{bmatrix} \begin{bmatrix} \dot{T}_1(t) \\ \dot{T}_2(t) \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix} + \begin{bmatrix} \dot{Q}_1(t) \\ \dot{Q}_2(t) \end{bmatrix} + c_p \begin{bmatrix} \dot{m}_1(t) \\ \dot{m}_2(t) \end{bmatrix} .^* \begin{bmatrix} T_c\mathbb{1} - T_1^m(t) \\ T_c\mathbb{1} - T_2(t) \end{bmatrix}. \tag{2.29}$$

Then, we set $\dot{T}_1(t) = \dot{T}_2(t) = 0$, $T_1(t) = T_1^m$, and $\dot{m}_2 = 0$ in (2.29), giving us

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} T_1^m \\ T_2^0 \end{bmatrix} + \begin{bmatrix} \dot{Q}_1 \\ \dot{Q}_2 \end{bmatrix} = \begin{bmatrix} -c_p\dot{m}_1^0.^*(T_c\mathbb{1} - T_1^m) \\ 0 \end{bmatrix}. \tag{2.30}$$

Equivalently, (2.30) can be written as two equations:

$$R_{11}T_1^m + R_{12}T_2^0 + \dot{Q}_1 = -c_p\dot{m}_1^0.^*(T_c\mathbb{1} - T_1^m) \tag{2.31}$$

and

$$R_{21}T_1^m + R_{22}T_2^0 + \dot{Q}_2 = 0. \tag{2.32}$$

From (2.32), it follows that $T_2^0 = -R_{22}^{-1}(R_{21}T_1^m + \dot{Q}_2)$. By plugging this into (2.31), we obtain

$$R_{11}T_1^m - R_{12}R_{22}^{-1}(R_{21}T_1^m + \dot{Q}_2) + \dot{Q}_1 = -c_p\dot{m}_1^0.^*(T_c\mathbb{1} - T_1^m),$$

which can easily be solved for baseline mass flow rate, yielding

$$\dot{m}_1^0 = (R_{11}T_1^m - R_{12}R_{22}^{-1}(R_{21}T_1^m + \dot{Q}_2) + \dot{Q}_1)./(-c_p(T_c\mathbb{1} - T_1^m)).$$

2.5.2 Numerical Values for Willard Airport Terminal Building

A three-dimensional model of the Willard Airport terminal building in Savoy, Illinois, was created by the Illinois Smart Energy Design Assistance Center (SEDAC) at the University of Illinois [43]. The model, shown in Fig. 2.11,

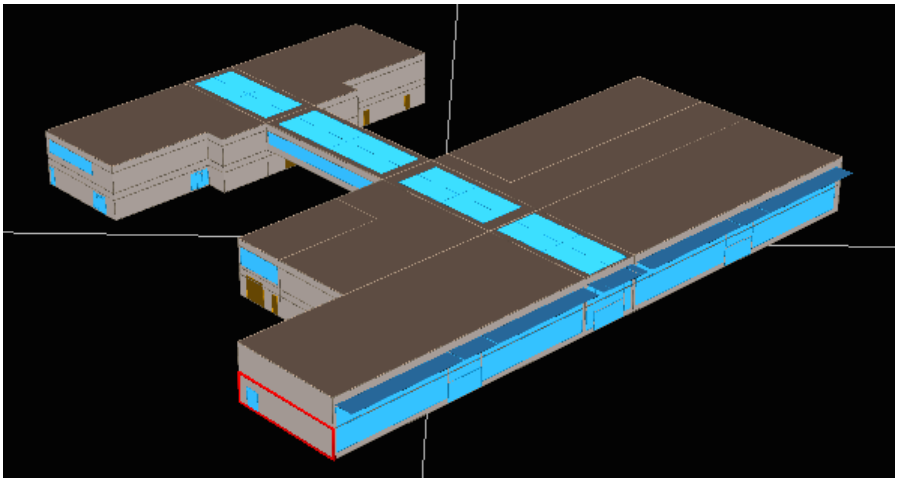

Figure 2.11: Three-dimensional model of Willard Airport terminal provided by Shawn Maurer, Andrew Robinson, and Todd Rusk, all affiliated with the Illinois Smart Energy Design Assistance Center (SEDAC) at the University of Illinois [43].

was created in eQuest, a software program designed to evaluate building energy performance [44]. The model comprises 41 zones, 19 of which are conditioned. Since it is a very detailed model of a real building, it makes for a compelling case study. We will next look at the development of some key parameters of the study; however, values for every parameter cannot be presented due to space restrictions.

Thermal Conductance

To construct the interior thermal conductance matrix, R_1, and exterior thermal conductance matrix, R_2, the material and geometry of the surfaces between each of the 41 zones and all the other zones and the exterior were considered.

Thermal Mass

One of the most important parameters for our study is the thermal mass M. The dimensions of each zone are known, so the volume of air can be

easily calculated. The volume multiplied by the density and specific heat capacity gives us the thermal mass. We assume the air is at a constant 25 °C temperature and 1 atm pressure for density calculations.

It is well known that solids have a higher specific heat capacity than gases; thus the walls, floors, and furnishings in a room usually have a higher thermal mass than the air. Because eQuest contains thickness, density, and specific heat values for the building materials, we can calculate their thermal masses. For interior walls, half the thermal mass was assigned to each of the two zones it separates. External surfaces can develop a significant temperature gradient across the insulation, making the effective mass smaller. To account for this, we divided the calculated mass by a factor of two to reach an effective amount.

For each zone, the thermal mass of the air is calculated. Then, the thermal mass of walls, floors, and ceiling is estimated by multiplying the square footage by the thermal mass per square foot of a representative zone.

Thermal Loads

Solar radiation values were recorded at the time of peak cooling load. These values were multiplied by a scaling factor to account for the time of day and day of year using standard insolation formulas [45]. However, this simplification does not account for the exact geometry of the building or the effect of cloud cover. Occupancy, lighting, and equipment loads are each updated hourly based on their regular weekday or weekend schedule.

Ambient, Limit, Supply Temperature

Ambient temperature (T_{oa}) data from 2013 in nearby Springfield, Illinois, is used due to its availability. Conditioned zone temperature limits were selected as 21.2 °C and 23.2 °C; these values are consistent with normal variations in temperature without the regulating controller. Historical data shows

some variation in supply temperature T_c, but an average of 15 °C was selected.

Other Scalar Parameters

Fan parameters and the chiller efficiency were estimated using least squares estimation against eQuest timeseries data. The authors believe that, for the purposes of this study, this calibration adequately compensates for other approximations.

2.5.3 Estimation Procedure Results

We next present the results obtained by using the battery model identification procedure described in Section 2.3 on the Willard Airport terminal building model. Three summer days were selected for the study. Numerical results are summarized in Table 2.3. We will first analyze the results for the time-invariant studies, followed by the time-varying studies.

Charge Rate Limit Estimation

We find $\overline{n}$ varies from 138 kW to 239 kW, and $\underline{n}$ varies from 46.8 kW to 148 kW. The sum of these values is always approximately 286 kW. This sum is driven by the difference between the upper and lower mass flow rate limits of the air handler units. The way this difference is distributed between the two rate limits depends on the baseline power. For example, on a cooler day, the relatively small baseline power results in $\overline{n} > \underline{n}$. Because of the asymmetry, this system could be more effectively utilized in a market setting that treats up and down regulation as two distinct services.

Capacity, Dissipation, and Initial Charge Estimation

The results in Section 2.4 show that for a time-invariant system results are independent of the type of input used. Step inputs are chosen for their

Table 2.3: Identified parameters

Date	Time-varying	$\overline{n}$ (kW)	$\underline{n}$ (kW)	a (Ms^{-1})	C (MWh)	x_0 (MWh)
10-Jun	no	138	148	8.92	0.580	0.228
10-Jun	yes	138	148	11.7	0.424	0.424
06-Jul	no	239	46.8	7.60	0.312	0.312
06-Jul	yes	239	46.8	6.86	0.318	0.317
26-Aug	no	140	145	7.92	0.633	0.147
26-Aug	yes	140	145	7.06	0.708	0.0174

simplicity.

Let $u_i(t) = k,\ t \geq 0$, with $k \in \mathbb{R}$. Sixteen values of k were chosen logarithmically-distributed between aC (which can be found using a search procedure similar to the one outlined in Algorithm 1) and $\overline{n}$. Figures 2.12 and 2.13 provide a plot of violation time versus input magnitude on June 10th and July 6th. If $a = 0$, we would expect a straight line with slope -1. The line curves upward for small inputs because there is more time for the effects of the dissipation to manifest themselves.

On June 10th, the parameters of the battery model obtained by our identification procedure are as follows: $a = 8.92 \times 10^{-6}\,\mathrm{s}^{-1}$, $C = 0.580\,\mathrm{MWh}$, and $x_0 = 0.228\,\mathrm{MWh}$. On July 6th, the identified parameters are $a = 7.60\,\mathrm{s}^{-1}$, $C = 0.312\,\mathrm{MWh}$, and $x_0 = 0.312\,\mathrm{MWh}$. We attribute the differences to changing environmental conditions (solar intensity, ambient temperature) and internal loads due to weekend versus weekday schedule. For both cases, even using our narrow range of acceptable temperatures, C is large compared to the charge rate limits. In fact, the virtual battery can supply maximum power for at least an hour without running out of charge, but a realistic regulation signal tends to alternate between charging and discharging with periods no longer than tens of minutes. Thus, we expect the charge rate limits to be the primary factor in determining the capability offer for the building.

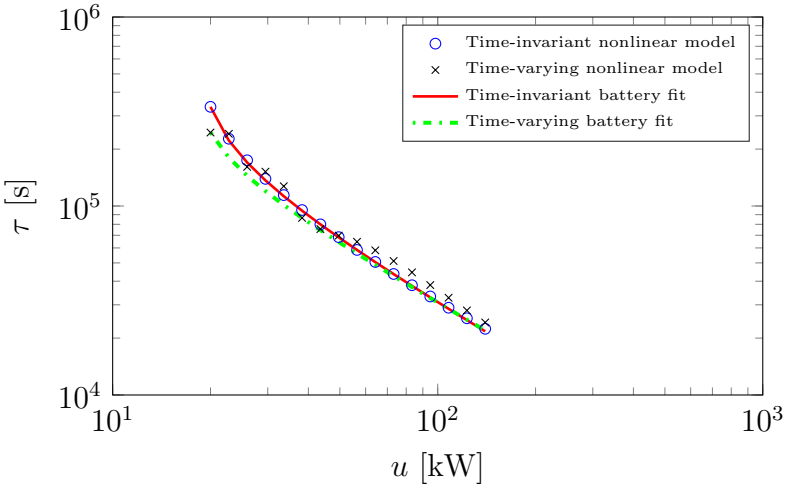

Figure 2.12: Experimental violation times and best fit battery model predictions for simulations starting June 10th.

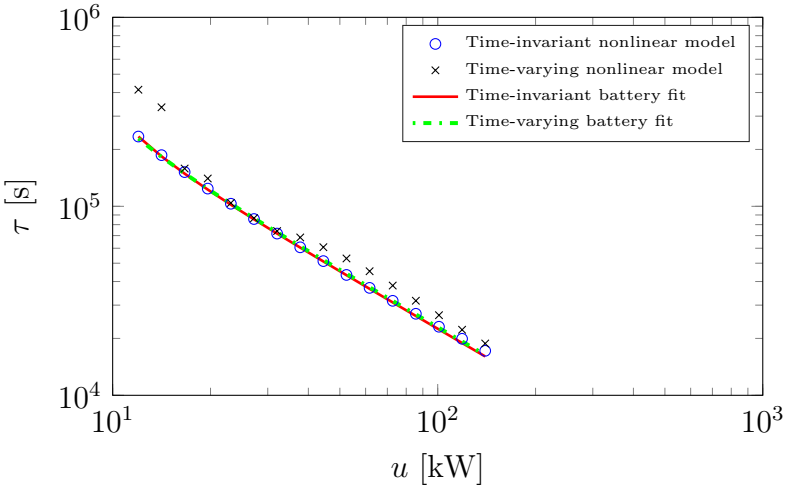

Figure 2.13: Experimental violation times and best fit battery model predictions for simulations starting July 6th.

2.5.4 Effect of Time-Varying Parameters

For the next set of studies, T_{oa}, $\dot{Q}_{\text{people}}$, $\dot{Q}_{\text{lighting}}$, $\dot{Q}_{\text{equipment}}$, and $\dot{Q}_{\text{solar}}$ are allowed to vary with time; in our previous work, these values were all assumed to be fixed. Figure 2.14 shows how this translates into time-varying power values, and Fig. 2.15 shows how that creates diurnal patterns in temperature and mass flow rate. This generalization is challenging to our model, and analysis brings a number of interesting points.

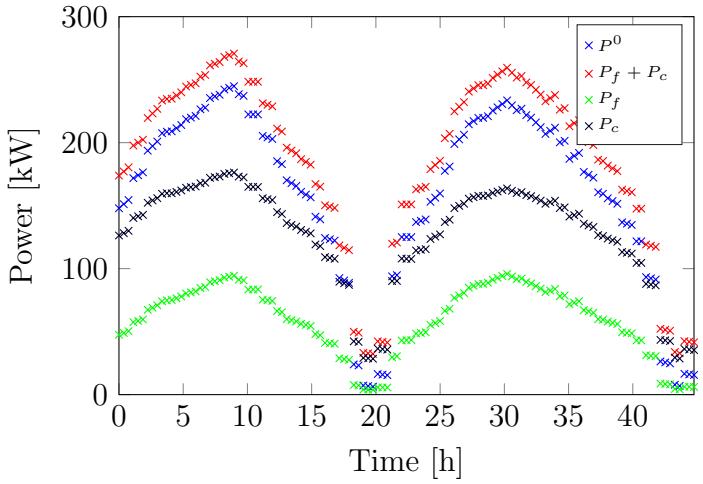

Figure 2.14: Power values for the time-varying simulation starting June 10th.

First, we note that for this study we assume knowledge of future values of the studied time-varying parameters. In practice, weather forecasts and historical thermal load data based on the regular airport schedule will need to be used. Future work could quantify the impact of uncertainty in predictions on output parameters. Further, as T_c and d_r may also vary, the procedure will need to be implemented in a receding-horizon manner, which will periodically update the identified battery parameters using new measurements to deal with the model-plant mismatch.

We note that the identified rate limits are identical to those in the time-invariant studies. This is because the rate limits are only identified at the first time step. As the baseline power varies, the rate limits will also change. A simple fix is to generalize the battery model to allow time-varying rate limits and identify these rate limits at each time step of the simulation.

Next, we compare the parameters a, C, and x_0. It would be natural to also allow a and C to vary with time, but this would add significant complexity and make the model more difficult to analyze. We note that the time-varying battery fit curves in Figs. 2.12 and 2.13 do not fit the data from the time-varying nonlinear model as well as the time-invariant battery fit curves for

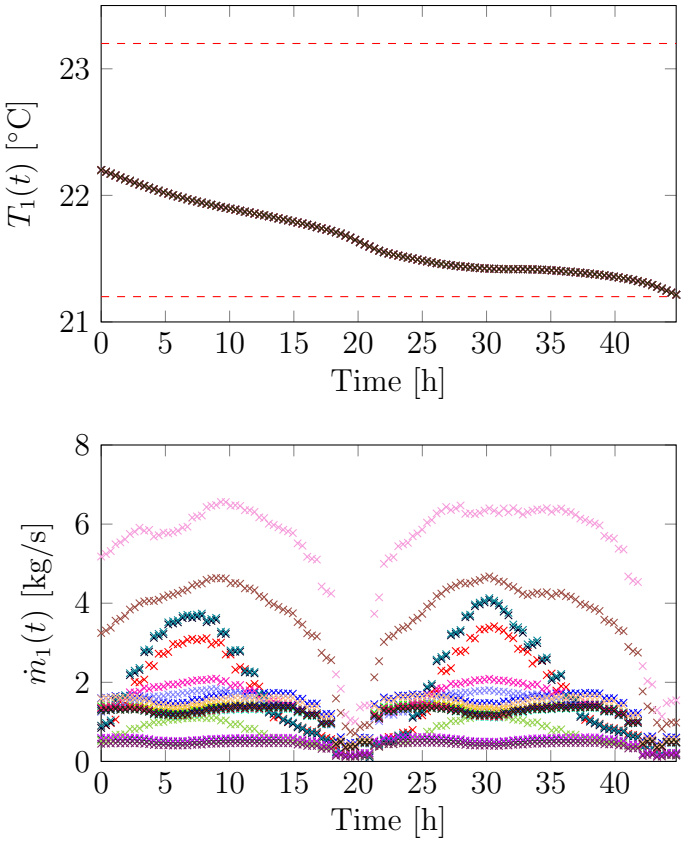

Figure 2.15: Variables for the time-varying simulation starting June 10th.

the time-invariant nonlinear model data; thus, using the battery model would lead to more conservative predictions. The user would have to decide if the accuracy/complexity trade-off is appropriate for the task at hand.

We also notice that the time-varying and time-invariant fit curves in Figs. 2.12 and 2.13 are quite similar, but the identified parameters in Table 2.3 can differ significantly. This occurs because the battery parameters can be sensitive to small changes in the violation time data. The dissipation parameter a is most apparent over long periods of time. Also, the effects of C and x_0 are relatively indistinguishable to our method over short amounts of time. To distinguish them, the behavior over long periods of time is also critical. Unfortunately, it is over longer periods of time that the time-varying and time-invariant parameters tend to diverge. For some purposes, using the time-invariant battery model may be deemed sufficient to model a time-

varying system. With sufficient data, studies can determine the functional relationship between the virtual battery parameters and exogenous variables such as time and ambient temperature, which can further improve the quality of the model.

2.5.5 Regulation Signal Input

In the final study, we illustrate the ability of a selected building to successfully follow a regulation signal that is within the capability characterized by the identified battery model without adversely impacting the indoor environment.

We utilize a normalized regulation signal $r(t)$, plotted in Fig. 2.16, which is chosen to be an example "RegD" signal published by PJM. In this example, the identification procedure is performed each hour. In the following hour, the commanded deviation from the baseline power is $\overline{n}(t)r(t)$ if $r(t) \geq 0$, and $\underline{n}(t)r(t)$ if $r(t) < 0$. This scenario approximates participation in a real-time market with instantaneous hourly clearing of separate zero-cost up-regulation and down-regulation capability offers. If the identified capacity limit were not large compared to the charge rate limits, the offer would be based on the more restrictive limit.

Figure 2.17 shows how the controller proposed in Section 2.1.3 enables the building to track the regulation signal and the effect on indoor temperatures. The temperature variations are small compared to the established bounds of 21.2 °C and 23.2 °C. Power consumption varies greatly from the baseline, with more up-regulation at night and more down-regulation at midday. In this example, all constraints are satisfied at each time step. However, unless exogenous parameters are time-invariant, it is possible for situations to arise where this is not true if $r(t)$ takes on unfortunate values and external variable change quickly. Possible mitigation techniques include obtaining more accurate forecasts, identifying battery parameters more frequently, and mak-

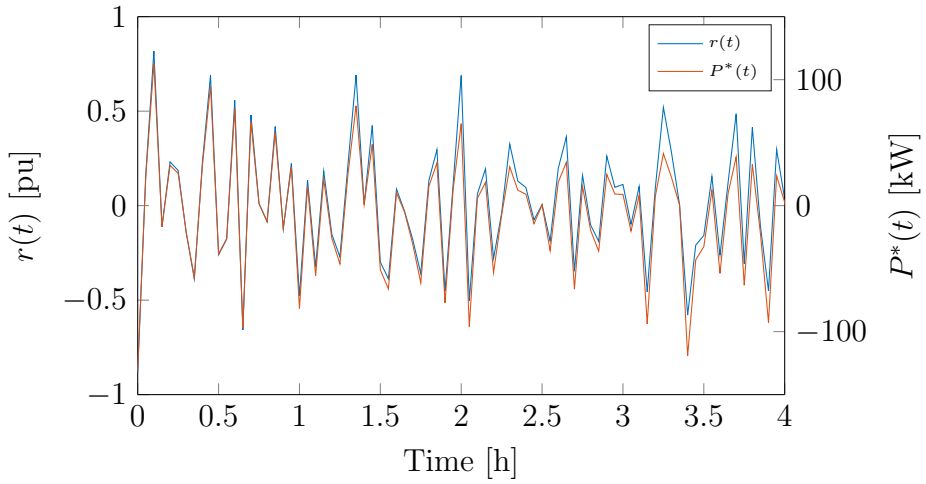

Figure 2.16: Normalized regulation signal $r(t)$ and commanded deviation from baseline power $P^*(t)$. During this period, $\overline{n}$ and $\underline{n}$ are nearly equal, making the plots similar. During peaks and nadirs in power consumption, $P^*(t)$ is less balanced.

ing more conservative capability offers (e.g., 90% of calculated maximum) to account for uncertainty.

2.6 Concluding Remarks

We have proposed a controller that allows for flexible loads to provide frequency regulation. We have introduced a method whereby the ability of the resulting closed-loop system to provide regulation can be accurately described by a simple, well understood battery model. Although the estimation method is an approximation, it was found to be effective on our University of Illinois Willard Airport test system.

Our case study revealed challenges that will be faced when applying the technique to real buildings. An ideal candidate building would have a large thermal mass with high, constant power consumption and the ability to consume much more or less power if required. Large thermal masses are common; however, arbitrary flexibility over power consumption is likely less abundant as excess capacity comes at a cost. Thus, for realistic buildings, rate lim-

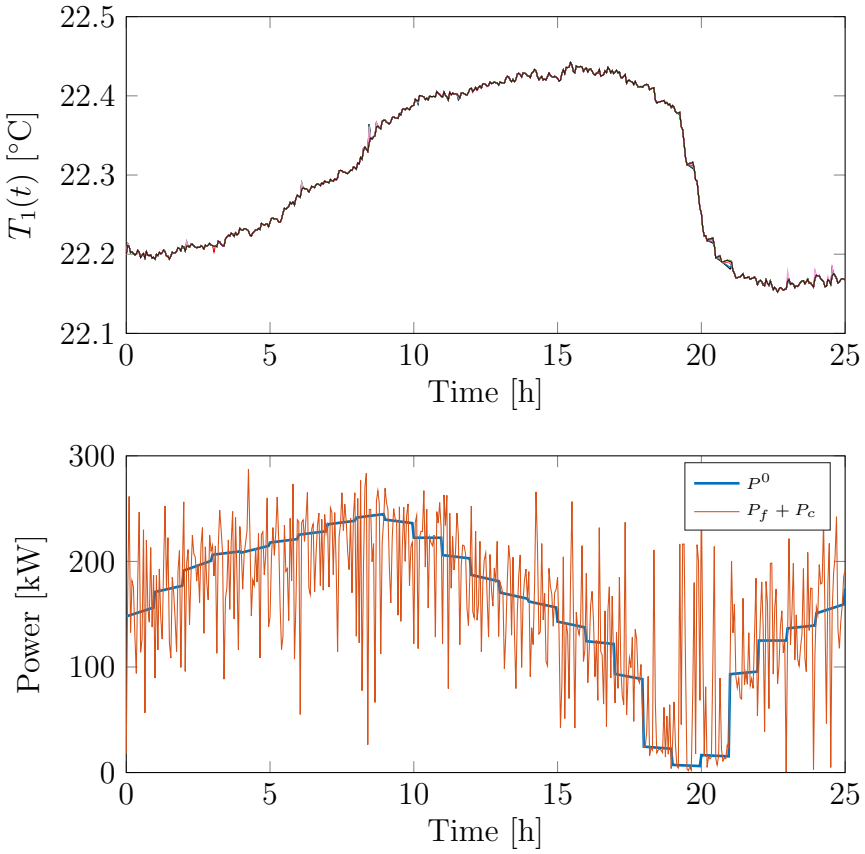

Figure 2.17: Effect of tracking regulation signal on temperature and power. Temperature values remain well within the bounds of 21.2 °C and 23.2 °C. Simulation starting June 10th.

its will primarily be the limiting factor in determining regulation capability, and these parameters will vary with baseline power. Asymmetric regulation markets are extremely helpful in allowing full use of the virtual battery.

CHAPTER 3

PREDICTIVE COORDINATION OF DISTRIBUTED ENERGY RESOURCES

In this chapter, we will introduce the dynamics and constraints of the generalized DER model, define the costs of utilizing the DERs, and formalize the aggregator coordination problem as an optimal control problem. Next, we propose solution procedures based on predictive control in various cases with different assumptions about forecasts and solution time horizons. Simulation results will highlight the importance of the forecasting procedure.

3.1 Problem Setting

In this section, we first introduce the model that describes the dynamics associated with the power delivery process of the different types of DERs considered in this chapter. We then capture the regulation cost associated with the DERs. Using these, we formulate the DER coordination problem faced by the aggregator.

3.1.1 DER Power Delivery Model

We assume the aggregator needs to coordinate various types of DERs, which could include small-rating conventional generators (commonly referred to as microturbines), commercial building HVAC systems, collections of PEVs or TCLs, and flexible industrial processes. We provide a single model—a generalization of the virtual battery model [39]—which can describe the behavior of any of these resources.

Let $P^i(t) = p^i(t) + p_0^i$ denote the power delivered by DER i at time t, where p_0^i is some nominal setting at which the DER is operating, and $p^i(t)$ is the amount of regulation power that this type of DER provides, and let $u^i(t)$ denote the rate of change of $p^i(t)$, i.e., $\frac{\mathrm{d}}{\mathrm{d}t}p^i(t)$. Also, let $X^i(t) = x^i(t) + x_0^i$ denote the DER energy level at time t, where x_0^i is some nominal energy level, and $x^i(t)$ is the variation in the DER energy level around x_0^i. Additionally, let $\overline{p}^i$ and -$\underline{p}^i$ denote the maximum and minimum values of $p^i(t)$ as determined by the charge rate limits of the DER (e.g., maximum power rating), and let $\overline{u}^i$ and $-\underline{u}^i$ denote the maximum and minimum values of $u^i(t)$ as determined by the DER ramping constraints (e.g., inertia). Finally, let C^i denote the limit on up and down variation in $x^i(t)$ around x_0^i as determined by capacity constraints (e.g., acceptable chemical charge or temperature range). Then, for the i^{th} DER, we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}p^i(t) &= u^i(t), \\ \frac{\mathrm{d}}{\mathrm{d}t}x^i(t) &= -a^i x^i(t) - p^i(t), \\ -\underline{u}^i \le u^i(t) \le \overline{u}^i, \quad -\underline{p}^i \le p^i(t) \le \overline{p}^i, \quad |x^i(t)| &\le C^i, \end{aligned} \tag{3.1}$$

where $a^i \ge 0$ captures the process of dissipation towards nominal energy, and $u^i(t)$ is controlled by the aggregator.

While we have considered symmetric constraints on $x^i(t)$, the formulation can be easily extended to the asymmetric case. Also, the dynamic model we adopt is consistent with those used in bulk power transmission systems to describe the regulation capabilities of units participating in AGC (see, e.g., [46]).

3.1.2 DER Coordination Problem Formulation

We assume that the aggregator does not own any DERs. To deliver the amount of frequency regulation stipulated through the clearing process of

the real-time market, the aggregator needs to coordinate the response of a collection of n heterogeneous DERs modeled as in Section 3.1.1.

We will assume that the DERs have agreed in advance to provide the service on behalf of the aggregator in exchange for some monetary compensation. For a given market-clearing price, in order to maximize its revenue, the aggregator needs to minimize its cost; thus, it needs to minimize the sum of the payments to the DERs and the penalty which it would incur if not able to follow the frequency regulation signal set by the RTO. In our formulation, the payments to the DERs are those associated with *power* and *energy* used for regulation provision. DERs may also receive a reservation payment based on *capacity* independent of whether or not they are used for service provision. From the perspective of this formulation, these would be sunk costs which would not influence the coordination scheme.

Let π_1^i denote the price per unit of power that the aggregator respectively pays DERs for providing power for both up and down regulation, and let π_2^i denote the price per unit of deviation from baseline energy. Also, let X denote the amount of power for up and down frequency regulation that the aggregator has offered in the real-time market. Additionally, let $\sigma X r(t),\ 0 \leq \sigma \leq 1$, where $r(t)$ is the value that the normalized regulation signal set by the RTO takes at time t, be the value of the signal that the aggregator needs to track at every time instant t; and let π^p denote the price per unit of power that the aggregator incurs as a penalty if it does not track the signal. Finally, let $[t_0, t_f]$ be the time interval over which the aggregator provides regulation service. Then, given (3.1), the DER coordination problem faced by the aggregator is to find functions u^i that minimize

$$J(u^1, \cdots, u^n) = \int_{t_0}^{t_f} L(p^1(t), x^1(t), \cdots, p^n(t), x^n(t), r(t))\, dt, \tag{3.2}$$

where

$$L(p^1(t), x^1(t), \cdots, p^n(t), x^n(t), r(t)) = \tag{3.3}$$
$$\pi^p |\sigma X r(t) - \sum_{i=1}^{n} p^i(t)| + \sum_{k=1}^{n} \left(\pi_1^i p^i(t) + \pi_2^i |x^i(t)|\right).$$

In (3.2), while t_0 is likely to correspond to the beginning of the period over which the aggregator needs to provide frequency regulation, t_f does not necessarily correspond to the time instant at which this period ends. In this regard, if the aggregator were to choose t_f to exactly coincide with the time at which the frequency regulation period ends, then it would maximize its revenue for this period. However, if the aggregator were to participate in subsequent periods, this strategy might not be optimal; thus, the aggregator might decide to look ahead and consider a longer time horizon to better position itself.

In (3.3), it is assumed there is a large penalty price, π^p, for error in tracking the regulation signal. A nonlinear imbalance penalty may be more accurate, but would greatly complicate the solution procedure. Regulation power is paid according to the amount of power used, which is negative for down regulation. Energy costs reflect the inconvenience cost of deviating from the baseline value, e.g., uncomfortable temperature, insufficient battery charge, or no hot water. This function could be generalized to include mileage payments, which account for increased maintenance costs due to cycling the equipment.

The normalized regulation signal, $r(t)$, is computed in real-time by the RTO based on the frequency error and inter-area power exchange errors (see, e.g., [46]); thus, this signal is not known a priori. This uncertainty adds a crucial complicating factor for the aggregator. In following section, we provide the top layer of a bilayer architecture that the aggregator can use to provide a solution to the DER coordination problem. In the top layer, all

costs are considered and regulation signal forecasts will be used. The bottom layer, which will be proposed in Chapter 4, will regulate around the top layer solution to minimize short-term tracking error.

3.2 Top Layer DER Coordination Scheme

We first discretize the DER coordination problem as defined in (3.1) – (3.3), and provide an exact solution for the case where the regulation signal is known a priori to the aggregator. We then discuss the effects of uncertainty and long time horizons.

3.2.1 Perfect Information, Fixed-Horizon

We will show that the DER coordination problem reduces to a linear program under perfect information with a fixed service interval. To this end, define $x(t) = [p^1(t), x^1(t), \cdots, p^n(t), x^n(t)]^T$, and $u(t) = [u^1(t), \cdots, u^n(t)]^T$; then, the differential equations in (3.1) can be written as

$$\dot{x}(t) = \widetilde{A}x(t) + \widetilde{B}u(t),$$

where

$$\widetilde{A}^i = \begin{bmatrix} 0 & 0 \\ -1 & -a^i \end{bmatrix} \quad \widetilde{B}^i = \begin{bmatrix} & \hat{e}^i & \\ 0 & \cdots & 0 \end{bmatrix}$$

$$\widetilde{A} = \begin{bmatrix} \widetilde{A}^1 & 0 & \cdots & 0 \\ 0 & \widetilde{A}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{A}^n \end{bmatrix} \quad \widetilde{B} = \begin{bmatrix} \widetilde{B}^1 \\ \widetilde{B}^2 \\ \vdots \\ \widetilde{B}^n \end{bmatrix},$$

with $\hat{e}^i$ being the unit row vector pointing in dimension i. This model can be replaced by a discrete-time state-space model of the form

$$x_k = Ax_{k-1} + Bu_{k-1},\ k = 1, \ldots, N, \tag{3.4}$$

where $N = (t_f - t_0)/\Delta T_1 \in \mathbb{N}$, $x_k = x(k\Delta T_1 + t_0)$, $u_k = u(k\Delta T_1 + t_0)$, $A = I + \widetilde{A}\Delta T_1$, and $B = \widetilde{B}\Delta T_1$. As is standard, the discretization error can be made negligible by choice of sampling time ΔT_1.

The constraints in (3.1) can also be compactly written in matrix form as follows:

$$E_x x_k \leq F_x, \quad E_u u_k \leq F_u, \tag{3.5}$$

where

$$\begin{aligned}
E_x &= \left[\hat{e}^1,\ -\hat{e}^1,\ \hat{e}^2,\ -\hat{e}^2,\ \cdots,\ \hat{e}^{2n-1},\ -\hat{e}^{2n-1},\ \hat{e}^{2n},\ -\hat{e}^{2n}\right]^T \\
F_x &= \left[\overline{p}^1,\ \underline{p}^1,\ C^1,\ C^1,\ \cdots,\ \overline{p}^n,\ \underline{p}^n,\ C^n,\ C^n\right]^T \\
E_u &= \left[\hat{e}^1,\ -\hat{e}^1,\ \cdots,\ \hat{e}^n,\ -\hat{e}^n\right]^T \\
F_u &= \left[\overline{u}^1,\ \underline{u}^1,\ \cdots,\ \overline{u}^n,\ \underline{u}^n\right]^T.
\end{aligned} \tag{3.6}$$

The cost functional in (3.2) can also be discretized as:

$$J(u) = \Delta T_1 \sum_{k=1}^{N} \left(Q_1 x_k + \|Q_2 x_k + R r_k\|_1\right), \tag{3.7}$$

where $r_k = r(k\Delta T_1 + t_0),\ k = 1, \ldots, N$, and

$$\begin{aligned}
Q_1 &= \left[0\hat{e}^1,\ \pi_1^1\hat{e}^1,\ 0\hat{e}^2,\ \cdots,\ \pi_1^n\hat{e}^{2n-1},\ 0\hat{e}^{2n}\right]^T \\
Q_2 &= \left[-\pi^p \textstyle\sum_{i=1}^{n} \hat{e}^{2i-1},\ 0\hat{e}^1,\ \pi_2^1\hat{e}^2,\ \cdots,\ 0\hat{e}^{2n-1},\ \pi_2^n\hat{e}^{2n}\right]^T \\
R &= \left[\pi^p \sigma X,\ 0,\ \cdots,\ 0\right]^T.
\end{aligned} \tag{3.8}$$

Combining (3.4) – (3.8), we can formulate the optimization problem

$$\begin{aligned} u_k^* = \quad & \underset{u}{\text{argmin}} \quad \Delta T_1 \sum_{k=1}^{N} \left(Q_1 x_k + \| Q_2 x_k + R r_k \|_1 \right), \\ & \text{subject to} \quad E_x x_k \leq F_x, \\ & \qquad\qquad\quad E_u u_k \leq F_u, \\ & \qquad\qquad\quad x_k = A x_{k-1} + B u_{k-1}, \end{aligned} \tag{3.9}$$

the solution of which can be used to solve the DER coordination problem as defined by (3.1) – (3.3).

Using the technique laid out in the Appendix, the optimization problem in (3.9) can be cast as a linear program of the form

$$\begin{aligned} & \underset{y}{\text{minimize}} \quad f^T y \\ & \text{subject to} \quad G y \leq h, \end{aligned} \tag{3.10}$$

where $f \in \mathbb{R}^{(3n+1)N}$, $y \in \mathbb{R}^{(3n+1)N}$, $G \in \mathbb{R}^{(3n+1)N \times (10n+2)N}$, $h \in \mathbb{R}^{(10n+2)N}$. This linear program can be solved using any of a number of well documented linear programming algorithms in the literature (see, e.g., [47, 48]).

3.2.2 Imperfect Information, Receding Horizon

To solve the DER coordination problem via (3.10), it is necessary to have complete information of the values that the regulation signal $r(t)$ takes for all $t = t_0 + k\Delta T_1,\ k = 1, \ldots, N$. Next, we propose an MPC-based solution to the DER coordination problem when $r(t)$ is not known in advance.

We use the subscript $k + l|l$ to denote an estimate, made at time $t_l = t_0 + l\Delta T_1$, of the value that a variable takes at time $t_{k+l} = t_0 + (k + l)\,\Delta T_1$, e.g., $n_{k+l|l}$ denotes the estimate of the regulation signal k steps ahead of time $t_0 + l\Delta T_1$ (this, and other notation used here was adopted from [49]). With this notation we can write an optimization program similar to (3.9), but with

an arbitrary starting point and no requirement of perfect knowledge of the future:

$$\begin{aligned} u_k^* = \quad & \underset{u}{\text{argmin}} \quad \Delta T_1 \sum_{k=1}^{N} \left(Q_1 x_{l+k|l} + \left\| Q_2 x_{l+k|l} + R r_{l+k|l} \right\|_1 \right), \\ & \text{subject to} \quad E_x x_{l+k|l} \leq F_x, \\ & \qquad\qquad\quad E_u u_{l+k|l} \leq F_u, \\ & \qquad\qquad\quad x_{l+k|l} = A x_{l+k-1|l} + B u_{l+k-1|l}, \end{aligned} \tag{3.11}$$

with E_x, F_x, E_u, and F_u as in (3.6); and Q_1, Q_2, and R as in (3.8). This optimization can also be rewritten in the form in (3.10).

Suppose the solution to (3.11) was calculated with $l = 0$. This is equivalent to (3.9) with an arbitrary forecast. This solution only depends on the current state $x_{0|0}$ and an estimate of future values to calculate the optimal control for the present time and next $N - 1$ time steps. The first calculated optimal control input, denoted $u^*_{0|0}$, is then applied. At time $t_0 + \Delta T_1$, the system state may not have evolved as predicted due to an inaccurate forecast of $r(t_0 + \Delta T_1)$, incorrect system parameters, or unmodeled disturbances, i.e., $x_{1|0} \neq x_{1|1}$, $r_{1|0} \neq r_{1|1}$. Thus, in order to obtain a better solution, this new information should be taken into account.

At the next time step we assign $l \leftarrow l + 1$ and update $x_{l|l}$ with new measurements, and the forecast of $r(t)$ with the latest information. We also assign $N \leftarrow N - 1$ to avoid making choices that would create higher costs for $t \leq t_f$ in exchange for even lower costs when $t > t_f$. The problem is solved again, giving us a new control plan taking the latest information into account. The first optimal control input from this optimization is applied at this time step, and so on. The procedure continues until $l = N - 1$. Then, at the following time step, we have that $t = t_f$, thus arriving to the final solution.

If the aggregator continues providing frequency regulation beyond time t_f

because it has cleared subsequent markets, it is not desirable for the aggregator to be left in a position where it cannot, or cannot profitably, meet future obligations. For this reason the aggregator would likely prefer to use a technique that takes a more far-sighted view. This can be done by performing the steps described in the previous paragraph without decrementing N. The procedure continues until some arbitrary time step $l = M$, $M > N$. This technique is known as receding horizon control because as time progresses, the optimization window is also extended. A receding horizon will be used in the numerical examples presented in the next section.

As future values of $r(t)$ are unknown, forecasts must be used. On a second by second basis the regulation signal does not change much, so in the short term we expect future values of the regulation signal to be similar to the present value. This gives rise to the persistence forecast

$$r_{k+l|l} = r_{l|l},\ k \geq 1. \tag{3.12}$$

However, the current value of the regulation signal tells us practically nothing about its value in the distant future; in this case it is best to predict the mean value. In the medium term it would be logical to interpolate between the present value and the mean. Two possible methods that take this into account are linear prediction

$$r_{k+l|l} = r_{l|l} \cdot \max\left(1 - \alpha_1 k \Delta T_1, 0\right),\ k \geq 1,\ \alpha_1 > 0, \tag{3.13}$$

and exponential prediction

$$r_{k+l|l} = r_{l|l} \cdot e^{-\alpha_1 k \Delta T_1},\ k \geq 1,\ \alpha_1 > 0. \tag{3.14}$$

Because the uncertainty of the forecasts increases with the prediction horizon, the objective function used by the predictive controller can be multiplied

by a factor which de-weights values that are further into the future, i.e.,

$$J(u) = \Delta T_1 \sum_{k=1}^{N} \left(e^{-\alpha_2 k \Delta T_1} \left(Q_1 x_{l+k|l} + \| Q_2 x_{l+k|l} + R r_{l+k|l} \|_1 \right) \right). \tag{3.15}$$

An "oracle" solution that assumes future values are known, i.e., $r_{k+l|l} = r_{k+l}$ is also considered for comparison purposes.

3.3 Case Studies

We next test the performance of the DER coordination architecture introduced in Section 3.2. We first describe the features of the dataset used in all case studies. Then, we present simulation results.

3.3.1 Dataset

We utilize "Normalized Dynamic and Traditional Regulation Signals" from PJM (the world's largest wholesale electricity market [50]) for the period January 1-18, 2013. This data is available at [42], and includes two regulation signals—RegD, a fast response signal, and RegA, a filtered version of RegD for slower ramping generators; both signals are updated every 2 s.

Figure 3.1 shows a representative segment of the aforementioned data. We choose the fast signal RegD for our studies because DERs are expected to have faster ramping rates than conventional generators, and because storage devices work best with a zero-mean signal. Average cost and capability data for the same period is calculated using data available at [51]. From January 1–18, 2013, the average cost of regulation capacity was \$14.3/MWh, and 947 MW of capacity was dispatched.

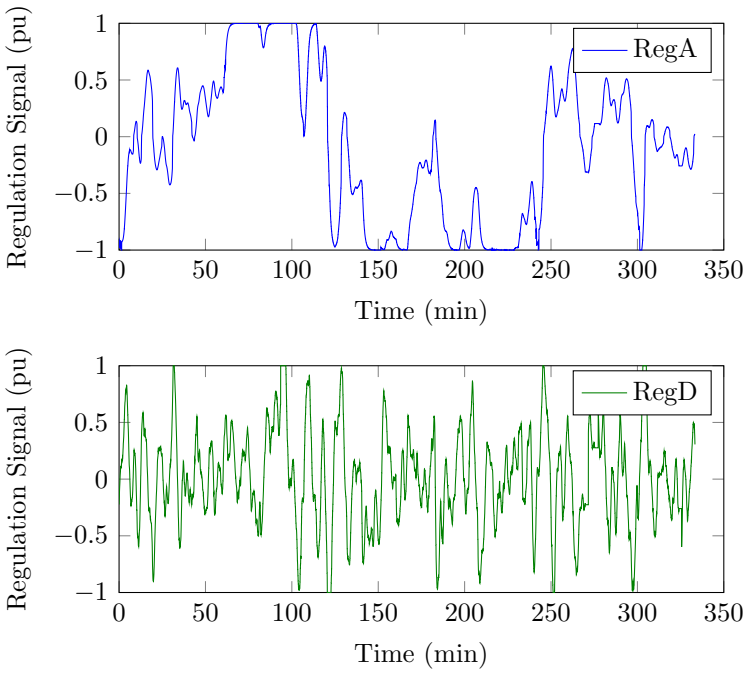

Figure 3.1: Segment of PJM regulation data. RegD is the fast dynamic response signal. RegA is a filtered version for slower responding units.

3.3.2 Base Case

A case study involving two units was investigated. The first unit has low ramp rate limits, but a low cost and high energy capacity. The second unit is able to ramp its power consumption more quickly, but is more expensive and has a lower energy capacity. The parameters for the studies are given in Table 3.1. The penalty price π^p was chosen to be ten times the PJM average capability clearing price. It was assumed the aggregator's dispatched capability made up about 2% of the total market. A time step, ΔT_1, of 20 s is used, and the prediction horizon T is set to 5 min. We use receding horizon control to calculate u^* for 60 min. The other parameters were selected to show different types of behavior that can arise.

A study was performed to find optimal values of α_1 and α_2 for each forecasting technique. Figure 3.2 illustrates the different results that arise based on the different forecasts. The total cost for each of the four forecasting techniques is given in Table 3.2. The oracle forecast has the lowest total cost

Table 3.1: Case study parameters

Parameter	Description	Value	Unit
π_1^1	Regulation Price	14.3	\$/MW
π_1^2	Regulation Price	42.9	\$/MW
π_2^1,π_2^2	Energy Price	0	\$/MWh
π^p	Imbalance Price	143	\$/MWh
$\overline{u}^1$, $\underline{u}^1$	Ramp Limit	0.04	MW/s
$\overline{u}^2$, $\underline{u}^2$	Ramp Limit	0.096	MW/s
$\overline{p}^1$, $\underline{p}^1$	Regulation Limit	11.9	MW
$\overline{p}^2$, $\underline{p}^2$	Regulation Limit	7.9	MW
C^1	Storage Energy Limit	0.45	MWh
C^2	Storage Energy Limit	0.15	MWh
a^1, a^2	Dissipation Constant	0	s^{-1}
σX	Regulation Signal Magnitude	18.9	MW
T	Prediction Horizon	5	min

Table 3.2: Base case total cost for different forecasting methods

Forecast Method	ΔT_1	Total Cost (\$)
Persistence	20 s	481.32
Linear	20 s	466.91
Exponential	20 s	468.91
Oracle	20 s	365.05

by a large margin. The results with the linear and exponential forecasting methods are both very similar, and, as expected, improve upon the results obtained with the persistence model. It is hypothesized that the forecasts could be further improved by utilizing the past data in addition to the present value of the regulation signal. This would allow forecasts that predict based on a short moving average, or even more complicated schemes.

3.3.3 Sensitivity

The optimal values of the forecast parameters were found using a gradient-based search procedure. Figure 3.3 plots the total operating costs versus the forecast parameters for the linear prediction method. We find that cost is much more sensitive to α_1, which affects the forecast itself, compared to α_2, which affects how much weight is given to forecasted values when calculating predicted costs.

3.4 Concluding Remarks

We have shown that the DER coordination problem that the aggregator is faced with can be cast as an optimal control problem. By coordinating the response of the DERs, the aggregator can sell this service in real-time regulation markets.

A related problem that is worthy of future exploration is the decision-making process that the aggregator uses under this framework to choose the capability and price to offer in the market under the uncertainty of DER parameters and regulation signal frequency content.

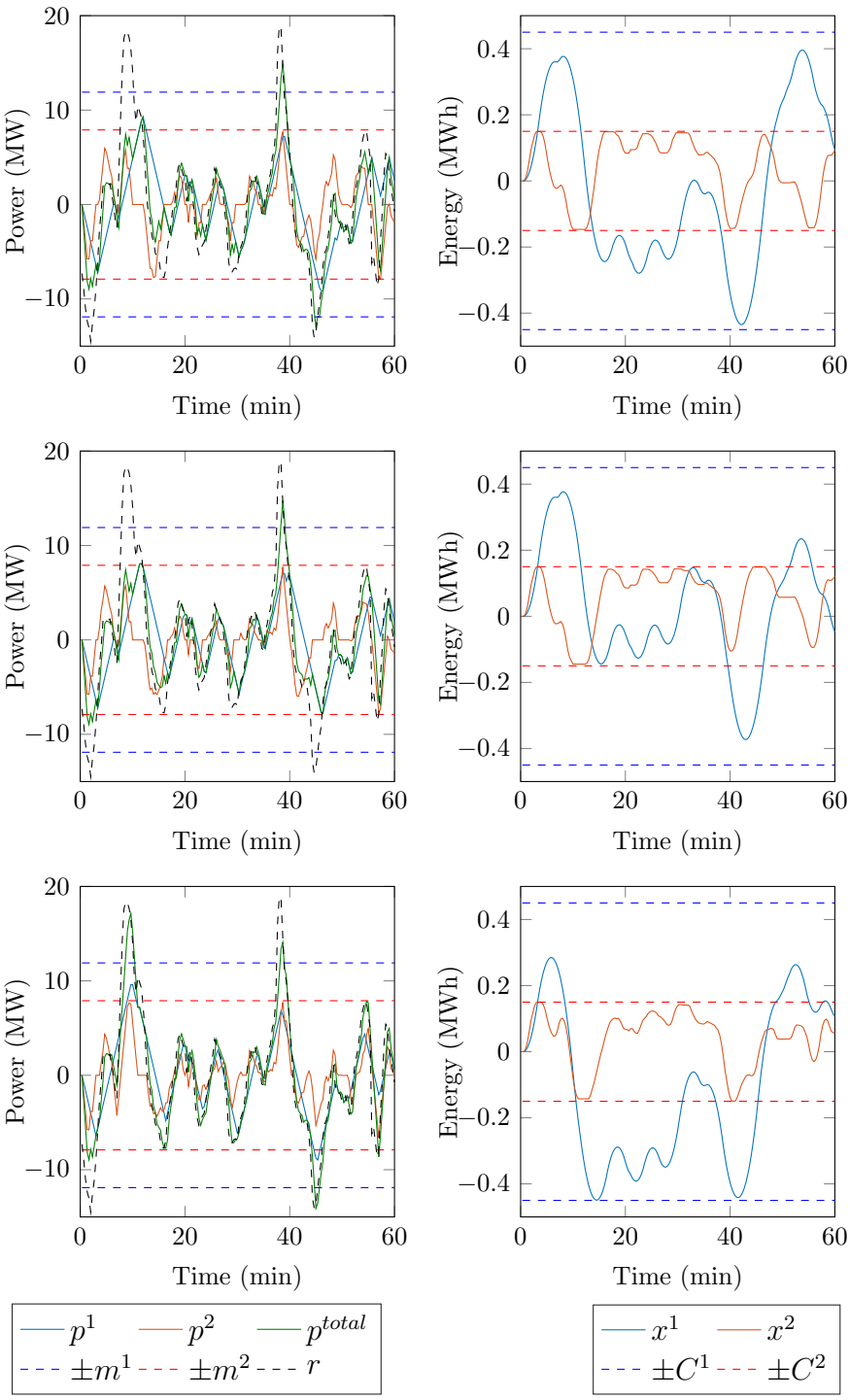

Figure 3.2: Numerical simulation results comparing forecasting strategies. Left to right: power, state of charge. Top to bottom: Persistence Forecast, Exponential Forecast, Oracle Forecast.

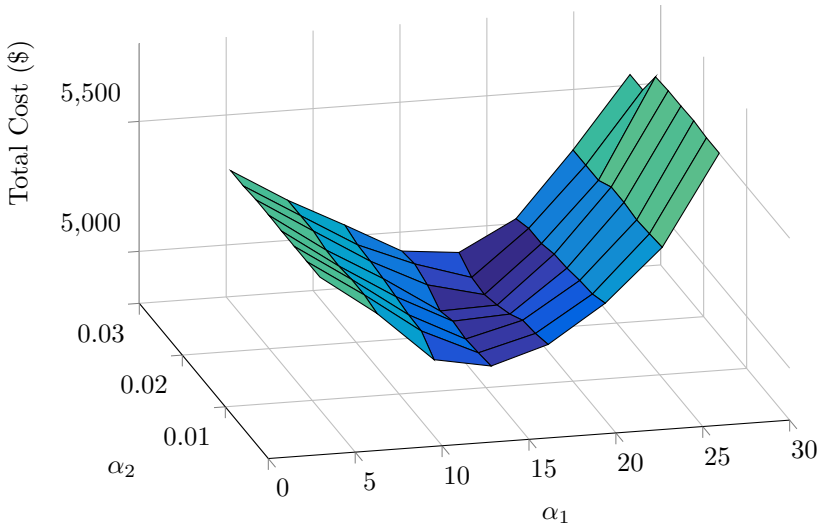

Figure 3.3: Sensitivity of operating costs to forecast parameters.

CHAPTER 4

BILAYER COORDINATION OF DISTRIBUTED ENERGY RESOURCES

In this chapter, we address the limitations of a single layer coordination strategy. We provide background on traditional AGC control; we then propose modifications to allow its use with constraint limited DERs inside of the aggregator framework. We show that this control can be integrated with the predictive control of Chapter 3, forming a bilayer control. Case studies show the performance of each layer individually and the combined bilayer control.

4.1 Background

In this section, we explore the need for a bottom control layer and develop the necessary preliminaries for formulating the control.

4.1.1 Effect of Uncertainty on Top Layer Control

In our framework, an aggregator participating in the frequency regulation market faces two primary sources of uncertainty in optimal coordination problem: uncertainty of parameters of the DERs and uncertainty in the regulation signal.

Consider a building which is modeled by a virtual battery as discussed in Chapter 2. To understand the uncertainty in the model parameters, it is first necessary to study how uncertainty in the building parameters and exogenous variables (e.g., ambient temperature, solar loads, internal loads) translate into changes in the battery model. This can be accomplished by

performing sensitivity studies using the procedures presented in Chapter 2. Historical data can be used to determine the probability density function of a realized value given a forecast some time period in advance. Distributions of battery parameters can then be found using transformation of random variables through a multi-variate function. This procedure is straightforward with sufficient data for the specific model in question, so we will assume the end result as known. In this work, we will treat the expected parameter value as the true value and deviations as disturbances to be rejected by the bottom control layer. A natural extension of this work would be to address this problem probabilistically.

Chapter 3 acknowledged uncertainty in the regulation signal and examined ad-hoc signal forecasting techniques. Because the forecasting techniques will be imperfect, receding horizon control was proposed in order to constantly use the latest measurements and forecasts. However, the framework does not have a closed-loop control scheme to account for disturbances between runs of the MPC optimization. In this chapter, we improve the control scheme to account for the uncertainty in resource parameters and regulation signal forecasts and address robustness, stability, and time-scale limitations of the top layer control.

4.1.2 Traditional Control

Traditional AGC utilizes an integral or PI controller with the objective of driving the ACE to zero [52]. Consider a power system balancing area with N generators. Let P denote the actual power interchange between the area and all connected areas, and let f denote the actual frequency in the area. Then, the ACE is given by

$$ACE = (P - P^{sch}) - b(f - f^{nom}), \tag{4.1}$$

where P^{sch} is the scheduled interchange power, f^{nom} is the nominal frequency in the area, and b is the area's bias factor.

To calculate P_i^{ref}, the power reference signal to be followed by unit i, PI control is used. Let P_i denote the current power generation by unit i, and let P_i^{ed} denote the optimal power generation at unit i as calculated by the economic dispatch procedure. Then, the controller variables z_1 and z_2, which in steady state will equal the total generation in the area, are

$$\frac{dz_2}{dt} = \eta_2 ACE + \sum_{i=1}^{N} \left(P_i - P_i^{ref} \right) \tag{4.2}$$

$$z_1 = \eta_1 ACE + z_2, \tag{4.3}$$

where η_1 and η_2 are controller gains. The reference signal for each generator is calculated as

$$P_i^{ref} = P_i^{ed} + \beta_i (z_1 - \sum_{j=1}^{N} P_j^{ed}), \tag{4.4}$$

with $\sum_{i=1}^{N} \beta_i = 1$, where participation factors β_i are based on the economics and operating constraints of each unit [53].

4.1.3 Adaptation for DER Control

To adapt the traditional scheme to be used by an aggregator, several changes must be made. First, the objective must change from minimizing the ACE to following a regulation signal command from the system operator. Second, commands must be issued in terms of deviation from baseline power rather than absolute power values. Third, we must respect the ramping, power, and energy constraints of the underlying resources. Finally, we must incorporate the top layer solution. All of these issues will be addressed in Section 4.2.

There are multiple advantages of adding an additional control layer. First, the stability of the top layer is difficult to analyze, whereas the faster control will be tractable. Second, an MPC-based solution can be slow; using a 2 s

time step, even the two DER system of Section 3.3 cannot be simulated in real-time by a desktop computer. A realistic problem would have a much higher order, which may make it too slow to handle the fast-changing regulation signal. A stochastic controller would require even more computation. An AGC-like controller is faster because of its simplicity. Finally, the closed-loop feedback in the fast controller compensates for parameter uncertainty and disturbances.

4.2 Bottom Layer Regulation Provision

We next formulate the second control layer, which provides closed-loop control so as to minimize the error that arises due to forecast error, plant-model mismatch, and the slower speed of the optimal control. Constraints are given special consideration, which removes the possibility of violations.

4.2.1 Unconstrained, Continuous Time Formulation

For clarity, we begin by introducing an unconstrained formulation. In this case, the continuous time DER power delivery model in (3.1) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}p^i(t) = u^i(t) \tag{4.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x^i(t) = -a^i x^i(t) - p^i(t), \tag{4.6}$$

where i indexes n DERs, x^i is the state of charge, a^i is the dissipation constant, and p^i is the power being charged or discharged.

As described in Section 4.1.2, conventional AGC, which is typically implemented using proportional-integral (PI) control, is used to coordinate participants in frequency regulation markets [53]. It is natural to apply a similar closed-loop control scheme to the problem at hand. Unlike predictive control, PI control is simple and can be performed very quickly. The controller

has a single state variable, z_k, which is proportional to the integral of the tracking error, i.e., the difference between the regulation signal, $\sigma X r(t)$, and the sum of regulation power provided by each DER. We will utilize PI control to control $\sum_{i=1}^{n} p^i(t)$ via $u^i(t)$ to follow $\sigma X r(t)$. For the integral part of the controller, we define a new state variable z, which is proportional to the integral of the regulation error. Its dynamics are

$$\frac{\mathrm{d}}{\mathrm{d}t} z(t) = \eta_2 \left(\sigma X r(t) - \sum_{i=1}^{n} p^i(t) \right), \tag{4.7}$$

where η_2 is the integral gain. The desired power is then

$$p^i(t) = \beta^i \left(\eta_1 \left(\sigma X r(t) - \sum_{i=1}^{n} p^i(t) \right) + z(t) \right), \tag{4.8}$$

where η_1 is the proportional gain, β^i is the participation factor, with $\sum_{i=1}^{n} \beta^i = 1$. Combining (4.5) and (4.8), we calculate $u^i(t)$ to be

$$u^i(t) = \frac{\mathrm{d}}{\mathrm{d}t} p^i(t) = \beta^i \left(\eta_1 \left(\sigma X \frac{\mathrm{d}}{\mathrm{d}t} r(t) - \frac{\mathrm{d}}{\mathrm{d}t} p(t) \right) + \frac{\mathrm{d}}{\mathrm{d}t} z(t) \right). \tag{4.9}$$

Eliminating $z(t)$ from (4.9) using (4.7) results in

$$u^i(t) = \beta^i \left(\eta_1 \left(\sigma X \frac{\mathrm{d}}{\mathrm{d}t} r(t) - \frac{\mathrm{d}}{\mathrm{d}t} p(t) \right) + \eta_2 \left(\sigma X r(t) - \sum_{i=1}^{n} p^i(t) \right) \right).$$

4.2.2 Unconstrained, Discrete Time Formulation

To discretize the formulations, we utilize the Forward Euler method, with subscript k denoting the time index. Equations (4.5)–(4.8) become

$$p^i_{k+1} = p^i_k + \Delta T_2 u^i_k \tag{4.10}$$

$$x^i_{k+1} = (1 - a^i \Delta T_2) x^i_k - \Delta T_2 p^i_k \tag{4.11}$$

$$z_{k+1} = z_k + \Delta T_2 \eta_2 \left(\sigma X r_k - \sum_{i=1}^{n} p^i_k \right) \tag{4.12}$$

$$p^i_{k+1} = \beta^i \left(\eta_1 \left(\sigma X r_k - \sum_{i=1}^{n} p^i_k \right) + z_{k+1} \right), \tag{4.13}$$

where $\Delta T_2 < \Delta T_1$ is the time step of the bottom layer control.

By combining (4.10) and (4.13), solving for u^i_k, and eliminating z_k using (4.12), we obtain

$$u^i_k = \eta_1 \left(\sigma X \frac{r_k - r_{k-1}}{\Delta T_2} - \sum_{i=1}^{n} \frac{p^i_k - p^i_{k-1}}{\Delta T_2} \right) + \eta_2 \left(\sigma X r_k - \sum_{i=1}^{n} p^i_k \right). \tag{4.14}$$

4.2.3 Constrained, Discrete Time Formulation

Now, we consider the constraints from (3.1),

$$-\underline{u}^i \le u^i(t) \le \overline{u}^i, \quad -\underline{p}^i \le p^i(t) \le \overline{p}^i, \quad |x^i(t)| \le C^i.$$

Clearly, (4.13) and (4.14) may no longer hold, as these values could violate these constraints. We introduce the superscript * to denote the value desired before considering constraints. Then, we can write

$$p^{i*}_{k+1} = \beta^i \left(\eta_1 \left(\sigma X r_k - \sum_{i=1}^{n} p^i_k \right) + z_{k+1} \right) \tag{4.15}$$

$$u^{i*}_k = \frac{p^{i*}_{k+1} - p^i_k}{\Delta T_2}. \tag{4.16}$$

For each pair of constraints in (3.1), we define an interval of feasible values for u_k^i. The ramping constraint interval straightforwardly limits the control input at the *current* time step:

$$S_1^i = \left[-\underline{u}^i, \overline{u}^i\right].$$

The interval associated with the power constraints ensures no control action is taken that would cause the power to exceed a limit at the *following* time step:

$$S_2^i = \left[\frac{-\underline{p}^i - p_k^i}{\Delta T_2}, \frac{\overline{p}^i - p_k^i}{\Delta T_2}\right].$$

The interval associated with the energy constraints is more difficult to calculate. The control input takes at least two time steps to affect the state of charge, and control action could potentially need to be taken even further in advance in order to avoid constraint violations. Thus, determining these limits is more difficult. The variables $\check{p}_{\check{k}^*}$ and $\hat{p}_{\hat{k}^*}$ indicate lower and upper bounds on power derived from energy capacity constraints; the exact procedure for calculating them will be given in the Section 4.2.4. For now, we define

$$S_3^i = \left[\frac{\check{p}_{\check{k}^*}^i - p_k^i}{\Delta T_2}, \frac{\hat{p}_{\hat{k}^*}^i - p_k^i}{\Delta T_2}\right]. \tag{4.17}$$

The constrained controller solution is the projection of the desired solution onto the intersection of the three feasible intervals, i.e., $S^i = S_1^i \cap S_1^i \cap S_3^i$. In other words,

$$u_k^i = \left[\frac{p_{k+1}^{i*} - p_k^i}{\Delta T_2}\right]^+, \tag{4.18}$$

where $[\cdot]^+$ indicates projection onto S^i. This ensures all constraints in (3.1) are satisfied at every time step.

4.2.4 Capacity Constraints

Here, we solve a backwards reachability problem to find the values $\hat{p}_{\hat{k}^*}$ and $\hat{p}_{\hat{k}^*}$ used in (4.17).

For a single DER, the dynamic equations in (3.1) can be replaced by a discrete-time state-space model of the form

$$X_k = AX_{k-1} + Bu_{k-1}, \tag{4.19}$$

where $X_k = [p(k\Delta T_2 + t_0),\ x(k\Delta T_2 + t_0)]^T$, $u_k = u(k\Delta T_2 + t_0)$, $A = I + \widetilde{A}\Delta T_2$, and $B = \widetilde{B}\Delta T_2$.

We consider the bound $x \leq C$. If $x = C$, we must enforce the constraint $p \geq -aC$. Because $\overline{u}$ limits how quickly we can increase p, if $p < -aC$ we must also make sure it is possible to ramp to $p = -aC$ before $x > C$. We can find the limits on u by starting at $\check{X}_0 = [-aC \quad C]^T$, and working the dynamics (4.19) backwards with $u_k = \overline{u}$ using the equation

$$\check{X}_{k+1} = A^{-1}\left(\check{X}_k - B\overline{u}\right),$$

with k increasing until $\check{p}_k \leq -\underline{p}$ or $\check{x}_k \leq -C$, where $\check{X}_k = [\check{p}_k \quad \check{x}_k]^T$. This procedure only needs to be done once; the values can be stored and used at each time step in the future.

Then, at each time step, calculate

$$\begin{aligned} \check{k}^* = \arg\min_k \quad & \check{x}_k - (x - \Delta T_2(p + ax)) \\ \text{subject to} \quad & \check{x}_k \geq (x - \Delta T_2(p + ax)). \end{aligned}$$

This solution will always exist and be unique. Although this may look like a computationally difficult optimization problem, it is not; since $\check{x}_k$ is monotonous, its value can be obtained via a lookup table. The limit on u

due to the constraint $x \leq C$ is then

$$\frac{\check{p}_{\check{k}^*} - p_k}{\Delta T_2}.$$

$\hat{p}_{\hat{k}^*}$ is calculated similarly, starting at $\hat{X}_0 = [aC \quad -C]^T$, and working the dynamics (4.19) backwards with $u_k = -\underline{u}$.

4.2.5 Final Bilayer Formulation

Let $p_{k+1}^{i\text{MPC}}$ denote the optimal power value for unit i at time step $k+1$ as calculated by the most recent solution from the MPC-based top layer control. Then, the final bilayer controller can be formulated as follows:

$$\begin{aligned}
z_{k+1} &= z_k + \Delta T_2 \eta_2 \left(\sigma X r_k - \sum_{i=1}^{n} p_k^i \right) \\
p_{k+1}^{i*} &= \beta^i \left(\eta_1 \left(\sigma X r_k - \sum_{i=1}^{n} p_k^i \right) + z_{k+1} \right) + p_{k+1}^{i\text{MPC}} \\
u_k^i &= \left[\frac{p_{k+1}^{i*} - p_k^i}{\Delta T_2} \right]^+ .
\end{aligned}$$

We reset the integral control, i.e., $z_k = 0$, when a new top layer solution is found. This control can also be used without the top layer by fixing $p_{k+1}^{i\text{MPC}} = 0$ and not resetting z_k.

Participation factor selection must balance optimality and complexity. The simplest possible values of β^i would be the constants $1/n$. The optimal values would be time-varying and would consider the costs, current values of state variables, and limits of each unit. The computation effort required to arrive at such a solution would likely be more effective if it were spent working on solving the MPC problem more quickly. A compromise solution would use a heuristic to update β^i infrequently, likely based on the slow optimization solution.

4.3 Case Studies

We compare the performance of the DER coordination architecture introduced in Section 3.2 with the new fast and bilayer controls from Section 4.2. Then, we explore the sensitivity of the solution to key parameters.

4.3.1 Base Case

The parameters from Table 3.1 will be used again for this study. Time steps ΔT_1 and ΔT_2 were set to 2 s and 20 s, respectively.

Figure 4.1 illustrates the behavior of the fast bottom layer controller, the slow top layer, and the bilayer control. We observe that the fast control better tracks the small variations in $r(t)$, whereas the slow control uses prediction to spend less time bounded by energy constraints. The bilayer controller combines these two advantages. As expected, results in Table 4.1 show that the top layer performed better than the bottom at equal time scales. If the MPC is constrained to run more slowly than the PI control, its performance decreases, but it still outperforms the bottom layer running ten times faster. However, combining the two methods results in cost values lower than the slow top layer, while requiring significantly less computation power than the fast top layer control.

Both the fast and slow control loops take measures to ensure solutions are feasible. However, feasibility problems are encountered when the time scales are mixed. The MPC problem can be given an initial condition which is feasible on the fast time scale, but unfeasible on the slow time scale. This must be resolved by requiring all fast control solutions to be slow time feasible, or by relaxing constraints on the slow time solution.

Table 4.1: Base case total cost for different control strategies

Control	Time-Step	Total Cost ($)
Bottom Layer	20 s	584.39
Bottom Layer	2 s	480.58
Top Layer	20 s	470.37
Top Layer	2 s	362.10
Bilayer	2 s/20 s	387.13

4.3.2 Sensitivity

Here, we explore the sensitivity of the control scheme total cost with regard to participation factors and controller gains. As in Chapter 3, base case controller parameters were optimized using a 12 hour long segment of the PJM regulation signal as training data. Figure 4.2 shows that the optimal participation factors require DER 1, which has a lower regulation price, to participate less than DER 2. However, DER 2 still has significant participation despite its higher price due to its faster ramping ability. Figure 4.3 illustrates the sensitivity of cost to controller gains. Even with a 12 h simulation period, these functions are non-convex, making optimization difficult. Multi-start methods were used to locate the global minimum.

4.4 Concluding Remarks

We have discussed a framework for an aggregator to coordinate the amount of power provided by a collection of heterogeneous DERs for providing up and down frequency regulation services. We have provided a bilayer control scheme that enables the aggregator to solve this problem by using slow but accurate predictive control techniques while also responding quickly to second-to-second variations in the regulation signal.

We note that in the top layer the controller may intentionally deviate from the regulation signal if the costs of utilizing DERs are too high compared to the penalty cost or in order to prepare for future regulation, but the bottom

control layer attempts to minimize the current regulation error at all costs. Even with a high penalty price, sub-optimal participation factors may cause the controller to deviate significantly from the optimal solution from the top layer. Thus, a balance must be struck between responding to fast variations and neglecting price information from the top layer solution. This can be done through choice of controller gains η_1 and η_2, which should be optimized using historical data.

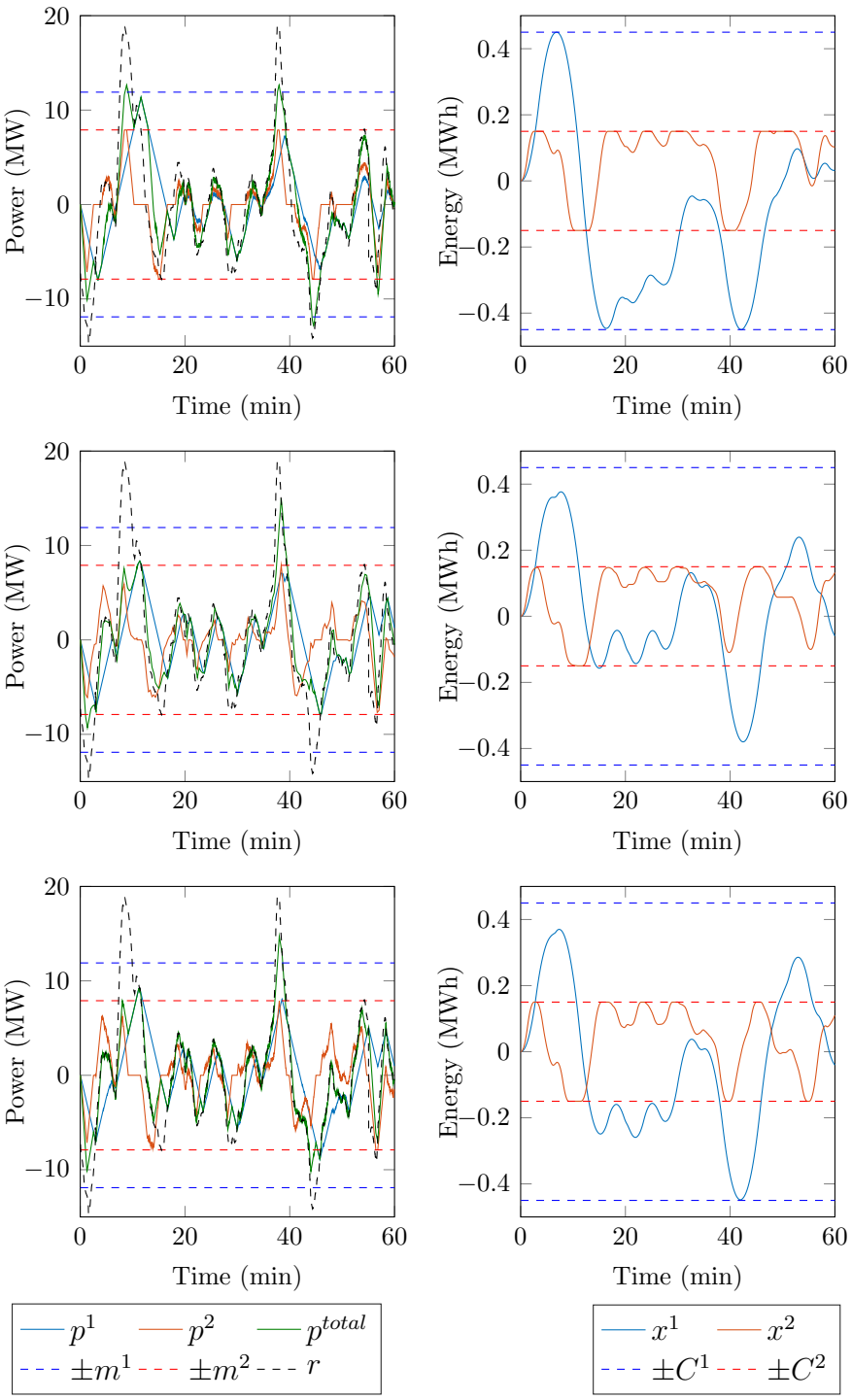

Figure 4.1: Numerical simulation results comparing control strategies. Left to right: power, state of charge. Top to bottom: Bottom layer control, Top layer control, Bilayer control.

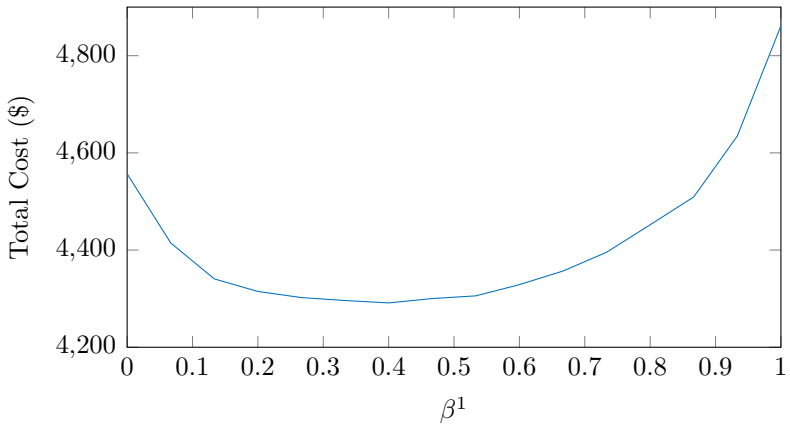

Figure 4.2: Sensitivity of operating costs to participation factors.

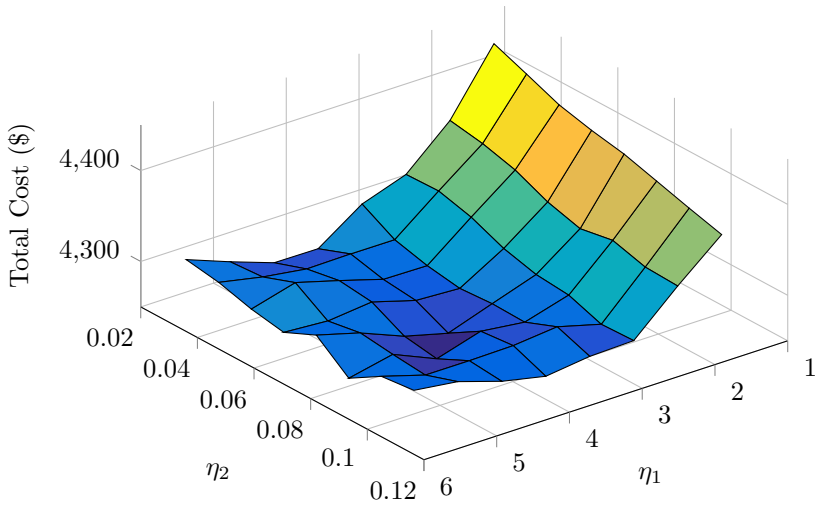

Figure 4.3: Sensitivity of operating costs to bottom layer controller parameters.

CHAPTER 5

CONCLUDING REMARKS

In this chapter, we revisit the contributions of the dissertation, and conclude with some final thoughts and observations.

5.1 Thesis Summary and Contributions

In this dissertation, we have addressed many of the challenges regarding the use of flexible loads to provide frequency regulation. In Chapter 1, we gave background on the ongoing changes that are making this idea possible and valuable, outlined the major parts of the framework, and examined related literature. In Chapter 2, we developed a model to capture the flexibility of loads through an abstraction that removes needless complexity and showed how to obtain its parameters. In the course of this development, we proposed a controller that enables resources to utilize their flexibility without compromising end function. The proposed methods were verified through simulation of a realistic model of an airport terminal building. Chapter 3 built a layer of control on top of these models. We proposed an aggregator entity which works with large numbers of DERs to enable them to participate in the frequency regulation market. We examined the costs of providing this service and proposed a controller to maximize profits. Chapter 4 improved the coordination strategy by introducing a second control layer based on traditional AGC and discussed how to incorporate its benefits while retaining the advantages of the predictive top layer control.

5.2 Conclusion and Future Work

From this dissertation and related literature, we conclude it is technically feasible and economically valuable to utilize flexible loads to provide frequency regulation. Further development is recommended in the area of detailed models of building and HVAC systems, with special attention to time constants, e.g., of the chiller [54], or through a second-order thermal zone model [55]. This work has used power balance equations to model the regulation service; it may also be of interest model network effects, where appropriate.

Further work could also be done on the internal control of various flexible loads with different modeling frameworks. For example, the virtual battery model could likely capture the flexibility of such systems such as aggregations of smaller deferable loads [56,57]. Studies could investigate the effect different control schemes have on the equivalent battery parameters.

Another remaining topic for investigation is determining how, within this framework, the aggregator chooses the capability and price to offer in regulation markets. A conventional generator has known heat rates and cost functions, but an aggregator represents a heterogeneous (in terms of both cost and performance) portfolio of units. Even if the aggregator is a price taker, the capability offer is not straightforward. Due to ramp and energy limits, the aggregator will have to offer a discounted aggregated capability. The optimal bidding strategy will likely take into account the statistical properties of the regulation signal to balance the desire for the biggest payments against the probability of imbalance penalties or even market disqualification. Recent papers that have addressed this problem in other frameworks include [34, 58–60]. The importance of this problem may vary depending on the reality faced by the aggregator. In practice, participating DERs may persist from day to day, with little variance in parameters. In this case, the optimal offer may be found through a simple adaptive bidding scheme. Or, perhaps, it may be vary more significantly, in which case having a rigorous

probabilistic approach would be justified to maximize the potential of each bid.

Although this framework has focused on large-scale power systems, many of the same concepts could benefit microgrids, e.g., in some industrial power systems. Utilizing flexible loads can bring about improvements in cost and emissions by decreasing the amount of regulation required by gensets.

APPENDIX

LINEAR PROGRAM TRANSFORMATION

The following definitions are used in the process of casting (3.9) as a standard linear program:

$$\mathcal{X} = \begin{bmatrix} x_1^T & x_2^T & \dots & x_N^T \end{bmatrix}^T$$
$$\mathcal{U} = \begin{bmatrix} u_0^T & u_1^T & \dots & u_{N-1}^T \end{bmatrix}^T$$
$$\mathcal{R} = \begin{bmatrix} Rr_1 \dots Rr_N \end{bmatrix} \Delta T_1$$

$$\mathcal{A} = \begin{bmatrix} A \\ A^2 \\ \vdots \\ A^N \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} B & 0 & \dots & 0 \\ AB & B & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ A^{N-1}B & \dots & AB & B \end{bmatrix}$$

$$\mathcal{E}_x = \begin{bmatrix} E_x & 0 & \dots & 0 \\ 0 & E_x & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & E_x \end{bmatrix} \quad \mathcal{E}_u = \begin{bmatrix} E_u & 0 & \dots & 0 \\ 0 & E_u & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & E_u \end{bmatrix}$$

$$\mathcal{Q}_1 = \begin{bmatrix} Q_1 & 0 & \dots & 0 \\ 0 & Q_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & Q_1 \end{bmatrix} \Delta T_1 \quad \mathcal{F}_x = \begin{bmatrix} F_x \\ \vdots \\ F_x \end{bmatrix}$$

$$\mathcal{Q}_2 = \begin{bmatrix} Q_2 & 0 & \dots & 0 \\ 0 & Q_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & Q_2 \end{bmatrix} \Delta T_1 \quad \mathcal{F}_u = \begin{bmatrix} F_u \\ \vdots \\ F_u \end{bmatrix}.$$

With these definitions, the problem can be written without the summation:

$$\begin{aligned} \underset{\mathcal{U}}{\text{minimize}} \quad & \mathbf{1}^T \mathcal{Q}_1 \mathcal{X} + \left\| \mathcal{Q}_2 \mathcal{X} + \mathcal{R} \right\|_1 \\ \text{subject to} \quad & \mathcal{E}_X \mathcal{X} \leq \mathcal{F}_X \\ & \mathcal{E}_U \mathcal{U} \leq \mathcal{F}_U \\ & \mathcal{X} = \mathcal{A} x_0 + \mathcal{B}\mathcal{U}. \end{aligned}$$

Introduce the variable Z_x to bound $\mathcal{Q}_2 \mathcal{X} + \mathcal{R}$, which removes the norm in the objective function, and the problem becomes

$$\begin{aligned} \underset{\mathcal{U}, Z_x}{\text{minimize}} \quad & \mathbf{1}^T \mathcal{Q}_1 \mathcal{X} + \mathbf{1}^T Z_x \\ \text{subject to} \quad & \mathcal{E}_X \mathcal{X} \leq \mathcal{F}_X \\ & \mathcal{E}_U \mathcal{U} \leq \mathcal{F}_U \\ & \mathcal{X} = \mathcal{A} x_0 + \mathcal{B}\mathcal{U} \\ & -Z_x \leq \mathcal{Q}_2 \mathcal{X} + \mathcal{R} \leq Z_x. \end{aligned}$$

Eliminating $\mathcal{X}$ by plugging in the dynamic equations gives us

$$\begin{aligned}
\underset{\mathcal{U},Z_x}{\text{minimize}} \quad & \mathbf{1}^T\mathcal{Q}_1\mathcal{A}x_0 + \mathbf{1}^T\mathcal{Q}_1\mathcal{B}\mathcal{U} + \mathbf{1}^T Z_x \\
\text{subject to} \quad & \mathcal{E}_x B\mathcal{U} \leq \mathcal{F}_x - \mathcal{E}_x\mathcal{A}x_0 \\
& \mathcal{E}_u\mathcal{U} \leq \mathcal{F}_u \\
& -Z_x \leq \mathcal{Q}_2\left(\mathcal{A}x_0 + \mathcal{B}\mathcal{U}\right) + \mathcal{R} \leq Z_x.
\end{aligned}$$

To create a single unknown variable and a single and inequality, we define

$$f = \begin{bmatrix} \mathbf{1}^T\mathcal{Q}_1\mathcal{B} & \mathbf{1}^T \end{bmatrix}^T, \quad y = \begin{bmatrix} \mathcal{U} & Z_x \end{bmatrix}^T,$$

$$G = \begin{bmatrix} -\mathcal{Q}_2\mathcal{B} & -I \\ \mathcal{Q}_2\mathcal{B} & -I \\ \mathcal{E}_u & 0 \\ \mathcal{E}_x\mathcal{B} & 0 \end{bmatrix}, \quad h = \begin{bmatrix} \mathcal{Q}_2\mathcal{A}x_0 + \mathcal{R} \\ -\mathcal{Q}_2\mathcal{A}x_0 - \mathcal{R} \\ \mathcal{F}_u \\ \mathcal{F}_x - \mathcal{E}_x\mathcal{A}x_0 \end{bmatrix}.$$

Note the term $\mathbf{1}^T\mathcal{Q}_1\mathcal{A}x_0$ is ignored as it is constant with respect to the decision variables. The problem is then in the form of a standard linear program (3.10).

REFERENCES

[1] "U.S.DoE — Smart Grid," http://www.oe.energy.gov/smartgrid.htm.

[2] "European technology platform for the electricity networks of the future," http://www.smartgrids.eu.

[3] D. S. Callaway and I. A. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 3, no. 1, pp. 434–442, 2012.

[4] C. Guille and G. Gross, "A conceptual framework for the vehicle-to-grid (V2G) implementation," *Energy Policy*, vol. 37, no. 11, pp. 4379–4390, 2009.

[5] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent, "A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service," in *Proc. of Allerton Conference on Communication, Control, and Computing*, Oct 2013, pp. 551–558.

[6] F. Schweppe, R. Tabors, J. Kirtley, H. Outhred, F. Pickel, and A. Cox, "Homeostatic utility control," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 3, pp. 1151–1163, May 1980.

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

[8] *Energy & Ancillary Services Market Operations*, PJM, December 2015, manual 11, revision 79.

[9] A. Kelman and F. Borrelli, "Bilinear model predictive control of a HVAC system using sequential quadratic programming," in *Proc. of IFAC World Congress*, Aug. 2011, pp. 9869–9874.

[10] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 189–198, Jan 2015.

[11] B. Kirby, "Frequency regulation basics and trends," Oak Ridge National Laboratory, Tech. Rep., 2004.

[12] B. Kirby, "Ancillary services: Technical and commercial insights," Wartsila North America Inc., Tech. Rep., 2007.

[13] P. Kokotovic and A. Haddad, "Controllability and time-optimal control of systems with slow and fast modes," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 111–113, Feb 1975.

[14] A. Venkat, I. Hiskens, J. Rawlings, and S. Wright, "Distributed MPC strategies with application to power system automatic generation control," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1192–1206, Nov 2008.

[15] M. Alizadeh and A. Scaglione, "Least laxity first scheduling of thermostatically controlled loads for regulation services," in *Proc. of IEEE Conference on Signal and Information Processing*, Dec 2013, pp. 503–506.

[16] B. Sanandaji, H. Hao, and K. Poolla, "Fast regulation service provision via aggregation of thermostatically controlled loads," in *Proc. of Hawaii International Conference on System Sciences*, Jan 2014, pp. 2388–2397.

[17] E. Vrettos and G. Andersson, "Combined load frequency control and active distribution network management with thermostatically controlled loads," in *Proc. of IEEE Conference on Smart Grid Communications*, Oct 2013, pp. 247–252.

[18] S. Meyn, P. Barooah, A. Bušić, and J. Ehren, "Ancillary service to the grid from deferrable loads: The case for intelligent pool pumps in Florida," in *Proc. of IEEE Conference on Decision and Control*, Dec 2013, pp. 6946–6953.

[19] Y. Lin, P. Barooah, and S. Meyn, "Low-frequency power-grid ancillary services from commercial building HVAC systems," in *Proc. of IEEE Conference on Smart Grid Communications*, Oct 2013, pp. 169–174.

[20] H. Hao, A. Kowli, Y. Lin, P. Barooah, and S. Meyn, "Ancillary service for the grid via control of commercial building HVAC systems," in *Proc. of IEEE American Control Conference*, June 2013, pp. 467–472.

[21] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn, "Ancillary service to the grid through control of fans in commercial building HVAC systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 2066–2074, July 2014.

[22] W. Mai and C. Y. Chung, "Economic MPC of aggregating commercial buildings for providing flexible power reserve," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2685–2694, Sept 2015.

[23] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental evaluation of frequency regulation from commercial building HVAC systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 776–783, March 2015.

[24] Y. J. Kim, D. H. Blum, N. Xu, L. Su, and L. K. Norford, "Technologies and magnitude of ancillary services provided by commercial buildings," *Proceedings of the IEEE*, vol. 104, no. 4, pp. 758–779, April 2016.

[25] I. Beil, I. Hiskens, and S. Backhaus, "Frequency regulation from commercial building HVAC demand response," *Proceedings of the IEEE*, vol. 104, no. 4, pp. 745–757, April 2016.

[26] J. Mathieu, M. Kamgarpour, J. Lygeros, G. Andersson, and D. Callaway, "Arbitraging intraday wholesale energy market prices with aggregations of thermostatic loads," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 763–772, March 2015.

[27] J. Mathieu, M. Kamgarpour, J. Lygeros, and D. Callaway, "Energy arbitrage with thermostatically controlled loads," in *Proc. of European Control Conference*, July 2013, pp. 2519–2526.

[28] M. Caramanis, I. Paschalidis, C. Cassandras, E. Bilgin, and E. Ntakou, "Provision of regulation service reserves by flexible distributed loads," in *Proc. of IEEE Decision and Control Conference*, Dec 2012, pp. 3694–3700.

[29] J. D. Kueck, B. J. Kirby, M. R. Ally, and C. K. Rice, "Using air conditioning load response for spinning reserve," Oak Ridge National Laboratory, Tech. Rep. ORNL/TM-2008/227, 2009.

[30] M. Sullivan, J. Bode, B. Kellow, S. Woehleke, and J. Eto, "Using residential AC load control in grid operations: PG&E's ancillary service pilot," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1162–1170, June 2013.

[31] J. Peppanen, M. J. Reno, and S. Grijalva, "Thermal energy storage for air conditioning as an enabler of residential demand response," in *Proc. of North American Power Symposium*, Sept 2014, pp. 1–6.

[32] S. Koch and G. Andersson, "Assessment of revenue potentials of ancillary service provision by flexible unit portfolios," in *Proc. of IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[33] E. Vrettos and G. Andersson, "Scheduling and provision of secondary frequency reserves by aggregations of commercial buildings," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 850–864, April 2016.

[34] E. Vrettos, F. Oldewurtel, and G. Andersson, "Robust energy-constrained frequency reserves from aggregations of commercial buildings," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–14, 2016.

[35] M. Maasoumy, B. Sanandaji, A. Sangiovanni-Vincentelli, and K. Poolla, "Model predictive control of regulation services from commercial buildings to the smart grid," in *Proc. of American Control Conference*, June 2014, pp. 2226–2233.

[36] M. Khalid and A. Savkin, "Model predictive control based efficient operation of battery energy storage system for primary frequency control," in *Proc. of International Conference on Control Automation Robotics Vision*, Dec 2010, pp. 2248–2252.

[37] T. Zhang and H. Gooi, "Hierarchical MPC-based energy management and frequency regulation participation of a virtual power plant," in *Proc. of Innovative Smart Grid Technologies Conference Europe*, Oct 2014, pp. 1–5.

[38] J. T. Hughes, A. D. Domínguez-García, and K. Poolla, "Virtual battery models for load flexibility from commercial buildings," in *Proc. of Hawaii International Conference on System Sciences*, Jan 2015, pp. 2627–2635.

[39] J. T. Hughes, A. D. Domínguez-García, and K. Poolla, "Identification of virtual battery models for flexible loads," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–10, 2016.

[40] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent, "A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service," in *Proc. of Allerton Conference on Communication, Control, and Computing*, Oct 2013, pp. 551–558.

[41] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[42] "Fast response regulation signal," http://www.pjm.com/markets-and-operations/ancillary-services/mkt-based-regulation/fast-response-regulation-signal.aspx.

[43] "Smart energy design assistance center," [Online] http://smartenergy.illinois.edu/.

[44] U.S. Department of Energy Office of Energy Efficiency & Renewable Energy Building Energy Software Tools Directory, "eQuest," [Online] http://apps1.eere.energy.gov/buildings/tools_directory/software.cfm/ID=575.

[45] G. M. Masters, *Renewable and Efficient Electric Power Systems; 2nd ed.* Hoboken, NJ: Wiley, 2013.

[46] A. Wood and B. Wollenberg, *Power Generation, Operation and Control*. New York: Wiley, 1996.

[47] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, 1st ed. Athena Scientific, 1997.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[49] J. Löfberg, "Minimax approaches to robust model predictive control," Ph.D. dissertation, Linköping University, Linköping, Sweden, 2003.

[50] "PJM markets fact sheet," https://www.pjm.com/~/media/about-pjm/newsroom/fact-sheets/pjms-markets-fact-sheet.ashx, January 2016.

[51] "PJM regulation zone preliminary billing data," http://www.pjm.com/markets-and-operations/market-settlements/preliminary-billing-reports/pjm-reg-data.aspx.

[52] A. Chakrabortty and M. Ilić, *Control and Optimization Methods for Electric Smart Grids*, ser. Power Electronics and Power Systems. Springer New York, 2011.

[53] D. Apostolopoulou, P. Sauer, and A. Domínguez-García, "Automatic generation control and its implementation in real time," in *Proc. of Hawaii International Conference on System Sciences*, Jan 2014, pp. 2444–2452.

[54] K. Deng, Y. Sun, S. Li, Y. Lu, J. Brouwer, P. G. Mehta, M. Zhou, and A. Chakraborty, "Model predictive control of central chiller plant with thermal energy storage via dynamic programming and mixed-integer linear programming," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 565–579, April 2015.

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

[56] A. Subramanian, M. Garcia, A. Domínguez-García, D. Callaway, K. Poolla, and P. Varaiya, "Real-time scheduling of deferrable electric loads," in *Proc. of American Control Conference*, June 2012, pp. 3643–3650.

[57] G. O'Brien and R. Rajagopal, "Scheduling non-preemptive deferrable loads," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 835–845, March 2016.

[58] F. Zhang, M. Tokombayev, Y. Song, and G. Gross, "Effective flywheel energy storage (FES) offer strategies for frequency regulation service provision," in *Proc. of Power Systems Computation Conference*, Aug 2014, pp. 1–7.

[59] M. Balandat, F. Oldewurtel, M. Chen, and C. Tomlin, "Contract design for frequency regulation by aggregations of commercial buildings," in *Proc. of Allerton Conference on Communication, Control, and Computing*, Sept 2014, pp. 38–45.

[60] D. Fooladivanda, C. Rosenberg, and S. Garg, "Energy storage and regulation: An analysis," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–11, 2015.