Coordination and Control of Distributed Energy Resources: Modeling and Analysis Techniques

by

Md Salman Nazir

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2020

Doctoral Committee:

Professor Ian Hiskens, Chair
Assistant Professor Johanna Mathieu
Professor Jeffrey Stein
Associate Professor Vijay Subramanian

Md Salman Nazir
mdsnazir@umich.edu
ORCID iD: 0000-0003-2663-1356

Dedication

To my loving wife, Asmita.

Acknowledgments

First, I wish to express my deepest gratitude to my PhD advisor, Professor Ian Hiskens, for his guidance through each stage of my PhD. Your insightful suggestions and detailed feedback have been extremely valuable. Your pursuit for knowledge is truly inspiring. I am also amazed by your eagerness, enthusiasm and ability to deal with numerous highly demanding objectives – be it related to research, your students, or your family – and be able to always excel!

Next, I would like to thank all my committee members, Professor Johanna Mathieu, Professor Jeffrey Stein and Professor Vijay Subramanian for their valuable feedback, and for many helpful discussions. I have been fortunate to work with Professor Mathieu on multiple projects - I have learnt a great deal working closely with you. Professor Stein inspired me to think deeply about how my research fits into the big picture and how my contributions fit into a single coherent theme. Professor Subramanian provided valuable feedback by helping me identify commonalities in modeling across different disciplines. I would also like to thank Professor Al-Thaddeus Avestruz, Professor Necmiye Ozay and Professor Roland Malhamé for many insightful discussions and for their helpful suggestions. I would like to extend my gratitude to Dr. Emiliano Dall'Anese and Dr. Andrey Bernstein for being my mentors during my internship at the National Renewable Energy Laboratory (NREL). Thanks to you, I had a highly intellectually stimulating experience and the ideas originated during my internship turned into a core component of this thesis. Finally, I would also like to pay my special regards to Professor Francisco Galiana and Professor François Bouffard for stirring up my interest in power systems during my undergraduate and master's degrees at McGill University - I will forever be thankful for your mentorship.

Additionally, for providing funding during my doctoral studies at the University of Michigan, I would like to sincerely thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the National Science Foundation (NSF) and the ARPA-E program of the U.S. Department of Energy (DOE).

I moved to Ann Arbor in Fall 2015, with my wife, Asmita and our then 4-month old daughter, Zuni. Very soon, we fell in love with Ann Arbor! We met many awesome friends and families and didn't realize how fast 4 years just passed by. I would like to thank my fellow graduate students and friends at the University of Michigan as well as friends from the Northwood Community – Aditya, Akshay, Ali, Amlin, Anna, Greg, Ioannis, Hanpyo, John, Jenny, Jing, Jonas, Joy, Kang, Leke, Milu, Mushfeq, Mouaz, Mike, Mengqi, Saki, Sijia, Shunbo, Steph, Sungyul, Sunho, Tanvir, Teddy, Xiaofan, Xin, Ziyou, Zunaid, and many others - as you can imagine, by no means can this list be comprehensive. Hanpyo, Kang and Shunbo, I will miss our discussions during the coffee breaks. Dear Dr. Alauddin Ahmed and Dr. Golam Mortuza, I am indebted to you for all your guidance and support during my PhD journey. I must also mention how easy it has been to navigate around numerous administrative tasks at the university, due to the kind support from Kristen, Anne, Judi, Kathy, Shelly and Andria – I really appreciate all your help.

Finally, I would like to thank my family members. First and foremost, I would like to thank my loving wife, Asmita – you have made the whole journey possible - no words are enough to express my gratitude for what you have done for me and for our family. Dear Zuni, thanks to you too, for your love, for your smiles, for spreading so much joy in our lives - life is beautiful due to you. Special thanks to my parents, Ammu, Mamoni, Baba and Abbu, for your blessings, unconditional love and support. Thanks to Moumita and Rayhan for extending your help whenever we needed and for filling our lives with joy with your surprise visits all the way from Tennessee. Thanks to Dr. Sheikh Ghafoor and Dr. Ambareen Siraj for many advices and for serving as role models. Thanks to Sanim and Shammi, and similarly, other family members who kept sending us well wishes and encouragements.

I am realizing that it's truly difficult to comprehensively thank all deserving individuals by their names – since our paths have crossed with so many of you. Hence, I must now simply state that I am truly grateful to Almighty for providing me with numerous sources of support and blessings – thanks for helping me successfully complete this wonderful journey.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Coordinated control of distributed energy resources (DERs), such as flexible loads, storage devices and solar photovoltaic inverters, can provide valuable services to the electricity grid by reducing peak demand, balancing renewables and avoiding voltage excursions. Aggregate control of thermostatically controlled loads (TCLs), such as air-conditioners, water-heaters and refrigerators, offers a promising way of accommodating significant DERs in power systems. This dissertation focuses primarily on modeling and analysis techniques for ensembles of TCLs. It also develops techniques for efficiently aggregating DER-based flexibility.

A wide-variety of load ensemble control techniques have been developed in the literature – with strategies including probabilistic switching signals, TCL set-point variation, and price-based signals. However, synchronization of TCL temperatures, oscillations in aggregate demand and bifurcations have been observed, which can lead to detrimental power- and voltage-related issues in the electricity grid. A detailed investigation is undertaken of a market-based transactive energy coordination (TEC) scheme, where TCL users submit bids for their energy demand and an aggregator clears the market to allocate energy among users. This study confirms the presence of such issues. To avoid these unintended consequences of load control, a Markov-chain-based state-transition model has been developed to capture the aggregate TCL dynamics under TEC. Using the state-transition model, a model predictive control scheme has been formulated to attain near-optimal control policies that maximize social welfare while limiting the possibilities of TCL synchronization and power oscillations. To further investigate unintended behavior arising from the control of load ensembles, a generalized hybrid dynamical system representation is developed to accurately capture the interactions between the continuous dynamics of loads and discrete control actions. This representation can capture diverse control-update intervals, from fifteen-minute intervals for economic dispatch problems to 2-10 seconds for frequency regulation services. Using this hybrid representation and modal analysis, it is shown that

synchronizing behavior in TCLs can be identified under a wide range of control schemes, such as probabilistic and priority-based switching, and TEC. A number of practical constraints, such as limited availability of TCLs for control and/or limited TCL parameter information, are considered to quantify performance bounds of load control schemes.

To compute the aggregate flexibility available from spatially distributed DERs, special convex sets known as homothets and zonotopes are employed. First, aggregation algorithms are developed assuming DERs are located at a single node of the network. The setting is then extended to spatially distributed resources by incorporating the network and power flow constraints. It is shown that network parameters and voltage limits often limit the flexibility that can be transferred from one node to its upstream or downstream neighbors. This flexibility model lends itself to several applications, including optimal power flow in distribution networks and efficient coordination of transmission and distribution systems.

Chapter 1.

Introduction

In recent years, power systems have been integrating significant amount of renewable generation sources. By the end of 2017, the global installed wind power capacity reached 540 GW, and solar, 400 GW [88]. In the U.S., total solar capacity reached 64 GW in 2018 and is expected to more than double over the next five years [6]. According to [28], the wind power capacity in the U.S. has almost tripled from 2008 to 2016. Similar trends are seen worldwide, with several European countries reaching unprecedented milestones. For example, on July 9 and 10, 2015, the wind power generated in Denmark reached up to 140% of the entire country's electricity demand [83]. While such rapid growth in renewables is remarkable, power systems is also experiencing new challenges in accommodating these resources due to their inherent variability and uncertainty.

In the electric power grid, generation must balance demand at all times. The net load curves, obtained by subtracting renewable generation from the electricity demand, have been changing considerably from the traditional. For example, net load curves can often undergo periods of fast-ramps and periods of over-generation [46]. Hence, system operators need flexible resources to manage such fast-ramps in net demand and mitigate the inherent uncertainty from intermittent renewable sources. The decorbonization of the electricity sector in the presence of significant renewables calls for increased participation from the demand-side using various *demand response* strategies [18], as well as better utilization of existing resources through enabling technologies. Otherwise the benefits of integrating renewables can be offset by increased operations of emitting generation plants.

Distributed energy resources (DERs), such as controllable loads, electric vehicles (EVs), storage devices and solar photovoltaic (PV) systems, are becoming increasingly prevalent in modern energy systems and are expected to continue their rapid growth over the coming

years. The flexibility offered by DERs can undoubtedly play a major role in improving the efficiency of the grid and in facilitating large-scale integration of renewables, hence is the topic of interest in this dissertation.

The work presented in this dissertation develops novel modeling, control and optimization techniques to integrate significant DERs in power systems. Detailed investigation into existing modeling techniques for thermostatically controlled loads (TCLs), such as air-conditioners (ACs), water-heaters and refrigerators, is performed to understand advantages and limitations of various approaches. For contributing towards control development, we presented simulation-based performance analysis of several existing control schemes and proposed novel control strategies to mitigate undesired synchronization and power oscillations. For analysis and comparison of TCL control schemes, a general hybrid system repsentation has been developed through model reformulation. For a large class of control schemes, we also explored if fundamental performance limitations may exist, for example due to the algorithm assuming a large number of controllable TCLs are available, where in fact a smaller number actually participates. With availability of controllable loads and other distributed resources, such as solar inverters and storage devices, the flexibility from DERs can be aggregated and effectively used in power system optimization problems. With this objective, we propose characterizing DER flexibility using convex polytopes and show how aggregations can be performed efficiently, with and without considering network models and power flow and voltage constraints. Thus, the common theme throughout this dissertation is to contribute towards development of efficient and scalable algorithms to facilitate integration of large scale DERs in power systems, and present analytical tools to enhance our understandings of capabilities and limitations of various control schemes.

1.1. Literature Review

The benefits of integrating the demand-side into power systems operation have been widely recognized since decades [17, 44, 66]. Alongside standard generation control actions, demand-side participation helps to lower costs, reduce emissions and meet security criteria. With modern communication and control technologies, loads can even undertake many tasks performed by traditional generators more efficiently. The storage available from a large populations of space heating and cooling systems, water heaters, and refrigerators, in the form of thermal gradients, offer significant flexibility to adjust their aggregate demand

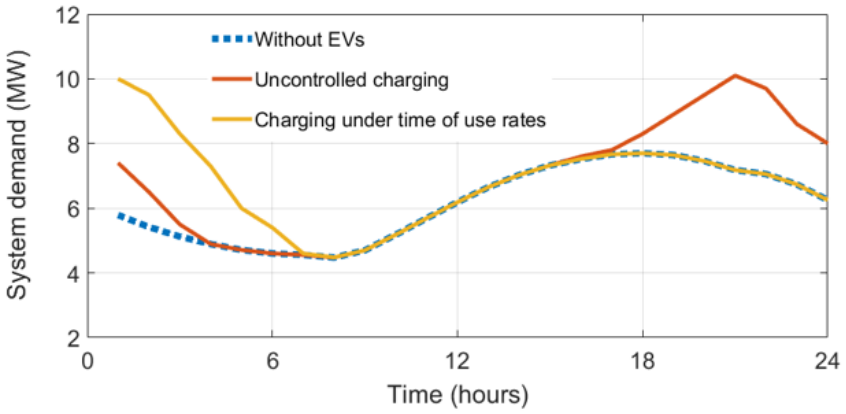

Figure 1.1: Impact of EV charging on net load in a distribution grid under uncoordinated and time-of-use charging (adapted from [26]).

without inconveniencing their users [18]. Development of efficient 'non-disruptive' control strategies guarantees customer comfort while unlocking flexibility for grid operators. A large number of papers therefore have shown how coordinated control of thermostatically controlled loads (TCLs), such as air-conditioner, water-heaters and refrigerators, can provide demand response services, such as balancing fluctuations from renewables, reducing peak demand, and providing frequency and voltage support [17, 18, 39, 40, 55, 70, 79, 104].

In the case of EVs, advanced coordination mechanism may soon become a necessity [18]. It is projected that, in the United States, by 2050 almost 30-40% of the transportation sector will primarily depend on electricity [27]. However, current distribution grids are not prepared to process the upcoming broad dissemination of EVs and cyber-physical technology due to possible new peaks in system demand (see Fig. 1.1) and voltage issues in distribution systems [26]. A fast charging EV consumes 60-90 kWh in just 30 minutes, which is almost twice the daily household consumption. Without coordination, as the number of EVs increase, their simultaneous (or synchronized) charging would lead to network congested, high electricity prices and have detrimental effect on distribution level transformers. Ad-hoc charging of EVs and price responsive loads can also induce instability in power consumption profiles and in electricity prices [18].

Various forms of coordination techniques are being developed and studied to enable efficient integration of DERs in the electric grid. Direct on/off control signals can be sent to specific load groups (for example, air-conditioners or water heaters) during specific

hours of the day [66], but such centralized methods may result in violations of customer comfort. Several works on TCLs have proposed variation of TCL temperature set-points for controlling the aggregate TCL demand [11, 17, 55]. Randomized control schemes have also been proposed where DERs switch on/off in response to probabilistic switching signals sent from distribution utilities or DER aggregators [30, 51, 70, 79]. Set-point variation and randomized switching schemes are forms of 'non-disruptive' ensemble control strategies since these aim to guarantee customer comfort. Depending on the controller performance objectives, frequent data exchange may be necessary. One the other hand, a price-based control paradigm known as Transactive Energy (TE) coordination has received significant attention [32, 43]. Strong support from the United States national laboratories and the department of Energy has led to multiple demonstration projects. Under the TE framework, owners of DERs, may participate in buying and selling energy based on their comfort versus price preferences. Through market clearing, specific consumption patterns are allocated. Network congestion can potentially be managed through such a market-based mechanism. Researchers have also proposed extending the TE framework to game theory-based coordination schemes, where participants behave strategically based on their local cost and comfort constraints while participating in load management programs. However, surprisingly little attention has been given to identifying unintended consequences that may arise under TE or other such market-based schemes. Upon detailed investigation, we observed that many direct control approaches, may also be prone to causing large fluctuations in aggregate power, exhibiting fast transients during operations, as well as causing new system peaks [80]. Hence, our work aims to present a systematic approach to analyze such behavior and design appropriate mitigation strategies.

The challenge we first need to address involves analyzing the effectiveness of modeling techniques for load aggregation. In that front, our aim is to gain understanding of various advantages and limitations of aggregate modeling techniques, especially when considering various forms of uncertainty and parameter heterogeneity. We focus primarily on physically-based models – originally developed by [20, 44, 66, 73]. In [66], Malhame and Chong showed that the aggregate dynamics of TCLs can be modeled using stochastic Fokker-Planck diffusion model, which are a set of coupled partial differential equations (CPDEs) with appropriate boundary conditions. More recently, in [18], Callaway showed how the CPDE model can be linearized for small perturbations around nominal operating points and developed a control TCL set-point control strategy using such linearization. [12] modeled the aggregate dynamics of TCLs by bi-linear PDEs and used finite-difference

based discretization technique to attain a state space model. Authors in [55] modeled the steady state aggregate temperature densities for ON/OFF states using Laplace transforms and applied it to develop controllers for set-point variation. In [51], a Markov chain based statistical modeling approach was presented to describe the evolution of the probability masses over "temperature bins". Each bin is defined by a specified temperature range and whether the TCLs in it are in their ON/OFF state. Such a model is often referred to as *bin models* or *state-bin models*. Authors in [70] used a similar model to design state estimators and develop control strategies for large population of TCLs. In [33], authors modeled temperature evolution over discretized temperature ranges using advection-diffusion PDEs and studied the model's numerical stability. For population model for water-heaters, [79] showed how to extend the bin model to account for stochastic water-draw of users and how bin models can be incorporated in day-ahead planning problems. A large body of recent work has been utilizing such bin models in designing new control techniques (see for example [21],[30], [76]). Our modeling and analysis framework in this dissertation is consistent with these existing models. However, we further extend these modeling approaches by seeking and establishing a general hybrid system framework for coordinating TCLs.

Note that while the aggregate dynamics of TCLs can be captured using continuous-time models, the control updates typically occur at relative slow discrete intervals (typically minutes range). Hence, a hybrid dynamical system representation accurately captures the overall continuous/discrete dynamics [63, 105] and allows us to systematically study and compare the performance of various control strategies.

Next, given the importance of DER flexibility to alleviate a number of operational challenges in the power grid [18, 40, 74], we investigate how the aggregate flexibility from DERs can be modeled. Power systems are in the process of accommodating an increased amount of distributed energy resources (DERs) – solar photovoltaic (PV) systems, energy storage systems, and controllable demand-side resources. Several recent works have therefore developed techniques to characterize the aggregate flexibility from DERs [2, 10, 13, 42, 52, 74, 75, 106]. In our work [78], we focus on flexibility-polytope based approaches, where such polytopes contain all feasible operating points of a population of DERs. Considering these polytopes in any optimization problem (such as optimal power flow and economic dispatch) eliminates the need to consider DER state dynamic equations, their power and energy limits and hence can provide significant computational advantage. However, for the storage-like DERs, these polytopes are high-dimensional (e.g. 24 dimensional for a 24-hour problem with hourly time steps), hence can be difficult to compute and typically

inner/outer approximations are sought [10, 74]. We find that while there are efficient techniques to compute outer approximating flexibility polytopes [10], inner-approximation algorithms [74, 106] may provide very conservative estimates. Hence, this dissertation builds upon and improves the existing inner-approximating techniques. Furthermore, for DERs distributed across different nodes of the power system, existing literature does not show how the network constraints limit the flexibility sets, hence will be addressed in this dissertation.

1.2. Contributions of this Dissertation

This dissertation focuses on developing novel modeling, control, analysis and optimization techniques to integrate significant DERs in power systems. The organization of the chapters and their main contributions are summarized below.

Chapter 2 reviews the mathematical models for capturing individual and aggregate TCL dynamics. In terms of modeling the aggregate dynamics, two different schools of thoughts exist in literature. In [66], authors showed that the aggregate dynamics of TCLs can be modeled using coupled PDEs, which have also been adopted in [17, 97]. On the other hand, recently much attention has been given to Markov-chain-based modeling techniques [51, 55, 70, 79] due to their tractability, scalability and lending to a linear state space representation which is convenient for control design. In this chapter, we summarize both modeling techniques and show their equivalence in the case of homogeneous TCLs. Since TCLs provide storage capabilities in form of thermal gradients, an alternate battery-equivalent model is also presented. The generalized battery-model is especially convenient for power system planning studies.

In **Chapter 3**, we investigated and developed techniques to incorporate noise and heterogeneity in the Markov-chain model, i.e. the *state-bin model*. We find that the state-bin models inherently lack in their ability to accurately capture the dynamics of heterogeneous TCLs, which has been resolved by decomposing the population into multiple homogeneous groups. Using sensitivity analysis, we also analyzed the impact of modeling error, noise and heterogeneity on aggregate behavior of loads.

Chapter 4 presents simulations of the market-based or Transactive energy coordination (TEC) and identifies several cases showing undesirable behavior, such as emergence of

temperature synchronization, exhibiting large oscillations in aggregate power and highly fluctuating demand and prices. To address these challenges associated with TEC, we develop an extended bin model (Markov-chain based representation) to the handle the market-based coordination. This allows predicting the behavior of DERs under TEC and optimizing their trajectories by solving multi-period optimization problems. An accurate mixed integer problem formulation and a relaxed quadratic programming formulation have been developed and tested. A case study shows that a population of TCLs can be managed economically while avoiding congestion in a distribution grid. Simulations also demonstrate that power oscillations arising from synchronization of TCLs can be effectively avoided.

Chapter 5 focuses on modeling and analysis of load control algorithms using a generalized hybrid systems framework. Detailed analysis of the eigen-modes of the discretized hybrid system reveals when the aggregate TCL dynamics can undergo limit cycle oscillations or exhibit period-adding bifurcation. Analytical bounds on parameter values can be established to ensure avoidance of undesirable large oscillations from the load ensemble. A candidate Lyapunov (energy) function is also derived and used to examine if the trajectories of the controlled TCL ensembles remain stable.

Chapter 6 focuses on quantifying and addressing performance limitations in TCL coordination when practical constraints are further imposed. Practical limitations typically appear due to hardware, software and communication restrictions, as well as limited availability of controllable TCLs. First, we establish performance limits for population of TCLs that are controlled via probabilistic dispatch for participating in fast energy balancing services. An analytical expression for expected RMS error due to probabilistic switching was developed. Then we extended our analysis to aggregate bin-based models for which switching probabilities may be non-uniform. The close agreement between analytical and simulation-based results demonstrated the validity of the performance bounds. It was shown that the RMS error in the fraction of TCLs switched varies with target levels as well as population size. Then, by using variance reduction techniques, new controllers have been developed to find the optimal switching probabilities.

In the existing literate on control of TCLs, parameters such as set-points and deadbands are typically assumed to be known, but in practice, TCL vendors and customers might not be willing to share set-point information, thus posing additional challenges for demand response aggregators. Using 'exploration-exploitation'-type strategies allows us to provide energy balancing services without violating users' comfort limits. We demonstrated the

tradeoffs between exploration and exploitation phases mainly by identifying key parameters that affect the performance of each phase.

In **Chapter 7**, techniques to compute and aggregate DER flexibility polytopes are developed. We primarily focus on computing inner-approximating polytopes to avoid any infeasible operating points. To improve the accuracy of the inner-approximations, we develop a novel technique mainly by considering unions of feasible intervals, *homothets* [106] and/or *zonotopes* [74], with are sub-classes of convex sets with special structures. We analyze the accuracy and performance tradeoffs compared to existing approaches. These polytopes can then be efficiently integrated in optimal power flow formulations, economic dispatch problems or used for sampling feasible points for simulation based studies. A case study demonstrating voltage control utilizing PV inveter flexibility sets is provided.

Chapter 8 utilizes the flexible polytopes of inverters and loads, developed in Chapter 7, to approximate the aggregate real and reactive power flexibility available at a distribution system substation. The net flexibility set at a substation needs to be formed by capturing flexibility available from different nodes on the distribution feeder. This is challenging since the network models and constraints will affect the amount of flexbility that can be transferred from different nodes to the substation. A major benefit of being able to characterize the net flexibility set at a substation is that transmission system operators can then better utilize the system resources without actually having to model any DERs at the distribution system. Numerical results are shown on a standard IEEE radial test feeders.

Finally, **Chapter 9**, a summary of the dissertation is presented, the main contributions of each chapter are summarized, and possible future research directions are discussed.

Chapter 2.

TCL Modeling

2.1. Chapter Introduction

This chapter primarily focuses on modeling of individual and aggregate TCL dynamics. Based on existing literature, first We discuss both continuous and discrete-time models for individual TCLs. In terms of modeling the aggregate dynamics of a TCL population, two main schools of thought exist in literature. In [66], authors showed that the aggregate dynamics of TCLs can be modeled using coupled Fokker-Planck partial different equations (PDEs), which have also been adopted in [17, 97]. However, more recently, much attention has been given in Markov-chain-based modeling techniques [51, 55, 70, 81], primarily due to their tractability, scalability and lending to a linear state space representation which is convenient for control design and state estimation. In this chapter, we summarize both modeling techniques. Then, following the work of [11], we show that for homogeneous TCLs, the finite difference discretization of the coupled PDEs [66] provide a reduced form aggregate model in a continuous-time form. Taking its discrete-time equivalent, we then obtain the transition matrix of the Markov-chain model [51, 55, 70]. Finally, since TCLs provide storage capabilities in form of thermal gradients, an alternate battery-equivalent model is also presented. The generalized battery-model is especially convenient for power system planning studies.

2.2. Individual TCL Dynamics

The temperature, $\theta_j(t)$ (in °C) dynamics of TCL j can be modeled using a first-order differential equation [44, 66],

$$\dot{\theta}_j(t) = \frac{1}{C_j R_j}\Big(\theta^{\text{amb}} - \theta_j(t) - m_j(t) P_j R_j\Big) \tag{2.1}$$

where C_j is the thermal capacitance (kWh/°C), R_j is the thermal resistance (°C/kW), P_j is the rate at which the TCL (here, an air-conditioner) absorbs heat when the load is turned on, θ^{amb} (°C) is the ambient temperature external to the TCL. The on/off state $m_j(t)$ is governed by the thermostatic switching law with some dead-band, δ^{db} (°C), around some user-specified set-point, θ_j^{set} (°C). Then, $\theta_j^{\min} = \theta_j^{\text{set}} - \delta^{\text{db}}/2$, $\theta_j^{\max} = \theta_j^{\text{set}} + \delta^{\text{db}}/2$, and

$$m_j(t) = \begin{cases} 0, & \text{if} \quad \theta_j(t) \le \theta_j^{\min} \\ 1, & \text{if} \quad \theta_j(t) \ge \theta_j^{\max} \\ m_j(t^-), & \text{otherwise.} \end{cases} \tag{2.2}$$

where t^- is used to represent limit from the left since $m(t)$ has discontinuities at the switching times [97]. Figure 2.1 shows the on/off behavior of an air-conditioner with varying temperature within the hysteresis dead-band.

The parameters of a TCL (air-conditioner), with their typical values are shown below:

Table 2.1: TCL (Cooling) Parameters.

Parameter	Meaning	Value	Units
h	Time step	10	seconds
θ^{set}	Temperature set-point	20	$^{\circ}C$
δ	Temperature dead-band width	0.5	$^{\circ}C$
θ^{amb}	Ambient air temperature	32	$^{\circ}C$
P	Energy transfer rate	14	kW
R	Thermal resistance	2	$^{\circ}C$/kW
C	Thermal capacitance	10	kWh/$^{\circ}C$
η	Coefficient of performance	2.5	

While the above model is deterministic, in reality TCL dynamics may be affected by several factors including changes in ambient temperature, occupant behavior and inaccuracies in thermal parameter values. Hence, such uncertainties can be accounted for by

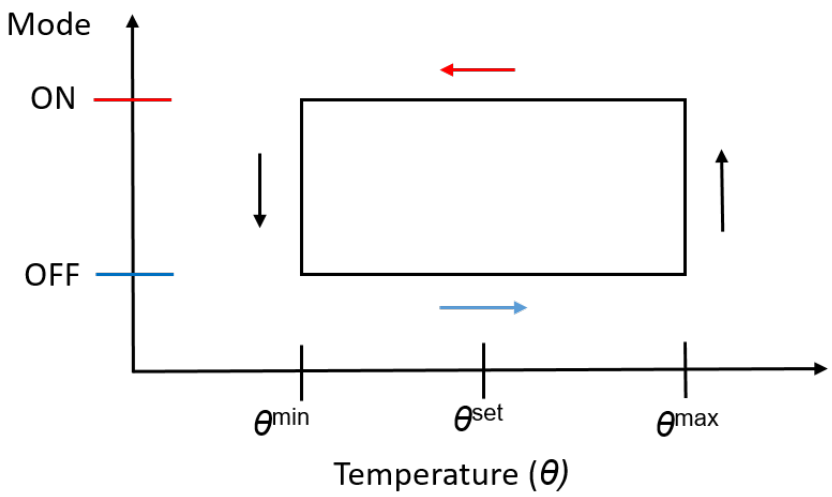

Figure 2.1: Temperature evolution and the associated on/off modes of an air-conditioner.

introducing a Brownian noise process, $w(t)$, with intensity parameter σ [66], to obtain

$$d\theta_j(t) = \frac{1}{C_j R_j}\Big(\theta^{\text{amb}} - \theta_j(t) - m_j(t) P_j R_j\Big) + \sigma dw(t). \tag{2.3}$$

Later, [73] developed a discrete-time representation for (2.4) as,

$$\theta_{j,(k+1)} = \tilde{a}_j \theta_{j,k} + (1 - \tilde{a}_j)(\theta^{\text{amb}} - m_{j,k} P_j R_j) + w_{j,k} \tag{2.4}$$

along with the following expression,

$$m_{j,(k+1)} = \begin{cases} 0, & \text{if} \quad \theta_{j,k} < \theta_j^{\min} \\ 1, & \text{if} \quad \theta_{j,k} > \theta_j^{\max} \\ m_{j,k}, & \text{otherwise.} \end{cases} \tag{2.5}$$

In (2.4), $\tilde{a}_j$ governs the thermal characteristics of the thermal mass and equals $\exp(-h/C_j R_j)$, where h is the time step (s). The on/off state of the TCL is determined by (2.5).

The temperature trajectories of a full population of N_{TCL} TCLs can be modeled by using N_{TCL} independent sets of equations (2.4) and (2.5). The population's total electrical

power consumed can be calculated as

$$P_k^{\text{tot}} = \sum_{j=1}^{N_{\text{TCL}}} \frac{m_{j,k} P_j}{\eta_j}, \tag{2.6}$$

where η_j is the coefficient of performance.

The steady-state power consumption, m_j^{o}, can be obtained using [55],

$$m_j^{\text{o}} = \frac{\theta^{\text{amb}} - \theta^{\text{set}}}{P_j R_j}. \tag{2.7}$$

For an AC unit, with the parameter values given in the Table 2.1, $m^{\text{o}} = 0.4286$ (in p.u.).

2.3. TCL Aggregate Model

For computational tractability, analysis and control design, there has been significant efforts to develop aggregate models to describe the dynamics of a TCL population. We describe two such approaches.

2.3.1. Markov-chain Model

Reduced-order aggregate plant models have been developed to make control design for a population of TCLs more tractable. In particular, discrete-state bin modeling approach has been proven to be very useful (see [51, 70, 81]). This modeling technique normalizes each TCL's temperature dead-band, so that the temperature of all TCLs can be commonly described on the same scale. As shown in Fig. 2.2, state bins are created by discretizing the normalized dead-band and dividing each discrete temperature range into two bins to differentiate ON/OFF fractions of TCLs. Here, $i = 1, 2, ..., N_B$ represent the OFF bins and $i = N_B + 1, N_B + 2, ..., 2N_B$ represent the ON bins. Let $x_{i,k}$ represent the fraction of the TCL population lying in bin i at time k. Then, the following LTI representation is used to describe the population-level autonomous dynamics as,

$$x_{k+1} = A x_k. \tag{2.8}$$

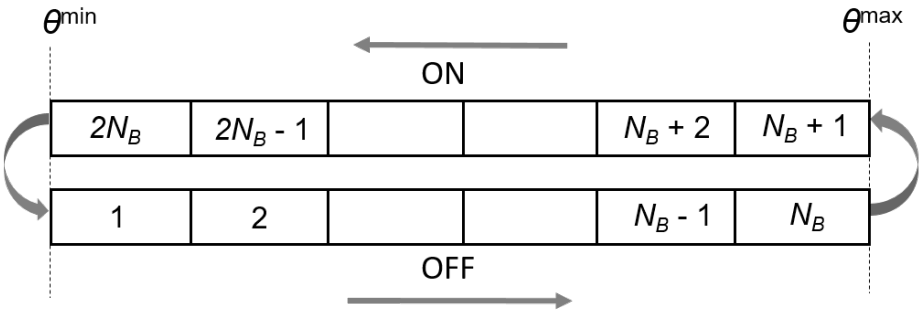

Figure 2.2: State bin transition model for a cooling TCL.

$$A = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,2N_{\mathrm{B}}} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,2N_{\mathrm{B}}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{2N_{\mathrm{B}},1} & p_{2N_{\mathrm{B}},2} & \cdots & p_{2N_{\mathrm{B}},2N_{\mathrm{B}}} \end{bmatrix}. \tag{2.9}$$

Matrix A is the transpose of a Markov transition matrix and each element of A-matrix, $p_{i,j}$, gives the proportion of bin j that transitions to bin i [51, 70]. Note that $0 \leq p_{i,j} \leq 1,\ \forall i, j$, and $\sum_{i=1}^{2N_{\mathrm{B}}} p_{i,j} = 1,\ \forall j$. Given $0 \leq x_{i,k} \leq 0$ and $\sum_{i=1}^{2N_{\mathrm{B}}} x_{i,k} = 1$, the properties of A-matrix ensure that $\sum_{i=1}^{2N_{\mathrm{B}}} x_{i,k+1} = 1$. The coefficients of A can be estimated using system identification techniques by running repeated simulations to find the average transition rates [51, 70]. Additional numerical recipes are also provided in Chapter 3 of this dissertation or in our work [81].

When external ON/OFF control is applied to the plant, a state space model can be used to describe the dynamics,

$$\begin{aligned} x_{k+1} &= A\, x_k + B\, u_k, \\ y_k &= C\, x_k, \end{aligned}$$

where $C = \bar{P}^{\mathrm{elec}}[\mathbf{0}^{1\times N} \quad \mathbf{1}^{1\times N_B}]$, with $\bar{P}^{\mathrm{elec}}$ being the power consumption (in kW) when a TCL is on. For a homogeneous population, $\bar{P}^{\mathrm{elec}} = \frac{P}{\eta}$. The entries of B depend on the choice of the control strategy [12, 51, 55, 70]. The movement of TCLs in response to control actions can be visualized using Fig. 2.2. Upon receiving switch ON requests, TCLs currently in an OFF bin, can move directly above (turning ON) to an ON bin, and vice versa for switch OFF requests.

2.3.2. Coupled PDE System

Assuming individual TCL's temperature and on/off state evolve by (2.3) and (2.2), it was shown in [66] that the aggregate dynamics of TCLs can be modeled using coupled Fokker-Planck equations. To express the couples PDEs, first, an extended temperature domain $[\hat{\theta}^{\min}, \hat{\theta}^{\max}]$ is divided in 3 zones, $k \in \{a, b, c\}$, as shown in Fig. 2.3. The system of PDEs are then given by,

$$\frac{\partial f_{0,k}(\theta,t)}{\partial t} = -\frac{\partial}{\partial \theta}\Big[(\alpha\theta+\beta) f_{0,k}(\theta,t)\Big] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial \theta^2} f_{0,k}(\theta,t), \quad k \in \{a,b\}, \tag{2.10a}$$

$$\frac{\partial f_{1,k}(\theta,t)}{\partial t} = -\frac{\partial}{\partial \theta}\Big[(\alpha\theta+\beta+\gamma) f_{1,k}(\theta,t)\Big] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial \theta^2} f_{1,k}(\theta,t), \quad k \in \{b,c\}, \tag{2.10b}$$

where $\alpha = \frac{1}{CR}$, $\beta = \frac{\theta^{\text{amb}}}{CR}$, $\gamma = \frac{P}{C}$. $f_{0,k}(\theta,t)$ and $f_{1,k}(\theta,t)$ are the probability densities of the TCLs (cooling devices) in off and on modes, respectively, over regions a, b, c.

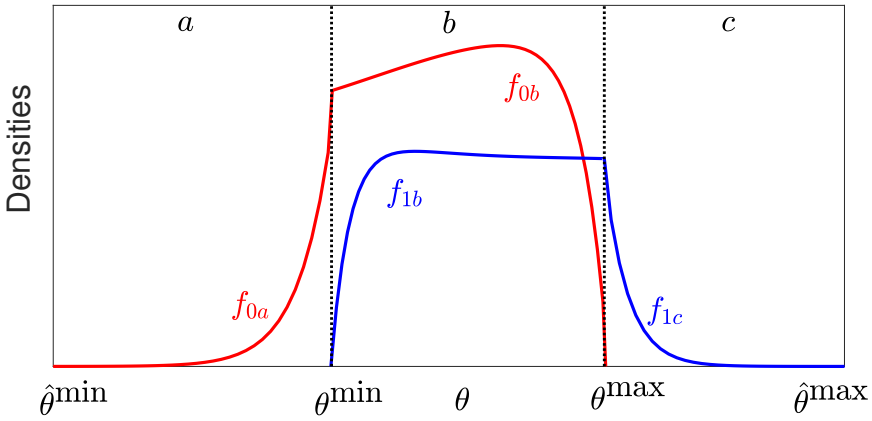

Figure 2.3: Densities by solving coupled PDE system.

Let $h_{i,k}$ be the probability flows, i.e. the integral over temperature (θ) coordinate of $\frac{\partial f_0(\theta,t)}{\partial t}$. Thus,

$$h_{0,k} = -(\alpha\theta+\beta) f_{0,k}(\theta,t) + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta} f_{0,k}(\theta,t), \tag{2.11a}$$

$$h_{1,k} = -(\alpha\theta+\beta+\gamma) f_{1,k}(\theta,t) + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta} f_{1,k}(\theta,t). \tag{2.11b}$$

The associated boundary conditions are given by:

$$f_{0b}(\theta^{\max}, t) = f_{1b}(\theta^{\min}, t) = 0, \tag{2.12a}$$
$$f_{0a}(\theta^{\min}, t) = f_{0b}(\theta^{\min}, t), \tag{2.12b}$$
$$f_{1a}(\theta^{\max}, t) = f_{1b}(\theta^{\max}, t), \tag{2.12c}$$
$$h_{1c}(\hat{\theta}^{\max}, t) = h_{0c}(\hat{\theta}^{\min}, t) = 0, \tag{2.12d}$$
$$h_{0a}(\theta^{\min}, t) = h_{0b}(\theta^{\min}, t) + h_{1b}(\theta^{\min}, t), \tag{2.12e}$$
$$h_{1c}(\theta^{\max}, t) = h_{0b}(\theta^{\max}, t) + h_{1b}(\theta^{\max}, t). \tag{2.12f}$$

Equation (2.12a) represents an absorbing boundary condition. Then, (2.12b)-(2.12c) ensure continuity of densities at $\theta^{\min}$ and $\theta^{\max}$. Similar to [97], conditions in (2.12d) represent artificially introduced reflecting boundaries designed to keep the state space for the temperature dynamics bounded. Thus probability mass remains confined to the interval $[\hat{\theta}^{\min}, \hat{\theta}^{\max}]$. Finally, (2.12e)-(2.12f) ensure conservation of probability. Additionally, one can enforce the integral of all $f_{i,k}$ ($i = 0, 1$ and $k = a, b, c$) over $\hat{\theta}^{\min}$ to $\hat{\theta}^{\max}$ to be one [17].

2.3.3. Computing a Reduced-Order Aggregate Model in the Noise-Free Case

In the noise-free case, no densities exist in regions a and c. Plugging in $\sigma^2 = 0$ in (2.10a) and (2.10b), we obtain

$$\frac{\partial f_{0,b}(\theta, t)}{\partial t} = -\frac{\partial}{\partial \theta}\Big[(\alpha\theta + \beta) f_{0,b}(\theta, t)\Big] \tag{2.13a}$$
$$\frac{\partial f_{1,b}(\theta, t)}{\partial t} = -\frac{\partial}{\partial \theta}\Big[(\alpha\theta + \beta + \gamma) f_{1,b}(\theta, t)\Big]. \tag{2.13b}$$

Similarly, the probability flows and the boundary conditions get simplified. Notice that (2.12e) and (2.12f) become

$$h_{0b}(\theta^{\min}, t) = -h_{1b}(\theta^{\min}, t), \tag{2.14a}$$
$$h_{0b}(\theta^{\max}, t) = -h_{1b}(\theta^{\max}, t), \tag{2.14b}$$

which says, at $\theta^{\max}$ boundary, all flows existing the off densities enter the on densities, and

at $\theta^{\min}$ boundary, all flows exiting the on densities enter the off densities.

Authors in [11] showed that for such PDE systems, finite-difference discretization can be applied to obtain an LTI representation for the evolution of $x(t)$ in the form of

$$\dot{x}(t) = \mathcal{A}x(t). \tag{2.15}$$

To construct matrix the continuous-time $\mathcal{A}$-matrix, consider N_B bins (temperature ranges) both in ON and OFF stages. Then, the bin width, Δ^{bw} is given by $\Delta^{\text{bw}} = \frac{\delta^{\text{db}}}{N_B}$. At each bin, indexed by $i = 1, ..., 2N_B$, there is a flux of TCLs entering it and a flux leaving it. Consider α_0 and α_1 to be the average OFF/ON rates of cooling TCLs. From (2.1), α_0 and α_1 can be approximated by [11, 55],

$$\alpha_0 = \frac{1}{CR}(\theta^{\text{amb}} - \theta^{\text{set}}), \tag{2.16}$$

$$\alpha_1 = \frac{1}{CR}(\theta^{\text{amb}} - \theta^{\text{set}} - PR) \tag{2.17}$$

Then, following the backward difference discretization procedure of [11], for bins $2 \leq i \leq N_B$ and $N_B + 2 \leq i \leq 2N_B$ we obtain

$$\dot{x}_i(t) = \frac{\alpha_0}{\Delta^{\text{bw}}}\Big(x_{i-1}(t) - x_i(t)\Big); \quad 2 \leq i \leq N_B, \tag{2.18a}$$

$$\dot{x}_i(t) = \frac{\alpha_1}{\Delta^{\text{bw}}}\Big(-x_{i-1}(t) + x_i(t)\Big); \quad N_B + 2 \leq i \leq 2N_B, \tag{2.18b}$$

and for the bins, $i = 1$ and $i = N_B + 1$, at $\theta^{\min}$ and $\theta^{\max}$ boundaries, we obtain

$$\dot{x}_1(t) = -\frac{\alpha_0}{\Delta^{\text{bw}}}x_1(t) - \frac{\alpha_1}{\Delta^{\text{bw}}}x_{2N_B}(t); \tag{2.19a}$$

$$\dot{x}_{N_B+1}(t) = -\frac{\alpha_0}{\Delta^{\text{bw}}}x_{N_B}(t) + \frac{\alpha_1}{\Delta^{\text{bw}}}x_{N_B+1}(t). \tag{2.19b}$$

Thus, the entries of the $\mathcal{A}$-matrix can be easily computed in the noise-free case. A small scale example of an $\mathcal{A}$-matrix is provided in Appendix A.

Note that, given a sampling time of t_s (seconds), a discrete-time equivalent A-matrix can be obtained directly from $\mathcal{A}$ by

$$A = \exp(\mathcal{A}t_s), \tag{2.20}$$

which can then be used in (2.8).

2.4. Battery Model Representation

An important analogy can be drawn between models of TCLs and storage systems. It has been shown in [40, 68] that the state dynamics of DERs, such as TCLs, electric vehicles and batteries, can be expressed using a *generalized battery model* (GBM).

Let $e_{j,k}$ be the j-th battery's energy state, i.e. its state of charge (SOC), at time k. Then, the evolution in $e_{j,k}$ can be modeled with the discrete-time difference equation,

$$e_{j,k+1} = a_j e_{j,k} + d_{j,k}, \tag{2.21}$$

where $e_j^{\min} \leq e_{j,k} \leq e_j^{\max}$, and $d_{j,k}$ is the power consumed, with $d_j^{\min} \leq d_{j,k} \leq d^{\max}$. The coefficient a_j is the j-th DER's energy dissipation rate, with $a_j \in (0, 1]$ for TCLs and HVAC systems, and $a_j \approx 1$ for storage devices [106].

Assume $e_{j,k}$ is normalized, i.e. $e_j^{\min} = 0$ and $e_j^{\max} = 1$. Considering $u_{j,k} \in \{0, 1\}$ is the discrete power on/off decision, and γ_j is the charging efficiency (i.e. the energy gain if the DER is on), we can write,

$$e_{j,k+1} = a_j e_{i,k} + \gamma_j u_{j,k}. \tag{2.22}$$

The GBM representation is especially convenient for power system studies since the states capture the amount of energy stored (kWh) and the power consumed (kW) and can be aggregated across different device types.

Chapter 3.

Reduced-order Aggregate Models in the Presence of Noise and Parameter Heterogeneity

3.1. Chapter Introduction

Large groups of TCLs can be controlled for providing various power systems services. However, controlling large groups of TCLs may cause their temperatures to synchronize, which might then lead to undesired behaviour of the total load (see [44], [54], [87]). Since aggregate models are commonly used to characterize TCL dynamics and to control their aggregate power (see [12], [70]), this chapter focuses on analyzing the ability of such models to capture total load behaviour, especially in the presence of noise and parameter heterogeneity. The need for analytical approaches to modeling is addressed.

TCLs are heterogeneous and their dynamic behavior is affected by several sources of uncertainty. Therefore, aggregate models need to capture those influences. Simulations have shown that heterogeneity and noise can add additional desirable damping to the system and the steady state is therefore reached faster (see [66], [17], [86], [33]). [86] provides an analytical model to approximate the influence of heterogeneity, specifically for the thermal capacitance parameter. [33] analyzes the impact of heterogeneity by considering different sub-groups of homogeneous TCLs. Several other studies have also used the bin model to consider probabilistic distributions of various parameters governing the TCL population (see [51], [70]). However, there remain ambiguities concerning the performance of bin models when considering heterogeneity and noise.

To model the system behavior using a bin model, it is necessary to find the Markov transition matrix that governs the population dynamics. While system identification (SI), based on sampling from Markov chains, may be used to identify the transition matrix (see [51], [70]), it can be computationally intensive for reasonably large systems. Moreover, the process must be repeated whenever parameters change. While an analytical approach was presented by [51], it ignored noise and made restrictive assumptions on parameter heterogeneity. Hence, the objective of this chapter is to provide an analytical (though approximate) approach to incorporate noise and heterogeneity into bin models. This provides a basis for analyzing their impact on the population performance.

The remainder of the chapter is organized as follows. Section 3.2 presents the model preliminaries, discusses the various ways of obtaining the aggregate model and highlights their advantages or disadvantages. Section 3.3 presents algorithms for computing the A-matrix taking into account noise and heterogeneity. Section 3.4 provides several analytical as well as simulation-based results on model performance, invariance properties of the bin model, and compares the influences of model error, noise and heterogeneity. Section 3.5 summarizes the main contributions.

3.2. Reduced Model Computation

3.2.1. Obtaining the A-matrix

Recall that matrix A in (2.8) is the transpose of the Markov transition matrix that is composed of the probabilities of transitioning from bin to bin (2.9). Two approaches to obtaining the A-Matrix are summarized below.

System Identification (SI) approach:

By simulating a system of TCLs using (2.4)-(2.5) and using full state information from all of the simulated TCLs, the number of TCL transitions from each starting bin to each ending bin can be counted. By collecting a large number of samples, a Monte-Carlo estimate of the Markov transition matrix can be obtained. To ensure conservation of probability mass, the resulting matrix must be normalized. Finally, the transpose of the identified Markov matrix is the desired A-matrix [70], [51].

The main difficulty related to the SI approach is that it can be computationally intensive,

especially when one aims to obtain a higher order A-matrix. With no noise or heterogeneity, we ran simulations with 1000 TCLs over 24 hours with $h = 10$ second time-steps and obtained the A-matrices for N_{B}=10, 20, 40, 80, 160, and 320 bins. The process of obtaining the A-matrix estimate with small N_{B} was relatively fast, for $N_{\mathrm{B}} = 10$ it took 50 seconds. However, the required time increased significantly for large N_{B}, with $N_{\mathrm{B}} = 160$ requiring 186 minutes.

Since the system identification process is parameter specific, obtaining the A-matrix for a different set of parameters requires repetition of the identification process. Also to obtain a good estimate using the SI-based approach, the sampling process needs to be exhaustive across diverse initial conditions. Otherwise, the identified A-matrix will perform poorly when it is applied to a population with an arbitrary initial condition.

Analytical approach:

The A-matrix can also be built by analytically computing the transition probabilities using the thermal dynamic equations (2.4)-(2.5) and numerically integrating over the temperature ranges of the sending and receiving bins. Authors in [51] proposed a technique that works for heterogeneity in the C parameter. However, this technique does not take into account noise and authors in [70] mention numerically solving for the transition probabilities may face difficulties when considering heterogeneity in parameters other than C. Hence, we believe that there is a need for alternate analytically based approaches to obtain the A-matrix considering noise and parameter heterogeneity, which is the primary focus of this chapter. Additionally, the intuition obtained from our proposed techniques will be applied in Section 3.4 to study bin model fidelity, the effects of bin widths, and to compare the influences of noise and heterogeneity.

3.3. Incorporating Noise and Heterogeneity in the Bin-Model

3.3.1. Incorporating noise

Initially, consider a homogeneous TCL population. Figure 3.1 shows two specific OFF bins, indexed n and $n + 1$, and their contents at two subsequent time period t_1 and t_2 (where $t_2 = t_1 + h$).

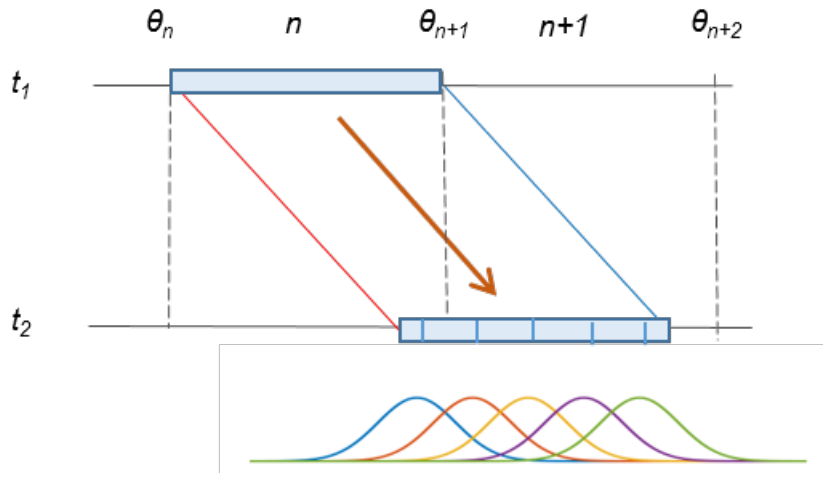

Figure 3.1: Top figure: TCLs propagating over a time-step; Bottom figure: Approximate probability distribution of TCLs at the end of a time-step.

Assume all TCLs lie *uniformly* inside bin n at time t_1, where the uniform assumption inside any bin is consistent with [51], [70]. Then, at time t_2, governed by (2.4)-(2.5), TCLs will transition to a range of bins.

With noise, $\mathbf{w} \sim N(0, \sigma_w^2)$, intuitively, the range of temperature reached at time t_2 should be wider. Depending on the noise variance, TCLs may reach the bins in different proportions to the noise-free and may also reach neighboring bins not reached previously. Thus, there is a need to compute the fractions that evolve to various bins given that they all start from the same bin. At time t_1, consider a specific sending bin n. For TCLs starting from bin n, let θ_n^{start} represent the initial temperature and θ_n^{end} represent the final temperature one time step ahead. Note that θ_n^{end} can be inside or outside bin n.

The probability of lying inside the receiving bin j, with boundaries θ_j and $\theta_{(j+1)}$, can be found from the fractions falling below and above bin j in the following manner,

$$\begin{aligned} &P(\theta_j \leq \theta_n^{\text{end}} \leq \theta_{j+1} | \theta_n^{\text{start}}) \\ &= 1 - \Big(P(\theta_n^{\text{end}} \leq \theta_j | \theta_n^{\text{start}}) + P(\theta_n^{\text{end}} > \theta_{j+1} | \theta_n^{\text{start}}) \Big). \end{aligned} \tag{3.1}$$

Instead of a continuous uniform distribution over the n-th bin, consider a discrete uniform distribution. The bin population is concentrated at M discrete values $\theta_{nl}^{\text{start}}, l = 1, ..., M$ that are evenly spaced across the width of the bin. The n-th bin now has M discrete *sub-bins*, each with height $1/M$. Inside each sub-bin, consider a TCL population

with the same parameters as considered for the bin.

Now, for TCLs starting from bin n and sub-bin l, let $\theta_{nl}^{\text{start}}$ and θ_{nl}^{end} represent the initial and the final temperature. Again, (2.4)-(2.5) dictate how the discrete pulses propagate forward in time. So, with $\mathbf{w} = 0$, the pulse at $\theta_{nl}^{\text{start}}$ will move to θ_{nl}^{end}. Next, with noise, $\mathbf{w} \sim N(0, \sigma_w^2)$, there will be a Gaussian distribution around each center, $\theta_{nl}^{\text{end}}, l = 1, ..., M$. This gives a Gaussian mixture distribution, i.e. a mixture of $N(\theta_{nl}^{\text{end}}, \sigma_w^2), l = 1, ..., M$ (as shown in Fig. 3.1).

The contribution to each receiving bin must be computed. This can be achieved by computing tail probabilities, according to $Q(\cdot) = 1 - \Phi(\cdot)$, where $\Phi(\cdot)$ is the Gaussian cumulative distribution function (CDF).

Starting from bin n, the probability of TCLs falling above the upper boundary of receiving bin j at time t_2 can be computed by summing over the tail probabilities of Gaussian distributions centered at $\theta_{nl}^{\text{end}}, l = 1, ..., M$ with variance σ_w^2,

$$P(\theta_n^{\text{end}} > \theta_{j+1} | \theta_n^{\text{start}}) = \frac{1}{M} \sum_{l=1}^{M} Q\Big(\frac{\theta_{j+1} - \theta_{nl}^{\text{end}}}{\sigma_w}\Big). \tag{3.2}$$

Similarly, the probability of the tail falling below the lower boundary is given by,

$$P(\theta_n^{\text{end}} \leq \theta_j | \theta_{nl}^{\text{start}}) = 1 - P(\theta > \theta_j | \theta_{nl}^{\text{start}}) \tag{3.3}$$

$$= 1 - \frac{1}{M} \sum_{l=1}^{M} Q\Big(\frac{\theta_j - \theta_{nl}^{\text{end}}}{\sigma_w}\Big). \tag{3.4}$$

This same process is repeated for every receiving bin, giving the entries of the transition matrix for sending bin n. The overall algorithm can be summarized as:

Algorithm I

For all sending bins $n = 1, ..., 2N_{\text{B}}$:

(i) For the specific sending bin n, divide the bin population evenly among M discrete temperature values at $\theta_{nl}^{\text{start}}, l = 1, ..., M$.

(ii) Using (2.4)-(2.5) and $\mathbf{w} = 0$, find the corresponding time-propagated values $\theta_{nl}^{\text{end}}, l = 1, ..., M$.

(iii) With $\mathbf{w} \sim N(0, \sigma_w^2)$, establish M normal distributions $N(\theta_{nl}^{\text{end}}, \sigma_w^2), l = 1, ..., M$.

(iv) Given the pre-specified temperature boundaries for receiving bin j, use (3.1)-(3.4) to compute the fractions of TCLs that propagate to each bin $j = 1, ...2N_{\mathrm{B}}$. □

When noise is high, it may cause transitions from n to $j < n$, i.e. in the opposite direction to normal. This must be handled carefully at the switching thresholds, bins 1 and $N_{\mathrm{B}}+1$.

TCLs falling below j=1 belong must remain off. If bins are not considered outside of the dead-band range, this small fraction of TCLs can be captured in bin $j = 1$. Similarly, TCLs starting in bin $N_{\mathrm{B}}+1$ and that experience a temperature rise must remain on and be captured in $N_{\mathrm{B}}+1$. However, for a more accurate model, specially when noise level is high, we suggest considering extra bins in 'OFF' state below $\theta^{\min}$ and similarly extra 'ON' bins above $\theta^{\max}$.

If noise is relatively low and the number of bins N_{B} is small (i.e. the bins are wide), we have observed that noise has limited effect on the A-matrix coefficients. This suggests that given a specific bin width, the A-matrix coefficients may be invariant to a certain level of noise. This property is analyzed further in Section 3.4.4.

3.3.2. Incorporating Parameter Heterogeneity by Mean-Variance Approach

In this section, Algorithm I will be modified to consider parameter heterogeneity rather than noise. Consider heterogeneity only in P, and assume P is distributed with $N(\mu_P, \sigma_P^2)$. Let σ_P can be expressed as a fraction σ_r of the mean value of P, i.e., $\sigma_P = \sigma_r \mu_P$.

Starting from temperature θ_t, a TCL's cooling and heating rates, $\alpha_{\mathrm{ON},t}$ and $\alpha_{\mathrm{OFF},t}$, are given by the expressions [17], [11],

$$\alpha_{\mathrm{ON},t} = \frac{\theta_t - \theta^{\mathrm{amb}} + PR}{CR} \tag{3.5}$$

$$\alpha_{\mathrm{OFF},t} = \frac{\theta_t - \theta^{\mathrm{amb}}}{CR}. \tag{3.6}$$

Given $P \sim N(\mu_P, \sigma_p^2)$, the expected value of $\alpha_{\mathrm{ON},t}$ can be computed using (3.5),

$$\mathrm{E}[\alpha_{\mathrm{ON},t}] = \mathrm{E}\Big[\frac{\theta_t - \theta^{\mathrm{a}} + PR}{CR}\Big] = \frac{\theta_t - \theta^{\mathrm{a}}}{CR} + \frac{\mu_P}{C} \tag{3.7}$$

and the variance of $\alpha_{\mathrm{ON},t}$ is,

$$\mathrm{Var}(\alpha_{\mathrm{ON},t}) = \mathrm{Var}\Big(\frac{\theta_t - \theta^{\mathrm{a}} + PR}{CR}\Big) = \frac{\sigma_P^2}{C^2}. \tag{3.8}$$

Hence, α_{ON} is distributed as $N(\frac{(\theta_t-\theta^{\mathrm{a}})}{CR} + \frac{\mu_P}{C}, \frac{\sigma_P^2}{C^2})$. From (3.6), $\alpha_{\mathrm{OFF},t}$ is independent of P.

Starting from a sending bin n and sub-bin l at time t_1, with $\theta_{nl}^{\mathrm{start}}, l = 1, ..., M$, use (2.4)-(2.5) with μ_P to obtain $\theta_{nl}^{\mathrm{end}}$. Similar to Section 3.3.1, we obtain a mixture of Gaussian distributions centered at $\theta_{nl}^{\mathrm{end}}, l = 1, ..., M$ and each with variance σ_P^2/C^2. Following the same procedure for computing the tail probabilities, using (3.1)-(3.4) but replacing σ_w^2 by σ_P^2/C^2, the fractions inside each receiving bin can be computed. Due to α_{OFF} being independent of P, the OFF-bin transition probabilities are identical to the homogeneous case. Only transitions through ON bins are affected by the variance term.

Heterogeneity in other parameters, such as $\theta^{\mathrm{a}}, C, R$ can be incorporated following a similar approach.

3.3.3. Incorporating Heterogeneity by Considering Groups

The previous section uses the mean and variance of the cooling and heating rates to obtain the A-matrix coefficients. Alternatively, since the aggregate response of a heterogeneous population is simply the combination of homogeneous TCLs, in this section, we model heterogeneity by dividing the population into N_G number of homogeneous groups, where g is the group index. The construction of homogeneous groups is elaborated by example in Section 3.4.5. Next, following the same approach as Algorithm I (with $\mathbf{w} = 0$), we can obtain the A-matrix for each group, $A_g, g = 1, 2, ..., N_G$. Then, the combined A-matrix can be readily constructed as below,

$$A = \begin{bmatrix} 0A_1 & [0] & \cdots & [0] \\ [0] & A_2 & \cdots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \cdots & A_{N_{\mathrm{G}}} \end{bmatrix}. \tag{3.9}$$

Note that since A_gs are independent, their computation and simulations can be performed in parallel. The aggregate response can be obtained by summing the weighted

response of all groups. In (3.9), we used $\mathbf{w} = 0$. Instead A-matrix can be computed at presence of noise (following Algorithm I or II). Let $\tilde{A}_g$ denote the A-matrix of group g at presence of noise. Then, we can replace all A_gs in (3.9) by $\tilde{A}_g$s, which then jointly accounts for heterogeneity and noise.

3.3.4. Comparison of the Mean-Variance-based and the Group-based Approaches

While the mean-variance based method presented in Section 3.3.2 requires $2N_B$ dimensional state space, in the group-based approach the dimension is further multiplied by N_G. While this adds computational burden, we argue that this is necessary to better approximate the aggregate dynamics of TCLs when considering parameter heterogeneity. To construct the A-matrix, the mean-variance approach in section 3.3.2 assumes within each bin, TCLs are uniformly distributed. However, in reality, TCLs transitioning with different heating/ cooling parameters shift to different bins over a time step. Hence, the *uniform distribution inside a bin* assumption leads to the model diverging from the true behavior. For an example, consider two groups of TCLs, each with $N/2$ TCLs. Assume, group-2's heating and cooling rates are two times faster than the group-1's rates. Thus, in one time-step, assume all TCLs in group-1 shift exactly by 1 bin, whereas all TCLs in group-2 shift exactly by 2 bins. We can see that, if all TCLs initially start at the same bin, they can only overlap in the same bin only at certain time-periods. For all other times, each group's TCLs lie in separate bins - which the model in Section 3.3.2 cannot distinguish. Simulation results further confirm our assumption and provide further justification on why the group-based approach lead to better approximations of the aggregate behavior.

3.4. Results and Analyses

3.4.1. Model Performance: Homogeneous Case without Noise

Consider the following values for TCL parameters: $h = 10$ seconds, $\delta = 0.5^{\circ}$C, $\theta^{\text{set}} = 20^{\circ}$C, $\theta^{\text{a}} = 32^{\circ}$C, $C = 10$ kWh/$^{\circ}$C, $R = 2^{\circ}$C/kW, $P = 14$ kW, and $\eta = 2.5$, [12], [17]. When a homogeneous, noiseless TCL population is simulated using (2.4)-(2.5), the power output shows undamped oscillations (the rectangular pulse shown in Fig. 3.2). Simulations with different numbers of bins N_{B} were carried out, with the responses shown in Fig. 3.2. It can

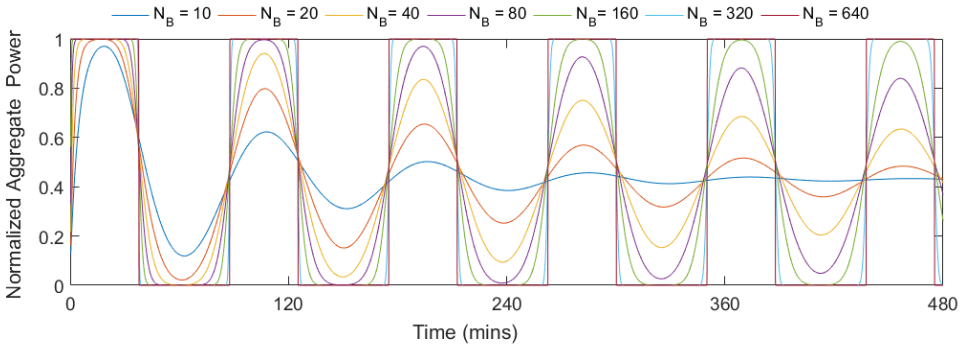

Figure 3.2: Simulation of homogeneous TCLs with varying N_{B}.

be seen that for $N_B = 320$ and 640, the bin model accurately matches the true rectangular pulse. In contrast, with N_{B} less than 100, the bin model (incorrectly) displays considerable damping.

The primary sources of model error are typically the discretization of time and temperature. Since the bin model is based on probabilistic transitions, low N_{B} cannot accurately capture the aggregate dynamics. Assume all (homogeneous) TCLs start at the same temperature $\theta^{\mathrm{start}} = \theta^{+}$ at t_1, and all are in the off state (i.e. in bin N_{B}). One time-step later, according to the transition matrix, $p_{N_{\mathrm{B}},N_{\mathrm{B}}}$ fraction stays at bin N_{B}, whereas the fraction $(1 - p_{N_{\mathrm{B}},N_{\mathrm{B}}})$ move to ON bins. In N_{B} time-steps, all ON bins will have received some fraction of the TCLs. Subsequently OFF bins start to fill. The aggregate power, given by the number of TCLs in the ON bins, will gradually reduce, rather than continuing to exhibit its true rectangular-pulse response. If the number of bins over the deadband range is low, damping appears sooner (as observed in Fig. 3.2).

Additionally, model performance depends on temperature initial conditions. If TCL temperatures are uniformly distributed over the dead-band, the power output is expected to have smoother transitions. Readers may refer to [11] where the performance of the bin model was similarly compared by varying N_{B}. Since initial conditions were dispersed, the resulting output power was triangular shaped, which was possible to match with lower N_{B} than that required for the case in Fig. 3.2.

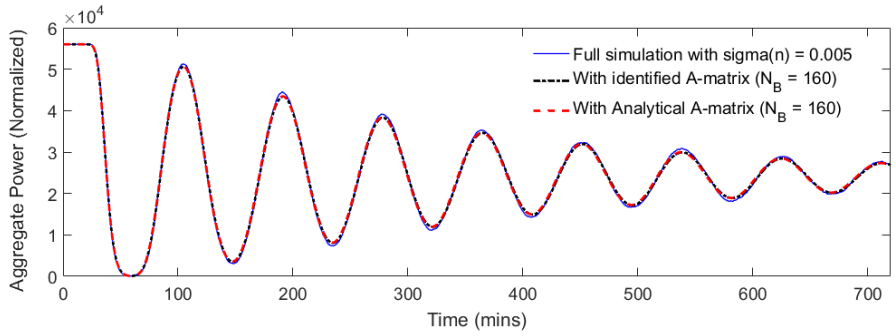

Figure 3.3: Simulation of TCLs with identified, analytical and full model with Gaussian noise.

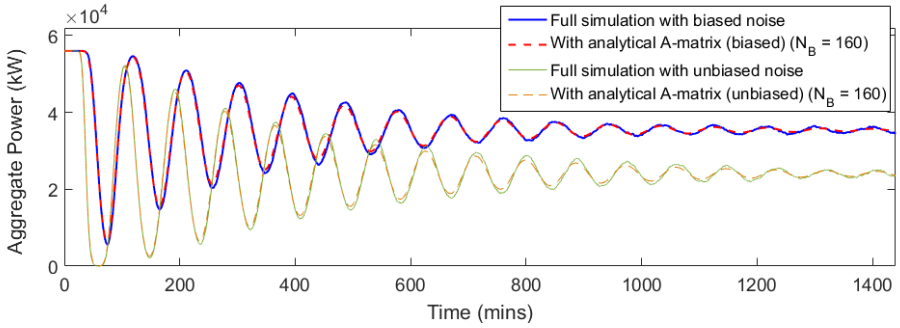

Figure 3.4: Simulation of TCLs with analytical and full model with biased and unbiased uniform noise.

3.4.2. Model Performance: Homogeneous Case with Noise

Let the noise variance have the value $\sigma_w = 0.005$. The analytical A-matrix for $N_{\mathrm{B}} = 160$ was obtained using Algorithm I. Also, 10,000 TCLs were simulated using (2.4)-(2.5) over 24 hours to obtain the identified A-matrix. Aggregate power outputs given by the analytical matrix and the identified matrix are compared in Fig. 3.3 with the output obtained using the full simulation. The aggregate power for all three representations agreed very closely. However, the analytical matrix was obtained in 2.5 seconds, whereas computation of the identified matrix required several hours.

Fig. 3.4 shows the performance of the analytical model when noise is uniformly distributed. Both unbiased (i.e. symmetric) and biased noise were considered. With positively-biased noise, the average aggregate power consumed is higher than when noise is unbiased. This is because positively-biased noise adds an upward temperature bias in (2.4), as air-

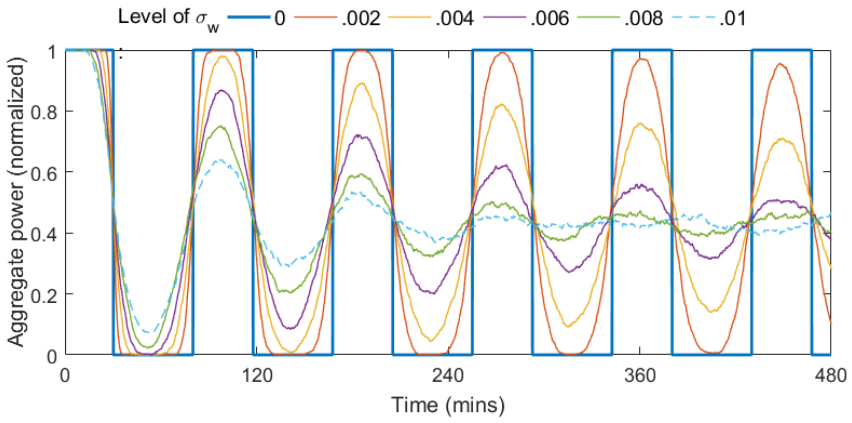

Figure 3.5: Aggregate power response with varying σ_w.

conditioners remain ON for longer and are OFF for a shorter time, i.e. the cooling rate is slower and the heating rate is faster.

3.4.3. Influence of Noise and Model Error on Damping

The influence of (Gaussian) noise on damping was analyzed by varying noise levels from $\sigma_w = 0$ to 0.01 in increments of 0.002. The results obtained from simulation of 10,000 TCLs using (2.4)-(2.5) are shown in Fig. 3.5. To compare the (true) decay of oscillations under the influence of noise against the (erroneous) decay due to model error (shown in Fig. 3.2), log decrement was used to quantify the decay rate. Using the first peak, D_1, and a subsequent peak after n periods, D_n, the log decrement is defined as $\delta_{\text{decay}} = \frac{1}{n}\ln(\frac{D_1}{D_n})$. The first and the third peaks were chosen for this investigation, with δ_{decay} computed for the cases shown in Fig. 3.2 and Fig. 3.5. The influence of model error and noise on decay rates is summarized in Fig. 3.6. The x-axis values, representing noise standard deviation and the number of bins, are normalized. The y-axis represents the log decrements in oscillation amplitudes.

To illustrate, compare the decays rates for $N_{\text{B}} = 40$ and $\sigma_w = 0.001$. In the first case, δ_{decay}=0.1 whereas $\delta_{\text{decay}} \approxeq 0$ for the second case. Hence, if a 40-bin model is used for a population with $\sigma_w = 0.001$, an erroneous damped response would be observed. The damping would be due to model error rather than the presence of noise. In contrast, using a 160-bin model for the same population would confirm $\delta_{\text{decay}} \approxeq 0$.

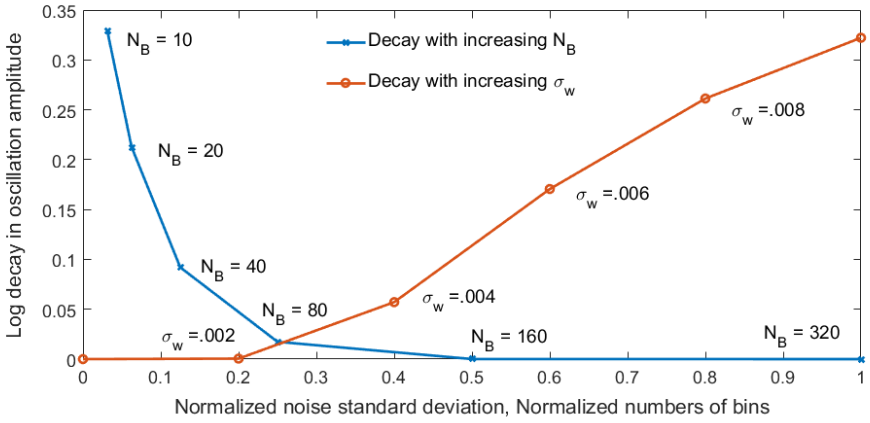

Figure 3.6: Damping due to varying N_{B}, σ_w.

Fig. 3.6 also shows that for $\sigma_w \leq 0.0025$ and $N_{\mathrm{B}} \leq 80$, the influence of model error on δ_{decay} is dominant. As the noise level rises above $\sigma_w = 0.0025$, noise has increasingly higher impact on damping. However, even at $\sigma_w \leq 0.006$, using a low order bin-model (e.g. N_{B} =10) would result in an apparent δ_{decay} that was higher than the actual value. Overall, this analysis demonstrates that a higher order model should be preferred to avoid the influence of model error on the damping of oscillations.

3.4.4. Noise Invariance in the Bin Model

Specifying N_{B} and δ determines the bin width $\Delta x = \delta / N_{\mathrm{B}}$. Fig. 3.1 shows that if bins are wide (N_{B} low), a small σ_w may not affect the bin transition probabilities - thus bins are *inherently tolerant* to some noise level. The probability of self-transition (i.e. sending and receiving bin are the same) is given by the diagonal elements of an A-matrix. This is shown in Fig. 3.7 for varying N_{B} and σ_w. (The diagonal element for the fifth bin is shown.) The x-axis represents noise as a percentage of the bin width. Each curve is for a specific value of N_{B}. For a particular N_{B}, when the curve is flat, the diagonal elements of the A-matrix are invariant to noise change. For example, when N_{B} = 10, 20, 40, 80, the A-matrix diagonal elements are almost invariant to noise increase until the noise level is quite high. In contrast, with large N_{B}, the diagonal elements are sensitive to low noise. Without noise, self-transition does not happen as bins are narrow. As noise level rises though, the left tail of the normal distribution (see Fig. 3.1) enters the bin, causing self-

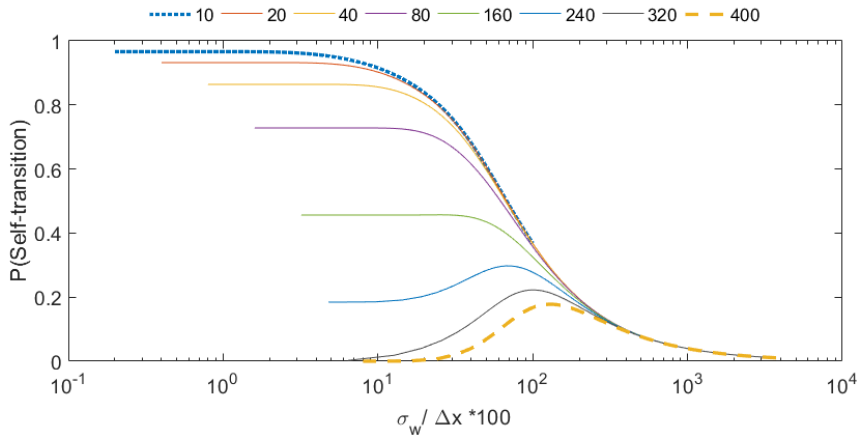

Figure 3.7: Probability of TCLs self-transitioning to a bin.

transition. Finally, ot can be seen that as $\sigma_w/\Delta x \times 100$ approaches 300 (i.e. $\sigma_w \approxeq 3\Delta x$), the self transition probabilities for all models converge. Similar convergence behavior was also observed for off-diagonal transitions. This implies that as σ_w exceeds $3\Delta x$, the noise level rather than the thermal dynamics tends to dominate the transition probabilities.

3.4.5. Model Performance Considering Parameter Heterogeneity

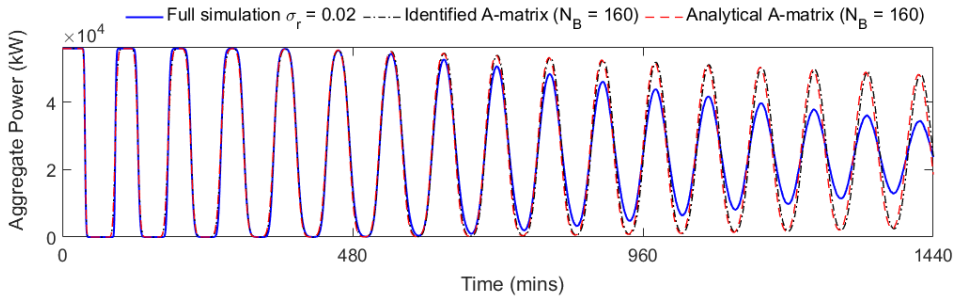

Figure 3.8: Simulation of TCLs with identified, analytical and full model with heterogeneity in P.

Consider a heterogeneous population of 10,000 TCLs. Assume P is distributed with $N(\mu_P, \sigma_P^2)$, where $\sigma_P = \mu_P \sigma_r$ with $\mu_P = 14$ kW, $\sigma_r = 0.02$. The analytical matrix was computed using a variation of Algorithm I. The identified matrix was obtained in the usual

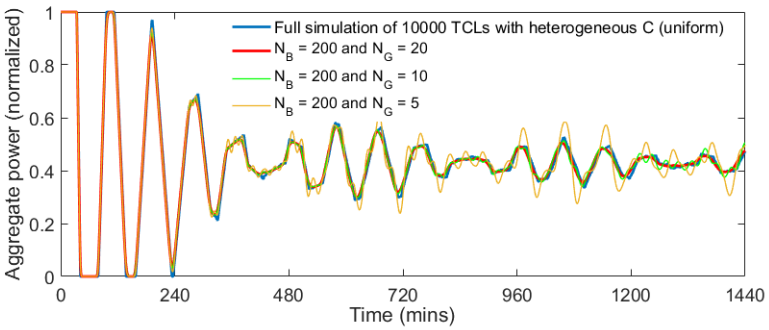

Figure 3.9: Aggregate power, heterogeneous C.

way. Fig. 3.8 shows that the aggregate power output from these two versions of the bin model matched exactly. When compared with the output from full simulation, results matched closely to around 620 minutes. Then, however, the output of the full simulation started to damp at a faster rate than the bin model. This deviation confirms that the bin model (obtained by sampling or by the mean-variance approach of section 3.3.2) has limitations in capturing heterogeneity, as discussed previously in section 3.3.4 and in [5]. This issue can be resolved by considering the group-based approach of section 3.3.3.

Consider, for example, heterogeneity in parameter C, with $C \sim U[9, 11]$ and $N_{\mathrm{G}} = 2$. The first group could represent $U[9, 10]$ and the second group $C[10, 11]$. Accordingly, two homogeneous groups could be considered, the first with $C = 9.5$ and the second $C = 10.5$. Fig. 3.9 provides a comparison between the full simulation of 10000 TCLs and groups with $N_{\mathrm{G}} = 5, 10, 20$. In all cases, $N_{\mathrm{B}} = 200$ to avoid damping due to model error. With both $N_{\mathrm{G}} = 10, 20$, aggregate power matches very closely, whereas oscillations are accentuated with $N_{\mathrm{G}} = 5$. Following a similar approach, heterogeneity in P was considered, with results presented in Fig. 3.10. In this case, more groups were required to match the full simulation. Interestingly, since the entire A-matrix building process is analytical, no dimensionality challenges arose. Also, note that the beating phenomenon is expected due to mixing of homogeneous groups that have different heating and cooling periods, [25]. The approaches based on system identification and Algorithm I both fail to identify this beating and instead converge to a steady state.

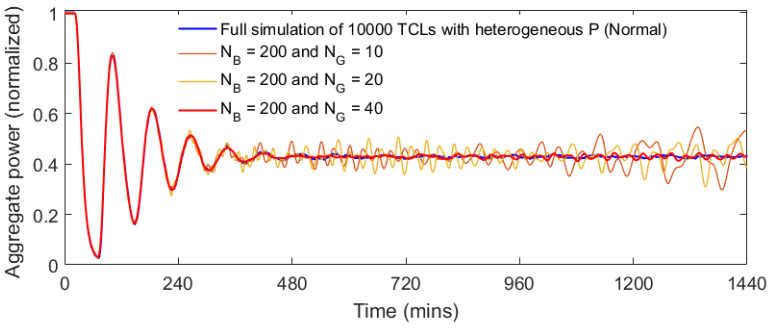

Figure 3.10: Aggregate power, heterogeneous P.

3.5. Chapter Conclusions

TCL aggregate models are widely in use yet there has been limited analysis of their ability to capture heterogeneity and noise. This work develops an analytical approach to incorporating noise and parameter heterogeneity into the A-matrix which governs the aggregate TCL dynamics. Large scale models that can take several minutes to hours using identification-based approaches can be obtained within seconds using the semi-analytical approaches proposed in this chapter. Using insights gained through this analytical modeling, the influence of noise and parameter heterogeneity on bin models has been studied. This showed how noise to bin-width ratio can indicate when variation of the noise level would have negligible impact on the model. Damping of power oscillations due to model error and noise was discussed. This highlighted the importance of choosing a sufficiently high number of bins to avoid damping due to model error. It was also shown that bin models are limited in their ability to capture heterogeneity. This can be addressed by decomposing the population into multiple homogeneous groups.

Chapter 4.

Modeling and Analysis of Transactive Energy Coordination

4.1. Chapter Introduction

Transactive energy (TE) coordination mechanism have been proposed as a framework for managing large numbers of distributed energy resources (DERs), such as thermostatically controlled loads (TCLs), energy storage and electric vehicles, in an electric grid [32, 43, 60]. Such schemes provide users with the flexibility to consume energy based on the price they are willing to pay. Several recent studies and demonstration projects have shown the applicability of such mechanisms to manage the aggregate demand of residential electric loads and commercial building heating, ventilation, and air conditioning (HVAC) systems [32, 38, 43, 50, 60, 61]. Applications include reduction of peak demand, provision of regulation services and congestion relief.

Under transactive (market-based) energy coordination, a population of distributed resources bids into the energy market and a certain level of demand is cleared, depending on the operating conditions of the grid. This process is primarily based on a double auction mechanism (illustrated in Fig. 4.2 and Fig. 4.3) [32]. A market mechanism proposed in [60] can attain economically efficient market outcomes while taking into account individual device dynamics. The applicability of this approach and the influence of different system parameters have been studied in [61]. While these market-based schemes can attain economically efficient outcomes, concerns remain regarding the impact on system stability. Previous analyses of electricity market dynamics [4, 89] have highlighted the possibility of power oscillations and highly volatile prices. In this chapter, we apply the bidding strategy

described in [32] (also in Section II of [60]) to show undesirable oscillatory response of a population of air-conditioners. Simulations suggest that several factors may contribute to load synchronization and power oscillations, including prolonged flat price signals followed by sharp changes in the price, feeder limits that are set too low and that are encountered periodically, and the form of user bid curves (results were originally published in our work [80]).

For generalization to other storage-like loads and electric vehicles (EVs), the second half of the chapter uses a generalized battery model, which is an equivalent representation for modeling TCL dynamics. Additionally, we incorporate *lockout conditions* to account for customers who may desire periods of uninterrupted supply. The possibility and nature of oscillations in the aggregate demand, given the presence of such lockout conditions, will be studied.

We then focus on developing a reduced-order aggregate model of DERs under market-based coordination, which is consistent with the modeling approach presented in Chapters 2 and 3. To accomplish this goal, a set Markov transition equations are developed over discrete ranges of DER energy and price levels (here referred to as "bins"). Lockout conditions are also incorporated. Eigenvalues of the resulting system matrix provide insights into system behavior. Although bin-based aggregate models have been used extensively for controlling aggregations of TCLs [51, 70, 79, 97], application to the transactive coordination framework has not previously been considered. Hence, a novel contribution of this modeling exercise is the extension of the bin-modeling approach to market-based coordination schemes for DERs. Furthermore, incorporating this model in a model predictive control (MPC) framework allows the calculation of optimal price signals for achieving desired DER energy usage. A reformulation of the transition equations ensures that the modified model remains linear for optimization-based market clearing strategies. However, a set of logical constraints are needed to model the market clearing behavior. Thus, a mixed integer formulation is obtained, providing a further contribution to aggregate modeling and market-based coordination of DERs. Additionally, relaxation of the integer model results in a quadratic program that is computationally more efficient. Measures to avoid synchronization are also incorporated in the MPC strategy.

Finally, we consider an application of MPC to manage a population of DERs in a distribution system. Several recent studies have shown that with increased penetration of controllable loads and storage devices, advanced congestion management techniques will be necessary to avoid simultaneous consumption from DERs and to limit payback from

unsatisfied demand [26, 36, 37, 45, 100]. Hence, we show that a population of TCLs can be managed economically while avoiding congestion in a distribution grid. Simulations also demonstrate that synchronization of TCLs and undesired power oscillations can be effectively avoided. The MPC formulation considers system level costs, operational constraints and constraints related to the DER population. Consequently, the generated price signals induce desirable behavior of the DER population. This addresses an important gap in transactive coordination as previous studies [32, 43, 60, 61] have taken price signals to be exogenous inputs rather than generated via feedback.

The rest of the chapter is organized as follows. Section 4.2 presents the TEC framework and section 4.3 presents various simulation results. Sections 4.4 and 4.5 present the bin-based aggregate model for DERs under market-based coordination (originally published in our work [76]). Section 4.6 presents the model predictive control (MPC) framework for controlling the DERs. Both a mixed-integer formulation and a relaxed quadratic program formulation are considered. Section 4.7 provides simulation results showing the performance of the aggregate model and the MPC controller. Section 4.8 concludes by summarizing our findings and discussing their implications.

4.2. Transactive coordination framework

The transactive control framework is based on a double auction mechanism [32]. Following the existing literature on the TC framework and modeling of the market clearing mechanism, subsequent work is based on the following assumptions: (i) A 'coordinator' is present to receive the bidding information from a population of devices and to send back the market clearing information. (ii) Each device is equipped with a smart thermostat that can measure the room temperature. It also has communication capabilities to exchange bid information with the coordinator. (iii) Before each market period, the device measures its room temperature, and submits a bid to the coordinator. The bid typically consists of the load power and the bidding price. (iv) The network flow constraints are ignored [60].

In a TC framework, every load submits a demand bid where it specifies its desired amount of energy demand over a specific interval. Note that to be consistent with the literature, market clearing intervals with *5-minute duration* have been considered. Hence, the bids are also based on average energy demand over 5-minute intervals.

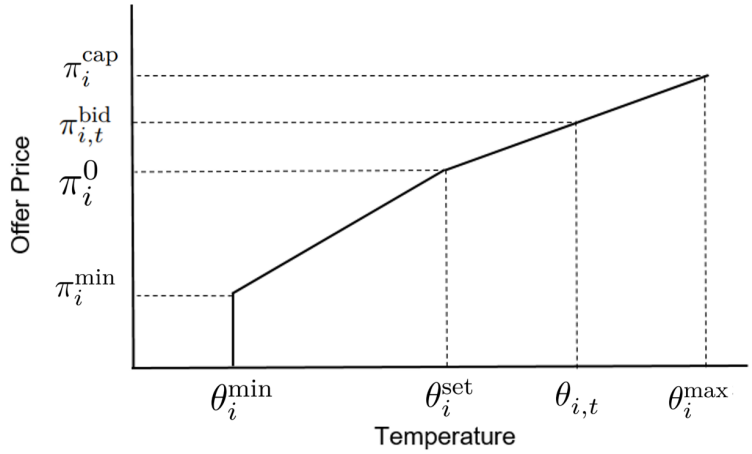

Figure 4.1: Demand side offer mapped to temperature.

4.2.1. Modeling TCL bids

Based on the above framework, let $\pi_{i,t}^{\text{bid}}$ denote the price bid of load i at time t and $d_{i,t}$ be its corresponding amount of energy demand over the next 5-minute period.

Fig. 4.1 shows how a TCL determines its bid [32, 61]. Here, an air-conditioner user bids π_i^0 if its temperature $\theta_{i,t}$ is at its set-point, θ_i^{set} (i.e. desired temperature level), with the offer varying if the temperature deviates from θ_i^{set}. Above a certain threshold θ_i^{max} the maximum bid is capped at π_i^{cap}. Similarly, below the threshold θ_i^{min} the TCL might not be willing to bid, so places $\pi_i^{\text{bid}} = 0$. Fig. 4.1 shows a piecewise linear mapping, with slopes γ_1 and γ_2 depending on if the temperature is above or below the set-point. Thus, the bid and temperature relation can be expressed as,

$$\pi_{i,t}^{\text{bid}} = \begin{cases} (\theta_{i,t} - \theta_i^{\text{set}})\gamma_1 + \pi_i^0, & \text{if} \quad \theta_{i,t} >= \theta_i^{\text{set}} \\ (\theta_i^{\text{set}} - \theta_{i,t})\gamma_2 + \pi_i^0, & \text{if} \quad \theta_{i,t} < \theta_i^{\text{set}} \\ 0, & \theta_{i,t} < \theta^{\text{min}} \\ \pi_i^{\text{cap}}, & \theta_{i,t} > \theta^{\text{max}}. \end{cases} \tag{4.1}$$

Since the bids are over 5-minute intervals, whereas TCLs have faster dynamics (few seconds), $\theta_{i,t}$ may be the latest measured temperature, or a predicted temperature (e.g. at 2.5 minutes ahead) based on its current on/off operating state, as detailed in [61].

4.2.2. Market Clearing Mechanism

Fig. 4.2 provides an overview of the market-clearing mechanism. Since in a transactive framework the grid is organized in a hierarchical way, the TCL nodes are connected to a distribution feeder, which clears an allowable demand level at a particular price. Initially the feeder broadcasts a base price, but adjusts that price if the feeder capacity constraint is exceeded.

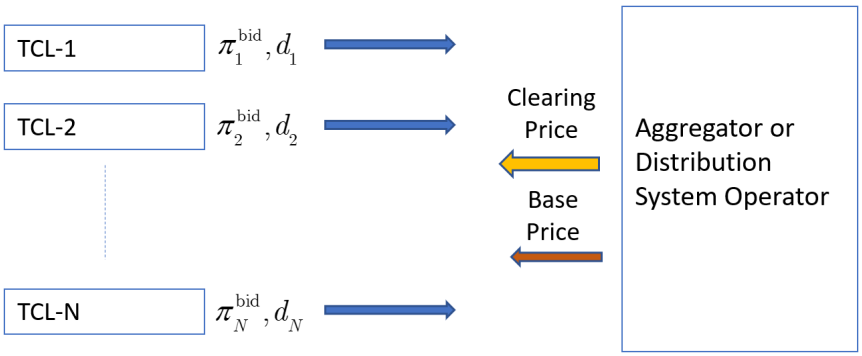

Figure 4.2: Transactive energy framework for TCLs.

Let π_t^{base} be the base price forecast at time t and d_t^{base} be the corresponding base aggregate demand. The clearing price π_t^{clr} and the cleared aggregate demand d_t^{clr} can be found at time t according to the following algorithm, keeping in mind that d_t^{clr} must satisfy the feeder capacity limit,

$$d_t^{\text{clr}} \leq d^{\text{Feeder}}. \tag{4.2}$$

The overall transactive control mechanism [32, 80] can be summarized as:

- Gather anonymous bids (price versus demand) and build an demand function (see Fig. 4.3).
- Using the aggregate demand function and the base price information for that time period π_t^{base}, obtain the corresponding base aggregate demand d_t^{base}.
- If $d_t^{\text{base}} < d^{\text{Feeder}}$ (see Fig. 4.3(a)), $d_t^{\text{clr}} = d^{\text{Feeder}}$. Set $\pi_t^{\text{clr}} = \pi_t^{\text{base}}$.
- If $d_t^{\text{base}} \geq= d^{\text{Feeder}}$ (see Fig. 4.3(b)), set $d_t^{\text{clr}} = d^{\text{Feeder}}$. Set π_t^{clr}.
- Each load compares its offer with π_t^{clr} and self-dispatches if $\pi_{i,t}^{\text{bid}} \geq \pi_t^{\text{clr}}$.

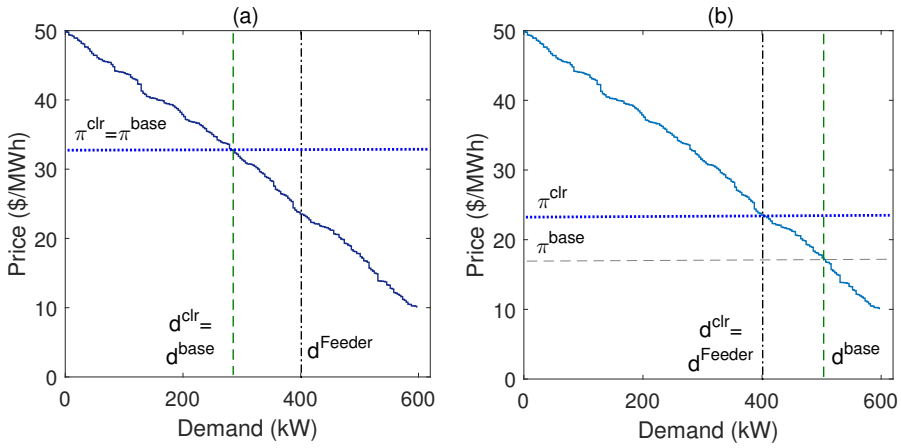

Figure 4.3: (a) Market clearing with feeder capacity not exceeded. (b) Market clearing with feeder capacity exceeded [32], [80].

4.2.3. Modified TCL Switching Logic

Under the transactive framework, the switching variable $m_{i,t}$ in (2.4) will be multiplied by an additional decision variable $v_{i,t}$, thus the overall expression becomes,

$$\theta_{i,t+h} = a_i\ \theta_{i,t} + (1 - a_i)(\theta^{\mathrm{a}} - m_{i,t}\ v_{i,t}\ \theta_i^{\mathrm{g}}) \tag{4.3}$$

where,

$$v_{i,t} = \begin{cases} 0, & \text{if} \quad \pi_{i,t}^{\mathrm{bid}} < \pi_t^{\mathrm{clr}} \\ 1, & \text{if} \quad \pi_{i,t}^{\mathrm{bid}} \geq \pi_t^{\mathrm{clr}}. \end{cases} \tag{4.4}$$

Here, $v_{i,t}$ can be thought of as an upper level decision variable, the TCL's response to a transactive incentive signal or a clearing price π_t^{clr}. If at any time $v_{i,t} = 1$ then the TCL simply follows its natural thermostat cycle. Note that under the above switching scheme,

- A TCL consumes power when $m_{i,t} = 1$, and $v_{i,t} = 1$.
- A TCL does not consume power when $m_{i,t} = 1$, $v_{i,t} = 0$.
- A TCL does not consume power when $m_{i,t} = 0$ (natural thermostat off mode).

4.3. Simulations

Consider a population of 1000 TCLs. Parameter values are similar to those used in [12, 17]. A base price is sent at 5-minute intervals. The coordinator sends the participants only the 5-minute ahead base price. Each load's bid levels are constructed with continuous offers, similar to Fig. 4.1. Bid levels can range between 10 to 50 \$/MWh. Each load has its own slopes γ_1 and γ_2 for its bid curve. Additionally, the feeder capacity constraint was set at 70% of the maximum power capacity of the TCLs (5600 kW for 1000 TCLs). Since the simulation of TCL temperature dynamics requires faster time steps, while market clearing occurs every 5 minutes, the TCL temperature dynamics were simulated using a time-step of $h = 10$ s, and the market mechanisms were simulated with 5-minute time-steps.

4.3.1. Oscillations Induced By Step Changes in Price Signals

Initial investigations considered the response of TCLs to sharp changes in the base price. The base price is initially 42 \$/MWh and stays at that level for 6 hours before suddenly dropping to 20 \$/MWh for a further 6 hours, and then finally to 9 \$/MWh for the remainder of the time. In reality, these price changes might correspond to sudden changes in background demand, such as an industrial load or electric vehicle charging.

Fig. 4.4 provides a prototypical example of TCL synchronization. The TCLs started with diverse initial temperatures. But because the base price remained high (at 42 \$/MWh) for a few hours, most of the TCLs did not initially consume power. (Their bids were not sufficiently high to be cleared.) However, within a few hours (around minute 200) their temperatures synchronized, and so do their bids. Later, as the base price drops to 20 \$/MWh, TCLs find the price level favorable and want to turn on. The aggregate demand reaches the feeder limit, causing the market clearing price to rise above the base price level. During minutes 480-720, the demand stays flat and TCL temperatures remain close to their set-point values. Next, at t=720 min, when the price drops to 9 \$/MWh, the TCLs find this low price even more favorable and many compete to consume power. Large oscillations in aggregate power are observed and the feeder limit is hit periodically. Thus, a step change in base price, especially to a low value, can induce large power oscillations. This is mainly due to TCL temperatures becoming synchronized during preceding periods of relatively high base prices.

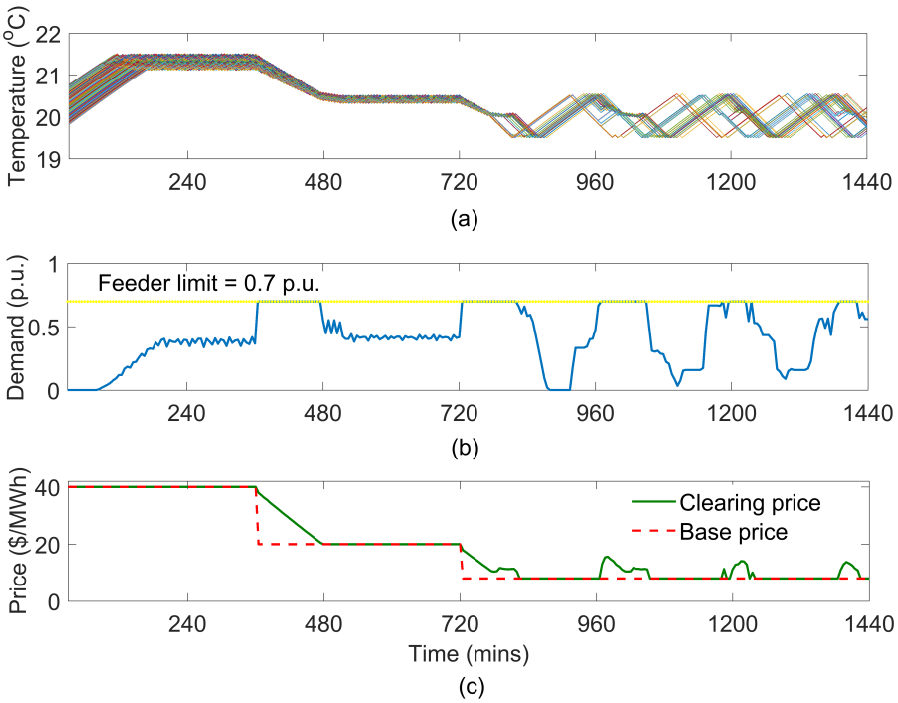

Figure 4.4: (a) Temperature evolution of individual TCLs, (b) 5-minute average aggregate demand, (c) base price and clearing price.

4.3.2. Fast Transients due to Temperature Synchronization and Price Fluctuations

Instead of large step changes in price, this case considered a price signal which fluctuates between 20 and 30 $/MWh. Behavior is shown in Fig. 4.5. We observe that this case triggers a highly fluctuating response in the 5-minute average TCL demand. Investigations suggest that variations in the TCL bids (as their temperatures change) relative to the base price cause these sharp transients in aggregate power levels.

This study assumed that the slopes of the bid curves, though heterogeneous, are not significantly different. Initially very few TCL bids were sufficiently high to be cleared. Hence, their temperatures rose to around 20.6°C. At this point, many placed sufficiently high bids and were subsequently cleared. If the base price remained unchanged, these TCLs would continue to consume power enabling their temperatures to reach the desired set-points. However, if the base price were to rise slightly, it would cause some TCLs to turn off since their bids become unfavorable. Conversely, if the majority of the TCLs were

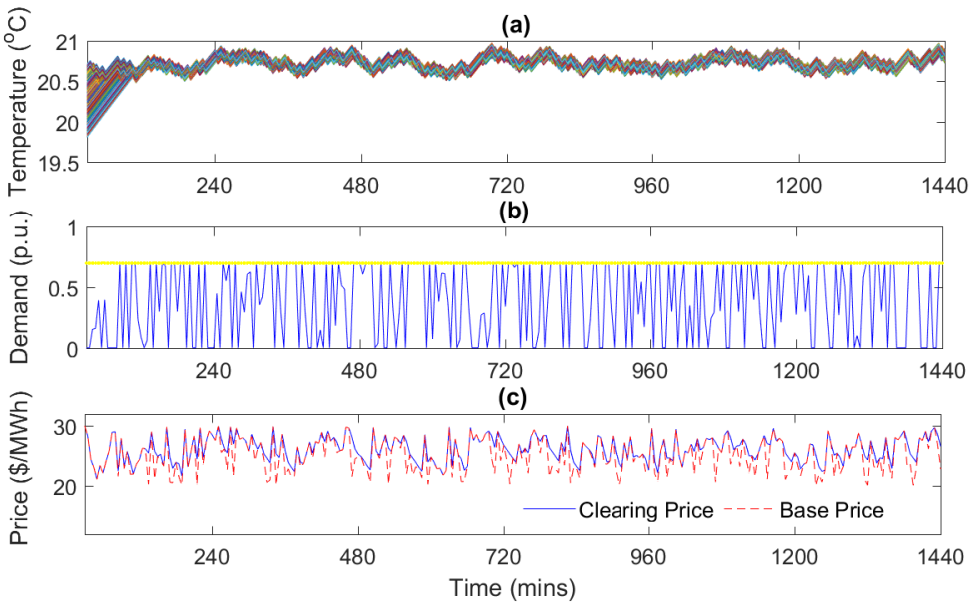

Figure 4.5: Temperature synchronization and highly fluctuating demand due to moderate fluctuations in base price.

off, then a small drop in the base price would lead to TCLs with similar bids being cleared and turning on. As their temperatures approach their set-points, they bid lower and at some point will no longer be cleared. Thus, these relative movements of the TCL bids (due to changes in their temperatures) compared to the base price levels may lead to significant fluctuations in the aggregate power, as shown in Fig. 4.5(b).

4.3.3. Oscillations Induced due to Feeder Capacity Constraint

Fig. 4.6 shows a situation where fast oscillations were induced due to the feeder capacity constraint. The base price signal in this case resembles a pulse train fluctuating between 14 and 24 \$/MWh. Every time the base price drops, TCLs switch on and the base aggregate demand of the TCLs reaches the feeder limit. For example, when the price drops to 14 \$/MWh at $t = 240$ min, all TCLs want to cool since their temperatures have risen considerably during the preceding high price period. However, if all TCLs turn on at the same time, the feeder limit will be violated. Following the mechanism described in Section 4.2.2, the clearing price is revised above the base price and therefore feeder limits are respected. However, as the clearing price approaches 14 \$/MWh, a specific pattern of fast

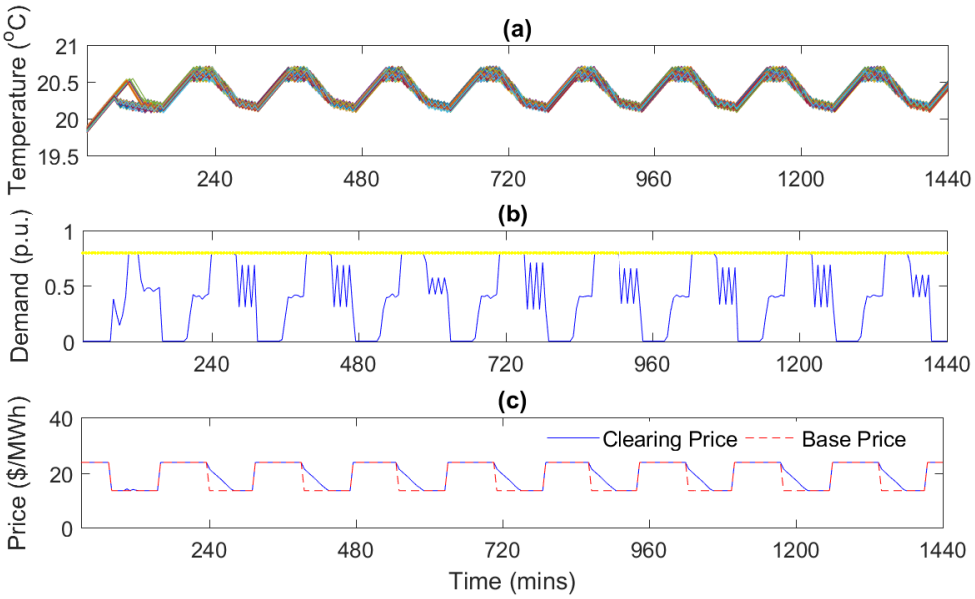

Figure 4.6: Fast oscillations due to groups of TCLs having synchronized temperatures.

oscillations emerges, as seen in Fig. 4.6(b).

By the time clearing prices approach 14 $/MWh, TCL temperatures are near their set-points so they offer low bids. However, a fraction of TCLs still bid higher than 14 $/MWh and are cleared. As these cool more, they bid lower and subsequently turn off. By that time, temperatures of a second group have risen such that their bids now exceed 14 $/MWh and they turn on. Thus, the most aggressive ones get cleared first, then the next group, and so on. Subsequently, as the base price rises again to 24 $/MWh, all loads turn off since they are unwilling to pay such a high price when their temperatures are already near their desired set-points. This behavior continues as long as the base price keeps oscillating.

4.3.4. Oscillations due to Subgroups of TCLs with Similar Bid Curves

This case shows that it is not necessary for all TCLs to be synchronized at the same temperature to cause power oscillations. It can be seen from Fig. 4.7 that groups of TCLs have synchronized temperatures, with TCLs within each group evolving in a similar manner. This then results in quasi-periodic behavior for the ensemble of loads. Besides large magnitude oscillations in power, the ensemble demand also displays jitter. The quasi-

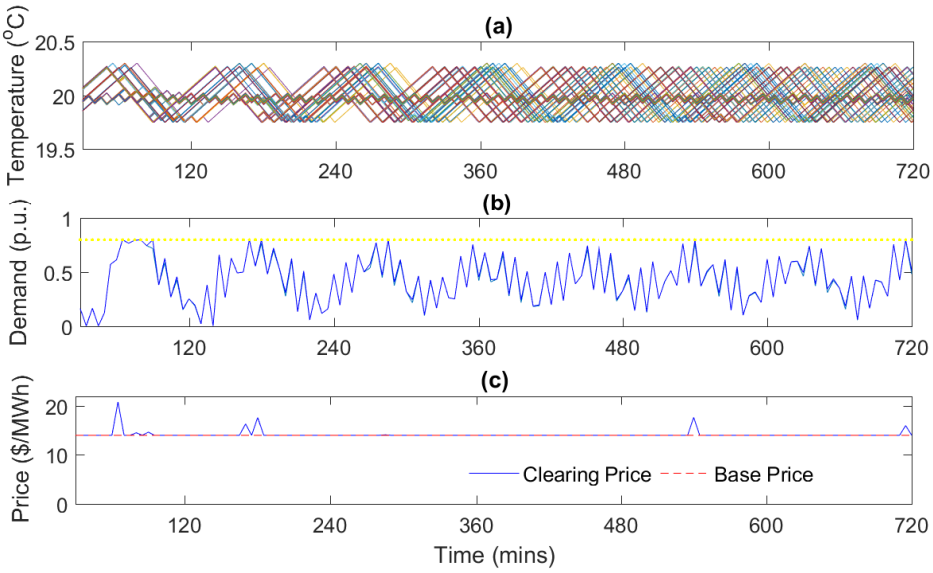

Figure 4.7: Fast power oscillations due to groups of TCLs having similar bids, leading to their synchronized temperatures.

periodic evolution of the ensemble results in mixing of oscillations of different frequencies.

4.4. TEC using a Generalized Battery Model Representation

Starting from this section and till the end of this chapter, we now generalize the TEC framework to consider other loads and stoage devices. As shown in Section 2.4, the state dynamics of DERs, such as TCLs, electric vehicles and batteries, can be expressed using a generalized battery model [40, 68]. Let $e_{j,k}$ be the j-th DER's energy state (i.e. SOC), at time k with $0 \leq e_{j,k} \leq 1$. $d_{j,k}$ is the power consumed, with $d_j^{\min} \leq d_{j,k} \leq d^{\max}$. $u_{j,k} \in \{0, 1\}$ is the discrete power on/off decision.

Next, let $\pi_{j,k}$ be the j-th DER's bid price at time k. Typically $\pi_{j,k}$ decreases as $e_{j,k}$ increases. Therefore, the bid function relating $\pi_{j,k}$ and $e_{j,k}$ can be expressed as,

$$\pi_{j,k} = \pi_j^{\max} - \beta_j e_{j,k}, \tag{4.5}$$

where $\pi_j^{\max}$ is the bid price when $e_{j,k} = 0$ and β_j represents the slope of the bid curve.

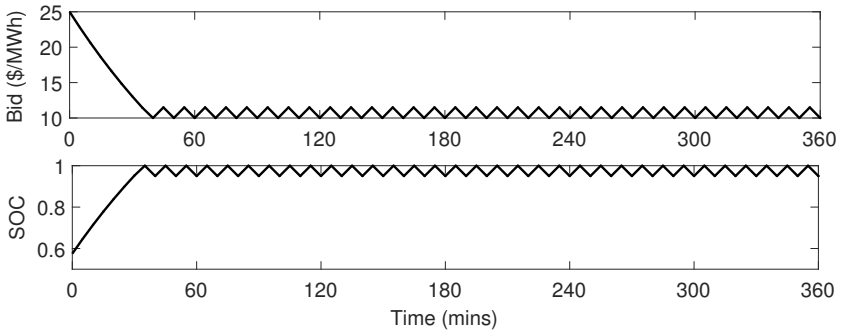

Figure 4.8: Fast cycling when no lockout mode is present.

The market-clearing mechanism remains the same as Section 4.2.2.

The response (dispatch decision) of an individual DER to a transactive incentive signal (price signal) π_k^{clr} is given by,

$$v_{j,k} = \begin{cases} 0, & \text{if} \quad \pi_{j,k} < \pi_k^{\text{clr}}, \\ 1, & \text{if} \quad \pi_{j,k} \geq \pi_k^{\text{clr}}. \end{cases} \tag{4.6}$$

In this case of market-driven decisions, $u_{j,k}$ can be replaced by $v_{i,k}$ in (2.22).

4.4.1. DER Switching Logic including Lockout Constraint

In (2.22) and (4.6), no restrictions have been imposed on how frequently $u_{j,k}$ can switch on/off. This may lead to fast cycling, which is often undesirable. Fig. 4.8 provides an illustration where a DER's bid price was initially above the clearing price so it charged. With time, the SOC increases and its bid price decreases. Eventually the DER will reach full SOC or its bid price will drop below the market clearing price. When that occurs, the DER switches off, which caused its SOC to drop and bid price to increase. Thus, its SOC and bid price oscillate within a narrow region. To avoid fast cycling when a DER is already fully charged, lockout constraints can be incorporated in the DER dispatch logic. For example, residential air conditioners employ hysteresis control to prevent fast cycling [80].

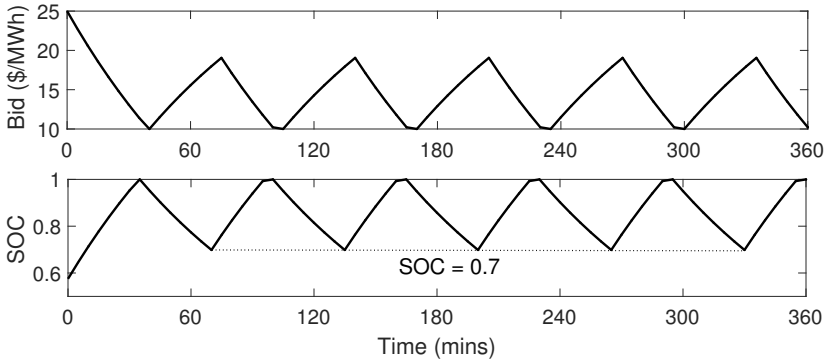

Figure 4.9: Prolonged on/off periods at presence of lockout mode.

Table 4.1: DER operating states.

Op. state	$m_{j,k}$	$v_{j,k}$	Outcome
1	1	1	Controllable and cleared
2	1	0	Controllable, but not cleared
3	0	0	Locked and turned OFF

Let $m_{j,k}$ denote the locked-out (off)/ not-locked out (controllable) operating state. Then,

$$m_{j,k+1} = \begin{cases} 0, & \text{if} \quad e_{j,k} \geq e_j^{\max}, \\ 1, & \text{if} \quad e_{j,k} < e_j^{\text{set}}, \\ m_{j,k}, & \text{otherwise}, \end{cases} \tag{4.7}$$

which implies once the SOC reaches $e_j^{\max}$, the DER enters a locked out mode until its SOC drops below a user-defined level, e_j^{set}. Thus, the modified generalized battery equation becomes,

$$e_{j,k+1} = a_j e_{j,k} + \gamma_j v_{j,k} m_{j,k}. \tag{4.8}$$

It follows that each DER has the three operating states given in Table 4.1. The effect of adding the lockout mode is shown in Fig. 4.9. Once reaching $e_{j,k} = e_j^{\max} = 1$ (full capacity), the DER remains off until its SOC falls below $e_{j,k} = 0.7$. Lockout conditions associated with $e_j^{\min}$ can be modeled in a similar way.

4.4.2. Simplified Aggregate Model

A simplified aggregate model is presented next to provide further intuition.

When u_k is relaxed to be a continuous variable, $0 \leq u_k \leq 1$, (2.22) can express the average dynamics of a homogeneous population of devices [106]. In the context of a large aggregate storage device, u_k represents the normalized power consumption at time k [49]. Alternatively, u_k can be interpreted as the probability of being ON (charging). To establish an initial (simple) model, assume there is a proportional relationship between the bid price and the charging decision (i.e. higher bids provide a higher probability of being charged). Then,

$$u_k = K_{\mathrm{p}} \pi_k, \tag{4.9}$$

where $K_{\mathrm{p}} \geq 0$ is a proportional constant, which implicitly reflects the DER's bid strength relative to the market clearing price. Let $K_{\mathrm{p}} = 1/\pi^{\max}$. Then $\pi_k = \pi^{\max}$ will give $u_k = 1$. Substituting (4.5) into (4.9) and then using (2.22) gives,

$$u_{k+1} = (a - \gamma\beta K_{\mathrm{p}}) u_k + \pi^{\max} K_{\mathrm{p}} (1 - a). \tag{4.10}$$

Define $K_{\mathrm{c}} = \pi^{\max} K_{\mathrm{p}} (1 - a)$ and $\alpha = (a - \gamma\beta K_{\mathrm{p}})$. Then,

$$u_{k+1} = \alpha u_k + K_{\mathrm{c}}, \tag{4.11}$$

which is a simple first order linear difference equation whose behavior is governed by underlying parameters γ, β and K_{p}. The solution to (4.11) can be written as,

$$u_k = \alpha^k u_0 + K_{\mathrm{c}} \sum_{j=0}^{k-1} \alpha^j, \tag{4.12}$$

with equilibrium given by,

$$u^* = \frac{K_{\mathrm{c}}}{1 - \alpha}, \quad e^* = \frac{\gamma \pi^{\max} K_{\mathrm{p}}}{1 - \alpha}, \quad \pi^* = \frac{\pi^{\max}(1 - a)}{1 - \alpha}, \tag{4.13}$$

where $\gamma, \beta, K_{\mathrm{p}} \geq 0$ and $\alpha \leq a$. Moreover, since $a < 1$ for any lossy battery, $\alpha < 1$. Stability conditions can easily be derived from (4.11). When $|\alpha| < 1$, the solution converges to the equilibrium u^* (i.e. the equilibrium is stable). When $0 < \alpha < a$ the solution is monotonic, whereas when $-1 < \alpha < 0$ the solution oscillates, with decreasing amplitude. Finally,

when $\alpha < -1$, the solution oscillates with increasing amplitude, resulting in instability. This may occur with small values of a or with large values of γ, β or K_{p}. Thus, the bid curve slope β (also mentioned in [106]), a, γ and K_{p} all influence stability conditions.

For a numerical example, consider two cases:

(i) $a = 0.9$, $\gamma = 0.1$, $\beta = 50$, $\pi^{\max} = 50$, $K_{\mathrm{p}} = 0.02$, and

(ii) $a = 0.7$, $\gamma = 0.25$, $\beta = 150$, $\pi^{\max} = 150$, $K_{\mathrm{p}} = 0.05$.

For the first case, $\alpha = 0.8$ and the solution converges to $u^* = 0.5$, $e^* = 0.5$ and $\pi^* = 25$. In the second case, $\alpha = -1.175$ and the solution oscillates with increasing amplitude, ultimately diverging. This suggests that a collection of highly lossy DERs which have fast charging rates (e.g. poorly insulated houses with large AC units) and aggressive bid functions can negatively impact system stability.

Thus, (4.11) provides valuable insights into the response of a collection of homogeneous devices under transactive coordination. However, it assumes that u_k is a continuous variable, $0 \leq u_k \leq 1$, whereas the on/off decisions for individual DERs are discrete. Moreover, the effects of lockouts and feeder limits cannot be captured using (4.11).

4.5. Aggregate Model under Market-based Coordination of DERs

It has been shown in [11, 51, 70, 79, 97, 104] that the natural dynamics of TCLs can be expressed in a bin model structure where each bin represents a temperature range and an on/off state. The evolution of the probabilities of TCLs lying in each bin can be described by,

$$X_{k+1} = AX_k, \tag{4.14}$$

where A is the transpose of the Markov transition matrix that is constructed from the probabilities associated with transitions between bins. It has been shown that the Markov transition matrix can be obtained analytically from the difference equations governing the TCL thermal dynamics, as showin in [70, 84] and in our earlier work [79, 81], or by applying system identification techniques [51, 70].

Existing bin-based aggregate models, however, do not capture the influence of transactive coordination on the behavior of DERs. There has also been limited attention given to

synchronization issues. Hence, an aggregate model that incorporates price will now be developed. Lockout constraints and other mechanisms that eliminate synchronization will also be incorporated. This bin model is then used within an MPC framework in Section 4.6.

4.5.1. Population Model

The aggregate bin model expressing the dynamics of DERs (e.g. TCLs, storage devices, electric vehicles) under transactive coordination depends both on SOC dynamic equations and user bid curves. For a homogeneous population, the storage dynamic equation (2.22) becomes,

$$e_{j,k+1} = ae_{j,k} + \gamma u_{j,k}, \tag{4.15}$$

and the relationship (4.5) between bid price and SOC is,

$$\pi_{j,k} = \pi^{\max} - \beta e_{j,k}, \tag{4.16}$$

where the maximum bid is $\pi^{\max}$ when $e_{j,k} = 0$ and the minimum bid is $\pi^{\min} = \pi^{\max} - \beta$ when $e_{j,k} = 1$. It is assumed that when $e_{j,k}$ reaches 1, the DER is automatically switched off.

Assume market clearing occurs every τ minutes, during which it is reasonable to assume there would be discrete shifts in the SOC levels and the price bids. Consider N_{B} bins (i.e. discrete price intervals) between $\pi^{\max}$ and $\pi^{\min}$, with each bin's width being $(\pi^{\max} - \pi^{\min})/N_{\mathrm{B}}$. The bins have indices $i = 1, ..., N_{\mathrm{B}}$. The bins are organized by decreasing price levels, with $\tilde{\pi}_0 = \pi^{\max}$ and $\tilde{\pi}_{N_{\mathrm{B}}} = \pi^{\min}$. The i-th bin's boundaries are $\{\tilde{\pi}_{i-1}, \tilde{\pi}_i\}$. For a homogeneous model, a direct mapping to the discrete SOC levels can be obtained. Let $\tilde{e}_i$, $i = 0, 1, ..., N_{\mathrm{B}}$, be the discrete SOC levels associated with $\tilde{\pi}_i$, so that $\tilde{e}_0 = e^{\min} = 0$ and $\tilde{e}_{N_{\mathrm{B}}} = e^{\max} = 1$.

Note that each bin also needs to consider the on/off and locked status outlined in Table 4.1. Hence, a three-stage bin model is proposed, as illustrated in Fig. 4.10, where bins are mapped to the operating states by defining the sets,

$$I_1 = \{1, ..., N_{\mathrm{B}}\} \qquad \text{for } m = 1, v = 1 \tag{4.17a}$$

$$I_2 = \{N_{\mathrm{B}} + 1, ..., 2N_{\mathrm{B}}\} \quad \text{for } m = 1, v = 0 \tag{4.17b}$$

$$I_3 = \{2N_{\mathrm{B}} + 1, ..., 3N_{\mathrm{B}}\} \text{ for } m = 0, v = 0. \tag{4.17c}$$

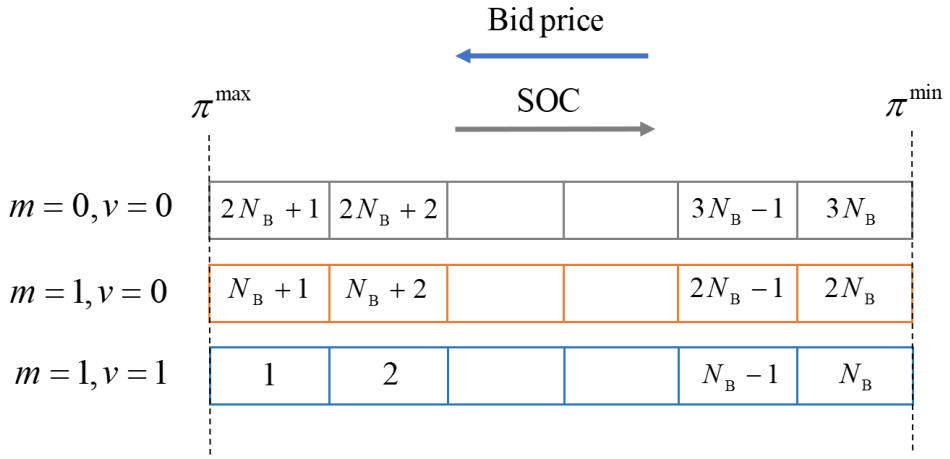

Figure 4.10: Bin model for a homogeneous population. Bid prices decrease from left to right whereas SOC levels increase. The three different operating states are marked on the left.

Let $x_{i,k} \geq 0$ be the fraction of DERs lying inside bin i at time k, where $i \in I_1$, I_2 or I_3, depending on the bid price (or the SOC level) and the operating state. Let $p_{j,i}$ be the probability of transitioning from bin j to bin i in one time step. Then the state transitions for the i-th bin can be written,

$$x_{i,k+1} = \sum_{j=1}^{3N_B} p_{j,i} x_{j,k}, \quad \forall i, k. \tag{4.18}$$

This can be expressed in matrix form (4.14) as,

$$A = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,3N_B} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,3N_B} \\ \vdots & \vdots & \ddots & \vdots \\ p_{3N_B,1} & p_{3N_B,2} & \cdots & p_{3N_B,3N_B} \end{bmatrix}^T. \tag{4.19}$$

Note that $0 \leq p_{j,i} \leq 1, \ \forall i, j$, and $\sum_{i=1}^{3N_B} p_{j,i} = 1, \ \forall j$.

Assume the clearing price is π_k^{clr}. Then, all TCLs in bins with bid prices greater than π_k^{clr} will be cleared. For a fixed value of π_k^{clr}, the transition probabilities $p_{j,i}$ can be estimated from a large number of samples by observing the evolution of DERs after the market clears at τ minutes.

The clearing price π_k^{clr} determines which bins are cleared. Hence, the $p_{j,i}$ are functions of π_k^{clr} and the A-matrix is time-varying for varying π_k^{clr}. The aggregate model can be expressed as,

$$X_{k+1} = A_k X_k. \tag{4.20}$$

If π_k^{clr} remains constant then $A_k = A, \forall k$. Section 4.7.2 illustrates the construction of the A-matrices, the connection between system behavior and the eigenvalues of A_k, and model performance.

4.5.2. Bin Model under Parameter Heterogeneity

When parameters are heterogeneous, bins are defined similarly to Section 4.5.1, except $\tilde{\pi}_0 = \max \pi_j^{\max}$ and $\tilde{\pi}_{N_{\text{B}}} = \min \pi_j^{\min}$. Compared to the homogeneous case, however, the discretization in SOC levels cannot be mapped directly to the bid price levels, unless additional bins are added to track both SOC levels and bid prices. The A-matrix elements can be obtained from a large number of samples. Alternatively, the dynamics of the heterogeneous population can be captured by considering clusters of homogeneous groups and their respective transition equations [1, 81].

4.5.3. Reformulation of the Transition Equations

While (4.20) can be used to simulate the aggregate behavior of DERs, the price signal must be known *a priori*. To determine the optimal price signal and resultant cleared demand in a multi-period optimization problem, the influence of price variation must be separated from the propagation dynamics (4.14). This can be acomplished by decomposing the state associated with each bin into ON and OFF fractions,

$$x_{i,k} = x_{i,k}^{\text{on}} + x_{i,k}^{\text{off}}, \quad x_{i,k}^{\text{on}}, x_{i,k}^{\text{off}} \geq 0, \quad \forall i \in I_1 \cup I_2, \forall k, \tag{4.21}$$

where $x_{i,k}^{\text{on}}$ and $x_{i,k}^{\text{off}}$ are decision variables determined by the optimization process. More specifically, the division between ON and OFF fractions of a bin is determined by a comparison between the bid price associated with that bin and the clearing price π^{clr}.

Given values for the decision variables $x^{\text{on}}_{i,k}$ and $x^{\text{off}}_{i,k}$, the states are reset according to,

$$x^{+}_{i,k} = x^{\text{on}}_{i,k} + x^{\text{on}}_{i+N_B,k}, \quad \forall i \in I_1, \forall k, \tag{4.22a}$$

$$x^{+}_{i,k} = x^{\text{off}}_{i,k} + x^{\text{off}}_{i-N_B,k}, \quad \forall i \in I_2, \forall k, \tag{4.22b}$$

where the '+' superscript indicates reset values. Note also that for the locked bins, $i \in I_3$, no DERs can be turned on. Hence,

$$x^{+}_{i,k} = x_{i,k} = x^{\text{off}}_{i,k}, \quad \forall i \in I_3, \ \forall k. \tag{4.23}$$

These reset equations can be expressed in matrix form,

$$X^{+}_{k} = B^{\text{on}} X^{\text{on}}_{k} + B^{\text{off}} X^{\text{off}}_{k}. \tag{4.24}$$

Starting from the reset state values, the evolution of the bin probabilities has a similar form to (4.14),

$$X_{k+1} = A X^{+}_{k}, \tag{4.25}$$

where A now captures the natural dynamics of DERs, as expressed by (4.15) and (4.16).

Finally, it should be noted that the structure of A ensures conservation of probability,

$$\sum_{i=1}^{3N_{\text{B}}} x_{i,k} = 1, \quad \forall k, \tag{4.26}$$

and $x_{i,k} \geq 0, \forall k$. Note that with this reformulation all equations remain linear. This is especially helpful for MPC design, as shown in Section 4.6.

An example with $N_{\text{B}} = 20$ (total 60 bins) is shown in Fig. 4.11. Bins lie in three sets $I_1 = \{1, ..., 20\}$, $I_2 = \{21, ..., 40\}$, and $I_3 = \{41, ..., 60\}$. The overall distribution, X_k at time k, is shown in Fig. 4.11(a). Recall from the bin definitions and from Fig. 4.10 that bins from left to right have decreasing bid prices. Also, while I_1, I_2, I_3 differ in their operating states, the bid prices at bins i, $i + N_{\text{B}}$ and $i + 2N_{\text{B}}$, where $i = 1, ..., 20$, are the same. Assume, $\pi^{\text{clr}}_{k} = \tilde{\pi}_7$. Thus, all DERs in bins 6, 7, 26 and 27 are cleared. The ON and OFF distributions, $X^{\text{on}}_{k}, X^{\text{off}}_{k}$, are shown in Figs. 4.11(b) and 4.11(c). By (4.24),(4.25), B^{on} and B^{off} will then act on X^{on}_{k} and X^{off}_{k}, respectively, to give X_{k+1}.

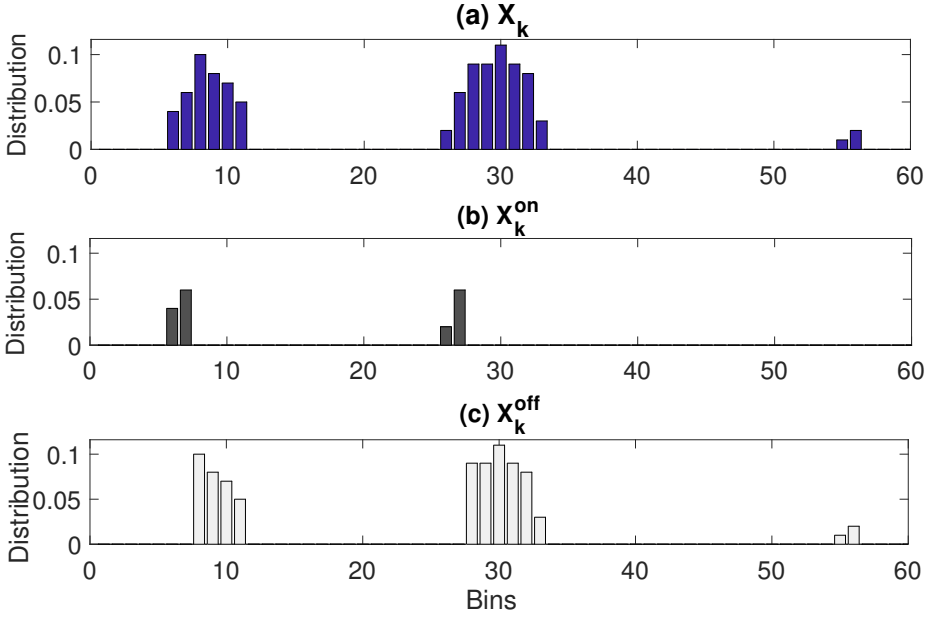

Figure 4.11: ON and OFF distributions over the bin space.

4.5.4. Logic Equations for Market Clearing

An arbitrary optimizer could choose the ON/OFF quantities from each bin as long as (4.21)-(4.26) were satisfied. However, in a transactive dispatch mechanism, DERs with higher bid prices than the clearing price are cleared. Hence, additional logic is required to simulate behavior under the transactive market clearing mechanism. For example, with the bin definitions of Fig. 4.10, bins from right to left have increasing bid prices, and so should have higher priority to turn on. To accomplish this, binary variables can be used.

Let $u^{\text{on}}_{i,k} \in \{0, 1\}$ be the binary on/off decision associated with clearing all TCLs in bin i, and consider the following set of equations:

$$0 \leq x^{\text{on}}_{i,k} \leq u^{\text{on}}_{i,k}, \quad i = 1, ..., 2N_{\text{B}} \tag{4.27a}$$

$$0 \leq x^{\text{off}}_{i,k} \leq u^{\text{off}}_{i,k}, \quad i = 1, ..., 3N_{\text{B}} \tag{4.27b}$$

$$u^{\text{on}}_{i,k} + u^{\text{off}}_{i,k} = 1, \quad i = 1, ..., 3N_{\text{B}} \tag{4.27c}$$

$$u^{\text{on}}_{i,k} \geq u^{\text{on}}_{i+1,k}, \quad i = 1, ..., N_{\text{B}} - 1 \tag{4.27d}$$

$$u^{\text{on}}_{i,k} = u^{\text{on}}_{i+N_{\text{B}},k}, \quad i = 1, ..., N_{\text{B}}. \tag{4.27e}$$

If $x^{\text{on}}_{i,k}$ DERs from bin i are to be chosen to be ON, we need $u^{\text{on}}_{i,k} = 1$, otherwise $u^{\text{on}}_{i,k} = 0$.

This is accomplished by (4.27a). Likewise, by (4.27b), $u^{\text{off}}_{i,k} \in \{0,1\}$ is used for choosing $x^{\text{off}}_{i,k}$. Ensuring that each bin has only ON or OFF DERs is achieved by (4.27c). (We assume that when a bin is cleared, all DERs in that bin turn on.) Next, (4.27d) ensures that bins with higher bid prices must be turned ON before bins with lower prices can be chosen. Finally, (4.27e) ensures that bins with the same bid prices, but different operating states I_1 and I_2 get cleared simultaneously. Note that the logic equations (4.27a)-(4.27e) are linear in $x^{\text{on}}_{i,k}$, $x^{\text{off}}_{i,k}$, $u^{\text{on}}_{i,k}$, and $u^{\text{off}}_{i,k}$.

4.6. Model Predictive Control Formulation

The bin-based aggregate model of DERs can be incorporated into a model predictive control framework to determine the DER schedule that gives minimum power supply cost over a finite horizon. Let the distribution network be supplied by power sources P^s_k, $s = 1, ..., N_{\text{s}}$ at time k. Each source has cost $C_s(P^s_k)$ which is typically quadratic. The overall demand of the network at time k is D_k, which is composed of controllable DER demand D^c_k and uncontrollable demand D^o_k of other loads. The cost minimization problem can be formulated as,

$$\min \quad \sum_{k=1}^{N_{\text{k}}} \sum_{s=1}^{N_{\text{s}}} C_s(P^s_k) \tag{4.28a}$$

$$\text{s.t.} \quad \sum_{s=1}^{N_{\text{s}}} P^s_k = D_k, \qquad \forall k \tag{4.28b}$$

$$D_k = D^c_k + D^o_k, \qquad \forall k, \tag{4.28c}$$

$$D^c_k = \sum_{i=1}^{3N_{\text{B}}} x^{\text{on}}_{i,k}, \qquad \forall k, \tag{4.28d}$$

$$x_{i,k=1} = x^{\text{ini}}_i, \qquad \forall i, \tag{4.28e}$$

$$D_k \leq d^{\text{Feeder}} \qquad \forall k, \tag{4.28f}$$

along with the transition equations (4.21)-(4.26) and logic constraints (4.27a)-(4.27e).

The objective function (4.28a) minimizes the total cost of supply. Supply-demand balance is enforced by (4.28b). The Lagrange multiplier associated with this constraint represents the electricity price, λ^{elec}_k. Total demand is established by (4.28c), with (4.28d) relating aggregate controllable load to the ON fractions of all bins. Initial conditions are

established by (4.28e), where X^{ini} is assumed to be known. Finally, (4.28f) ensures the feeder limit is not violated. Individual capacity and ramp limits for each supplier may also be incorporated.

Additionally, to avoid arbitrarily setting $u^{\text{on}}_{i,k} = 1$, a cost $\tilde{C}_i(u^{\text{on}}_{i,k})$ is included in the objective,

$$\sum_{k=1}^{N_{\text{k}}} \sum_{s=1}^{N_{\text{s}}} C_s(P^{\text{s}}_k) + \mu_{\text{w}} \sum_{k=1}^{N_{\text{k}}} \sum_{i=1}^{3N_{\text{B}}} \tilde{C}_i(u^{\text{on}}_{i,k}), \tag{4.29}$$

where μ_{w} is a tuning parameter. Note, $\tilde{C}_i(u^{\text{on}}_{i,k}) = 1, \forall i, k$ has been used in this chapter.

Due to the presence of both continuous and binary variables, the overall formulation is a mixed integer programming (MIP) problem. A mixed integer linear program (MILP) can be obtained by approximating the suppliers' quadratic cost functions with piecewise linear segments.

4.6.1. Measures to Avoid DER Synchronization

Synchronization of DERs occurs when a large fraction of DERs lie within a narrow range of bins. Since synchronized DERs typically respond to control update signals in a similar manner, in the presence of synchronized DERs, the control schemes under consideration may cause large oscillations in the aggregate demand and high volatility in the clearing prices [80]. Hence, to limit DER synchronization, additional measures should be taken.

Let $b^{\max}$ be the maximum allowable fraction in any bin. Then, the following constraint can be enforced,

$$x_{i,k} \leq b^{\max}, \quad \forall i, k. \tag{4.30}$$

Alternatively, an additional cost term can be added in the objective function as,

$$\sum_{k=1}^{N_{\text{k}}} \sum_{s=1}^{N_{\text{s}}} C_s(P^{\text{s}}_k) + \mu_{\text{w}} \sum_{k=1}^{N_{\text{k}}} \sum_{i=1}^{3N_{\text{B}}} \tilde{C}_i(u^{\text{on}}_{i,k}) + \mu_{\text{s}} \sum_{i=1}^{3N_{\text{B}}} (x_{i,k} - b^{\text{avg}})^2, \tag{4.31}$$

where the last term represents the cost of synchronization, and, for computational convenience, has been expressed here by a quadratic function. b^{avg} can be set equal to $1/(3N_{\text{B}})$ and μ_{s} can be adjusted to distribute DERs widely over the bin space. The value of μ_{s} will guide how effectively DER synchronization can be avoided, and may cause the solution of the optimization problem to be less economic than obtained by considering (4.28a) only.

4.6.2. QP Formulation

Instead of using binary variables in (4.27a)-(4.27e) and (4.29), a relaxed QP formulation can be established by assigning costs per bin, $\widehat{C}_i(x^{\text{on}}_{i,k})$, that increase with increasing bin index. For example, $\widehat{C}_i(x^{\text{on}}_{i,k}) = w_o^i x^{\text{on}}_{i,k}, \forall i \in I_1$ and $\widehat{C}_i(x^{\text{on}}_{i,k}) = w_o^{i-N_{\text{B}}} x^{\text{on}}_{i,k}, \forall i \in I_2$, where $w_o > 1$. (The simulations in Section 4.7 use $w_o = 3$.) The modified objective function becomes,

$$\sum_{k=1}^{N_{\text{k}}} \sum_{s=1}^{N_{\text{s}}} C_s(P^{\text{s}}_k) + \mu_{\text{w}} \sum_{k=1}^{N_{\text{k}}} \sum_{i=1}^{3N_{\text{B}}} \widehat{C}_i(x^{\text{on}}_{i,k}) + \mu_{\text{s}} \sum_{i=1}^{3N_{\text{B}}} (x_{i,k} - b^{\text{avg}})^2, \tag{4.32}$$

where μ_{w} is a tuning parameter.

The overall formulation consists of the objective (4.32), constraints (4.28b)-(4.28f) and transition equations (4.21)-(4.26). All constraints become linear and the objective function remains quadratic. Overall, we obtain an efficient QP form.

4.6.3. Transactive Dispatch Rule

Once the above problem has been solved, the indices of the ON bins can be recovered from $x^{\text{on}}_{i,k}$ or $u^{\text{on}}_{i,k}$. Define $i^{\max}_k$ as the largest index among the ON bins at period k (for $i \in I_2$ subtract N_{B} from bin indices). For the MIP solution,

$$i^{\max}_k = \max_i \{u^{\text{on}}_{i,k} = 1\}, \quad \forall k. \tag{4.33}$$

Since u^{on} does not appear in the QP formulation, in that case a bin is considered to be cleared when $x^{\text{on}}_{i,k}$ is larger than a threshold ζ,

$$i^{\max}_k = \max_i \{x^{\text{on}}_{i,k} \geq \zeta\}, \quad \forall k. \tag{4.34}$$

From the SOC-bin mapping of Fig. 4.10, the price associated with $i^{\max}_k$ becomes the clearing price π^{clr}_k, that is broadcast to all the DERs for time period k. Upon receiving this price, all DERs in bins $i \leq i^{\max}_k$, (i.e. bins with higher bid prices) should dispatch,

$$u_{j,k} = \begin{cases} 1, & \text{if} \quad \pi_{j,k} \geq \tilde{\pi}_{i^{\max}_k}, \\ 0, & \text{if} \quad \pi_{j,k} < \tilde{\pi}_{i^{\max}_k}. \end{cases} \tag{4.35}$$

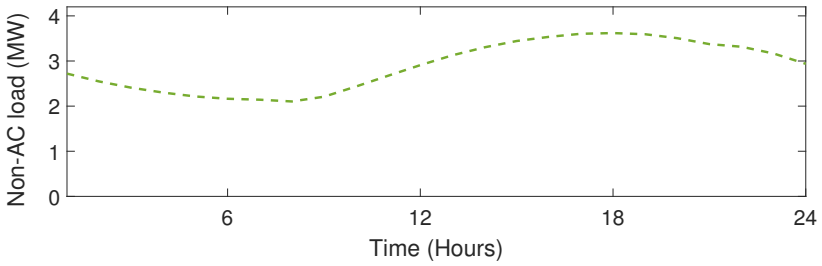

Figure 4.12: Aggregate demand profile excluding air-conditioner (AC) demand.

4.6.4. Practical Considerations

Assume all DERs are managed by a single aggregator in a distribution system. To obtain X^{ini}, which is required for (4.28e), an aggregator could collect local measurements from DERs typically near the end of an MPC horizon. Each DER's on/off, locked status and its associated bid price are used to construct X^{ini}. Since only aggregate information is needed, each DER sends updates anonymously. Assuming a distribution system operator (DSO) solves the MPC, the aggregator would send X^{ini} to the DSO. The DSO also has access to the A-matrix of (4.25), which the aggregator estimates separately. The DSO solves the MPC and sends the aggregator the clearing prics $\pi_k^{\text{clr}}, k = 1, ..., N_{\text{k}}$. The price π_k^{clr} can be revealed to DERs either just for the next market clearing interval or for several periods ahead, depending on communication bandwidth availability. Since MPC already accounts for the feeder limit, the two-step market clearing process described in Section 4.2.2 is not necessary. However, if the mechanism in Section 4.2.2 is followed, individual DER bids additionally need to be collected at every market clearing interval.

4.7. Simulation Results

4.7.1. Data

The case study considers a typical distribution system serving predominantly residential loads (1,473 customers) located in Austin, TX [26]. The system peak demand was recorded at 7.77 MW. Energy usage analysis of real data from 88 single-family houses in the Mueller neighborhood of Austin, from July 2012 to June 2013, was undertaken in [85]. It was found that approximately 47% of the peak household demand was consumed by air-conditioner

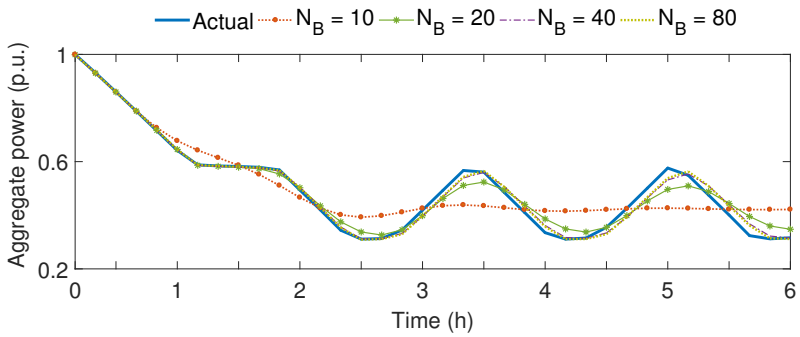

Figure 4.13: Aggregate demand profiles, for varying N_B, with clearing price at 10 \$/MWh (TCL initial temperatures uniform within 19 to 21°C).

(AC) units during summer peak days. Based on this analysis and hourly demand data from the Electric Reliability Council Of Texas (ERCOT) [29], the non-AC demand profile of Fig. 4.12 was estimated.

Following the calculation procedure in [17], for a 176 m^2 house with 3 ton (approximately 10.55 kW) AC [85] and $\eta_j = 3.5$, the TCL parameters are approximated as, $P_j^{\text{elec}} = 3$ kW when on, $R_j = 2.84$°C/kW, and $C_j = 7.04$ kWh/°C. With each AC consuming approximately 3 kW on average, the maximum instantaneous AC load can be up to 4.4 MW, which would result in a peak demand exceeding 8 MW.

4.7.2. Aggregate Model Performance

To study the performance of the aggregate model (4.20), first the A-matrix coefficients need to be obtained. Consider a homogeneous population with $\beta_j = 40$ \$/MWh, $\pi_j^{\max} =$ \$50, $\forall j$. With $\tau = 10$ min and known price signals, 1000 TCLs were simulated. For TCLs originating in a specific sending bin, we can find the range of bins reached by TCLs at the end of 10 minutes. Repeating for all bins and normalizing these quantities, the transition probabilities (thus, the A-matrix) for a known price signal were obtained.

First, consider $\pi^{\text{clr}} = 10$ \$/MWh. All TCLs have bids just above the clearing price, hence can get cleared. The aggregate power consumed by 1000 TCLs over 6 hours has been shown in Fig. 4.13. Note that the initial temperatures of TCLs were distributed uniformly between 19 to 21°C. With $\tau = 10$ minutes, the aggregate behavior was also simulated using bin models of various orders, N_B =10, 20, 40, 80. With $N_B = 10$ the aggregate demand profile deviates significantly from the actual. The profiles obtained with the other models

matched the actual reasonably well. Fig. 4.14 shows the performance for varying N_{B} when $\pi^{\mathrm{clr}} = 30$ \$/MWh. Again, the profile obtained using the $N_{\mathrm{B}} = 10$ model deviates from the actual.

Next, the A-matrix obtained for 10 \$/MWh was used to simulate the aggregate demand with initial temperatures of TCLs distributed uniformly between 19.8 to 20.2°C. Profiles are shown in Fig. 4.15. Since most TCLs turned on/off almost at the same time, the oscillation amplitudes were larger compared to those in Fig. 4.13.

With the price signal fixed at $\pi^{\mathrm{clr}} = 30$ \$/MWh, the aggregate demand reaches a constant level, whereas with $\pi^{\mathrm{clr}} = 10$ \$/MWh, oscillatory behavior has been observed. This is because with a low clearing price, TCLs enter the locked mode when their temperatures reach 19°C and again become controllable when temperatures exceed 19.6°C. Analyzing the eigenvalues of the A-matrices, we observed that for the A-matrix with $\pi^{\mathrm{clr}} = 30$ \$/MWh, the eigenvalues have only real parts. For the A-matrix with $\pi^{\mathrm{clr}} = 10$ \$/MWh, pairs of complex eigenvalues exist, suggesting an oscillatory response. Ideally under $\pi^{\mathrm{clr}} = 10$, a homogeneous population exhibits undamped oscillations, whereas the eigenvalues of the A-matrix suggest damped oscillations. This discrepancy exists due to modeling error from the discretization [12, 51, 81]. However, as shown in Figs. 4.13 and 4.15, with sufficiently large N_{B} the actual behavior can be tracked closely for several hours, which is suitable for control purposes [12, 81].

The effect of changing the market clearing interval τ is analyzed next. Consider $\tau = 1$, 10, 30 and 60 minutes. The A matrices, with $N_{\mathrm{B}} = 40$, were obtained for each case under $\pi^{\mathrm{clr}} = 10$ \$/MWh. As shown in Fig. 4.16, with $\tau = 10$ min, the profile obtained by (4.20)

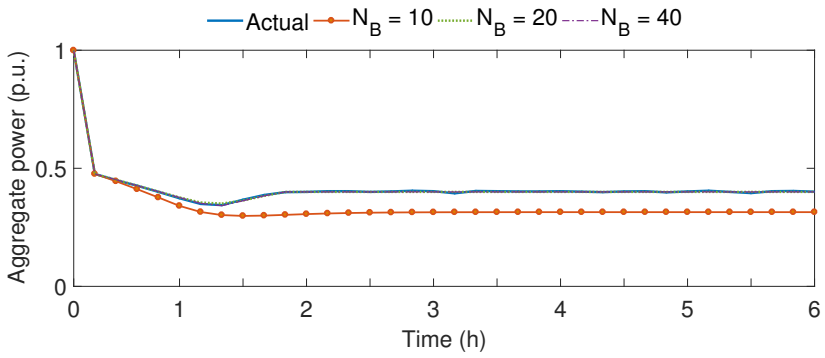

Figure 4.14: Aggregate demand profiles, for varying N_{B}, with clearing price at 30 \$/MWh (TCL initial temperatures uniform within 19 to 21°C).

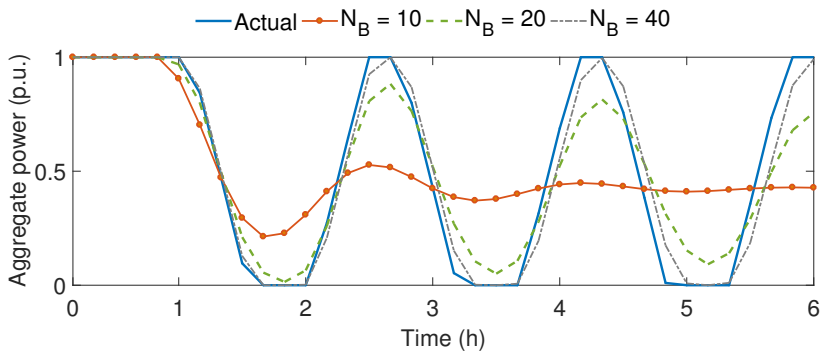

Figure 4.15: Aggregate demand profiles, for varying N_B, with clearing price at 10 \$/MWh (TCL initial temperatures uniform within 19.8 to 20.2°C).

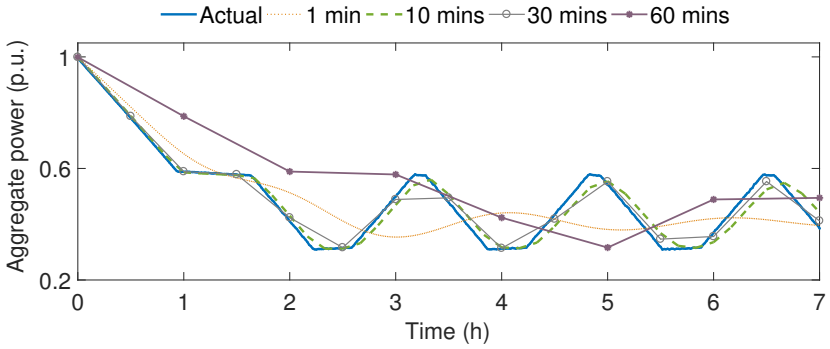

Figure 4.16: Aggregate demand profiles, for varying market clearing intervals, with $\pi^{clr} = 10$ \$/MWh (Initial temperatures uniform within 19-21°C).

deviates from the actual. This is because with a smaller τ, the small changes in temperature (or bids) can only be accurately captured by using large N_B [11, 70, 81], and using $N_B = 40$ is not sufficient. The profile obtained with $\tau = 10$, 30 minutes match the actual behavior reasonably well. With $\tau = 60$ minutes, the profile again deviates significantly. During the duration of 60 minutes, many TCLs reach their temperature thresholds and change state. Hence, the 60 minute bin model could not capture the intra-hour power consumption dynamics.

Table 4.2: Performance Comparison

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Set up							
MPC type	MIP	MIP	MIP	MIP	QP	QP	QP
Horizon, N_{k}	12	12	18	18	18	18	18
$b^{\max}$	1	0.25	1	1	0.25	0.15	0.25
Tuning parameter, μ_{s}	0	0	0	1000	1000	1000	1000
Tuning parameter, μ_{w}	1	1	1	1	200	200	200
Results							
Average system demand, $\overline{D}_k$, in MW	4.82	7.10	4.74	6.08	4.68	4.74	4.68
Average TCL demand, $\overline{D}^{\mathrm{c}}_k$, in MW	1.30	3.87	1.28	2.62	1.21	1.28	1.21
Peak system demand, $\widehat{D}_k$, in MW	5.20	8.00	5.19	8.00	5.20	6.78	5.21
RMSE (normalized), in %	0.82	1.19	0.79	0.72	3.22	7.88	3.16
$\bar{\lambda}^{\mathrm{elec}}_k$ (average electricity price), in \$/MWh	34.1	47	33.7	40.4	33.4	33.7	33.4
$\lambda^{\mathrm{elec-}}_k$, $\lambda^{\mathrm{elec+}}_k$ (min, max), in \$/MWh	30.2, 36.5	38.5, 50	30, 35.9	26.7, 50.2	30.7, 34.8	29.4, 35.8	30.8, 34.6
Maximum TCLs in a single bin, in %	0.41	0.25	0.39	0.16	0.25	0.15	0.25
Bin spread at $X_{N_{\mathrm{k}}}$	5	13	5	12	12	11	12

4.7.3. MPC Performance

Consider $\tau = 10$ min, and $N_{\mathrm{B}} = 20$. Cost of supply, $C^s_k(P^{\mathrm{s}}_k) = 10P^{\mathrm{s}}_k + 2.5(P^{\mathrm{s}}_k)^2$. The substation feeder limit is set at 8 MW.

For a storage based system, deciding only based on current or near-term situations may lead to significant reduction of the feasible operating region in future periods [79, 84]. Hence, we look ahead several periods in the MPC. Consider $N_{\mathrm{k}} = 18$. With $\tau = 10$ min, the MPC looks ahead a 3 hour window (2 hours in cases 1 and 2). Additionally, let $\sum_{k=1}^{N_{\mathrm{k}}} D^c_k \geq 1.21 N_{\mathrm{k}}$, where 1.21 MW is an average aggregate TCL demand allowed to avoid the depletion of the aggregate SOC.

Recall that MPC requires the A-matrix. To obtain the coefficients of A, distribute TCLs uniformly over all bins. For TCLs originating from each sending bin, find the range of bins covered at the end of $\tau = 10$ minutes. Normalizing these quantities gives the transition probabilities. Recall that this matrix is not a function of the clearing price, hence is valid under any clearing price signal.

Both the QP and MIP problems were programmed in MATLAB and YALMIP [65]. QP problems (cases 5-8) were solved using Quadprog, whereas the MIPs (Cases 1-4) were

solved using Gurobi. For validation, in each case we ran the full dispatch of 1473 TCLs to test the performance of the MIP/QP solutions. Due to space limitations, only results for tests starting from hour 18 have been presented.

Test parameters and results are summarized in Table 4.2. The average system demand, $\overline{D}_k$, average TCL demand, $\overline{D}_k^{\mathrm{c}}$, and system peak demand $\widehat{D}_k$ are shown. Also, the average, minimum and maximum price of electricity, $\bar{\lambda}_k^{\mathrm{elec}}$, $\lambda_k^{\mathrm{elec-}}$, and $\lambda_k^{\mathrm{elec+}}$, were recorded. The error between the scheduled and the actual system demand during the dispatch process is captured by the root mean square error (RMSE) (here, normalized by the system peak capacity of 8 MW). Initially, temperatures of TCLs were distributed uniformly within 20 to 21°C. To measure synchronism, maximum TCLs in a single bin (%) in $X_k, \forall k$, and the bin spread (i.e. the number of bins with non-zero quantities of TCLs) in $X_{N_{\mathrm{k}}}$ were recorded. The following observations are made,

- In cases 1 and 3, $b^{\max} \leq 1$ (i.e. all TCLs can be in a single bin) and $\mu_{\mathrm{s}} = 0$, hence the risk of synchronization was not accounted for. As a result, $X_{N_{\mathrm{k}}}$ was narrowly distributed over 5 bins only. Also, the maximum fraction of TCLs in a single bin reached approximately 40%.
- Measures to avoid synchronization were taken by setting $b^{\max} \leq 0.25$ in case 2, and by $\mu_{\mathrm{s}} = 1000$ in case 4. Maximum TCLs lying in a single bin decreased considerably and wider $X_{N_{\mathrm{k}}}$ were obtained. However, $\overline{D}_k^{c}$, $\widehat{D}_k$ and $\bar{\lambda}_k^{\mathrm{elec}}$ were significantly higher in cases 2 and 4.
- In case 5, applying the QP solution, the obtained $\overline{D}_k^{\mathrm{c}}$, $\widehat{D}_k$ and $\bar{\lambda}_k^{\mathrm{elec}}$ were similar to the values in case 3. The demand profiles (and $\lambda_k^{\mathrm{elec}}$) obtained in cases 3 and 5 are shown in Fig. 4.17. However, the RMS error was higher in case 5. This is because the QP does not need to enforce strict on/off decisions for the entire bin, whereas the MIP does. Under QP, contents in a bin can be fractionally chosen to be on. However, during dispatch, all TCLs in that bin are cleared due to receiving the same price signal. This causes the actual profile to slightly deviate from the predicted one.
- In case 6, when a stricter $b^{\max}$ limit was imposed than in case 5, $\overline{D}_k^{\mathrm{c}}$, and $\bar{\lambda}_k^{\mathrm{elec}}$ slightly increased. RMSEs and $\widehat{D}_k$, however, increased noticeably. Again, this was mostly due to the QP solutions favoring fractional ON quantities in a bin in order to meet lower $b^{\max}$, which eventually led to higher error during the dispatch.
- In case 7, noise (uniform $[-0.02, 0.02]$, in °C/min) was introduced in the model, affecting

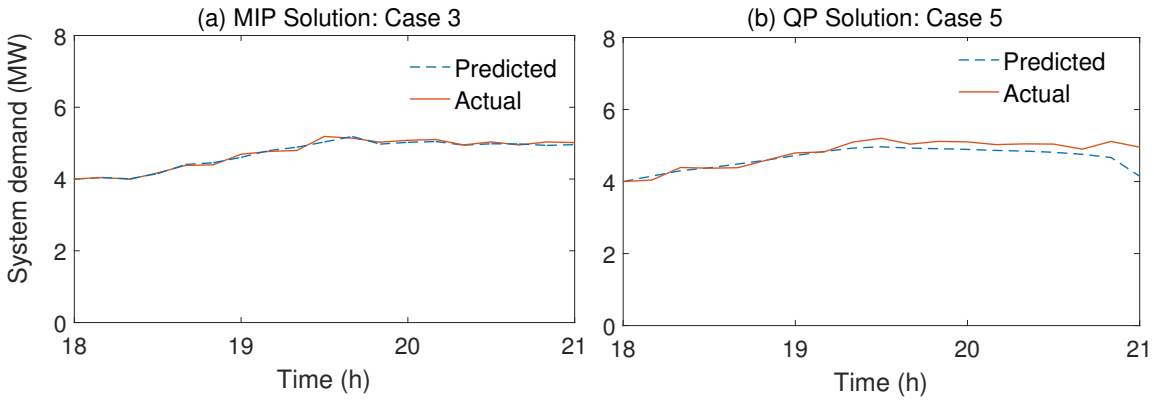

Figure 4.17: Predicted and actual system demand profiles using (a) MIP in case 3, and (b) QP in case 5.

the temperature dynamics of the TCLs. The bin model was identified under noisy data. During the dispatch process, the RMSE did not increase, rather results were comparable to case 5.

- In case 8, bid slopes were heterogeneous (uniform $[36, 44]$, in \$/MWh). While the cost of the solution (i.e. $\bar{\lambda}_k^{\text{elec}}$) remained nearly the same, the RMSE increased slightly. This can be attributed to the bin model's reduced accuracy to deal with parameter heterogeneity. Model performance could worsen further when considering heterogeneity in other TCL parameters, as discussed in our work [81]. This could be better modeled by using clusters of homogeneous groups [81], [1].

Additionally, it was noted that the average time taken to solve the MPCs were 95 s for MIPs with $N_{\text{k}} = 18$, 23 s for MIPs with $N_{\text{k}} = 12$, and less than 2 s for all QP problems.

4.7.4. Discussions

Comparing the MIP and QP solutions in Table 4.2, in general, we observe that the average TCL demand, peak demand and average electricity prices with QP were lower compared to the results obtained by the MIP. However, the QP solutions typically lead to higher error (i.e. RMSEs values). To reduce the error during the dispatch process, alternative dispatch schemes could be sought in future work. Discriminatory prices or incentive signals could be sent to the TCLs. For example, each TCL could receive a clearing price that is slightly perturbed by noise. This could potentially reduce the RMSE, however, detailed

investigation should be carried out to analyze the effectiveness and the fairness of such schemes.

To reduce the likelihood of an oscillatory response and to lower error during the dispatch, we included $\mu_s > 0$ in the objective or imposed b^{max}. Due to the large degree of freedom when using the QP formulation, the above strategies typically did not increase the cost of the solution (except in case 6). The more restrictive MIP form typically led to higher cost solutions, especially for high μ_s or low b^{max}.

Our framework was shown to effectively relieve congestion at a substation feeder by looking ahead several hours. While for slowly varying systems, such as EVs and commercial building HVAC systems, hourly time steps have been used in several recent studies [8, 36, 37, 59], 10 minute price signals have been used to capture the dynamics of TCLs in our case. Future work could apply our proposed framework in a rolling horizon setting and for other power system applications. One natural extension could be to deal with network congestion. Recent work investigates distribution locational marginal prices (LMPs) to alleviate network congestion while considering dynamics of EVs [8, 59] and HVACs [36, 37]. Compared to these approaches, our method has the advantage that it lets users choose their individual bid functions and the aggregator makes decisions utilizing the bin-based aggregate model. One could consider aggregating loads at different nodes of the network, which will then allow computing the nodal LMPs and the optimal incentive signals for each aggregation.

4.8. Chapter Conclusions

Recent studies have shown that transactive coordination of DERs, such as TCLs, batteries, and EVs, can lead to undesirable power oscillations. In this chapter, we analyzed the causes of such oscillations and identified different factors that can affect the aggregate DER dynamics. The user defined bid slopes, preference for setting locking conditions, price signals sent to DERs for coordinating their responses, and imposing feeder limits, all can affect the natural charging and discharging cycles of DERs, hence can lead to load synchronization and undesired power oscillations. To address these issues, we developed a bin-based DER aggregate model under transactive coordination. The trade-offs when using varying model orders and market intervals were analyzed. With reformulation of the transition equations, we showed how the model can be incorporated in an MPC framework.

The MPC can then be solved to find optimal price signals that should be sent to the DERs for governing their aggregate responses in a desired manner. An accurate MIP formulation, and a relaxed QP formulation have been developed and tested. Simulation results compared the performance under several different scenarios by varying the initial conditions, penalty for synchronization, noise and heterogeneity. Future work could involve extending our approach to study hierarchical coordination among system operators, DSOs and aggregators. Schemes that can consider non-uniform incentive signals could also be investigated. Chapter 7 continues further investigation into the various aspects presented in this chapter using a unifying hybrid systems approach.

Chapter 5.

Hybrid Systems Approach to Modeling and Analysis of TCL Coordination

5.1. Chapter Introduction

In this Chapter, we present a hybrid dynamical systems framework to model both the continuous and discrete dynamics of load ensembles. Analysis of eigenmodes of the underlying discrete-time system provide insights into situations under which control of load ensembles may lead to undesirable, highly nonlinear behavior such as synchronization, sustained oscillations and bifurcations. An analysis technique based on Lyapunov (energy) functions for switched systems has also been presented. Simulations are presented to support the theory and enhance our understanding regarding the range of possible behavior that the overall system may exhibit.

Various load coordination techniques have been proposed in the literature including randomized switching, temperature set-point variation, and Transactive energy coordination. However, such controls may sometimes lead to undesirable phenomenon such as synchronization of TCLs and large fluctuations in aggregate demand of loads [17, 76, 93, 94]. These oscillations may cause new peaks in system demand and result in voltage violations in distribution systems. For an example, consider a distribution system [9]. With synchronization of TCLs, large fluctuations occur in the aggregate demand of TCLs, hence in the overall distributions system's demand. These fluctuations then may cause frequent voltage violations, as shown in Fig. 5.1.

Temperature synchronization of TCLs a well-known problem, especially in the context of demand response (DR) events where applying a prolonged 'off' signal to TCLs causes

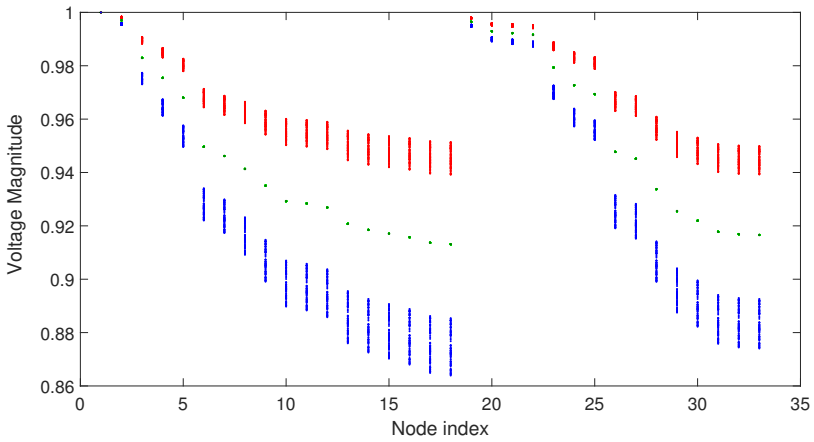

Figure 5.1: Voltage magnitudes in a 33 bus distribution feeder [9]. Voltage violation were caused by large oscillations from a synchronized TCL population.

loss of natural diversity in TCLs. Consequently at the end of the DR event, new peaks and oscillations are observed [66, 93]. However, synchronization of TCL temperatures is less studied and understood under advanced coordination mechanisms, such as market-based coordination, set-point variation and randomized switching, where control signals may be updated regularly with update intervals varying from few seconds to minutes ranges. A systematic analysis of such cases is therefore the main focus of our work.

Since the ensemble behavior of hysteresis-based loads is challenging to model, simulation-based studies are often undertaken to show the possibility of oscillations and characterize the damping due to noise and heterogeneity. In [18], oscillatory behavior was observed in the aggregate demand when simulating a large number of price-responsive electric vehicles. In [80, 94], simulations show that under market-based coordination of TCLs, a sequence of price signals can induce synchronization and large oscillations in the aggregate TCL demand.

Some recent work [24, 34, 71] provide analytical results characterizing the behavior of TCLs in the presence of noise and heterogeneity. However, the synchronizing and oscillatory behavior observed in [53, 80] cannot be explained fully without resorting to modeling both the continuous-time behavior of loads as well as the discrete events due to the control actions that occur at slower intervals. Hence, the reset-based hybrid systems model presented in this chapter can provide intuition into complex system behavior which is not available under simplifying modeling assumptions that are common in the existing

literature.

The chapter presents an analytical investigation of conditions under which temperature synchronization of TCLs may appear and large oscillations in the aggregate demand of a load ensemble may emerge, which reflects our work [77]. For understanding oscillatory behavior with TCLs, under a given control strategy, it is important to understand if the control has a tendency to synchronize the TCLs, and to what degree. We show that eigenmode analysis can (i) guarantee if synchronization will appear, (ii) the rate at which it would appear. To accomplish this, the dynamics of a controlled TCL population, under a given strategy, is expressed using a reset-based hybrid system. This allows us to study the problem as a parameter-dependent eigenvalue problem. Using the hybrid representation, the aggregate autonomous dynamics of TCLs is captured using a continuous-time model while the control updates (e.g. price or set-point updates) are assumed to occur at slower discrete intervals (minutes range). Then, the eigenvalues and steady-state distributions of a discretized system are analyzed to explain the effects of control input variations, check if synchronization is imminent, and find bounds on demand fluctuations. A set of indices to automatically measure synchronization has also been proposed. Then, for both priority- or market-based control and randomized switching, we show how control update intervals and levels of control variations can influence periodic behavior, synchronization, and/or damping of oscillations. For a given initial condition, modal analysis can also lead to reduced computational needs.

The remainder of the chapter is organized as follows. Section 5.2 describes the modeling and control framework. Section 5.3 presents modal analysis to characterize the systems. Section 5.4 proposed a set of indices to automatically measure synchronization. Section 5.5 derives a Lyapunov function to analyze stability. Section 5.6 provides numerical examples. Finally, section 5.7 concludes by summarizing our findings and discussing their implications.

5.2. Hybrid System Representation

5.2.1. Controlled Dynamics as a Reset-based Hybrid System

In earlier work, state space models are used for TCL control design, where control is applied at every time-step. However, the control updates may typically occur at relatively slow discrete intervals, whereas the aggregate dynamics of TCLs can be accurately captured using

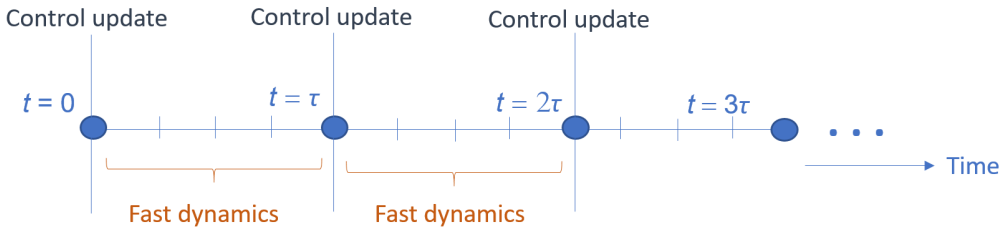

Figure 5.2: Hybrid systems overview.

continuous-time models. Hence, to capture the time-sclae separation, the overall continuous/discrete dynamics of TCLs coordinated by switching signals can be more generally expressed as:

$$\dot{x}(t) = \mathcal{A}x(t), \qquad \text{(continuous dynamics)} \tag{5.1a}$$

$$x(t) = B(t)x(t^{-}), t = 0, \tau, 2\tau, ... \text{ (discrete switchings)} \tag{5.1b}$$

$$y(t) = Cx(t), \qquad \text{(output)} \tag{5.1c}$$

In (5.1b), we assume that the discrete switchings occur only at $t = \tau, 2\tau, ...$, where τ is the switching interval, and the effect of control actions is captured by the reset matrix, $B(t)$. The entries of $B(t)$ are dependent on the chosen control strategy. Sections 5.2.2 and 5.2.3 show B-matrices for randomized switching and for market- or priority-based strategies.

While the aggregate dynamics of TCLs can be captured using continuous-time models, the control updates typically occur at relative slow discrete intervals (minutes range) (see Fig. 5.2.

Assuming $\mathcal{A}$ is time-invariant, and A is it's discrete-time equivalent, from one reset event to the next, the evolution in $x(t)$ at $t = 0, \tau, 2\tau, ...$ can be described using,

$$x(t + \tau) = B(t)A^{\tau}x(t), \; t = 0, \tau, 2\tau, ... \tag{5.2}$$

In discrete-time, (5.2) can also be written as,

$$x^{+}_{k+1} = B_k A^{\tau} \, x^{+}_{k}, \quad k = 0, 1, 2, ... \tag{5.3}$$

where k is the index of each discrete interval and x^{+} represents the TCL distribution immediately after reset.

Similarly, the evolution in x_k^-, the TCL distributions right before resets, can be described in the following manner,

$$x_{k+1}^- = A^\tau B_k \, x_k^-, \quad k = 0, 1, 2, ... \tag{5.4}$$

The hybrid model (5.1a)-(5.1c) is general enough to support a variety of control strategies, such as randomized switching, market-based coordination and set-point variation. Depending on the strategy, mainly the reset maps and the switching intervals would vary.

5.2.2. Randomized Switching

Sending probabilistic switching signals to increase/decrease power consumption has frequently been considered in TCL literature [51, 70, 79, 97]. Consider a proportional control logic, where all off/on bins receive a command signal to shift a fixed fraction of a bin's quantity to a corresponding on/off bin. For power increase, i.e. shifting a fixed fraction, f^+ from the off bins to on bins, the reset equations can be written as,

$$x_i^+ = (1 - f^+)x_i^-; \quad 1 \leq i \leq N_B; \tag{5.5a}$$

$$x_i^+ = x_i^- + f^+ x_{2N_B - i + 1}^-; \quad (N_B + 1) \leq i \leq 2N_B. \tag{5.5b}$$

For power decrease, i.e. shifting a fixed fraction, f^- from the on bins to off bins, the reset equations can be written as,

$$x_i^+ = x_i^- + f^- x_{2N_B - i + 1}^-; \quad 1 \leq i \leq N_B; \tag{5.6a}$$

$$x_i^+ = (1 - f^-)x_i^-; \quad (N_B + 1) \leq i \leq 2N_B. \tag{5.6b}$$

The reset matrices, B^+ associated f^+, and B^- associated f^-, can be readily obtained from the above. The column sums for B^+ and B^- equal to 1 to ensure probability conservation.

5.2.3. Market- or Priority-based Scheme

As detailed in our earlier work [76], market-based or Transactive techniques for coordinating TCLs can be incorporated in the aggregate model using reset equations. Assume that TCLs that reach at higher temperatures are willing to pay increasingly higher prices to

turn on than the ones already at cooler temperatures [32, 76]. Upon broadcast of a price signal, TCLs with offers above the market-price will clear. In the bin-model, this means (i) TCL price offers increase from lower temperature bins to higher bins, and (ii) on/off bins at the same temperature level have the same offer price, hence are cleared simultaneously. A clearing price $\pi^{\text{clr}}(t)$, thus, determines which bins are *cleared* (i.e. allowed to turn ON to consume power). Note that the above coordination mechanism is conceptually similar to the 'priority-stacking' scheme [40, 70] where bins near the edges get progressively higher priority to switch on/off depending on if they are near the upper/ lower limit of the dead-band range.

Assume each bin, i, (see Fig. 5.3) has a corresponding price level π_i. Assume a market-clearing price, π^{clr}. Let b^{clr}, with $b^{\text{clr}} \in \{1, .., N_B\}$, be the clearing bin index associated with π^{clr}. Then, for $i \in \{b^{\text{clr}} + 1, .., N_B\}$, $\pi_i \geq \pi^{\text{clr}}$. Hence, all such bins get cleared. On the other hand, bins $i \in \{1, .., b^{\text{clr}}\}$ do not get cleared. The ON bins, corresponding to each OFF bin behave similarly because $\pi_i = \pi_{2N_B - i + 1}$ for $i = 1, .., N_B$. Hence, the reset equations can be written as,

$$x_i^+ = x_i^- + x_{2N_B - i + 1}^-; \quad 1 \leq i \leq b^{\text{clr}}, \tag{5.7a}$$

$$x_i^+ = 0; \quad (b^{\text{clr}} + 1) \leq i \leq N_B, \tag{5.7b}$$

$$x_i^+ = x_i^- + x_{2N_B - i + 1}^-; N_B + 1 \leq i \leq (2N_B - b^{\text{clr}}), \tag{5.7c}$$

$$x_i^+ = 0; \quad (2N_B - b^{\text{clr}} + 1) \leq i \leq 2N_B. \tag{5.7d}$$

Equations (5.7a)-(5.7d) form the B-matrix. The column sum of B is always equal to 1 to ensure probability conservation. In a market-based coordination framework, the clearing prices $\pi^{\text{clr}}(t)$ can vary with time t. For different clearing prices $\pi^{\text{clr}}(t)$, the corresponding clearing bins, $b_{\text{clr}(t)}$ are known. Hence, the corresponding $B(\pi^{\text{clr}}(t)) = B(b^{\text{clr}}(t))$ can be computed using (5.7a)-(5.7d). Assuming each market interval has a duration of τ minutes, (5.1b) becomes

$$x(t) = B(b^{\text{clr}}(t))x(t^-), t = 0, \tau, 2\tau, ... \tag{5.8}$$

In the following section, we develop an eigenvalue-based approach to analyze how system behavior changes as TCL parameters change and under different control strategies.

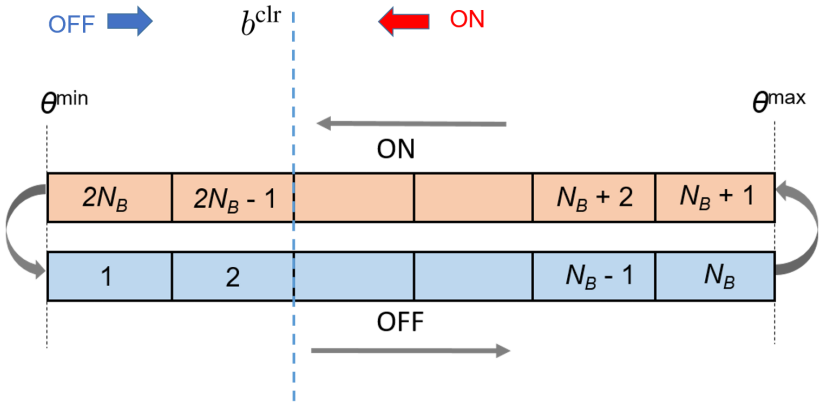

Figure 5.3: Bin-based model under market-based coordination.

5.3. Analysis using Eigenmodes

5.3.1. Modal Decomposition

In this section, the evolution in $\mathbb{A} = A^{\tau}B$ and the system behavior will be studied by using modal analysis. Assume matrix $\mathbb{A}$ is diagonalizable and is a $M \times M$ matrix, where $M = 2N_B$. Let $\lambda_1, \lambda_2, ..., \lambda_M$ be eigenvalues of $\mathbb{A}$-matrix and $v_1, v_2, ..., v_M$ be the corresponding eigenvectors, i.e. $\mathbb{A}v_i = \lambda_i v_i$. Since A is the transpose of a Markov transition matrix, and for A and B probability is conserved, $\mathbb{A}$ is also the transpose of a transition matrix and has one eigenvalue at 1 (largest) [58]. Hence, $1 = |\lambda_1| \geq |\lambda_2| \geq .... \geq \lambda_M$.

Let the initial vector, $x_0 = c_1 v_1 + ... + c_M v_M$, where $\mathbf{c} = V^{-1}x_0$. Here, matrix $V = [v_1\ v_2\ ...\ v_M]$, i.e. contains the eigenvectors. Then,

$$x_k = \mathbb{A}^k x_0 = \mathbb{A}^k(c_1 v_1 + ... + c_M v_M) \tag{5.9a}$$

$$= c_1 \mathbb{A}^k v_1 + ... + c_M \mathbb{A}^k v_M \tag{5.9b}$$

$$= c_1 \lambda_1^k v_1 + ... + c_M \lambda_M^k v_M \tag{5.9c}$$

$$= \sum_{i=1}^{M} c_i \lambda_i^k v_i. \tag{5.9d}$$

This means that x_k is simply the weighted sum of the independent temporal evolution of the eigenmodes themselves. Note that, by default, these modes are orthogonal. c_i, λ_i, v_i can be complex depending on the structure of matrix $\mathbb{A}$. A complex valued eigenmode would contribute to x_k, in conjunction with its complex conjugate (which is also an eigenmode

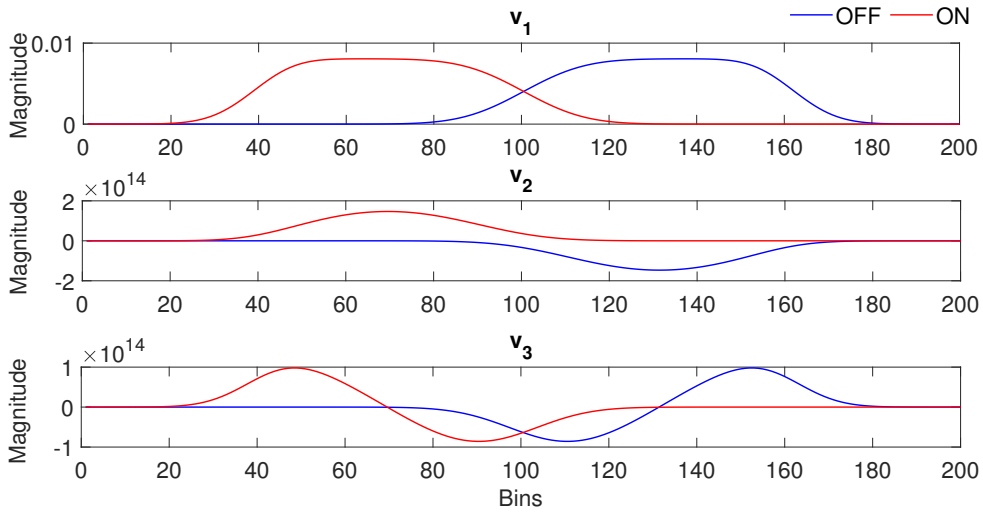

Figure 5.4: Mode shapes for the first three modes.

of matrix $\mathbb{A}$) as, $c_i\lambda_i^k v_i$ and $c_i^*\lambda_i^* v_i^*$.

The typical mode shapes (eigen-vectors) associated with $\lambda = 1$ and the next two dominant modes (with $|\lambda| < 1$) of a typical $\mathbb{A}$-matrix of TCLs is shown in Fig. 5.4. The first mode gives the steady-state distribution and the other modes influence the transient behavior. While different initial conditions may lead to complicated dynamics of TCLs, (5.9) allows representing the dynamics as a linear combination of the system's modes. When the initial conditions are close to, but not equal to, one of the modes, the behavior should be close to, but not equal to, the behavior of that mode alone (i.e., the other modes are only minimally excited). If initial conditions lie on only one of the modes, the behavior will be dictated by that mode alone.

5.3.2. Convergence Rate Analysis

To study synchronization, it is also essential to study how rapidly TCL state may synchronized given a particular coordination strategy. Therefore, here we seek properties relating the convergence rate and eigenvalues of $\mathbb{A}$.

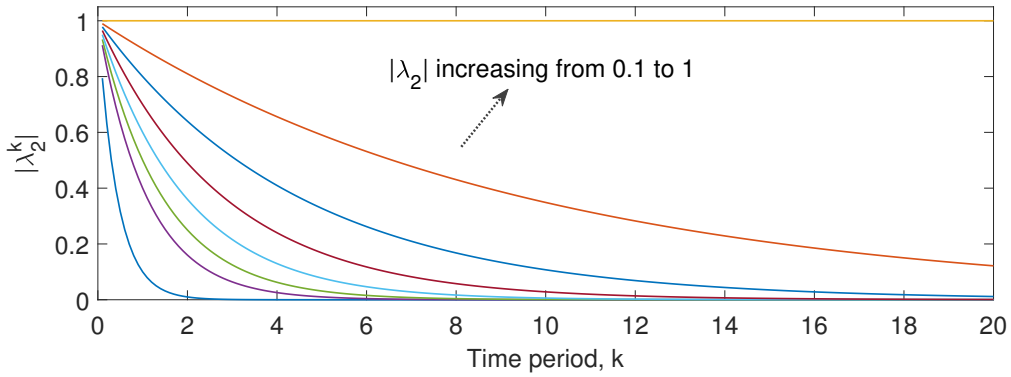

Figure 5.5: Eigenvalue convergence, shown by considering λ_2 from 0.1 to 1 and plotting λ_2^k, k = 0,1,...,20.

From (5.9), we can write,

$$\begin{aligned}\mathbb{A}^k x_0 &= c_1 v_1 + c_2 \lambda_2^k v_2 + ... + c_n \lambda_n^k v_n \\ &= c_1 (v_1 + \frac{c_2}{c_1} \lambda_2^k v_2 + ... + \frac{c_n}{c_1} \lambda_n^k v_n). \end{aligned} \tag{5.10}$$

If $|\lambda_2| < 1$, $\mathbb{A}^k x_0 \to c_1 v_1$ since $\lambda_2^k \to 0$ as $k \to \infty$. Thus, x_k converges to a multiple of eigenvector v_1. The convergence is geometric with ratio $\frac{|\lambda_2|}{|\lambda_1|} = |\lambda_2|$, as illustrated in Fig. 5.5 for a range of $|\lambda_2|$ values with evolution in k. In certain cases, λ_2 = -1 (as will be shown in Section 5.6). Then, λ_2^k will not approach 0, instead x_k will evolve as a linear combination of contributions from λ_1 and λ_2.

Let $x_{\lambda_1} = c_1 v_1$, i.e. the steady-state eigenvector v_1 normalized to sum to 1 for TCL ensembles. From (5.10), observe that the difference between x_{λ_1} and an arbitrary initial state x_0 goes to 0 at least as fast as λ_2^k goes to 0. In (5.10), let the contribution of λ_2 after k time steps be $\lambda_{2,k} = \lambda_2^k$. Taking log on both sides of $\lambda_{2,k} = \lambda_2^k$ we obtain, $k = {}^{\log(\lambda_{2,k})}/_{\log(\lambda_2)}$. To obtain $\lambda_{2,k} = \epsilon$, a small number so that it becomes negligible (e.g. 10^{-5}), we then need

$$k(\epsilon) = \frac{\log(\epsilon)}{\log(\lambda_2)}. \tag{5.11}$$

Let $\epsilon = 10^{-5}$. For $\lambda_2 = 0.5$, we obtain $k = 16.61$, and for $\lambda_2 = 0.2$, $k = 7.61$. Thus, in a limited number of time-steps, the contribution of λ_2 vanishes. For all $|\lambda_i| < |\lambda_2|$, $i = 3, ..., M$, λ_i^k would vanish even faster. Thus, the second largest eigenvalue can be used

to estimate the rate of convergence to the steady-state TCL distribution. Since $\mathbb{A}$ can be shown to be the transpose of a Markov transition matrix, the convergence behavior can also be studied using the spectral gap, γ^*, of matrix $\mathbb{A}$, where $\gamma^* = 1 - |\lambda_2|$, hence large gaps indicate faster convergence [58]. Section 5.6 will present several numerical examples for controlled TCL populations.

5.3.3. Variations in Aggregate Power

Fixed Reset Conditions

Assume B is fixed. The dynamics of x^- and x^+ is then governed by $\mathbb{A} = A^\tau B$. The steady-state of x^-, denoted by $x^-_{\lambda_1}$, is given by the scaling v_1 of $\mathbb{A}$. The steady-state $x^+_{\lambda_1}$ can be obtained by $x^+_{\lambda_1} = Bx^-_{\lambda_1}$. Thus, at resets, the absolute change in power consumed by TCLs can be obtained using

$$|y^+ - y^-| = |C(x^+_{\lambda_1} - x^-_{\lambda_1})| \tag{5.12a}$$

$$= |C(B - I)x^-_{\lambda_1}|. \tag{5.12b}$$

Variable Reset Conditions

Similarly, variable reset conditions can also be considered. For example, it is straightforward to extend to the case of periodic reset signals where two reset maps B_1 and B_2 are are applied alternatively during reset events. In this case, the dynamics of x^+_k can be written in either of the two forms,

$$x^+_{k+1} = B_1 A^\tau B_2 A^\tau \, x^+_k, \quad k = 0, 1, 2, ... \tag{5.13a}$$

$$x^+_{k+1} = B_2 A^\tau B_1 A^\tau \, x^+_k, \quad k = 0, 1, 2, ... \tag{5.13b}$$

The distributions corresponding to $x^+_{\lambda_1}$ are given by the eigen-vectors of $(B_2 A^\tau B_1 A^\tau)$ and $(B_1 A^\tau B_2 A^\tau B_1)$. Knowing $x^-_{\lambda_1}$ and $x^+_{\lambda_1}$, the changes in aggregate power consumed can be computed without resorting to simulations. The above technique can also be extended to study other combinations of control sequences.

5.4. Synchronization Index

While it can be easy to detect full or partial synchronization of TCL temperature distributions by visual inspection, it is not a formal technique. Availability of standard and easy to interpret indices would allow to systematically measure the level of synchronization in a given TCL distribution, and indicate if actions should to be taken to mitigate. For example, appropriate feedback or penalty terms to suppress the synchronizing tendency of TCLs can be considered. Hence, few measures are proposed below. First, recall that a bin represents a temperature range and its associated on/off state. The temperature distribution, x^θ, from the TCL distribution x over the bin space can be reconstructed using,

$$x_i^\theta = x_i + x_{2N_B - i + 1} \; i = 1, ..., N_B. \tag{5.14}$$

This is because i and $2N_B - i + 1$, $i = 1, ..., N_B$ represent the same temperature range. Now using x^θ, the following synchronization indices are proposed.

Maximum Fraction of TCLs in a Bin

Given a distribution x^θ over N_B bins, S_1 $(0 \leq S_1 \leq 1)$ is given by,

$$S_1 = \max\{x_i^\theta, i = 1, ..., N_B\} \tag{5.15}$$

Bin Spread

Define $\mathcal{X}_\epsilon$ as a set of indices of bins containing TCL fractions above some threshold, ϵ. Hence,

$$\mathcal{X}_\epsilon = \{i : x_i^\theta \geq \epsilon, i = 1, ..., N_B\} \tag{5.16}$$

Then, a measure for bin spread is given by the cardinality of $\mathcal{X}_\epsilon$, i.e. $|\mathcal{X}_\epsilon|$. By normalizing we obtain, S_2 as

$$S_2 = 1 - \frac{|\mathcal{X}_\epsilon|}{N_B}. \tag{5.17}$$

Note that $0 \leq S_2 \leq 1$. A small S_2 value indicates the distribution is more widely spread over the temperature bins, whereas larger values would indicate synchronization.

Bin Range

Note however that S_2 as a measure of the bin spread is still not indicative of whether fractions of TCLs are lying in adjacent bins or are spread apart. Hence, range of $\mathcal{X}_\epsilon = \{\min \mathcal{X}_\epsilon, \max \mathcal{X}_\epsilon\}$ should also be considered. We obtain index S_3 as,

$$S_3 = 1 - \frac{\max \mathcal{X}_\epsilon - \min \mathcal{X}_\epsilon}{N_B}. \tag{5.18}$$

Again, $0 \leq S_3 \leq 1$, and smaller values of S_3 mean TCL distributions are more widely spread over bins, whereas larger values indicate synchronization.

(d) Combined Metric: Indices $0 \leq S_1, S_2, S_3 \leq 1$ can be measured and reported separately. However, since all are normalized quantities, consider a combined index, $\widehat{S}$ ($0 \leq \widehat{S} \leq 1$) as,

$$\widehat{S} = S_1 S_2 S_3. \tag{5.19}$$

Alternatively, $S_t = \sum_i \omega_i S_i / \sum_i \omega_i$, where $\omega_i, i = 1, 2, 3$ are user-defined weights. While more sophisticated measures can also be considered, this chapter will consider the listed three - more advanced indices could be a topic of future work. The above indices, together with convergence rates from eigenvalues and spectral gaps, provide detailed insights on whether a given control strategy will induce synchronization, to what degree and at what rate.

5.5. Stability Analysis using Lyapunov Functions

In this section, the use of Lyapunov (energy) function based method will be investigated for the hybrid system (5.1a)-(5.1b). First, a candidate Lyapunov function must be derived for (5.1a) assuming the $\mathcal{A}$-matrix in (5.1a) is time-invariant.

Consider the quadratic Lyapunov function,

$$V(x(t)) = x(t)^T P x(t), \tag{5.20}$$

where $P = P^\top$ and $P > 0$ (positive definite).

Fig. 5.6 illustrates the behavior of a system, which consists of 2-states and a stable $\mathcal{A}$ matrix, under reset actions. The state trajectories for two different initial conditions are shown, along with the Lyapunov function level sets. Since the original system is stable, the

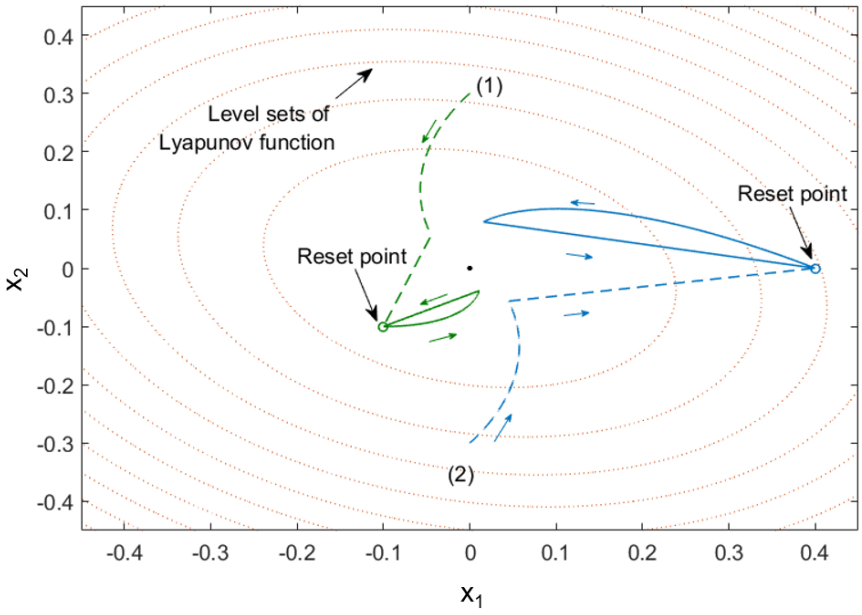

Figure 5.6: Level sets of Lyapunov function and state trajectories with resets.

trajectories approach origin. However, we see that reset actions take them away from the origin. Eventually, the trajectories converge to stable limit cycles. A similar analysis needs to be performed for (5.1a)-(5.1b). However, complications may arise since the dimension of the state-space is typically very large and $\mathcal{A}$ is semi-stable (with one eigenvalue at zero).

Generally, P-matrix in (5.20) is obtained by solving, $\mathcal{A}^\top P + P\mathcal{A} = -Q$, where $Q = Q^\top$ and $Q > 0$ [48]. However, in our case, the $\mathcal{A}$-matrix governing the autonomous behavior of a TCL population has an eigenvalue at zero, hence is semi-stable. Due to this, standard numerical recipes to obtain a feasible P typically fail. However, since the eigenvalue at zero is structural (ensures probability conservation), the n $(= 2N_B)$-order model can be reduced to $n-1 (= 2N_B - 1)$ order model. The $(n-1)$-order model provides a stable $\mathcal{A}$-matrix, hence we can easily obtain $\tilde{P}$, the Lyapunov matrix corresponding to the reduced $\mathcal{A}$-matrix. Using $\tilde{P}$, we provide a procedure to then obtain a P matrix associated with the original system.

5.5.1. State Reduction and Transformations

For $\dot{x} = \mathcal{A}x$, from the state constraint $\sum_{i=1}^{n} x_i = 1$, x_n can be written as,

$$x_n = 1 - \mathbf{1}^\top \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \tag{5.21}$$

Therefore,

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} = \mathcal{A}_{(1:n-1,1:n-1)} \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)} x_n \tag{5.22}$$

$$= \mathcal{A}_{(1:n-1,1:n-1)} \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)} - \mathcal{A}_{(1:n-1,n)} \mathbf{1}^\top \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \tag{5.23}$$

$$= \left(\mathcal{A}_{(1:n-1,1:n-1)} - \mathcal{A}_{(1:n-1,n)} \mathbf{1}^\top \right) \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)}. \tag{5.24}$$

Let $\tilde{x}$ denote the reduced state vector and $\tilde{\mathcal{A}}$, the corresponding reduced system matrix. Thus, (5.24) can be written as, $\dot{x} = \tilde{\mathcal{A}}\tilde{x} + \mathcal{A}_{(1:n-1,n)}$.

Consider the eigenvector v corresponding to the zero-eigenvalue of $\mathcal{A}$. Scale the elements

of v such that $\sum_{i=1}^{n} v_i = 1$.

$$\mathcal{A}v = 0 \tag{5.25}$$

$$\Rightarrow \quad \mathcal{A}_{(1:n-1,1:n-1)} \begin{bmatrix} v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)} v_n = 0 \tag{5.26}$$

$$\Rightarrow \quad \mathcal{A}_{(1:n-1,1:n-1)} \begin{bmatrix} v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)} - \mathcal{A}_{(1:n-1,n)} \mathbf{1}^\top \begin{bmatrix} v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} = 0 \tag{5.27}$$

$$\Rightarrow \quad \left(\mathcal{A}_{(1:n-1,1:n-1)} - \mathcal{A}_{(1:n-1,n)} \mathbf{1}^\top \right) \begin{bmatrix} v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} + \mathcal{A}_{(1:n-1,n)} = 0. \tag{5.28}$$

To transform $\tilde{x}$ into x, we can write,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = K \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} + K_0, \tag{5.29}$$

$$\Rightarrow \quad x = K\tilde{x} + K_0, \tag{5.30}$$

where

$$K = \begin{bmatrix} \mathbf{I}^{(n-1)\times(n-1)} \\ -\mathbf{1}^{1\times(n-1)} \end{bmatrix}, \quad K_0 = \begin{bmatrix} \mathbf{0}^{n-1} \\ 1 \end{bmatrix},$$

Notice that K shown above is not a square matrix. Hence, we obtain a square matrix, K_1, by appending a zero column (has no effect on the dynamics) as

$$K_1 = \begin{bmatrix} K & 0 \\ -\mathbf{1}^{1\times(n-1)} & 0 \end{bmatrix}. \tag{5.31}$$

5.5.2. A Candidate Lyapunov Function

Next, to obtain a P-matrix for the Lyapunov function corresponding to $\mathcal{A}$, first solve $(\tilde{\mathcal{A}}^\top \tilde{P} + \tilde{P}\tilde{\mathcal{A}}) = -\mathbf{I}$. Then, applying the similar operation to convert $\tilde{x}$ to x, we can obtain a P matrix of dimension $(n \times n)$ as,

$$P = \frac{P_1 + P_1^\top}{2}, \tag{5.32}$$

where

$$P_1 = \begin{bmatrix} \tilde{P} & 0 \\ -\mathbf{1}^{1\times(n-1)} & 0 \end{bmatrix} \tag{5.33}$$

Finally, assume x_s is the steady-state eigenvecotr of $\mathcal{A}$, s.t. $\mathcal{A}x_s{=}1x_s$. Then, $V(x)$, using the P from (5.32), can be written as,

$$V(x) = (x - x_s)^\top P(x - x_s), \tag{5.34}$$

For a more detailed derivation, please refer to the Appendix D.

5.5.3. Impact of Resets on Lyapunov Function's Value

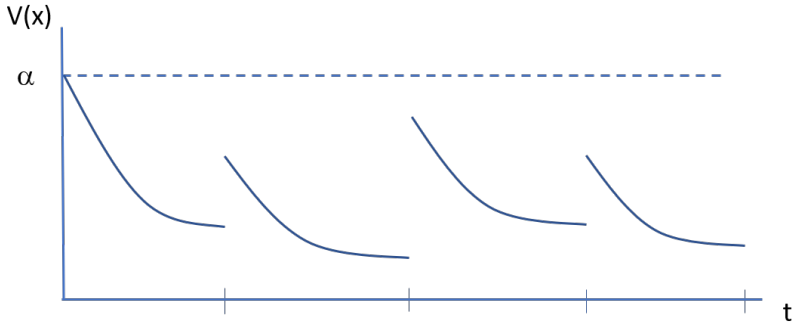

Figure 5.7: Value of Lyapunov function with state resets.

Fig. 5.7 illustrates the effect of the variable reset maps (B varies) on Lyapunov function values. For a particular reset map, it can be investigated if it is possible to ensure $V(x(t)) \leq \alpha$, for all $t \geq 0$, where α is a constant ($\alpha \geq 0$ and $\alpha \in \mathbb{R}$). This guarantees that states are convergent and reset actions would not induce instability [63]. The value of α can be chosen appropriately based on the duration of the switching interval and the specific control objective. It is also possible to specify α to obtain invariance conditions [48], i.e. for a range of Bs, M can be a positively invariant set. Define M as, $M = \{x \; s.t \; V(x) \leq \alpha\}$. In other words, it will ensure that as long as Bs (for example, price signals) are chosen from a permissible set, the state trajectories are guaranteed to remain within set M. Numerical examples are provided in the simulations section.

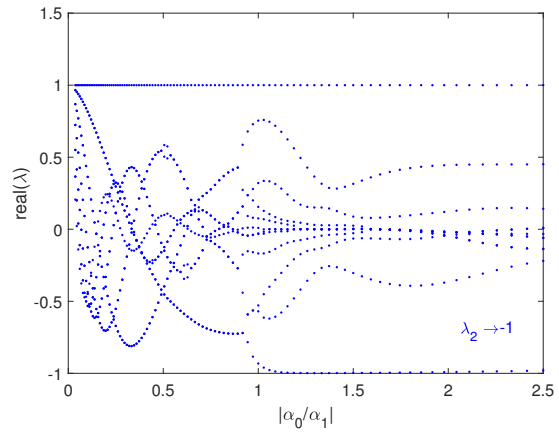

Figure 5.8: Real parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (a).

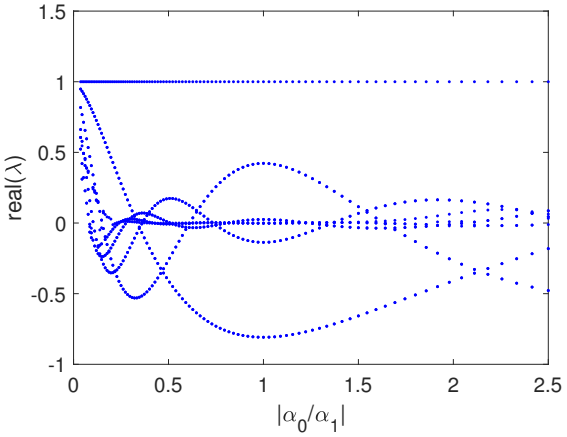

Figure 5.9: Real parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (b).

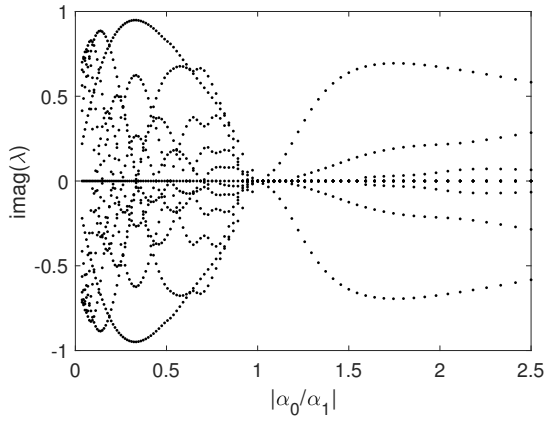

Figure 5.10: Imaginary parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (a).

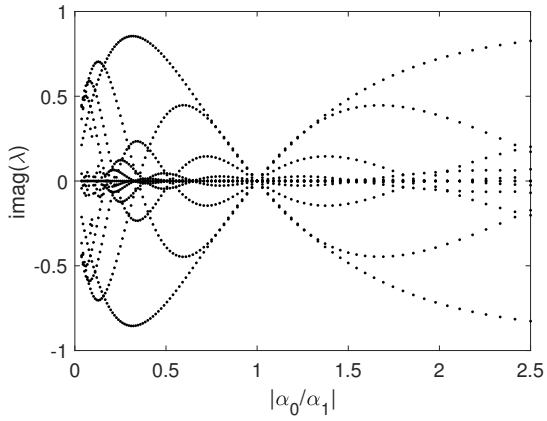

Figure 5.11: Imaginary parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (b).

5.6. Simulations

5.6.1. Influence of Parameters on System Behavior

Consider $N_B = 200$. To obtain, $\mathcal{A}$, the average heating and cooling rates in (2.16) and (2.17) are obtained using $P = 14$ kW, $R = 2°\text{C/kW}$, and $C = 10$ kWh/°C, $\theta = 20°\text{C}$ and $\theta^{\text{amb}} = 32°\text{C}$ (unless specified otherwise). A-matrix is obtained using $t_s = 1\text{s}$ in (2.20). Matrix B for market-based switching is calculated using b^{clr} and for randomized switching (RS) using f^+ and f^-.

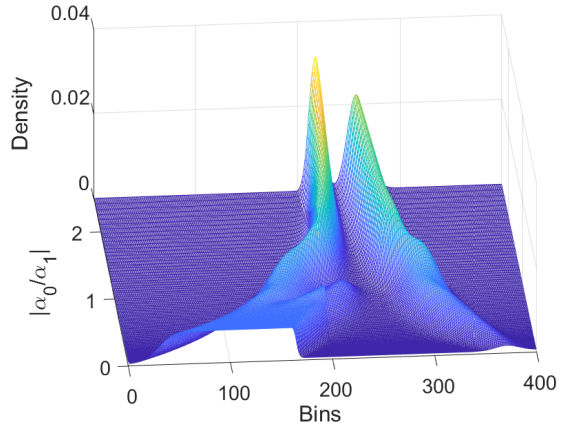

Figure 5.12: $x^-_{\lambda_1}$ distributions as a function of $|\alpha_0/\alpha_1|$ in case (a).

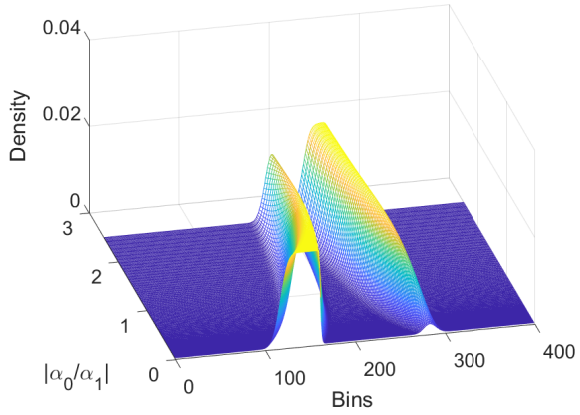

Figure 5.13: $x^-_{\lambda_1}$ distributions as a function of $|\alpha_0/\alpha_1|$ in case (b).

To study how the system behavior changes with changes in TCL parameters, we analyze their effect on eigenvalues and eigenvectors of $\mathbb{A}$. Four cases are considered, (a) $\tau = 30$ mins and $b^{\text{clr}} = 0.8N_B$, (b) $\tau = 10$ mins and $b^{\text{clr}} = 0.8N_B$, (c) $\tau = 30$ mins and $b^{\text{clr}} = 0.5N_B$, and (d) $\tau = 10$ mins and $b^{\text{clr}} = 0.5N_B$. For case (a), we vary θ^{amb} from 21 to 40°C. For each θ^{amb}, we compute α_0 and α_1, construct the A-matrix, and apply B to obtain $\mathbb{A}$ (see (5.4)). The changes in real and imaginary parts of the eigenvalues are shown in Fig. 5.8 and Fig. 5.10, as a function of $|\alpha_0/\alpha_1|$ (here, which changes due to changing θ^{amb}). Similarly, we repeat the experiment for cases (b) and (c). For case (b), the changes

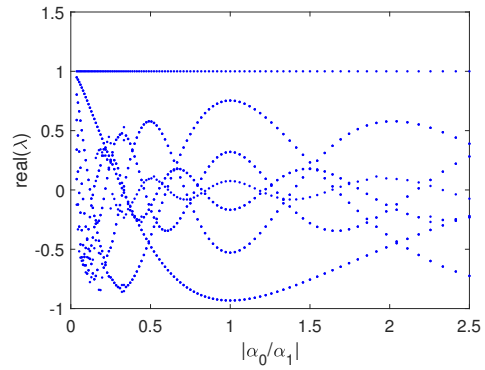

Figure 5.14: Real parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (c).

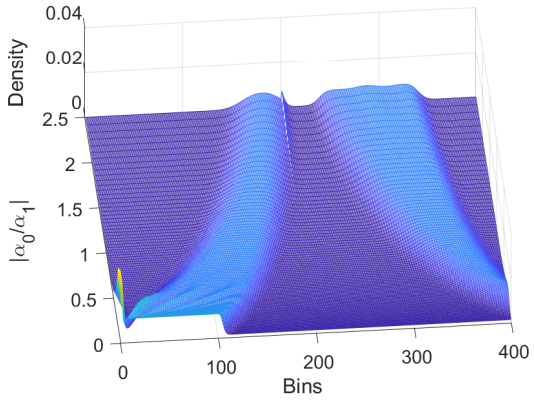

Figure 5.15: $x^-_{\lambda_1}$ distributions as a function of $|\alpha_0/\alpha_1|$ in case (c).

in real and imaginary parts of the eigenvalues are shown in Fig. 5.9 and Fig. 5.11. For case (c), the changes in real eigenvalues are shown in Fig. 5.14. For case (c), the behavior of the imaginary parts is similar to case (b), hence not repeated. Results for case (d) were also similar to case (b), hence not shown.

In case (a), we observe that there exist two eigenvalues with absolute value = 1 will emerge since a second eigenvalue reaches -1. This suggests that, for these parameter values, depending on the initial condition, both modes may persist and then instead of converging to a single steady state distribution (corresponding to the first mode), the states will go through limit cycle oscillations. Only when the initial condition, x_0, does not span the non-decaying modes associated with λ_2, x_k will converge to the stead-state associated with λ_1, in all other cases, x_k will evolve as a linear combination of the two non-decaying modes. From Figs. 5.10-5.11, we also notice how the imaginary parts of the eigenvalues disappear when $|\alpha_0/\alpha_1|$ approaches 1, indicating structural changes in the system behavior. Two eigenvalues with modulus 1 suggests period of 2. In other words, any x_0 spanned just by the eigenvectors corresponding to the +1 and -1 eigenvalues will be returned to exactly after two steps.

Fig. 5.16 shows the values of the synchronization index $\hat{S}$ (5.19) for cases (a)-(d). For τ = 30 mins, we observe that the index increases in (a), as $|\alpha_0/\alpha_1|$ increases. This can also be explained from the series of $x^-_{\lambda_1}$ distributions shown for each cases in Figs. 5.12-5.15. For narrower distributions the index value indeed was higher, demonstrating usefulness of the index.

Next, we performed a similar analysis for randomized switching (RS) based coordination, where TCLs are asked to increase and subsequently decrease power consumption. Three cases are considered, (i) $f^+ = 0.001, f^- = 0.001, \tau = 5$ mins; (ii) $f^+ = 0.001, f^- = 0.001, \tau = 10$ mins; and (iii) $f^+ = 0.0001, f^- = 0.0001, \tau = 5$ mins. Construct B^+ using f^+ and B^- using f^-. Then, the evolution in x^-_k is then captured by $\mathbb{A} = (AB^+)^\tau(AB^-)^\tau$ and $\mathbb{A} = (AB^-)^\tau(AB^+)^\tau$. The resulting synchronization indices are shown in Fig. 5.17. The indices were much higher for small values of $|\alpha_0/\alpha_1|$. This suggests, when θ^{amb} is low, α_0 is much smaller than α_1 (cooling rate is much faster since heating is slow due to low ambient temperature). In this case, under a symmetric and fixed up-down power request signal, the temperatures synchronize rapidly. In case (ii), with longer time duration of τ = 10 mins, this is even more severe. The synchronization level was slightly lower in case (iii) due to f^+ and f^- being smaller. The eigenvalue plot associated with case (i), Fig. 5.21, also reveals different behavior near small values of $|\alpha_0/\alpha_1|$. When $|\alpha_0/\alpha_1|$ exceeded $\sim$0.4, the

synchronization index approached 0, ensuring dispersed distributions and safe operating conditions.

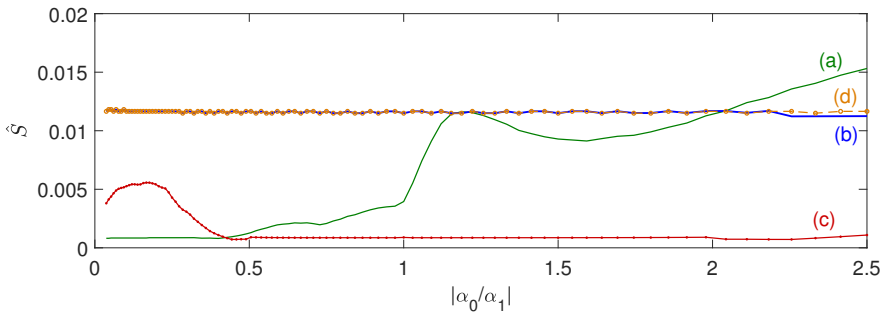

Figure 5.16: Comparing the synchronization index values for cases (a)-(d) under market-based coordination.

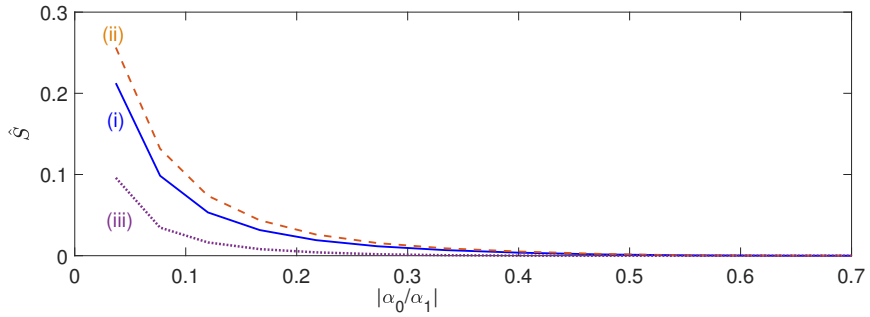

Figure 5.17: Comparing the synchronization index values for cases (i)-(iii) under randomized switching signals.

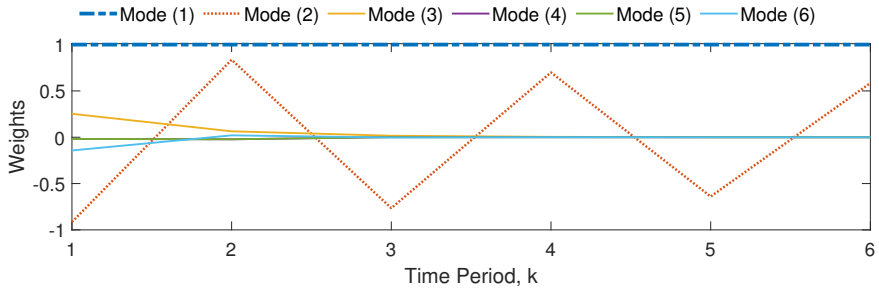

Figure 5.18: Evolution in modal weights at discrete time intervals.

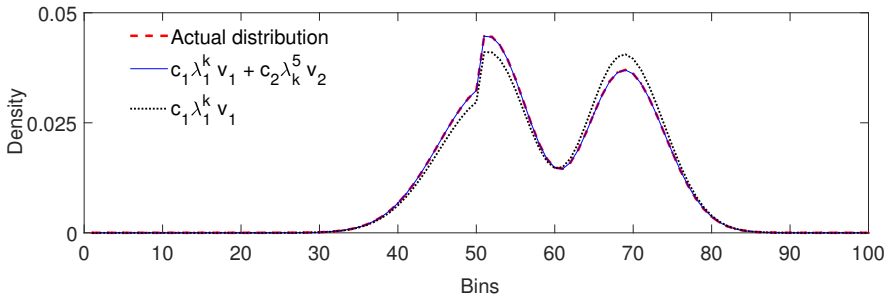

Figure 5.19: At time period $k = 5$, actual TCL distribution vs. approximate distributions obtained using only 2 modes (dashed line), only 1 mode (dotted line).

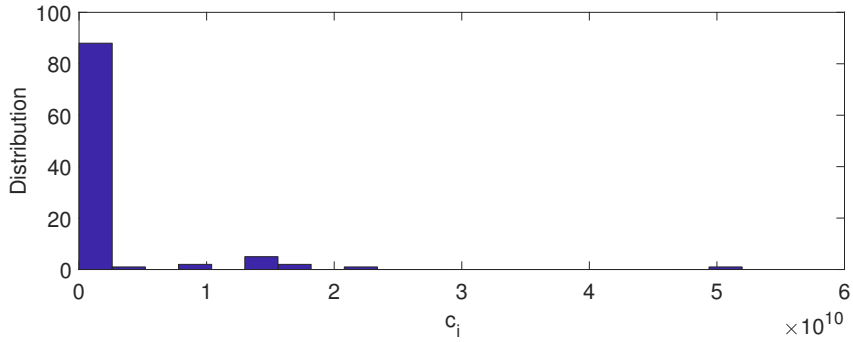

Figure 5.20: Histogram of modal weights for a random initial distribution

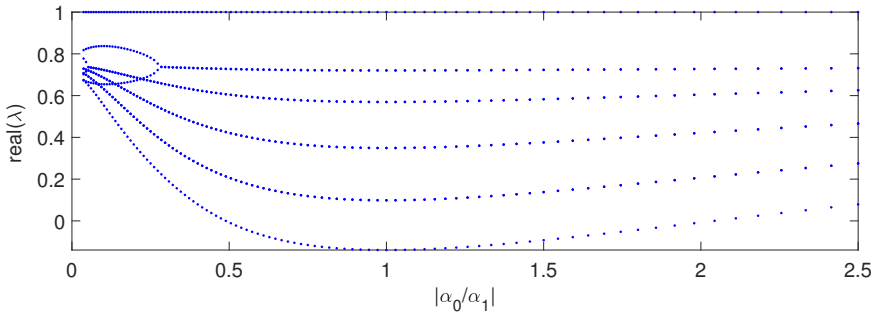

Figure 5.21: Real parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (i) (randomized switching).

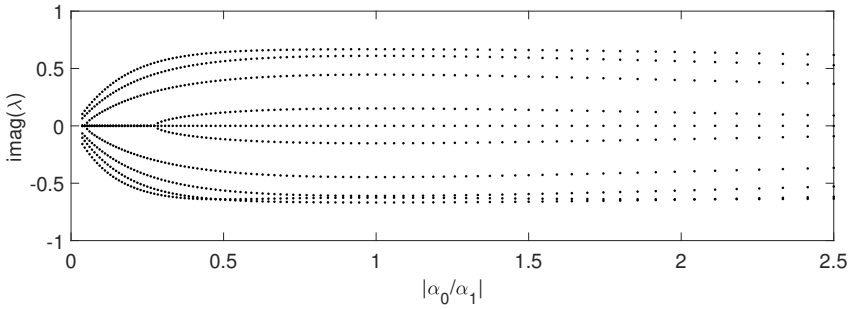

Figure 5.22: Imaginary parts of eigenvalues as a function of $|\alpha_0/\alpha_1|$ in case (i) (randomized switching).

5.6.2. Dominant Modes and Convergence

In (5.9d), assuming a sorted list of eigenvalues, the first few modes are often referred to as the dominant modes. We chose x_0 to be uniformly distributed over $2N_B = 100$ bins. Then, we simulated to obtain x_k at $k = 0, 1, ..., 5$. In Figure 5.19, we compared the actual x_k against different number of modes summed to give approximations. In this case, just the first and second modes were sufficient to obtain almost the exact distribution, whereas using just the first mode resulted in some error. The evolution in the i-th mode's weights, i.e. $w_i = c_i\lambda_i^k$ is shown in Fig. 5.18 for the first 6 modes, with $\lambda_1 = 1$ and $\lambda_2 = -0.87$. We see that the evolution in the 2nd mode's weight is oscillatory and does not die out rapidly, hence is important to consider. To further analyze, in Fig. 5.20, we plotted the histogram of all 100 elements of $\mathbf{c}$, which indicates how x_0 is composed of the different modes. We see that only a limited number of modes contribute to x_0. Additionally, among such modes, many have small eigenvalues, hence their contributions are negligible. Therefore, the use of modal analysis and modal coordinates can provide significant computational advantage. Instead of simulating TCL dynamics using $2N_B \times 2N_B$ matrices, we can obtain the behavior using a limited number of dominant modes. This is a topic of future research.

5.6.3. Variation in Aggregate Power Consumed

Assume TCLs are coordinated based on price signals in a double-auction market [76]. We applied periodic step changes in price signals, by varying b^{clr}, and observed different forms of oscillations in aggregate demand. Fig. 5.23 shows how a periodic price signal with small step changes can induce large amplitude oscillations in the aggregate demand. The

variations at resets matched the predicted value of 0.83 obtained via the method described in section 5.3.3. The eigen-structure associated with the discretized hybrid system thus describes the dynamics that underpins this power variation at resets. Similarly, for a variety of control signals, the variations in output power can be predicted through the modal analysis.

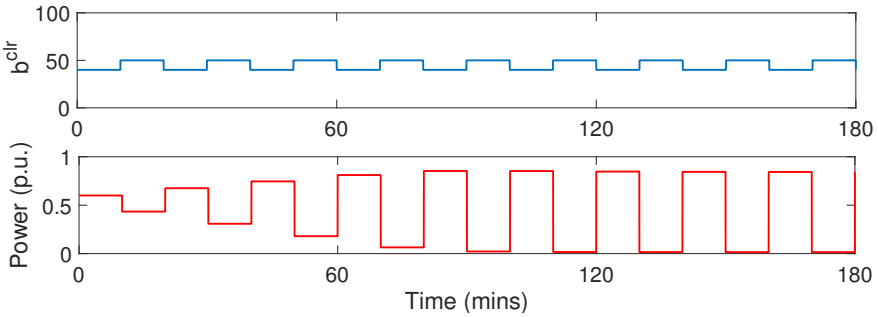

Figure 5.23: Large fluctuations in demand induced by a periodic price signal (mapped to b^{clr}).

5.6.4. Impact of Resets on Lyapunov Function Values

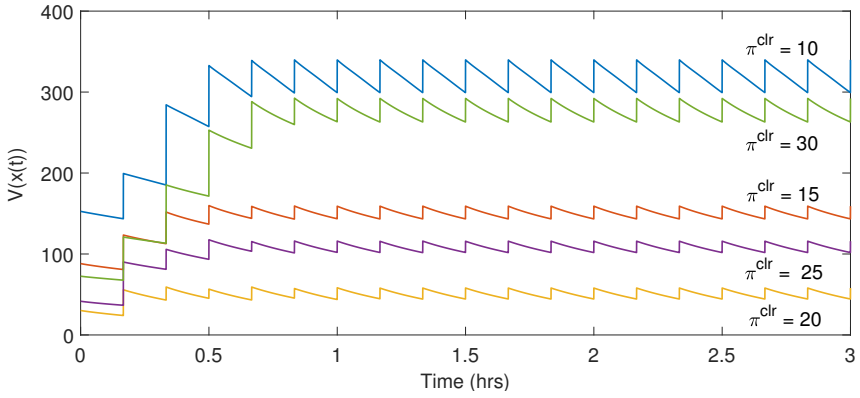

Figure 5.24: $v(x(t))$ values under different $\pi^{\mathbf{clr}}$.

Fig. 5.24 shows $v(x(t))$ values under different π^{clr}. While starting from the same initial condition, we see that as long as π^{clr} stays within 16 to 25 \$/MWh, the energy gain stays relatively small. However, as prices fall below or rise above this range, the energy gain

through resets increase rapidly. This suggests that such state resets at high or low price levels can cause the states to move drastically. As discussed before, since x^- and x^+ reach steady-state, the Lyapunov function values at resets can also be computed directly from the steady-state eigenvectors and results agreed with simulations.

5.7. Chapter Conclusions

This chapter proposed an analytical framework to explore conditions under which temperature synchronization of TCLs may appear and large oscillations in aggregate demand of load ensembles may emerge. We show that eigen-structure analysis can (i) identify if synchronization will appear, and (ii) determine the rate at which it would appear. To accomplish this, the dynamics of a controlled TCL population, under a given strategy, is expressed using a reset-based hybrid system. This allows us to study behavior as a parameter-dependent eigenvalue problem. Using the hybrid representation, the time-scale between the ensemble dynamics and control actions is systematically captured. Then, the eigenvalues and steady-state distributions of the discretized system explain whether the control will induce synchronization. Under priority- or market-based control and randomized switching, we showed that control parameters and update intervals can influence periodic behavior, synchronization, and/or damping of oscillations. For a given initial condition, use of dominant modes led to reduced computational needs. The spectral gap of the transition matrix was used to estimate the convergence rate.

We also derived a Lyapunov (energy) function for the autonomous $\mathcal{A}$-matrix governing the TCL dynamics. The impact of reset actions on the Lyapunov function's value were studied and it was showed that the system remains stable under reset actions. The Lyapunov function method can be used to establish invariance conditions under which states remain bounded within a predefined level-set while control actions can be varied indifferently.

The insights developed here can be used to quickly assess benefits and limitations of control techniques. Future work will involve comparing other control techniques such as set-point variation and will add additional operational constraints such as lockouts.

Chapter 6.

Quantifying and Addressing Performance Limitations in TCL Coordination

6.1. Chapter Introduction

While there is ongoing research to develop novel TCL control schemes – the purpose of this chapter is to provide underlying theoretically based performance limits for a large class of commonly used control schemes. Practical limitations also appear due to hardware, software and communication restrictions, which are often ignored in standard literature. For example, once an air-conditioner is turned off, the compressor must stay off for several minutes - thus causing periods of 'lock-outs'. In the existing literate on control of TCLs, parameters such as set-points and deadbands are typically assumed to be known, but in practice, TCL vendors and customers might not be willing to share set-point information, thus posing additional challenges for demand response aggregators. This chapter explores the impacts of considering these practical constraints and contributes to developing insights on how novel controllers can be designed to address these challenges.

A body of research has shown that populations of electric loads can be controlled to provide power system services, such as contingency reserves or real-time energy balancing (see [18]). Thermostatically controlled loads (TCLs), such as water heaters and air-conditioners, are especially suitable for these services because each load has some thermal energy storage capacity. A TCL's inherent energy storage allows for slight shifts in its power consumption patterns while still maintaining its temperature within the desired range. By aggregating

thousands of TCLs together, these slight shifts can be coordinated to make a population's total power consumption track a desired signal. Moreover, a TCL population can respond more quickly than a conventional thermal plant because each individual TCL can turn on or off almost instantaneously. For this reason, TCLs are especially suitable for providing real-time energy balancing services, such as regulation or load following.

To achieve energy balancing with TCLs, researchers generally control either the TCLs' thermostat set-points, or the TCLs' on/off operating states. While temperature set-point based control has been shown to be effective (e.g. [17]), by definition it alters user settings, which could result in a reduction of user comfort. In contrast, on/off switching can be implemented by changing the TCLs' on/off states while temperatures lie within the user defined temperature bands, thus ensuring user comfort (see [70]).

To reduce communication requirements, many on/off control schemes broadcast the same signal to all TCLs in the form of a "probability of switching". Upon receiving the signal, each TCL interprets it and independently decides whether to switch. As an example, suppose that to achieve a certain power target, 5% of the "off" TCLs need to be switched on; to achieve this, a switching probability of 0.05 could be broadcast. Each off TCL would draw a random number between 0 and 1 and switch if the number was less than 0.05. Given a large enough TCL population, the number that switch probabilistically is close to (but not guaranteed to be equal to) the desired, deterministic number. Thus, there is some inherent error due to the probabilistic switching, which we believe has not been studied in existing TCL literature. Hence, this chapter aims to determine a relationship between probabilistic switching error and the tracking performance of a population of TCLs. In regulation or load following services, a population's tracking performance can be quantified by the root mean square (RMS) error between the population's total power consumption and the reference power signal. We provide an analytical method to calculate the expected RMS tracking error that is solely due to probabilistic switching, which then serves as a performance bound.

A number of recent papers have used probabilistic switching in various contexts. For instance, [70] developed control schemes for reducing tracking error when state measurements are not fully available. The control scheme of [56] aims to minimize tracking error for a TCL-aggregator system that is subject to communication delays. In [21], authors developed an adaptive control scheme for mitigating the effects of population wide disturbances on the tracking performance. Author's earlier work [79] proposed a predictive scheme to track intra-hour optimal temperature distributions for meeting hourly unit commitment

objectives. [104], [90], [64], [47], [102], [69], [16] and [97] have also employed probabilistic switching based techniques for distributed control of TCLs to attain diverse objectives. In the literature there has been an implicit assumption that, when controlling a large population of TCLs, probabilistic switching will have negligible impact on tracking error; however, our analysis suggests otherwise. For example, the expected RMS error, solely due to probabilistic switching, in turning on (/off) 20% of a TCL population is 0.4%, for 10,000 TCLs, and 1.26% for 1,000 TCLs. We assert that knowing these performance bounds on tracking errors is valuable because they provide analytical benchmarks, based on the population sizes and the control schemes, against which empirical results can be compared. Additionally, it inspires the design of alternate switching schemes which can overcome the limitations of probabilistic switching.

To the best of our knowledge, there has been no prior study (except our work originally published in [82]) that establishes a performance bound on RMS tracking error for probabilistically dispatching TCLs for power system services. The primary contribution of this work is an analytical development of such performance bound, which is done by invoking properties of the binomial distribution. We extend our analysis to cases where switching commands can be distributed uniformly or non-uniformly across the TCL dead-band. Applying variance reduction techniques, we showed that it is possible to obtain lower RMS error by non-uniform switching.

Next, we deal the issue of unknown set-points. For non-disruptive control, i.e., control that respects user preferences, of thermostatically controlled loads (TCLs) such as air-conditioners and water-heaters, knowledge of TCLs' most recent set-points and their deadbands is essential. While smart thermostats can provide such information, the set-points and their adjustment strategies are typically protected by the TCL vendors. Without set-point and deadband information, demand response aggregators cannot effectively control such resources without potentially causing discomfort/inconvenience to users. Therefore, in this work, we design controllers that are agnostic to user's set-point changes and the underlying vendor-specific adjustment mechanisms, but the controllers learn user set-points over time. We assume TCL set-points are variable and unknown, but can be identified using our proposed algorithm. In the case of a smart residential thermostat, set-points of a TCL can change frequently depending on the behavior and preferences of users. Therefore, the key idea behind developing a set-point agnostic controller lies in using exploration-exploitation-type strategies, where a TCL controller takes turns to discover the new set-points versus providing services to the power grid (e.g. track regulation signals). With proper adjust-

ment of exploration and exploitation phases, and due to randomization, we show that the control strategies employed here would not noticably alter user consumption patterns, hence remain non-disruptive. A smart thermostat thus can participate effectively in grid services, and generate additional revenue for users or demand response aggregators. At the same time, the grid can also benefit from such fast dynamic resources.

6.2. Probabilistic Switching Schemes for TCLs

6.2.1. Uniform Probabilistic Dispatch

In this work, we assume that there is a centralized controller, operated by a load aggregator, that is able to communicate with every TCL. Our analysis is agnostic to how the centralized controller determines how many TCLs should switch in the next time step, but it must use some form of probabilistic dispatch as the control command. Probabilistic dispatch is the use of commands that consist of one or more "probabilities to switch". Each TCL receives the command and locally makes the decision whether to switch.

A uniform probabilistic dispatch is a scalar command, here denoted γ, that applies uniformly to all TCLs that are available to switch in the desired direction. If N_{OFF} number of TCLs are in their OFF states, and N^{+}_{SW} number of TCLs need to be switched ON, the uniform switching probability of switching from OFF to ON, γ^{+}, is defined as,

$$\gamma^{+} = N^{+}_{\text{SW}}/N_{\text{OFF}}. \tag{6.1}$$

The probability of switching from ON to OFF, γ^{-}, can be similarly defined.

Upon receiving the switching probability, γ, and the direction to switch (+ or −), each TCL available to switch in the desired direction generates a random number, r_i, from the uniform distribution between 0 and 1, and switches if $r_i \leq \gamma$.

Because prior work assumes the dispatch of TCLs should be non-disruptive to the user, a TCL should only switch when it is within its user set temperature range. To incorporate non-disruptive, probabilistic dispatch within the plant, we can write a TCL's ON/OFF

state switching logic as follows,

$$m_{i,(k+1)} = \begin{cases} 0, & \text{if} \quad \theta_{i,(k+1)} \leq \theta_i^{\min} \vee (m_{i,k} = 1 \wedge r_{i,k} \leq \gamma_k^-) \\ 1, & \text{if} \quad \theta_{i,(k+1)} \geq \theta_i^{\max} \vee (m_{i,k} = 0 \wedge r_{i,k} \leq \gamma_k^+) \\ m_{i,k}, & \text{otherwise.} \end{cases} \tag{6.2}$$

6.2.2. Non-Uniform Probabilistic Dispatch

When the state-bin model is used and state measurements or estimates are known, a control signal can be designed to target TCLs in different bins differently. This is what we refer to as non-uniform dispatch. For instance in [70], a controller is designed to switch TCLs that would otherwise soon switch naturally; this strategy benefits air conditioners by preventing their compressors from cycling too frequently. In [79], TCLs in different bins were sent different switching probabilities in order to track optimal target distributions.

For example, a non-uniform dispatch could protect compressors from short cycling by switching TCLs ON only when in the upper half of the temperature dead-band, and OFF only when in the lower half. This dispatch strategy would have switching probability γ_j^+ specified for each OFF bin j, such that

$$\gamma_j^+ = \begin{cases} \frac{N_{\mathrm{SW}}^+}{(N_{\mathrm{B}}/2)} \times \frac{1}{x_j N_{\mathrm{TCL}}} & j \in [N_{\mathrm{B}}/2 + 1, N_{\mathrm{B}}] \\ 0 & j \in [1, N_{\mathrm{B}}/2]. \end{cases} \tag{6.3}$$

where x_j is the fraction of TCLs in bin j. In the first case of (6.3), the first quotient is equal to the number of TCLs that need to switch per target bin, and the second quotient converts this number into a switching probability.

6.3. Analysis of Error due to Probabilistic Switching

In this section, we determine the performance limits under uniform and non-uniform probabilistic dispatch. First we find the expected RMS error of the fraction of TCLs that have switched under uniform dispatch; we then extend this analysis to the non-uniform case. Leveraging these results, we derive a performance bound on the RMS error of the population's power. We conclude this section by comparing results obtained analytically and by simulations.

In the following analysis, the characteristics of the error due to switching TCLs ON is identical to that due to switching TCLs OFF. We simplify our derivations by considering broadcasting a generic switching probability γ to a population of available TCLs of size N_{A}. Thus, we omit the ON/OFF subscripts for N_{ON} and N_{OFF} and instead use N_{A} to represent the number of TCLs that are available for switching (i.e. if TCLs are to be switched ON then $N_{\mathrm{A}} = N_{\mathrm{OFF}}$). The superscripts on γ^+ and γ^- are also omitted.

6.3.1. RMS Error of Fraction Switched: Uniform

In this section we derive an expression for the RMS error of the fraction of TCLs that switch, given a desired fraction γ. The error is defined as deviation from γ and the mean of the error is taken over K time steps. In order to arrive at a simple expression, we assume that the uniform switching probability is constant across all time steps (i.e. $\gamma_k = \gamma \ \forall k$) and the number of available TCLs is constant across all time steps (i.e. $N_{\mathrm{A},k} = N_{\mathrm{A}} \ \forall k$).

A TCL's switching decision can be modeled as a Bernoulli random variable, Z, with probability of "success" (i.e. switching) equal to γ. At each time step, the random variable Z is independently sampled N_{A} times, once by each TCL, resulting in the samples $\{Z_1, Z_2, ..., Z_{N_{\mathrm{A}}}\}$. The sample mean of Z is defined as

$$\bar{Z} = \frac{1}{N_{\mathrm{A}}} \sum_{n=1}^{N_{\mathrm{A}}} Z_n. \tag{6.4}$$

Note that the sample mean of Z serves as an estimator for the expectation of Z, i.e. γ. Moreover, the sample mean is an unbiased estimate of γ because

$$\mathrm{E}[\bar{Z}] = \frac{1}{N_{\mathrm{A}}} \sum_{n=1}^{N_{\mathrm{A}}} \mathrm{E}[Z_n] = \frac{1}{N_{\mathrm{A}}} \sum_{n=1}^{N_{\mathrm{A}}} \gamma = \gamma. \tag{6.5}$$

If we define the random variable Y to be the number of TCLs switched in a given time step, then

$$Y = \sum_{n=1}^{N_{\mathrm{A}}} Z_n = N_{\mathrm{A}} \bar{Z}. \tag{6.6}$$

Because $\bar{Z} = Y/N_{\mathrm{A}}$, the sample mean $\bar{Z}$ can also be interpreted as the fraction of TCLs switched.

Since Y is equal to the sum of independent, identically distributed Bernoulli random

variables, it constitutes a binomial distribution; in this case $Y \sim B(N_\mathrm{A}, \gamma)$ with mean $N_\mathrm{A}\gamma$ and variance $N_\mathrm{A}\gamma(1-\gamma)$. Using the properties of expectation, we find the variance of the sample mean $\bar{Z}$ to be

$$\mathrm{Var}(\bar{Z}) = \mathrm{Var}\left(\frac{Y}{N_\mathrm{A}}\right) = \frac{\mathrm{Var}(Y)}{N_\mathrm{A}^2} = \frac{\gamma(1-\gamma)}{N_\mathrm{A}}. \tag{6.7}$$

The mean square error (MSE) of the mean estimator $\bar{Z}$ can be decomposed into two components – bias and variance – (see [19]), such that

$$\mathrm{MSE}(\bar{Z}) = \left(\mathrm{E}[\bar{Z}] - \gamma\right)^2 + \mathrm{Var}(\bar{Z}). \tag{6.8}$$

Because $\bar{Z}$ has been shown to be an unbiased estimate of γ (see (6.5)), this expression simplifies to

$$\mathrm{MSE}(\bar{Z}) = \mathrm{Var}(\bar{Z}). \tag{6.9}$$

By application of (6.7) it further equates to

$$\mathrm{MSE}(\bar{Z}) = \frac{\gamma(1-\gamma)}{N_\mathrm{A}}. \tag{6.10}$$

Thus, we have that

$$\mathrm{RMSE}(\bar{Z}) = \sqrt{\gamma(1-\gamma)/N_\mathrm{A}}. \tag{6.11}$$

If we send probabilistic dispatches to N_A TCLs for each of K time steps, then there will be K instances of $\bar{Z}$. We define the vector of these instances to be $\bar{z} = [\bar{z}_1\ \bar{z}_2\ ...\ \bar{z}_K]$. The RMSE of $\bar{z}$ can be computed as

$$\mathrm{RMSE}(\bar{z}) = \sqrt{\sum_{k=1}^{K} (\bar{z}_k - \gamma)^2 / K}. \tag{6.12}$$

As K becomes very large, the RMS error computed using the instances (6.12) converges to the RMS error shown in (6.11), hence,

$$\mathrm{RMSE}(\bar{z}) \xrightarrow[K\to\infty]{} \sqrt{\gamma(1-\gamma)/N_\mathrm{A}}. \tag{6.13}$$

Thus, we have shown that the RMS error of the fraction of TCLs switched over K time

steps converges to $\sqrt{\gamma(1-\gamma)/N_{\mathrm{A}}}$ in the limit. In practice, a switching command γ will not be constant for an arbitrarily large number of time steps; hence, $\sqrt{\gamma(1-\gamma)/N_{\mathrm{A}}}$ can be thought of as the expected RMS error or the "expected performance" of the population in response to a known γ over any number of steps.

6.3.2. RMS Error of Fraction Switched: Non-Uniform

As discussed in Section 6.2.2, non-uniform dispatch can be used in conjunction with a state bin model of the population. In this section, we analyze the RMS error in the fraction switched due to non-uniform switching probabilities and compare it to that found for uniform dispatch.

Consider the case in which TCLs are being switched ON from OFF bins. Now, instead of sending the same γ to TCLs in all OFF bins, each OFF bin j is sent a separate switching probability γ_j. We are again interested in finding the population variance of the fraction of TCLs that switch from all bins, i.e. $\mathrm{Var}(\bar{Z})$. With Y representing the number of overall TCLs switched and Y_j representing the number of TCLs that switch from bin j, $Y = \sum_{j=1}^{N_{\mathrm{B}}} Y_j$. Then, with N_{A} available TCLs, $\mathrm{Var}(\bar{Z})$ can be expressed as,

$$\mathrm{Var}(\bar{Z}) = \mathrm{Var}(Y/N_{\mathrm{A}}) = \mathrm{Var}(\sum_{j=1}^{N_{\mathrm{B}}} Y_j/N_{\mathrm{A}}). \tag{6.14}$$

Next, let us define n_j as the number of TCLs in bin j and w_j as the jth bin's proportion of the total number of TCLs available for switching. Then, $w_j = n_j/N_{\mathrm{A}}$ and the sum of w_j for all j is 1. Thus, we can decompose the variance in $\bar{Z}$ as,

$$\mathrm{Var}(\bar{Z}) = \mathrm{Var}(\sum_{j=1}^{N_{\mathrm{B}}} w_j Y_j/n_j). \tag{6.15}$$

Because TCL switching decisions are independent, switching decisions per bin are also independent. Hence, all Y_j are independent and identically distributed (i.i.d.). Thus, from the properties of variance for i.i.d. random variables we have

$$\mathrm{Var}(\bar{Z}) = \sum_{j=1}^{N_{\mathrm{B}}} \mathrm{Var}(w_j Y_j/n_j) = \sum_{j=1}^{N_{\mathrm{B}}} w_j^2 \ \ \mathrm{Var}(Y_j/n_j). \tag{6.16}$$

Note that, Y_j/n_j is an unbiased estimate of γ_j, $\forall j$ (similar to Section 6.3.1). Hence, substituting the population variance calculated in (6.7) for each variance term in (6.16) and using $w_j = n_j/N_{\mathrm{A}}$, we obtain,

$$\mathrm{Var}(\bar{Z}) = \sum_{j=1}^{N_{\mathrm{B}}} \frac{n_j^2}{N_{\mathrm{A}}^2} \frac{\gamma_j(1-\gamma_j)}{n_j} = \sum_{j=1}^{N_{\mathrm{B}}} \frac{n_j \gamma_j(1-\gamma_j)}{N_{\mathrm{A}}^2}. \tag{6.17}$$

The above expression (6.17) then serves for obtaining the expected RMS error of the fraction of TCLs switched under non-uniform switching probabilities. Note that for the case of uniform switching probability per bin, $\gamma_1 = \gamma_2 = ... = \gamma_{N_{\mathrm{B}}} = \gamma$, the population variance is equivalent to that shown in (6.7). In the case of TCLs being switched OFF from ON states, similar analysis holds.

By comparing the derived formulas for the variance of $\bar{Z}$ under uniform and non-uniform dispatch, we observe that the variance of $\bar{Z}$ that results from uniform dispatch can be reduced by using a non-uniform dispatch scheme, while still switching the same number of TCLs. This observation has been further elaborated by examples in Section 6.3.4 and has motivated the design of controllers proposed in Sections 6.4.1 and 6.4.2.

6.3.3. RMS Error of Population's Power

RMS error can also be defined in terms of the aggregate power consumed versus desired, as a percentage of the steady state power consumption of the TCL population (e.g. [70], [51].

To obtain a simple analytical relation for RMS error that is analogous to (6.13), but in terms of power, assume that errors are only due to probabilistic switching. Let $\Delta P_k^{\mathrm{actual}}$ be the actual change in aggregate power in k, $\Delta P_k^{\mathrm{des}}$ the desired change in aggregate power in k, P^{ss} the steady state aggregate power consumption. Let P^{ON} be the average power consumed by the ON TCLs. We assume that TCLs are homogeneous; hence, $P^{\mathrm{ON}} = P^{\mathrm{rate}}/\eta$. Again, we assume that γ is constant for all time steps, and thus $\Delta P_k^{\mathrm{des}} = \Delta P^{\mathrm{des}}$ for all k.

Then, with N_{A} TCLs available for switching, the switching probability, γ, and the desired change in aggregate power, ΔP^{des}, are related by,

$$\Delta P^{\mathrm{des}} = \gamma N_{\mathrm{A}} P^{\mathrm{ON}}. \tag{6.18}$$

At a particular time k, the fraction actually switched $\bar{z}_k$, and the change in actual power consumed, $\Delta P_k^{\text{actual}}$, are related by,

$$\Delta P_k^{\text{actual}} = \bar{z}_k N_{\text{A}} P^{\text{ON}}. \tag{6.19}$$

The RMSE in tracking the aggregate power can be obtained as,

$$\text{RMSE}(\Delta \mathbf{P}^{\text{actual}}) = \sqrt{\sum_{k=1}^{K} \left(\Delta P_k^{\text{actual}} - \Delta P^{\text{des}}\right)^2 / K}. \tag{6.20}$$

Plugging (6.18) and (6.19) into (6.20), and after simplifying, we obtain,

$$\text{RMSE}(\Delta \mathbf{P}^{\text{actual}}) = N_{\text{A}} P^{\text{ON}} \sqrt{\sum_{k=1}^{K} (\bar{z}_k - \gamma)^2 / K}. \tag{6.21}$$

Finally, as K becomes very large, using (6.12) and (6.13), we obtain,

$$\text{RMSE}(\Delta \mathbf{P}^{\text{actual}}) \xrightarrow[K \to \infty]{} N_{\text{A}} P^{\text{ON}} \sqrt{\gamma(1-\gamma)/N_{\text{A}}}. \tag{6.22}$$

The above serves as an analytical expression for the expected RMS error, in terms of power, for a group of homogeneous TCLs responding to the same γ. To obtain the RMS error normalized by the steady state aggregate power consumption, the right-hand side of (6.22) must be divided by P^{ss}.

We reiterate that (6.22) only captures the error due to probabilistic switching. There could be other sources of error during tracking, for example, due to natural switchings at the dead-band boundaries. One can use an estimator to account for natural switchings, but depending on the estimation technique and availability of state information, performances could vary. In our work, we do not focus on designing estimators since irrespective of the performance of the state estimation techniques used, the error due to probabilistic switching is inherent and hence serves as a performance bound.

6.3.4. Simulation Verification of Performance Bound

To verify the performance bound derivations, we perform simulations of a TCL population responding to probabilistic dispatch. In particular, we consider the case of uniform switching probabilities and analyze the RMS error in the fraction of TCLs switched. For

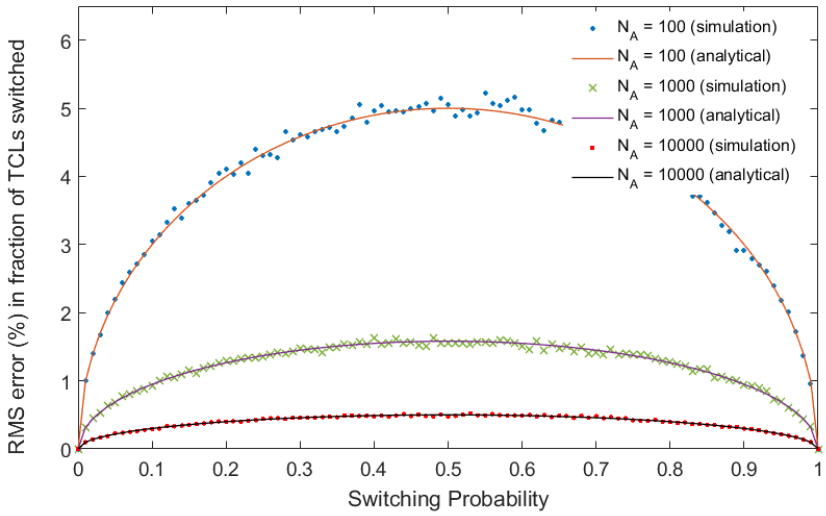

Figure 6.1: RMSE in fraction of TCLs switched with varying population size, obtained by simulations and analytically

the TCL population, parameters are according to [12]. Simulations were performed with N_{TCL} copies of (2.4) and (6.2) to collect error samples using the following steps: (a) Initialize the population at set-point temperature, θ_{set}, where all TCLs are in OFF state, hence $N_{\text{A}} = N_{\text{TCL}}$; (b) Send γ; (c) Every TCL performs a self-dispatch by drawing a random number; (d) Measure the aggregate number of ON TCLs and the error; (e) To obtain RMS error, repeat steps (a)-(d) K times, where K is a large number (here set to 1000) and compute RMSE using the sample based definition from (6.12).

Fig. 6.1 shows the RMS error for varying switching probabilities and with varying size of TCL population available for switching, $N_{\text{A}} = 100$, 1000 and 10000. In Fig. 6.1, results obtained via simulation are plotted with distinct points, and analytical results are plotted as lines. We see that results obtained theoretically and by simulations agree closely.

We observe that the RMS error varies with γ and peaks at $\gamma = 0.5$, which is a known property of the variance of a binomial random variable. In simulations with TCLs, we can therefore expect that, depending on the value of γ, different levels of variability will be observed. Next, in Fig. 6.2, the effect of varying the number of TCLs that are available (N_{A}) on the RMS error is shown for three values of γ. This plot was generated from the analytical expression (6.11). With increased N_{A}, RMS error decays with rate $1/\sqrt{N_{\text{A}}}$.

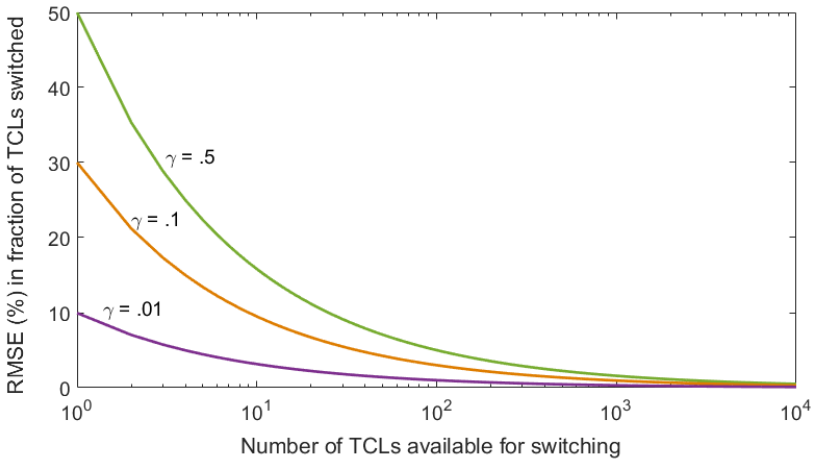

Figure 6.2: RMS error vs. numbers of TCLs available for switching

6.4. Controller Design for Variance Reduction

6.4.1. Optimization Formulation

In this section we develop a control law that reduces the error due to probabilistic switching. To do this, we leverage the extra degrees of freedom provided by non-uniform dispatch. We use the following objective function to determine the optimal γ_j values for each bin:

$$\min_{\gamma} \ \mathrm{Var}(\bar{Z}) = \sum_{j=1}^{N_\mathrm{B}} \frac{n_j \gamma_j (1-\gamma_j)}{N_\mathrm{A}^2} \tag{6.23}$$

This objective function is equivalent to minimizing $N_\mathrm{A}^2 \mathrm{Var}(\bar{Z}) = \sum_{j=1}^{N_\mathrm{B}} n_j \gamma_j (1-\gamma_j)$ since N_A is known and fixed at a given time. In turn, because $n_j = w_j N_\mathrm{A}$, we can rewrite the objective function in terms of w_j to obtain the following optimization problem,

$$\begin{aligned} \underset{\gamma}{\text{minimize}} \quad & \sum_{j=1}^{N_B} w_j N_\mathrm{A} \gamma_j (1-\gamma_j) \\ \text{subject to} \quad & 0 \le \gamma_j \le 1, \quad j = 1, \ldots, N_\mathrm{B}; \\ & \sum_{j=1}^{N_B} \gamma_j w_j = \gamma \ . \end{aligned} \tag{6.24}$$

The first constraint is to ensure that each γ_j is bounded within 0 and 1, and the second constraint is to ensure the fractions switched from different bins add up to the desired fraction.

The problem is non-convex. We implemented it using MATLAB's *fmincon* function which provided solutions within tolerance limits.

6.4.2. Variance Reduction along with Prioritized Switching

Different bins may be prioritized based on their proximity to the dead-band boundaries. This is useful to avoid short-cycling (i.e. frequent switching at dead-band boundary) of TCLs (e.g. [70]). It also reduces error due to natural cycling, as will be shown in case studies. It can be done by assigning a weight w_j^P per bin j. For example, here we assume w_j^P has a value proportional to its bin index j. Then, TCLs approaching natural switching boundaries will switch with higher priority if we modify the objective function of (6.24) to include the $w_j^P \gamma_j$ term, as expressed below,

$$\underset{\gamma}{\text{minimize}} \quad \sum_{j=1}^{N_B} w_j N_{\text{A}} \gamma_j (1-\gamma_j) - w_j^P \gamma_j \tag{6.25}$$

also subject to the same constraints as (6.24).

6.5. Application to Tracking

In this section, the results of two tracking scenarios are shown for three different controllers. *Controller 1* uses a uniform switching probability for TCLs in all bins. *Controller 2* uses non-uniform switching to reduce variance, as per Section 6.4.1, and *Controller 3* uses non-uniform switching to reduce variance as well as assign bin priorities, as per Section 6.4.2. In both non-uniform controllers the number of bins used is 40.

For tracking a power reference signal, P^{des}, a standard control law will be tested which was used for benchmarking by [70]. Given the measured aggregate power at time step k is P_k^{actual}, the uniform switching probability γ_{k+1} can be computed using the following,

$$\gamma_{k+1} = \frac{P_{k+1}^{\text{des}} - P_k^{\text{actual}}}{P^{\text{ON}}} * \frac{1}{N_{\text{A},k}} \tag{6.26}$$

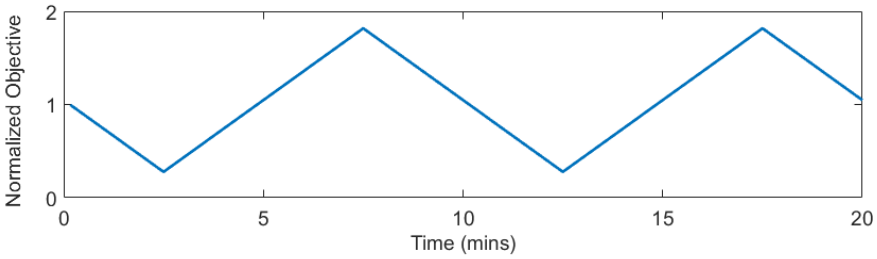

Figure 6.3: Reference triangular tracking signal (normalized)

Table 6.1: RMSE (%), in fraction of TCLs switched and in aggregate power tracking, due to probabilistic switching

RMS Error (%)	N_{TCL}	Controller		
		1	2	3
in fraction switched	1000	1.21	0.47	0.57
	10000	0.36	0.14	0.14
in power consumed	1000	1.17	1.67	0.46
	10000	0.39	1.59	0.14

The first quotient in (6.26) calculates the number of TCLs that need to switch, and the second quotient converts this number into a fraction with respect to the number of TCLs that are available to switch. In Controller 1, the uniform switching probability that is broadcast at time step k is γ_k, calculated using (6.26). In Controllers 2 and 3, the equality constraint in the respective optimization problems is replaced with $\sum_{j=1}^{N_B} \gamma_{j,k} w_{j,k} = \gamma_k$, where again γ_k is calculated using (6.26).

6.5.1. Scenario 1: Tracking a Triangular Reference Signal

In this section, results for the three controllers tracking a triangular reference signal (shown in Fig. 6.3) are presented. The reference signal is over 20 minutes. A simulation time step of 10 seconds was used.

Table 6.1 compares the RMS error (%) in fraction of TCLs switched and in aggregate power tracking due to probabilistic switching comparing three controllers, with $N_{\text{TCL}} = 1000$ and $N_{\text{TCL}}= 10{,}000$.

As percentage of the steady state error, with $N_{\text{TCL}} = 1000$, the RMSE for uniform

Table 6.2: RMSE (%), in fraction of TCLs switched and in aggregate power tracking, due to probabilistic switching

RMS Error (%)	N_{TCL}	Controller		
		1	2	3
in fraction switched	1000	0.22	0.16	0.17
in power consumed	1000	0.29	0.32	0.23

switching was 1.17% while the lowest achieved by non-uniform switching was 0.46% (by Controller 3). Using Controller 2 resulted in higher error in tracking the power. Since Controller 2 may often choose TCLs near the dead-band boundaries to switch, over-ride from natural switching contributes to the higher error. Controller 3 is less prone to this error since it switches TCLs with higher priority as they go closer to the boundaries.

6.5.2. Scenario 2: Tracking a Market Signal

In this section, we simulate a population of 1,000 TCLs tracking a reference signal (see Fig. 6.4), which was used by [70]. A simulation time step of 2 seconds has been used. Table 6.2 shows that Controller 2 obtains lowest RMS error in fractions switched, while priority based Controller 3 obtains the lowest RMS error in tracking the power. Results are consistent with Scenario 1.

6.5.3. Discussion on Tracking Results

We have found our RMS error results to be consistent with those presented in the literature. In [70], RMS errors of the power consumed by a tracking TCL population range from 0.26-10% for populations of 1,000 to 10,000 TCLs. As expected, our results are lower than than those from [70] because of our assumptions that the TCL population is homogeneous and that measurements from the TCLs are perfect and always available. Additionally, our non-uniform switching based controllers send the optimal dispatch fraction to each bin, which lowers the error due to probabilistic switching. Furthermore, our priority based Controller 3 reduces error due to natural switching. Additionally, comparing our two scenarios, the second scenario showed lower RMS errors, which was mainly due to low switching probabilities required to track a relatively conservative tracking objective.

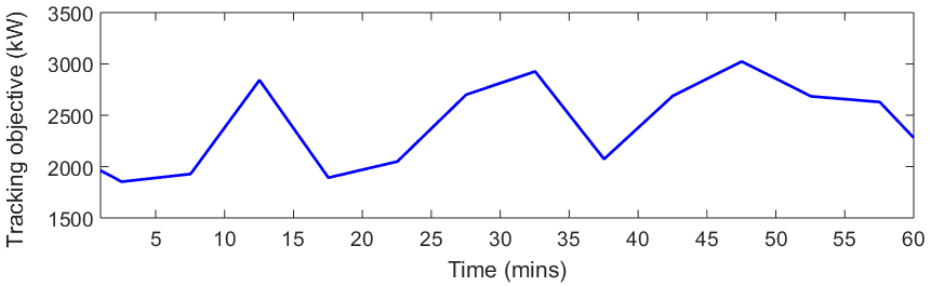

Figure 6.4: Market signal for Scenario 2

Applying the minimum variance non-uniform switching scheme reduces the error due to probabilistic switching, but may result in partial temperature synchronization. Hence, it is important to take these tradeoffs into account when designing controllers. In our simulations, we did not encounter any case with zero RMS error. However, it is theoretically possible to attain such cases. This would happen when the population dynamics evolve deterministically and a minimum variance controller is able to find optimal dispatch strategies to drive the RMS error due to probabilistic switching to zero. Furthermore, applying a priority-based controller (similar to Controller 3) would ensure over-rides due to natural switching would not result in any error. For more examples and detailed discussions, please refer to our work [82].

6.6. TCL Operations with Unknown Set-points

In this section, it is further considered that the information regarding a TCL's latest set-point is not readily available to a demand response aggregator. To capture the consequences of not knowing the set-point, first, the information flow among the TCL's internal thermostat, an external TCL controller unit (TCU) and the AC compressor is shown in Fig. 6.5. Here, we consider the case of a cooling TCL (air-conditioner).

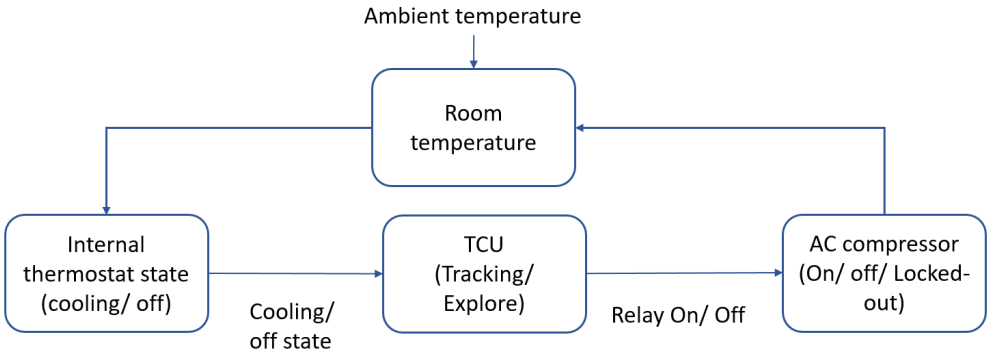

Figure 6.5: Information flow among the TCL's internal thermostat, the controller unit (TCU) and the AC compressor.

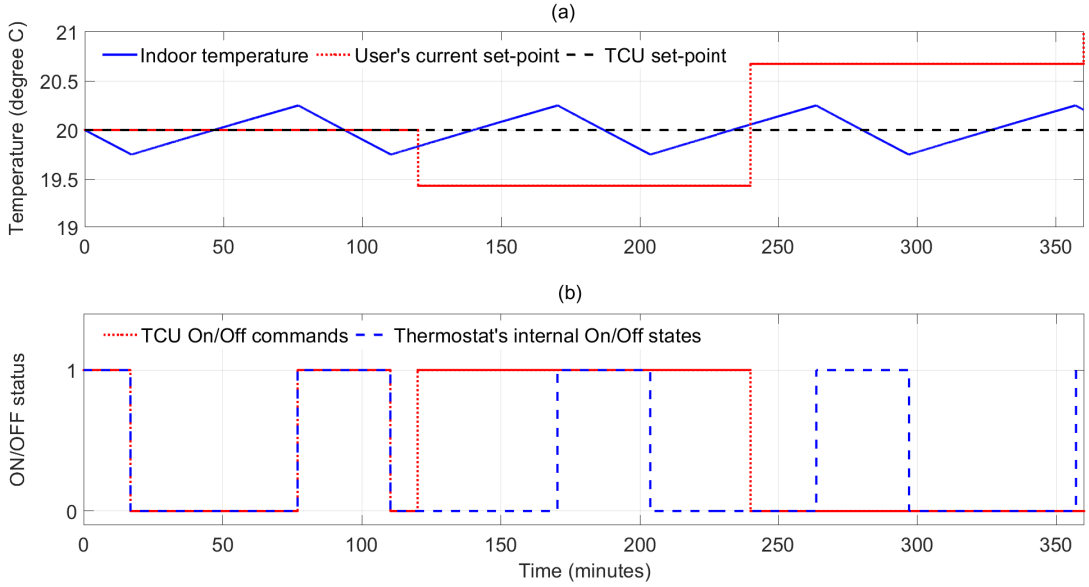

Figure 6.6: (a) Temperature and set-point evolution, (b) Thermostat's cooling/ off modes with and without TCU.

6.6.1. Assumptions

The following assumptions are made about the operations of a TCU and the internal thermostat of the AC.

- A1. The TCU is aware of the m_k status (cooling/ off) of the thermostat.
- A2. The TCU has no access to the temperature set-points of the AC user (i.e. the learning thermostat's set-points).
- A3. When thermostat is in the 'cooling' (ON) mode, TCU can issue off/on commands. Similarly, when the thermostat is in the 'OFF' mode, TCU can issue on/ off commands.
- A4. When switched OFF, the compressor must remain off (locked-out) for at least 5-minutes before turning the compressor back on.
- A5. When turned ON, must remain ON (locked-out) for at least 10-30 seconds.
- A6. At any time, user's may choose to opt out for several hours and during the opt-out period the thermostat follows its natural operations.

A single TCL's dynamics was simulated under the above assumptions. The temperature, set-points and on/off status are shown in Fig. 6.6. The TCU unit assumes the set-point is at 20^oC and does not update it. The user, however, updates the set-points at $t = 120$ min, and again at $t = 240$ min. Due to these changes, under normal operations, the internal thermostat of the AC would have turned on at $t = 120$ min, and off at $t = 240$ min. However, TCU, unaware of the set-point changes, keeps switching the AC within its current deadband around 20^oC. This will cause customer dissatisfaction since the control is not non-disruptive. This problem is addressed next by developing an 'explore-exploit' controller.

6.7. Set-point Agnostic Explore-Exploit Control Design

The variables are listed below,

μ_k TCU's active/ inactive status at time k.

ζ_k TCU's operating mode (exploration/ tracking) at k.

u_k^F TCU's command at k for switching the relay.

$\theta_k^{\text{set,F}}$ TCU's knowledge of the thermostat set-point at time k.

τ_k^{explore} Variable to indicate when the TCU should switch to an exploration phase.

The parameters are given by,

λ_k^{10} Switching probability to turn on from off state at time k.

λ_k^{01} Switching probability to turn off from on state at time k.

$\tau^{\text{Lock,0}}$ Locked time duration when turned off.

$\tau^{\text{Lock,1}}$ Locked time duration when turned on.

Additionally, the TCU can impose strict temperature boundaries, θ^{MIN} and θ^{MAX}, (e.g. 19 and 24^{c}C - wider than the natural deadband) as a safety mechanism.

Let $\theta^{\text{MIN}} = \min \theta_k$, i.e. equal to the lowest historically observed temperature for the user (or, a user-defined value). Similarly, let $\theta^{\text{MAX}} = \max \theta_k$, i.e. equal to the highest historically observed temperature for the user (or, user-defined value).

Additionally, the user can opt out of the control cycle or over-ride for 8 hours. This sets the TCU inactive, $\mu_k = 0$ for the entire time window, and the TCL simply operates without any TCU interaction.

6.7.1. Control Design

When a TCU is in an inactive state ($\mu_k = 0$), the thermostat follows its natural operations. On the other hand, an active TCU ($\mu = 1$) has two operating modes,

$$\text{Operating Mode} = \begin{cases} \text{Exploration}, & \text{if} \quad \zeta_k = 1 \\ \text{Tracking}, & \text{if} \quad \zeta_k = 0 \end{cases} \tag{6.27}$$

The exploration phase is to identify the set-point of the thermostat, whereas the tracking phase is to exploit the TCL's flexibility for regulation services.

Exploration Phase

Assume $\zeta_k = 1$. During the exploration phase, based on if a thermostat starts to call for cooling from a turned off state (or vice versa), the TCU can detect if the TCL's current temperature has hit one of the temperature boundaries. Hence, it updates τ_k^{explore} and $\theta_k^{\text{set,F}}$ in the following manner,

$$\tau_k^{\text{explore}} = \begin{cases} 0\ , & \text{if} \quad (m_k = 0\ \&\ m_{k-1} = 1)\ OR\ (m_k = 1\ \&\ m_{k-1} = 0) \\ \tau_{k-1}^{\text{explore}} + 1, & \text{otherwise.} \end{cases} \tag{6.28}$$

$$\theta_k^{set,F} = \begin{cases} \theta_k - \Delta/2, & \text{if} \quad m_k = 0\ \&\ m_{k-1} = 1 \\ \theta_k + \Delta/2, & \text{if} \quad m_k = 1\ \&\ m_{k-1} = 0 \\ \theta_{k-1}^{\text{set,F}}, & \text{otherwise.} \end{cases} \tag{6.29}$$

At initialization, $\theta_0^{\text{set,F}}$ can be set to the latest TCL temperature (measured), θ_k.

Immediately upon updating $\theta_k^{\text{set,F}}$, TCU sets $\zeta_k = 0$. TCU continues to belief that the user is maintaining its current set-point at or near $\theta_k^{\text{set,F}}$, but as time passes, this confidence reduces. Hence, the need to initiate exploration rises. This is described mathematically as follows,

$$\rho_k = \begin{cases} 0\ , & \text{if} \quad (m_k = 0\ \&\ m_{k-1} = 1)\ OR\ (m_k = 1\ \&\ m_{k-1} = 0) \\ 1 - \exp(-b\tau_k^{\text{explore}}), & \text{if } \zeta_k = 0 \end{cases} \tag{6.30}$$

where τ_k^{explore} is a design variable which imposes certain threshold time for the exploration phase and b is a constant. Note that any time a set-point update is detected, τ_k^{explore}, ρ_k and ζ_k are reset to 0, otherwise τ_k^{explore} grows at each step and ρ_k approaches 1. Eventually, when $\rho_k \geq 1-\epsilon$, where $\epsilon \geq 0$ and can be chosen appropriately, $\zeta = 1$. This can be described as follow,

$$\zeta_k = \begin{cases} 0\ , & \text{if} \quad (m_k = 0\ \&\ m_{k-1} = 1)\ OR\ (m_k = 1\ \&\ m_{k-1} = 0) \\ 1 & \text{if } \rho_k \geq 1 - \epsilon \end{cases} \tag{6.31}$$

The algorithm for set-point exploration can be summarized as follows,

Algorithm 1: TCL Set-point Exploration

S1. If $\rho_k \geq 1 - \epsilon$, set $\zeta_k = 1$.

S2. If $\zeta_k = 1$ and $m_k = 1$ (in cooling mode), set $u_F = 1$. If $\theta_k > \theta^{\text{MIN}}$, maintain until $m_k = 0$ when a new set-point is detected by (6.29). Else if $\theta_k \leq \theta^{\text{MIN}}$, set $u_F = 0$ and continue until a new set-point is found. Set $\zeta = 0, \tau_k^{\text{explore}}$ as soon as a new set-point is found.

S3. If $\zeta_k = 1$ and $m_k = 0$ (in off mode), set $u_F = 0$. If $\theta_k < \theta^{\text{MAX}}$, maintain until $m_k = 1$ (i.e a new set-point is detected by (6.29)). Else if $\theta_k \geq \theta^{\text{MAX}}$, set $u_F = 1$ and continue until a new set-point is found. Set $\zeta = 0, \tau_k^{\text{explore}}$ as soon as a new set-point is found.

Tracking Phase

The tracking phase considered here is similar to the probabilistic switching scheme described earlier in this chapter. The switching probabilities, λ^{01} and λ^{10}, are computed by a demand response aggregator and broadcast to the TCLs, who perform local randomized draws to turn on/ off.

6.7.2. Controller Performance

Fig. 6.7 provides an example of the explore-exploit control scheme. The temperature and set-point evolution are shown in the top figure and the various operating modes, cooling/off and explore are shown in the bottom figure. We see that, due to exploration, the TCU is able to successfully identify the updated set-points.

The expected number of switchings over a time period can be estimated from the ratio of the exploration to exploitation time, the expected (average) changes in the temperature set-points, the ON/OFF lockout times and the ON/OFF switching probabilities broadcast to the TCL. Results are shown in Table D.1 and Figures 6.7, 6.8 , 6.9, D.2, D.3 and D.1.

Key observations include,

- The TCL's set-points can be sought successfully following the proposed exploration-exploitation strategy (shown in Fig. 6.7, and also in Figures D.2, D.3 and D.1). Lockouts have also been taken into account.

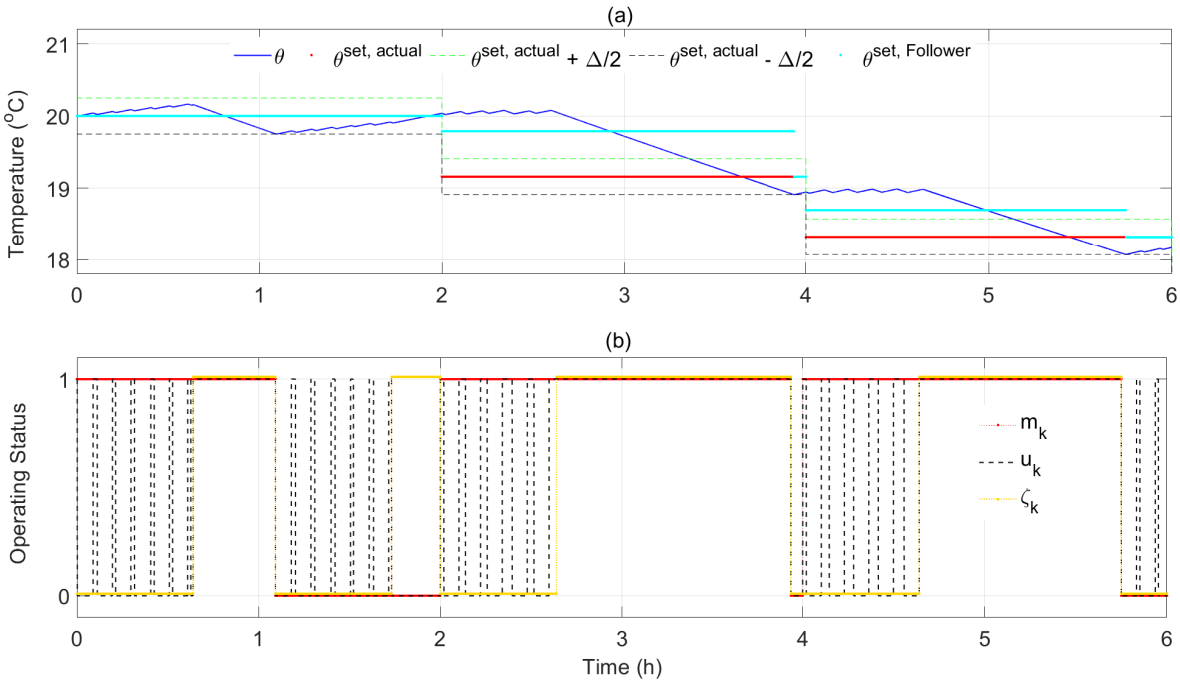

Figure 6.7: (a) Temperature and set-point evolution, (b) Thermostat's internal cooling/ off mode and TCU's on/off/explore modes.

- By adjusting the threshold time to initiate exploration, the fraction of time spent in exploration versus performing tracking can be chosen (shown in Fig. 6.8).
- The average number of TCL switchings (per hour) reduces with increase in the time spent in the exploration phase (shown in Fig. 6.9). Longer time spent in exploitation (/ tracking) phase however can lead to the TCU's temperature set-points remaining away from the actuals on a more frequent basis.
- By increasing the switching probabilities during tracking, the average number of switching (per hour) increases.
- Decreasing the lockout-time increases the average number of switching (per hour).

6.7.3. Aggregate Modeling using Extended State-Bin Model

The availability of the state variables and equations allow us to develop various optimization and control problem formulations. As discussed before, state-bin models, proposed

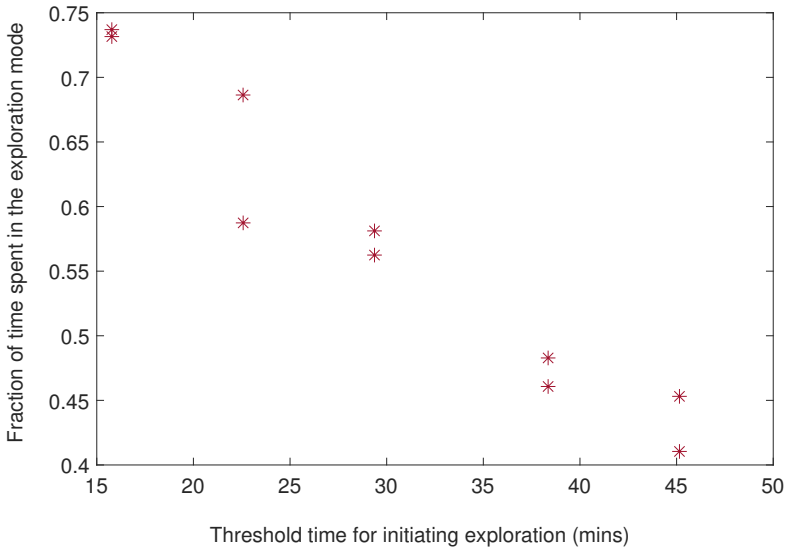

Figure 6.8: Fraction of time spent in exploration due to varying threshold times to initiate exploration.

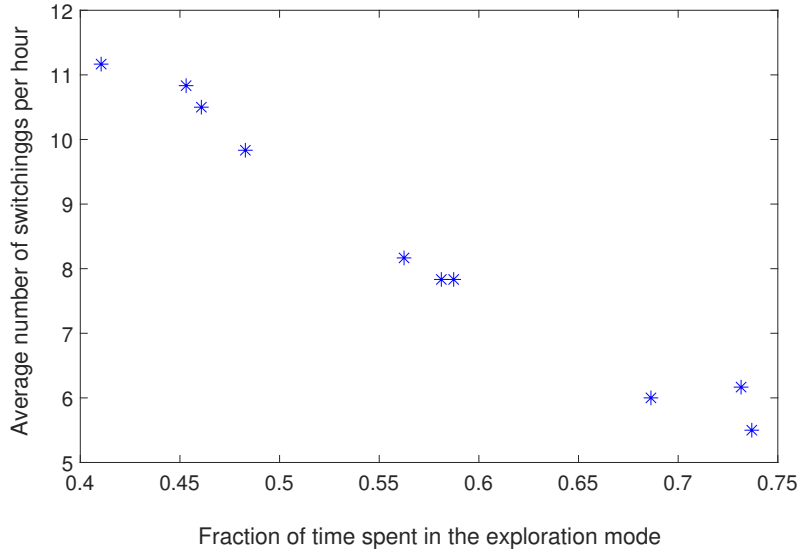

Figure 6.9: Average number of TCL switchings versus fraction of time spent in exploration.

and applied in various recent studies [11, 51, 55, 70] can be modified to incorporate the additional lock-out and explore states.

One approach is to consider a wider range than the current deadband, i.e., bins outside $\theta^{\min}, \theta^{\max}$ of current set-point (see Fig. 6.10). Then, incorporate, per ON/OFF bin, the additional modes of operations, i.e. lockout and explore. This is illustrated in Fig. 6.10. Note that bin-models with timer-locking mechanism have also been considered in [97, 104], and in our recent work [76]. Similarly, the state space and the system matrices can be extended to incorporate the exploration mode. This has not been elaborated further in this dissertation, but is a topic of future interest.

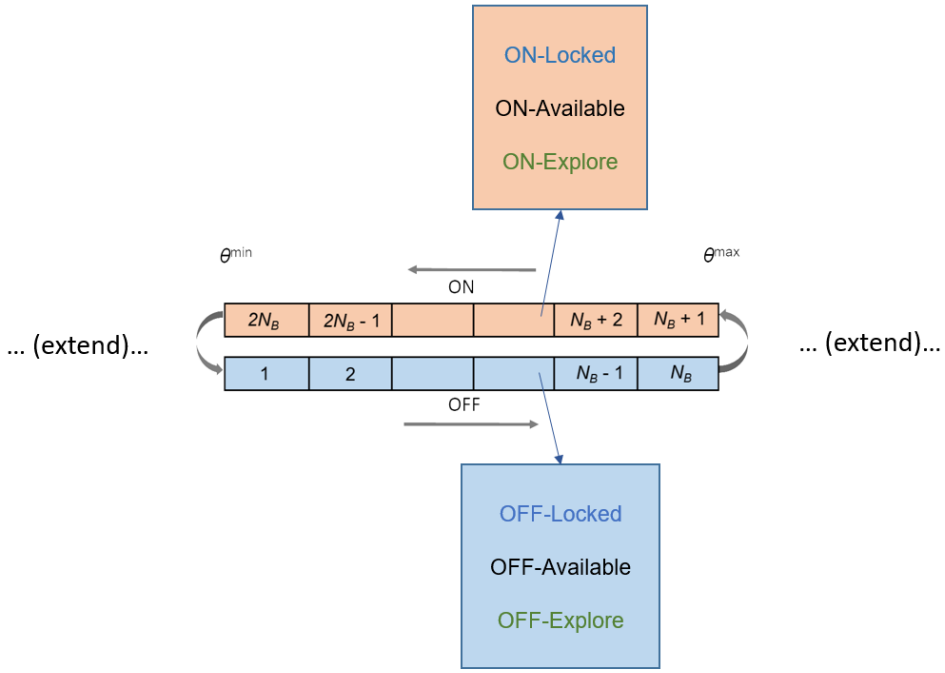

Figure 6.10: Extended bin model with lockouts and exploration modes.

6.8. Chapter Conclusions

In this chapter, we showed that both analytical and simulation-based results provide important insights regarding the performance of various TCL control schemes. We established performance limits for population of TCLs that are controlled via probabilistic dispatch. An analytical expression for expected RMS error due to probabilistic switching was first developed for a population of TCLs, all receiving the same switching probability. Then we extended our analysis to aggregate bin-based models for which switching probabilities

may be non-uniform. The close agreement between analytical and simulation-based results demonstrated the validity of the performance bounds. It was shown that the RMS error in the fraction of TCLs switched varies with target levels as well as population size. Then, by using variance reduction techniques, new controllers have been developed to find the optimal switching probabilities for each bin. In the scenario with a triangular reference signal, the minimum variance controller obtained 1.5-2 times lower RMS error than the uniform dispatch. The analysis and techniques used in this study can be used to estimate the least attainable RMS tracking error, in expectation, for a given reference signal and under a specific control law.

Next, to deal with the issue of unknown set-points, we introduced and tested an Explore-Exploit controller. We showed that the smart thermostat vendor or its users then do not need to share the TCL set-points. Using 'exploration-exploitation'-type strategies allows us to provide frequency regulation without potentially causing discomfort/inconvenience. We demonstrated the tradeoffs between exploration and exploitation phases mainly by identifying key parameters that affect the performance of each phase. The learning algorithm (or the user's set-point schedule) can remain unaffected while applying the exploration-exploitation algorithm, thus using smart thermostats for frequency regulation can be offered seamlessly as an add-on service by demand response aggregators.

Chapter 7.

Aggregation of Flexibility from DERs using Polytopes

7.1. Chapter Introduction

Power systems are in the process of accommodating an increased amount of distributed energy resources (DERs) – solar photovoltaic (PV) systems, energy storage systems, and controllable demand-side resources just to mention a few. As outlined in Chapter 1, the flexibility from DERs can be leveraged to alleviate a number of operational challenges in the power grid – for example, to address voltage regulation issues – and to aid system-level operations by realizing the emerging vision of virtual power plants. To address this, key is to characterize the aggregate flexibility from DERs [2, 10, 13, 42, 52, 74, 75, 106].

A general framework for characterizing DER flexibility is presented in [39], where methods to compute the aggregate flexibility using the *Minkowski sum (M-sum)* are also described. *M-sum*, by definition, is a set-theoretic approach. A set that contains all feasible operating points of a DER can be called its *flexibility set*. For example, for two DERs, let $\mathcal{X}_1$ and $\mathcal{X}_2$ represent their flexibility sets. Then, the aggregate flexibility set is given by the M-sum as follows,

$$\mathcal{X}_1 \oplus \mathcal{X}_2 = \{x_1 + x_2 | x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\}.$$

This is illustrated by an example in Fig. 7.1.

M-sum can be computed accurately by summing all the vertices of given polytopes [62, 91]; however, such approaches are not computationally feasible due to the exponential

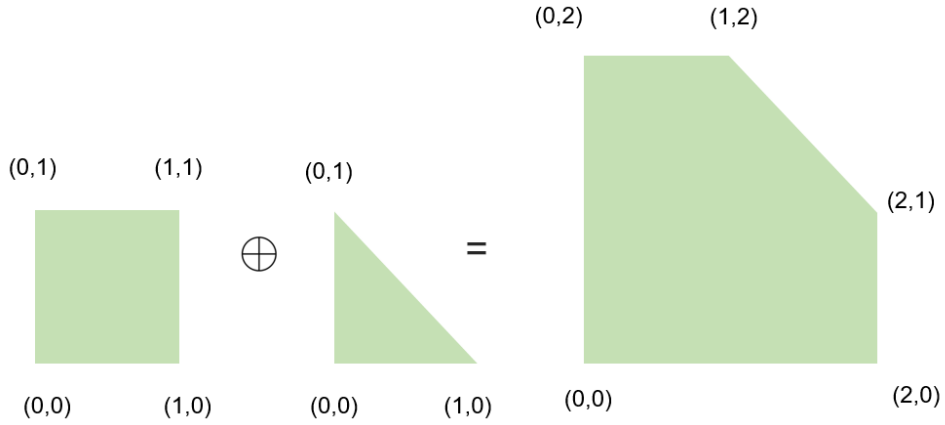

Figure 7.1: M-sum of two sets.

growth in complexity for large number of devices [10, 74, 101]. Hence, several works in the literature sought efficient algorithms to compute the M-sum. In [10], the authors provide an algorithm to compute an outer approximation of the M-sum; however, outer approximations might include infeasible points, which is undesirable especially if utilized in optimization settings. The authors in [74] present an algorithm to compute an inner approximation of the M-sum by using *zonotopes*. A zonotope is a convex polytope with some special properties: (i) its symmetric about an origin, and (ii) it is representable as the M-sum of finite line segments [31, 74]. A simple example of a zonotope is an M-dimensional unit hyper-cube.

Zonotopes have also been used for computing M-sum widely in literature due to their features which allow easily summing them to obtain the M-sum [3, 31, 74, 75]. One major limitation is that zonotopes are centrally symmetric polytopes, hence if original resource polytopes are not symmetric (as will be shown for the inverter case), approximating those using single symmetric polytopes might lead to a conservative estimate of the aggregate flexibility.

As an alternative to zonotopes, the authors in [106] present an algorithm to compute both inner and outer approximations of M-sum by using *homothets*. Homothets are convex sets that are dilation and/or translation of a prototype set (a given/ user-defined convex set). Given a prototype set, it can be scaled and translated to fit inside (for inner approximation) or just outside (for outer approximation) of given resource polytopes. However, choosing an appropriate prototype for a heterogeneous population is challenging and arbitrary choices

may lead to conservative estimates of the aggregate flexibility.

This chapter aims to extend zonotope- and homothet-based approaches for computing M-sums. The focus is on finding inner approximations of the M-sum so that the feasibility of control actions is guaranteed. The flexibility provided by inverter-interfaced devices as well as controllable loads is considered. To that end, a polytopic representation of the feasible operating region of an inverter is presented first. For certain special cases, we provide analytical expressions for the inner approximation of the M-sum by leveraging homothet-based representations.

However, as the level of heterogeneity increases, these analytical expressions might provide highly conservative estimates. Therefore, we propose to employ multiple homothets (here, axis-aligned boxes, which are also zonotopes by default [74]) per device polytope, and show how to efficiently perform the M-sum computation.

Our approach consists of: (i) a decomposition procedure to find a number of homothet-based sub-polytopes per device polytope; and (ii) performing the M-sum computation from the union of such sub-polytopes. We provide asymptotic guarantees on the accuracy of the approximation, which is generally difficult to achieve for non-vertex based M-sum algorithms. Since the number of unions grows exponentially [95, 101] with the number of devices and the number of sub-polytopes per device, techniques to limit the computational complexity of the methods are explored. The accuracy versus complexity trade-offs are investigated.

The rest of the chapter is organized as follows. Section 7.2 provides an overview of flexible operating region of inverter-interfaced devices and controllable loads, such as pumps, variable speed drives, electric vehicles (EVs) and thermostatically controlled loads (TCLs). A discretization technique to obtain a convex flexibility polytope is also presented. Section 7.3 describes homothet-based approach to obtain the M-sum and proposes simple analytical expressions for its inner approximation. Section 7.4 presents a union-based M-sum algorithm, along with homothet-based decomposition technique. Section 7.5 illustrates the effectiveness and accuracy of our techniques through numerical results. Finally, Section 7.6 presents a use case where we solve an optimal power flow problem subject to PV inverters' flexibility polytope.

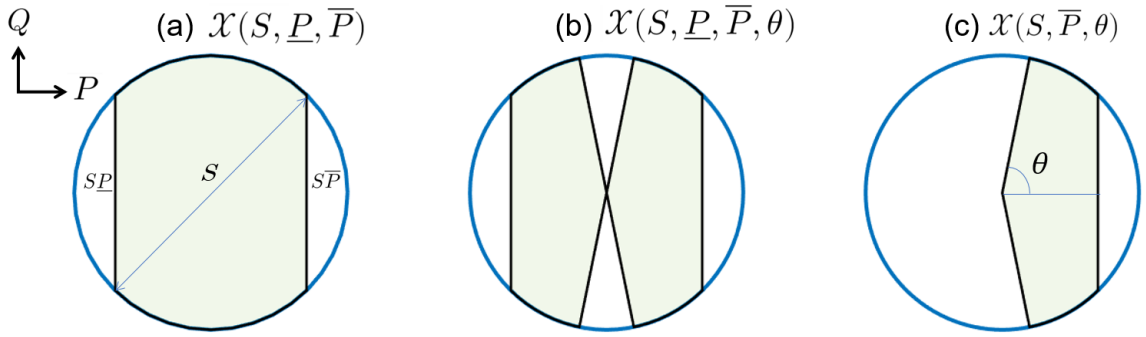

Figure 7.2: Feasible sets of inverters.

7.2. Flexibility Characterization

Before performing the M-sum for a large population of DERs, we first need to characterize the flexibility sets of inverter-based devices and controllable loads.

7.2.1. PV-Inverter Flexibility Set

Let $\mathcal{X} \subseteq \mathbb{R}^2$ be the set that contains the inverter's real and reactive power operating points, $x = [P, Q]^T, x \in \mathbb{R}^2$. Then, $\mathcal{X}$ can be written as (see, e.g., [22])

$$\mathcal{X}(S, \underline{P}, \overline{P}) = \{(P, Q) : S\underline{P} \leq P \leq S\overline{P}, Q^2 \leq S^2 - P^2\}. \tag{7.1}$$

Here, for a PV system, S is the apparent power rating of the inverter, $\underline{P} = 0$ and $\overline{P}$ (normalized w.r.t S) is the available power based on solar irradiance. For a storage device interfaced with an inverter, S is the inverter's rating; $\underline{P}$ and $\overline{P}$ (normalized w.r.t S) are the minimum and maximum real power available at a specific time. Note that $\mathcal{X}(S, \underline{P}, \overline{P})$ is a convex set. Additionally, to enforce a minimum power factor of $\cos(\theta)$, the following constraint can be included:

$$|Q| \leq \tan(\theta) P; \tag{7.2}$$

let $\mathcal{X}(S, \underline{P}, \overline{P}, \theta)$ denote the resulting set, and note that it is generally non-convex. However, for PV systems, since $\underline{P} = 0$, it is convex; with a slight abuse of notation we let $\mathcal{X}(S, \overline{P}, \theta)$ denote this latter set. Fig. 7.2 illustrates these feasible sets - $\mathcal{X}(S, \underline{P}, \overline{P})$ in (a),

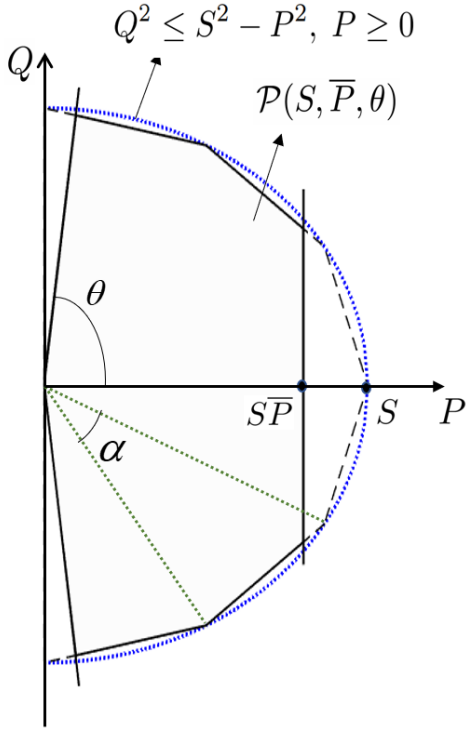

Figure 7.3: Flexibility polytope for a photovoltaic inverter

$\mathcal{X}(S, \underline{P}, \overline{P}, \theta)$ in (b) and $\mathcal{X}(S, \overline{P}, \theta)$ in (c). For the rest of this chapter, we will only focus on the convex cases (a) and (c).

7.2.2. Inverter Flexibility Polytope

Polytopes can generally be expressed with vertices (V-rep) or with half-space constraints (H-rep). The H-rep is useful for optimization purposes [10, 106]. Hence, a polytopic representation of $\mathcal{X}$ will be developed in this section.

Definition 7.1. *Let* $\mathcal{P} = \{Ax \leq b\}$*, where* $x = [P, Q]^T \in \mathbb{R}^2$*,* $A \in \mathbb{R}^{m \times 2}$*, and* $b \in \mathbb{R}^m$*. If* $\mathcal{P} \subseteq \mathcal{X}$*, then* $\mathcal{P}$ *is an* inner approximation *to* $\mathcal{X}$*.*

To obtain A, b, first we inscribe an N-sided polygon inside the circle $P^2 + Q^2 = S^2$. Assume N is even and $N \geq 4$. The angle formed between two successive vertices of the polygon can be found as, $\alpha = \frac{2\pi}{N}$.

The set of vertices $\{(P_j, Q_j)\}_{j=1}^{N}$ can be found as

$$P_j = S\cos((j-1)\alpha), \quad j = 1, 2, ..., N, \tag{7.3}$$

$$Q_j = S\sin((j-1)\alpha), \quad j = 1, 2, ..., N. \tag{7.4}$$

Additionally, by convention, $P_{N+1} = P_1$ and $Q_{N+1} = Q_1$. From these vertices, the slopes m_j can be computed as,

$$m_j = \frac{(Q_{j+1} - Q_j)}{(P_{j+1} - P_j)}, \quad j = 1, 2, ..., N. \tag{7.5}$$

Then, the constraint set for the H-rep of $Q^2 \leq S^2 - P^2$ can be obtained as,

$$(Q - Q_j) \leq m_j(P - P_j), \quad j = 1, 2, ..., \frac{N}{2}, \tag{7.6}$$

$$-(Q - Q_j) \leq -m_j(P - P_j), \quad j = \frac{N}{2} + 1, ..., N. \tag{7.7}$$

Overall, for $\mathcal{X}(S, \underline{P}, \overline{P})$, we obtain the polytope

$$\mathcal{P}(S, \underline{P}, \overline{P}) = \{(P, Q) : S\underline{P} \leq P \leq S\overline{P}, \text{ (7.6) and (7.7)}\}$$

Similarly, for $\mathcal{X}(S, \overline{P}, \theta)$, we obtain

$$\mathcal{P}(S, \overline{P}, \theta) = \{(P, Q) : 0 \leq P \leq S\overline{P}, \text{ (7.2), (7.6) and (7.7)}\}$$

The H-rep of $Q^2 \leq S^2 - P^2$ and the resulting inverter polytope, $\mathcal{P}(S, \overline{P}, \theta)$, are shown in Fig. 7.3. In this case, since $\mathcal{P}(S, \overline{P}, \theta)$ spans two-quadrants, only the positive half-circle needs to be considered.

Note that, by construction, $\mathcal{P}$ is a convex polytope and for any finite N, $\mathcal{P} \subset \mathcal{X}$. Let $A_{\mathcal{P}}$ denote the area of $\mathcal{P}$ and $A_{\mathcal{X}}$ denote the area of the entire feasibility set $\mathcal{X}$. Then, the area ratio can be defined as $\eta := \frac{A_{\mathcal{P}}}{A_{\mathcal{X}}}$.

Proposition 7.1. *Consider $\mathcal{X}(S, \underline{P}, \overline{P})$, with $\underline{P} = -1$, $\overline{P} = 1$, or $\mathcal{X}(S, \overline{P}, \theta)$ with $\overline{P} = 1$ and $\theta = \pi/2$. Then, $\eta = \frac{sin\alpha}{\alpha}$, where $\alpha = 2\pi/N$ and $N \geq 4$, and as $N \to \infty$, $\eta \to 1$.*

Proof: Consider $\mathcal{X}(S, \underline{P}, \overline{P})$. Assume $\underline{P} = -1$ and $\overline{P} = 1$. Then, the area ratio, η, can be

expressed as,

$$\eta = \frac{A_{\mathcal{P}}}{A_{\mathcal{X}}} = \frac{0.5NS^2\sin(2\pi/N)}{\pi S^2} = \frac{sin(\alpha)}{\alpha}. \tag{7.8}$$

By application of the L'Hopital's rule, $\lim_{\alpha\to 0}\frac{sin(\alpha)}{\alpha} = \lim_{\alpha\to 0} cos(\alpha) = 1$. Therefore, $\eta \to 1$ as $\alpha \to 0$ (i.e. $N \to \infty$).

When considering $\mathcal{X}(S, \overline{P}, \theta)$, with $\overline{P} = 1, \theta = \pi/2$, the third fraction in (7.8) will additionally have 0.5 multiplied both in numerator and denominator (due to half-circle), which cancel out. Thus, the ratio $\eta = \frac{sin(\alpha)}{\alpha}$ holds, and again $\eta \to 1$ as $\alpha \to 0$. ■

Proposition 7.2. *Consider $\mathcal{X}(S, \underline{P}, \overline{P})$, with $\underline{P} > -1$, $\overline{P} < 1$, or $\mathcal{X}(S, \overline{P}, \theta)$ with $\overline{P} < 1$ and $0 \leq \theta < \pi/2$. Then, for N sufficiently large, $\eta \approx 1$.*

Proof: Without loss of generality, assume $S = 1$, as shown in Fig. 7.4. α is the angle formed by fitting an N-sided polygon inside $P^2 + Q^2 = 1$. Thus, the area of BCED = (area of sector OBDE - area of ΔOBE), is the approximation error.

The area of OBDE = $\frac{\alpha}{2}$ and the area of ΔOBE = $\frac{\sin\alpha}{2}$. By inclusion of ΔOBE inside region OBDE, $\frac{\alpha}{2} \geq \frac{\sin\alpha}{2}$.

However, for α sufficiently small ($\alpha << 1$ rad), $\alpha \approx \sin\alpha$ (small angle approximation). Thus, $\frac{\alpha}{2} \approx \frac{\sin\alpha}{2}$, i.e. area of BCED ≈ 0.

Consider $\underline{P} = -1$ and $\overline{P} < 1$. Assume $\overline{P}$ lies between A and B. Now, the area of CED is the approximation error when the sector OBE is approximated by ΔOBE and constraint $P \leq \overline{P}$ is added. For $\overline{P} < 1$, area of CED < the area of BCED. Hence, with α sufficiently

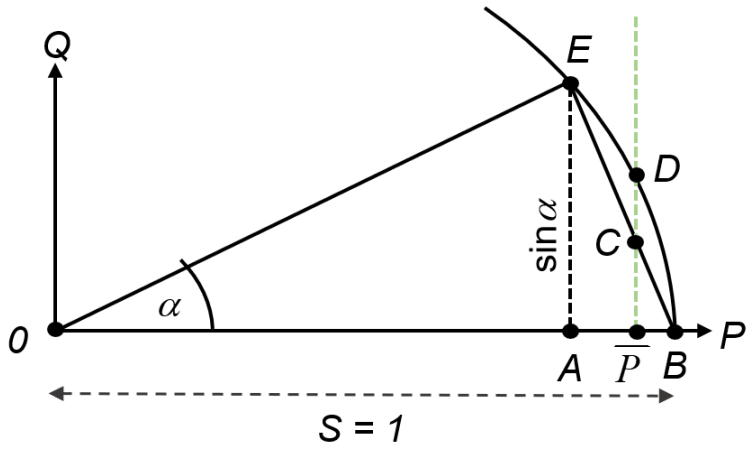

Figure 7.4: Discretization of circle and adding constraint.

small, area of BCED ≈ 0, hence, the area of CED ≈ 0. We can show the same considering $\underline{P} > -1$. Combining these, we obtain, $A_{\mathcal{P}} \approx A_{\mathcal{X}}$, i.e. $\eta \approx 1$, for N sufficiently large.

It is also easy to verify that the same holds when considering $\mathcal{X}(S, \overline{P}, \theta)$, with $\overline{P} < 1, 0 \leq \theta < \pi/2$. ■

Applying Proposition 7.1 with $N = 6$ yields $\eta = 0.83$; with $N = 12$ yields $\eta = 0.95$; and with $N = 24$ yields $\eta = 0.99$. Hence, $N = 24$ will be used in our simulations below.

7.2.3. Flexibility from Controllable Loads

We also consider controllable loads, such as variable speed drives and pool pumps. Their feasible set $\mathcal{X}$, considering only real power, can simply be written as

$$\mathcal{X}(\underline{P_{\mathrm{L}}}, \overline{P_{\mathrm{L}}}) = \{P_{\mathrm{L}} : \quad \underline{P_{\mathrm{L}}} \leq P_{\mathrm{L}} \leq \overline{P_{\mathrm{L}}}\}. \tag{7.9}$$

where P_{L} is denotes the real power consumed (kW) by the load, with minimum and maximum quantities given by $\underline{P_{\mathrm{L}}}$ is the minimum, and $\overline{P_{\mathrm{L}}}$.

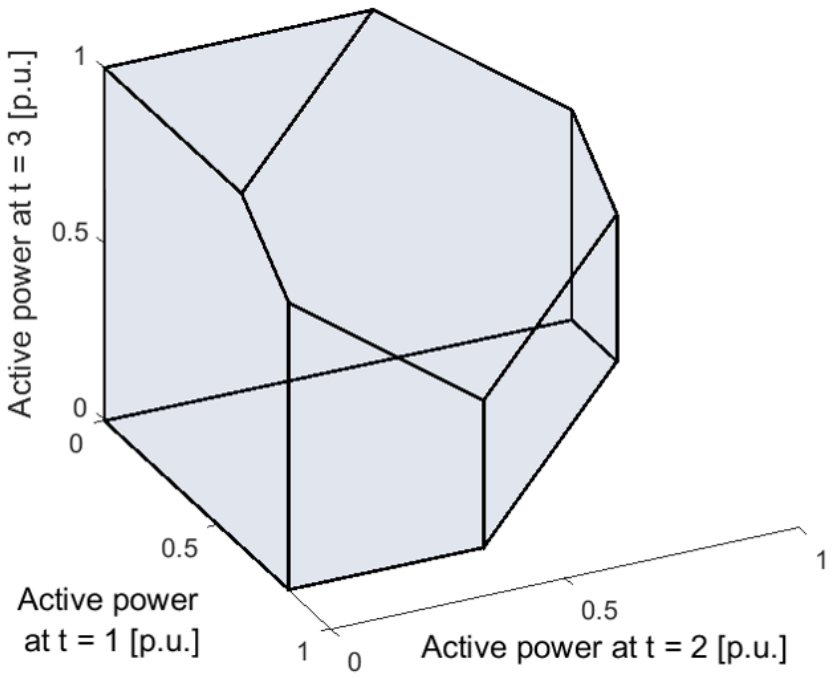

Figure 7.5: Flexibility polytope of a storage-like load (e.g. air-conditioner, EV).

For storage-like DERs, such as TCLs and EVs, the internal states (e.g. temperature, state of charge) also evolve with time. Consider M time-intervals, indexed by $k = 1, 2, ..., M$. Let e_k be the DER's normalized energy state, i.e. its state of charge (SOC), and $P_{\mathrm{L}k}$ be the real power consumed by the DER at time k. Then, using a generalized battery model, the dynamics of e_k can expressed as, $e_{k+1} = ae_k + \gamma P_{\mathrm{L}k}$, where $a \in (0, 1]$ is the energy dissipation rate, and γ is the charging efficiency [39, 106]. Knowing the initial SOC, e_o, the overall feasible set $\mathcal{X}$ can be written as [10],

$$\mathcal{X}(\underline{P_{\mathrm{L}}}, \overline{P_{\mathrm{L}}}, a, \gamma, e_0) = \Big\{[P_{\mathrm{L}k}]^{\mathrm{T}} : \underline{P_{\mathrm{L}}} \leq P_{\mathrm{L}k} \leq \overline{P_{\mathrm{L}}}, \\ 0 \leq a^k e_o + \sum_{t=1}^{k} a^{k-t} \gamma P_{\mathrm{L}(t-1)} \leq 1, = 1, 2, ..., M\Big\}. \tag{7.10}$$

with $x = [P_{\mathrm{L}k}]^{\mathrm{T}} \in \mathbb{R}^M$. Figure 7.5 presents an illustration of the flexibility polytope of a storage-like load. Refer to [10, 106] for more details on (7.10) and on how to include discharging (e.g. vehicle to grid applications) and ramp constraints.

7.3. Aggregation by Minkowski Sum

In this section, we describe how the flexibility from DERs of the same type can be aggregated using M-sum. Consider a population of n_d devices, with indices $i = 1, 2, ..., n_d$. Let $\mathcal{X}_i$ denote the feasible set of device i. The aggregate flexibility, $\mathcal{X}^{\mathrm{Aggr}}$, can be found by computing the M-sum of $\mathcal{X}_i$s as,

$$\mathcal{X}^{\mathrm{Aggr}} := \mathcal{X}_1 \oplus \mathcal{X}_2 \oplus ... \oplus \mathcal{X}_{n_d} = \oplus_{i=1}^{n_d} \mathcal{X}_i. \tag{7.11}$$

where $\oplus$ denotes the M-sum. For computation of M-sum, applying (7.11) is, however, not efficient, especially when n_d is large [101]. Hence, zonotopes [31, 74] and homothet-based [91, 106] approaches have been shown to be useful. The applicability of homothets for our case will be presented next.

7.3.1. Homothets and Minkowski Sum

Given a compact convex set $\mathcal{X}_0$, $\beta_i \mathcal{X}_0 + t_i := \{x \in \mathbb{R}^2 : x = \beta_i \zeta + t_i, \zeta \in \mathcal{X}_0\}$ is a homothet of $\mathcal{X}_0$, where $\beta_i > 0$ is a scaling factor and t_i is a translation factor [91]. $\mathcal{X}_0$ can

be referred to as a prototype set [106].

Homothets are useful for computation of M-sums due to the following property [91, 106],

$$\oplus_{i=1}^{n_d} (\beta_i \mathcal{X}_0 + t_i) = \sum_{i=1}^{n_d} \beta_i \mathcal{X}_0 + \sum_{i=1}^{n_d} t_i. \tag{7.12}$$

Hence, if $(\beta_i \mathcal{X}_0 + t_i) \subseteq \mathcal{X}_i$, then,

$$\oplus_{i=1}^{n_d} (\beta_i \mathcal{X}_0 + t_i) \subseteq \oplus_{i=1}^{n_d} \mathcal{X}_i. \tag{7.13}$$

For example, consider $\mathcal{X}_i(\underline{P_{\mathrm{L}i}}, \overline{P_{\mathrm{L}i}})$, defined in (7.9). Take $\mathcal{X}_0 = \mathcal{X}(0, 1)$. We need to find β_is and t_is for all $\mathcal{X}_i$. Assuming $\underline{P_{\mathrm{L}i}}, \overline{P_{\mathrm{L}i}} \geq 0$, we see that $\mathcal{X}_i(\underline{P_{\mathrm{L}i}}, \overline{P_{\mathrm{L}i}}) = \beta_i \mathcal{X}_0 + t_i$ with $\beta_i = (\overline{P_{\mathrm{L}i}} - \underline{P_{\mathrm{L}i}})$ and $t_i = \underline{P_{\mathrm{L}i}}$, for each $\mathcal{X}_i$. For example, for $\mathcal{X}_i(3, 10)$, $t_1 = 3$ and $\beta_1 = 7$. Then, applying (7.12), the aggregate flexibility from n_d controllable loads, is simply given by, $\mathcal{X}^{\mathrm{Aggr}} = \sum_{i=1}^{n_d} \beta_i \mathcal{X}_0 + \sum_{i=1}^{n_d} t_i = \mathcal{X}(\sum_{i=1}^{n_d} \underline{P_{\mathrm{L}i}}, \sum_{i=1}^{n_d} \overline{P_{\mathrm{L}i}})$.

7.3.2. Special Cases: Aggregate Flexibility from Inverter-interfaced Devices

Under certain conditions, the properties (7.12) and (7.13) of homothets lead to simple analytical expressions for the M-sum of inverter-interfaced devices.

For example, assume only the rated power of the inverters vary, while $\underline{P_i}$ and $\overline{P_i}$ are homogeneous. This situation can appear commonly when a collection of inverters have different ratings, S_i, but undergo similar solar irradiance conditions, which could be due to their geographic proximity. Their aggregate flexibility can be obtained by Theorem 7.1.

Theorem 7.1. *Consider $\mathcal{X}_i(S_i, \underline{P_i}, \overline{P_i})$, where $\underline{P_i} = \underline{P_0}$ and $\overline{P_i} = \overline{P_0}$ for $i = 1, ..., n_d$. The aggregate flexibility set is then given by*

$$\oplus_{i=1}^{n_d} \mathcal{X}_i(S_i, \underline{P_i}, \overline{P_i}) = \mathcal{X}\left(\sum_{i=1}^{n_d} S_i, \underline{P_0}, \overline{P_0}\right).$$

Proof: For two inverters, assume $S_1 \neq S_2$, $\underline{P_1} = \underline{P_2} = \underline{P_0}$ and $\overline{P_1} = \overline{P_2} = \overline{P_0}$. Let $m_{2,1} > 0$ be the ratio of the rated powers of the two inverters, i.e. $m_{2,1} = S_2/S_1$. Choose prototype set, $\mathcal{X}_0 = \mathcal{X}_1$. Then, $\mathcal{X}_2 = m_{2,1} \mathcal{X}_0$. By (7.13), we can write,

$$(\mathcal{X}_1 \oplus \mathcal{X}_2) = (1 + m_{2,1})\mathcal{X}_0.$$

Similarly for n_d devices, with $\mathcal{X}_0 = \mathcal{X}_1$ we obtain,

$$\mathcal{X}^{\text{Aggr}} = (1 + m_{2,1} + ... + m_{n_d,1})\mathcal{X}_0.$$

where $m_{i,1} = S_i/S_1, i = 2, ..., n_d$.

Thus, substituting $m_{i,1}$, we verify that $\mathcal{X}(\sum_{i=1}^{n_d} S_i, \underline{P_0}, \overline{P_0}) = \mathcal{X}^{\text{Aggr}}$. ■

Corollary 7.1. *Consider $\mathcal{X}_i(S_i, \overline{P_i}, \theta_i)$, where $\overline{P_i} = \overline{P_0}$ and $\theta_i = \theta_0$ for $i = 1, ..., n_d$. The aggregate flexibility set is given by $\mathcal{X}(\sum_{i=1}^{n_d} S_i, \overline{P_0}, \theta_0)$.*

The proof is similar to the proof of Theorem 7.1.

Next, consider heterogeneous S_i, $\overline{P_i}$ and θ_i for $\mathcal{X}_i(S_i, \overline{P_i}, \theta_i)$. In this case, Theorem 7.2 applies.

Theorem 7.2. *Consider $\mathcal{X}_i(S_i, \overline{P_i}, \theta_i)$, $i = 1, ..., n_d$, where $S_i, \overline{P_i}, \theta_i$ are heterogeneous. Let $S_0 = \min_i S_i$, $\overline{P_0} = \min_i \overline{P_i}$, and $\theta_0 = \min_i \theta_i$ $(0 \le \theta_i \le \pi/2)$. Then,*

$$\mathcal{X}(n_d S_0, \overline{P_0}, \theta_0) \subseteq \oplus_{i=1}^{n_d} \mathcal{X}_i(S_i, \overline{P_i}, \theta_i).$$

Moreover, strict equality holds if and only if all the parameters are homogeneous.

Proof: Consider n_d devices with feasible sets $\mathcal{X}_i(S_i, \overline{P_i}, \theta_i)$. Let $S_0 = \min_i (S_i), \overline{P_0} = \min_i (\overline{P_i},), \theta_0 = \min_i (\theta_i)$ and obtain $\mathcal{X}_0(S_0, \overline{P_i}, \theta_0)$. In this case, we obtain $\mathcal{X}_0 = \mathcal{X}_1 \cap \mathcal{X}_2 \cap ... \cap \mathcal{X}_{n_d}$. Hence, $\mathcal{X}_0 \subseteq \mathcal{X}_i, \forall i$. Approximate every $\mathcal{X}_i$ by $\mathcal{X}_0$, which gives $\beta_i = 1$, $t_i = 0$ in (7.13). Hence, $\sum_{i=1}^{n_d} \beta_i = n_d$. Therefore,

$$\mathcal{X}(n_d S_0, \overline{P_0}, \theta_0) = n_d \mathcal{X}_0 \subseteq \mathcal{X}^{\text{Aggr}}.$$

In the homogeneous case, $\mathcal{X}_0 = \mathcal{X}_i, \forall i$, thus strict equality holds. ■

Example 1. Take $n_d = 100$ inverters characterized by $\mathcal{X}(S_i, \overline{P_i}, \theta_i)$. Feasible sets of three heterogeneous inverters is shown in Fig. 7.6.

In the first case, consider, $S_i = 1, \theta_i = 1.45$ rad, $\forall i$. $\overline{P_i}$ is distributed uniformly between 0.75 and 1. By Theorem 7.2, using $\min_i \overline{P_i}$, the prototype set and the approximate M-sum were obtained. Next, we formed the polytopes for the n_d inverters using the procedure

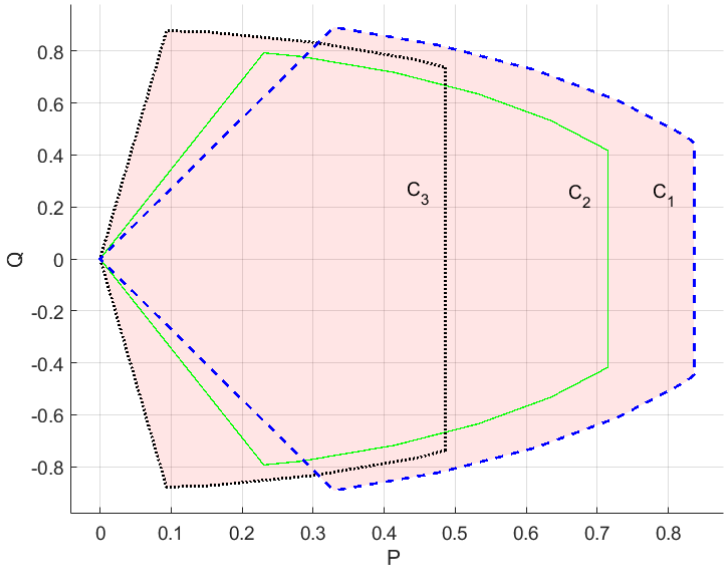

Figure 7.6: Feasible sets of three heterogeneous inverters.

described in section 7.2.2. Using the MPT toolbox [41], the actual M-sum and its area were computed. The ratio of the area of approximate aggregate M-sum polytope to the area of the true M-sum aggregate polytope was 0.90. Note that the MPT toolbox obtains M-sum by performing vertex sums. Hence, the computation time is very long (25 minutes). One the other hand, the M-sum computation using Theorem 7.2 is algebraic (with $\min_i \overline{P_i}$ known), hence is readily available and scalable to any population size.

Next, we repeated the above case study assuming that $\overline{P_i}$ is distributed uniformly between 0.5 to 1. In this case, the area ratio was found to be ≈ 0.71. Finally, a population was considered with all parameters being heterogeneous, S_i uniformly distributed within [0.75,1], $\overline{P_i}$ within [0.75,1] and θ_i within $[1.27,\pi/2]$ rad. In this case, the area ratio drastically reduced to 0.29. Thus, with increased level of heterogeneity, applying Theorem 7.2 may lead to very conservative estimates. ■

The accuracy of the M-sum approximated by using homothets depends on the choice of the prototype set. Typically, as the level of heterogeneity increases, the accuracy worsens considerably. Hence, in the next section, we present an approach to consider multiple homothets per $\mathcal{X}_i$ and show that the inner approximation of the M-sum can approach the true M-sum.

7.4. Union-based Minkowski Sum

For a heterogeneous population, the shapes of the flexibility sets may vary considerably. Hence, choosing a single prototype set $\mathcal{X}_0$ may be limiting and result in a conservative estimate of the M-sum. To address this, in this section, we show how to decompose each polytope into a union of homothetic sub-polytopes. The M-sum can then be computed by applying the distributivity property of M-sum, as elaborated in the next sub-section.

Our union-based approach can also be motivated by the optimization applications as follows. Given a collection of convex compact subsets $\{\mathcal{X}_\omega\}_{\omega=1}^{n_\Omega}$ of $\mathbb{R}^M$, consider the union $\mathcal{X} := \bigcup_{\omega=1}^{n_\Omega} \mathcal{X}_\omega$. Let $f : \mathbb{R}^M \to \mathbb{R}$ be a convex function, and consider the optimization problem,

$$\text{(P0)} \quad \min_{x \in \mathcal{X}} \; f(x). \tag{7.14}$$

It is clear that (P0) is equivalent to:

$$\text{(P1)} \quad \min_{\omega \in \{1,\ldots,n_\Omega\}} \; \min_{x \in \mathcal{X}_\omega} \; f(x). \tag{7.15}$$

In this case, if $\mathcal{X}$ represents the aggregate flexibility set, we avoid computing the overall M-sum; instead, we find the optimal solution from the candidate solutions obtained from solving multiple sub-problems.

7.4.1. Distributivity Property of Minkowski Sum

Let each $\mathcal{X}_i$, the set for the i-th of the n_d DERs, be expressed by n_i sub-sets. Let $W_i = \Big\{(i,j) : j = 1, ..., n_i\Big\}$. Then,

$$\mathcal{X}_i = \cup_{\omega \in W_i} \mathcal{X}_\omega. \tag{7.16}$$

Also, let Ω be the Cartesian product of all W_i, i.e. $\Omega = W_1 \times ... \times W_{n_d} = \Big\{(\omega_1, ..., \omega_{n_d}) : \omega_i \in W_i, \forall i = 1, ..., n_d\Big\}$. Then, by the distributivity property of M-sum [91],

$$\oplus_{i=1}^{n_d} \mathcal{X}_i = \oplus_{i=1}^{n_d} \Big(\cup_{\omega \in W_i} \mathcal{X}_\omega \Big) = \cup_{(\omega_1,\ldots,\omega_{n_d}) \in \Omega} \Big(\oplus_{i=1}^{n_d} \mathcal{X}_{\omega_i} \Big). \tag{7.17}$$

Because for each of the n_d DERs, one can choose from n_i sub-sets, the cardinality of Ω is $n_\Omega = \prod_{i=1}^{n_d} n_i$. Note that while (7.17) holds with equality, one may choose subsets $\bar{\Omega} \subseteq \Omega$ and obtain,

$$\cup_{(\omega_1,\ldots,\omega_{n_d})\in\bar{\Omega}} \left(\oplus_{i=1}^{n_d} \mathcal{X}_{\omega_i} \right) \subseteq \oplus_{i=1}^{n_d} \mathcal{X}_i, \tag{7.18}$$

i.e. an inner approximation of $\oplus_{i=1}^{n_d} \mathcal{X}_i$ is obtained. For example, consider $\mathcal{X}_1$ and $\mathcal{X}_2$ and assume $n_1 = n_2 = 2$. Then, by (7.17),

$$\mathcal{X}_1 \oplus \mathcal{X}_2 = (\mathcal{X}_{(1,1)} \oplus \mathcal{X}_{(2,1)}) \cup (\mathcal{X}_{(1,1)} \oplus \mathcal{X}_{(2,2)}) \cup (\mathcal{X}_{(1,2)} \oplus \mathcal{X}_{(2,1)}) \cup (\mathcal{X}_{(1,2)} \oplus \mathcal{X}_{(2,2)}).$$

Of course, $(\mathcal{X}_{(1,1)} \oplus \mathcal{X}_{(2,2)}) \subseteq (\mathcal{X}_1 \oplus \mathcal{X}_2)$.

Finally, expressing every subset of $\mathcal{X}_i$ as a homothet of the same prototype set $\mathcal{X}_0$, from (7.17) and (7.12) we obtain,

$$\cup_{(\omega_1,\ldots,\omega_{n_d})\in\Omega} \oplus_{i=1}^{n_d} (\beta_{\omega_i} \mathcal{X}_0 + b_{\omega_i}) \subseteq \oplus_{i=1}^{n_d} \mathcal{X}_i. \tag{7.19}$$

The challenges associated with union-based M-sum include: (1) Optimally partitioning a given polytope into convex sub-polytopes, (2) Analyzing the trade-offs between computational complexity and accuracy with increasing n_i and n_d. To efficiently handle these, a decomposition algorithm is proposed next.

7.4.2. Homothet-based Polytope Decomposition (HPD)

The key idea here is to decompose each of the given M-dimensional convex polytopes $\mathcal{P} := \{x : Ax \leq b\}$ into a number of homothets. Consider axis-aligned boxes. Let the lower and upper boundaries of a box in each axis be given by ${x_k}^-, {x_k}^+$, where $k = 1, ..., M$ and $x \in \mathbb{R}^M$. Here, ${x_k}^- \leq x_k \leq {x_k}^+$. Thus, an aligned box is denoted by $B(x^-, x^+)$ (or, succinctly by B).

To ensure we obtain homothets, define a prototype box, $\hat{B}^0$. The choice of $\hat{B}^0$ can be arbitrary; for example, one could consider a square in $\mathbb{R}^2$, or a hypercube in $\mathbb{R}^M$. In our case, we find $\hat{B}^0$ by solving for the largest volume box [99] that fits in a representative polytope, $\mathcal{P}^0$, chosen from the n_d given polytopes. Let the lengths of the edges of $\hat{B}^0$ be, $d_k^0 = (x_k^+ - x_k^-),\ k = 1, 2, ..., M$. Then, the ratios edge lengths w.r.t d_1^0 are, $r_{1,k}^0 = \frac{d_1^0}{d_k^0}$. $k = 2, ..., M$. We require all boxes, B, must be homothets of $\hat{B}^0$.

Given $\mathcal{P}$ and $\hat{B}^0$, in order to find a homothet, B, with maximum volume, such that $B \subseteq \mathcal{P}$, the following problem must be solved,

$$(P2) \max_{x^+,x^-} \prod_{k=1}^{M} (x_k^+ - x_k^-) \tag{7.20a}$$

$$\text{s.t. } A^+x^+ - A^-x^- \leq b, \tag{7.20b}$$

$$x_k^- \leq x_k^+, \quad k = 1, 2, ..., M, \tag{7.20c}$$

$$(x_1{}^+ - x_1{}^-) = r_{1,k}^0 (x_k{}^+ - x_k{}^-), \; k = 2, ..., M, \tag{7.20d}$$

where $A_{ij}^+ = \max\{0, A_{ij}\}$ and $A_{ij}^- = \max\{0, -A_{ij}\}$, with i, j being the row and column indices of A^+, A^- [99]. Note that the objective function (7.20a) can be replaced by $\sum_{k=1}^{M} \log(x^+ - x^-)$, which will be a convex problem [75]. Constraint (7.20d) ensures B will be a homothet of $\hat{B}^0$. Next, we show how (P2) can be used in a multi-stage algorithm for decomposing $\mathcal{P}$ into a number of homothets.

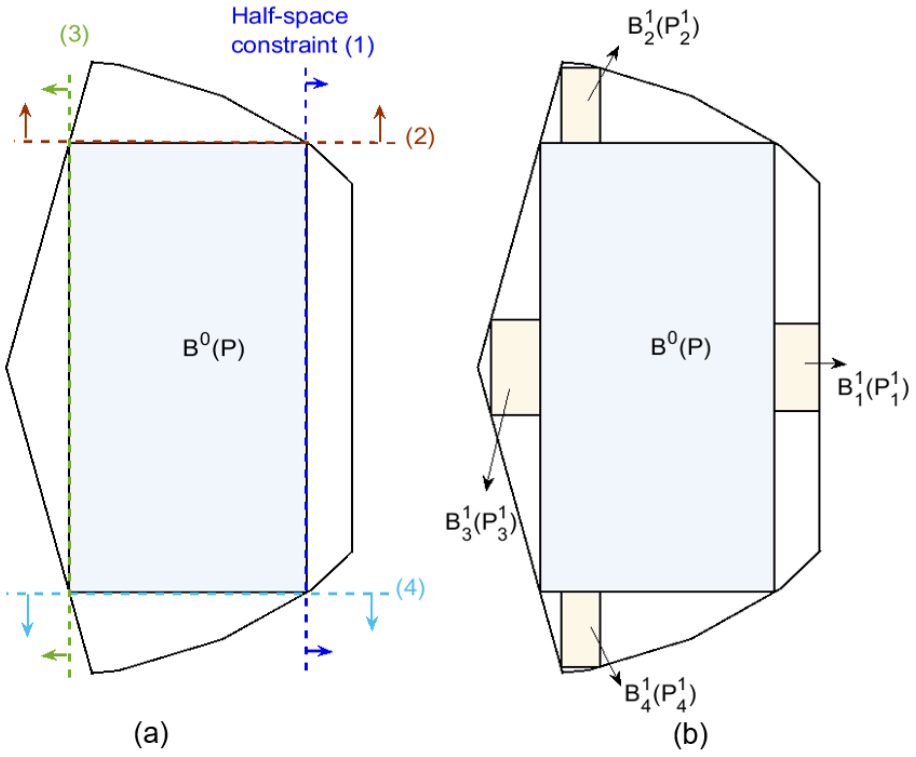

Figure 7.7: (a) Homothet-based polytope decomposition: $B^0(\mathcal{P})$ obtained in $s = 0$ and the half-space constraints originating from $B^0(\mathcal{P})$. (b) Decomposition after completion of stages, $s = 0$ and $s = 1$.

Let n_s represent the total number of stages of the proposed algorithm and $s = 0, 1, ..., n_s$ denote the stage index. At $s = 0$, (P2) is solved for a given polytope $\mathcal{P}$ to obtain $B^0(\mathcal{P})$, i.e. B is an outcome of the polytope. Here, $B^0(\mathcal{P}) = \beta_0 \hat{B}^0 + c_0$, i.e. a homothet of $\hat{B}^0$. Fig. 7.7(a) shows $B^0(\mathcal{P})$ inside $\mathcal{P}$.

Next, at $s = 1$, additional homothets will be sought in each region outside $B^0(\mathcal{P})$, but inside $\mathcal{P}$. In general, since each B lies in $\mathbb{R}^M$, it has $2M$ half-space constraints, indexed by $j = 1, ..., 2M$. As shown in Fig. 5.7, $B^0(\mathcal{P})$ is defined by four half-space inequalities in $\mathbb{R}^2$. Each region outside $B^0(\mathcal{P})$, but inside $\mathcal{P}$, can be defined using the half-space inequalities of $B^0(\mathcal{P})$, except the sign of the inequalities must be reversed, as illustrated in Fig. 5.7(a). By construction, each region outside $B^0(\mathcal{P})$, but inside $\mathcal{P}$ is convex and compact. Let $\mathcal{P}^s_\sigma$ denote an updated polytope at stage s, where $\sigma \in \mathbb{R}^s$ is a vector that contains the half-space indices for all the stages $1, ..., s$. To illustrate, $\mathcal{P}^1_{[j]}$ denotes the updated polytope obtained as the intersection of $\mathcal{P}$ and the half-space inequality j of $B^0(\mathcal{P})$. More generally,

$$\mathcal{P}^{s+1}_{[\sigma,j]} = \mathcal{P}^s_\sigma \cap \text{half-space inequality } j \text{ of } B^s_\sigma.$$

The solution of (P2) corresponding to $\mathcal{P}^s_\sigma$ is denoted $B^s_\sigma(\mathcal{P}^s_\sigma)$. The homothet-based polytope decomposition (HPD) concept is further illustrated in Fig. 5.7(b) with stages $s = 0$ and $s = 1$ solved. The decomposition can continue up to n_s stages. The algorithm is summarized below.

Algorithm (HPD):

S1. $s = 1$. Starting with $\mathcal{P}$, compute $B^0(\mathcal{P})$.

For $j = 1, ..., 2M$

$\sigma = [j]$

$\mathcal{P}^s_\sigma = \mathcal{P} \cap$ half-space inequality j of B^0

Compute $B^s_\sigma(\mathcal{P}^s_\sigma)$.

End

S2. For all $\sigma \in \{1, ..., 2M\}^s$

For $j = 1, ..., 2M$

$\mathcal{P}^{s+1}_{[\sigma,j]} = \mathcal{P}^s_\sigma \cap$ half-space inequality j of B^s_σ

Compute $B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]})$.

End

End

S3. $s = s + 1$. If $s < n_s$ repeat **S2**, else stop.

7.4.3. Convergence of Polytope Decomposition and M-Sum

Following the HPD algorithm, at every stage, new regions will be covered by solving (P2), unless the entire polytope has already been covered.

Observation 7.1. *If at the end of stage s, $(\mathcal{P} - \cup_{\forall s,\sigma} B^s_\sigma(\mathcal{P}^s_\sigma)) \neq \emptyset$, then at least for one constraint j of $B^s_\sigma(\mathcal{P}^s_\sigma$, $vol(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]}))) > 0$.*

Recall that $B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]})$ is a box obtained at stage s that follows from constraint j of $B^s_\sigma(P^s_\sigma)$. Then, the following holds.

Observation 7.2. $vol(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]}))) \leq vol(B^s_\sigma(P^s_\sigma))), \forall j$.

The above holds because $B^s_\sigma(P^s_\sigma)$ is the largest volume box inside P^s_σ obtained by (P2). Hence, $\text{vol}(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]}))) > \text{vol}(B^s_\sigma(P^s_\sigma)))$ would contradict (P2). After completion of the decomposition phase for the n_d polytopes, the approximate M-sum polytope, $\mathcal{P}^{\text{Aggr}}$, can be obtained using (7.19). Then, the following asymptotic result holds.

Proposition 7.3. $\mathcal{P}^{Aggr} \to \mathcal{X}^{Aggr}$ *as* $s \to \infty$.

Proof: To prove, first according to (7.11), we need to show, as $s \to \infty$, $\mathcal{P}_i \to \mathcal{X}_i$, where i denotes the index of the DER polytope. Dropping the subscript i for brevity, for an arbitrary DER polytope, assume $\text{vol}(\mathcal{P}) = \text{vol}(\cup_{\forall s,\sigma} B^s_\sigma(\mathcal{P}^s_\sigma)) + \delta$, where $\delta \in \mathbb{R}$. First consider $\delta < 0$. Then, $\text{vol}(\cup_{\forall s,\sigma} B^s_\sigma(\mathcal{P}^s_\sigma)) > \text{vol}(\mathcal{P})$, but this is a contradiction since $\cup_{\forall s,\sigma} B^s_\sigma(\mathcal{P}^s_\sigma) \subseteq \mathcal{P}$. Next, consider $\delta > 0$. This implies $(\mathcal{P} - \cup_{\forall s,\sigma} B^s_\sigma(\mathcal{P}^s_\sigma)) \neq \emptyset$, and for at least one constraint j of $B^s_\sigma(\mathcal{P}^s_\sigma)$, we can find a region $\mathcal{P}^{s+1}_{[\sigma,j]}$, where applying HPD results in $\text{vol}(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]})) > 0$. Let $(\text{vol}(\mathcal{P}^s_\sigma) - \text{vol}(\mathcal{P}^{s+1}_{[\sigma,j]})) \leq \epsilon$, where $\epsilon \in \mathbb{R}$ and $0 \leq \epsilon \leq \delta$. By Observation 7.1, $\text{vol}(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]})) \leq \text{vol}(B^s_\sigma(P^s_\sigma))$, $\forall j$. Also, by construction, $\text{vol}(\mathcal{P}^{s+1}_{[\sigma,j]}) \leq \text{vol}(\mathcal{P}^s_{[\sigma]}), \forall j$ and $\text{vol}(B^{s+1}_{[\sigma,j]}(\mathcal{P}^{s+1}_{[\sigma,j]})) \leq \text{vol}(\mathcal{P}^{s+1}_{[\sigma,j]}), \forall j$. Thus, as $s \to \infty$, $\epsilon \to 0$ and hence $\delta \to 0$.

Thus, $s \to \infty$, $\mathcal{P} \to \mathcal{X}$, and holds for any $\mathcal{P}_i$ and $\mathcal{X}_i$.

Therefore, from (7.19), it follows that, $s \to \infty$, $\mathcal{P}^{\text{Aggr}} \to \mathcal{X}^{\text{Aggr}}$. ■

Remark 7.1. *The HPD algorithm and the union-based M-sum computation procedure using the axis-aligned boxes are general to any dimension.*

7.4.4. Practical Considerations

While the proposed algorithm can guarantee asymptotic convergence to the true M-sum, considering a large number of sub-polytopes for n_d devices can be computationally challenging. Hence, a number of strategies can be considered.

The HPD algorithm, presented in section 7.4.2, stops after completing n_s stages. Alternatively, the stopping condition can be based on the volume of B^s_σ. After reaching a certain threshold, one can stop because all subsequent boxes will be smaller (by Observation 7.2).

As discussed earlier, for many applications, it could be sufficient to utilize (7.15). Hence, instead of computing the entire M-sum, one can use a set of candidate polytopes and solve (7.15). Section 7.5 provides detailed examples on how to efficiently choose such candidates. In $\mathbb{R}^2$, one can also consider computing the convex hull (C-hull) of the aggregate boxes to obtain a single M-sum approximation polytope, which will also be shown in section 7.5.1. Section 7.5 provides numerical examples, considering inverter polytopes in $\mathbb{R}^2$ and storage-like loads in $\mathbb{R}^6$, to analyze the performance of the proposed schemes and discusses trade-offs.

7.4.5. Simplification for Axis-Aligned Boxes

The choice of axis-aligned boxes leads to a further simplification due to applicability of interval analysis techniques [72]. For the i-th DER in $\mathbb{R}^M$, with M being the dimension of the space, define an interval $I(x^-_{i,k}, x^+_{i,k})$, $k = 1, 2, ...M$, such that $x_{i,k} \in I(x^-_{i,k}, x^+_{i,k})$ implies $x^-_{i,k} \leq x_{i,k} \leq x^+_{i,k}$. Then, $\sum_{i=1}^{n_d} x^-_{i,k} \leq x^{\text{Aggr}}_k \leq \sum_{i=1}^{n_d} x^+_{i,k}$, where $x^{\text{Aggr}}_k = \sum_{i=1}^{n_d} x_{i,k}$ for all $x_{i,k} \in I(x^-_{i,k}, x^+_{i,k})$, $i = 1, 2, ..., n_d$ and $k = 1, 2, ...M$. In $\mathbb{R}^M$, we obtain $B(\sum_{i=1}^{n_d} x^-_i, \sum_{i=1}^{n_d} x^+_i)$, where $x^-_i, x^+_i \in \mathbb{R}^M$. $B(\sum_{i=1}^{n_d} x^-_i, \sum_{i=1}^{n_d} x^+_i)$ is by default an inner approximation to $\mathcal{X}^{\text{Aggr}}$. The decomposition procedure described in HPD remains exactly the same, except, we can relax (7.20d). The convergence results discussed in section 7.4.3 trivially extend to the case of applying interval arithmetic on axis-aligned boxes.

7.5. Numerical Results

7.5.1. Performance of Union-based M-sum for Inverters

Consider four different inverters, $\mathcal{X}_i(S_i, \overline{P_i}, \theta_i), i = 1, 2, 3, 4$, with parameters (A) (1, 0.9,π/2 rad), (B) (1, 0.8, 1.37 rad), (C) (1, 0.6, 1.37 rad), and (D) (1, 0.3,π/2 rad).

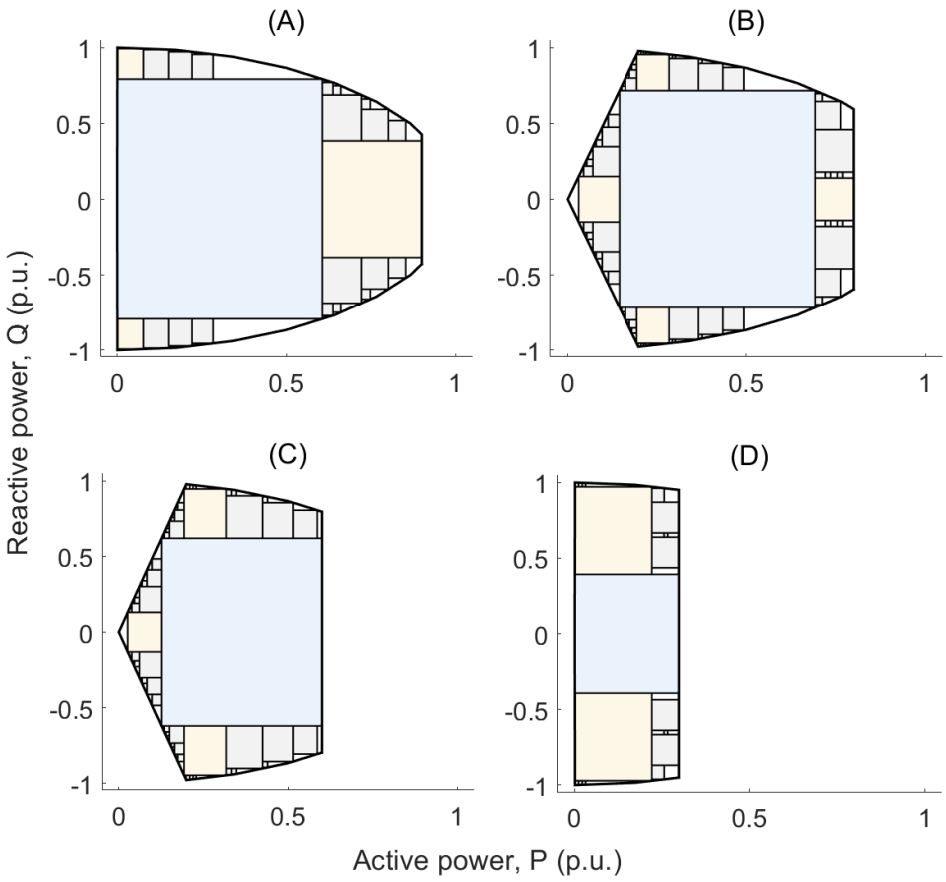

Figure 7.8: Decomposition of four inverter polytopes using the homothet-based polytope decomposition algorithm.

The results obtained by applying the HPD algorithm to each of the four inverters are shown in Fig. 7.8. We used CVX [35] to solve (P2). All computations were performed on a computer with Intel Core i-5 3.20 GHz processor with 8 GB RAM.

The area ratios (approximated area divided by the area of the i-th inverter polytope) after completion of each stage are given in Table 7.1. From the area ratios, on average, the total area captured after stage, $s = 0$ was 58 %, whereas after $s = 4$ was 95 %. The average time taken to complete decomposition up to each stage s is shown in Fig. 7.9. Up to $s = 1$, the average computation time was only 6.7s, whereas the area ratios averaged 79%, a 21 % increase from the case of $s = 0$.

As discussed in sections 7.4.1 and 7.4.4, instead of computing the entire M-sum, we

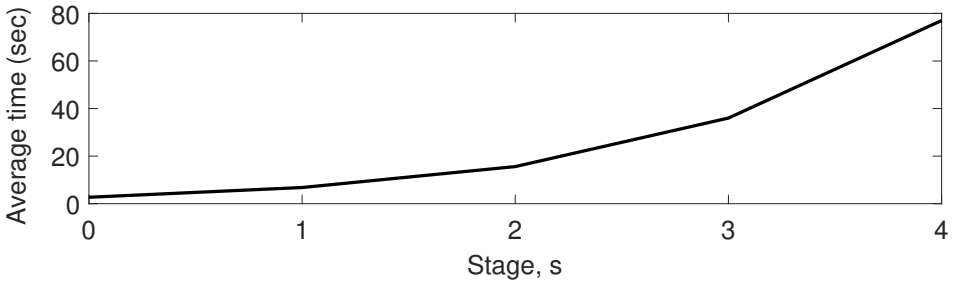

Figure 7.9: Average time for decomposition up to stage s.

consider a limited set of *candidate* boxes, which in this case correspond fist selecting only stage 0 and stage 1 boxes for each polytope. Secondly, for computing the M-sums, instead of considering all combinations of unions, consider : $B_0^{Aggr} = \oplus_{i=1}^{n_d} B_i^0(\mathcal{P}_i)$, $B_{\tilde{\sigma}}^{Aggr} = \oplus_{i=1}^{n_d} B_{i,\tilde{\sigma}}^1(\mathcal{P}_{i,\tilde{\sigma}}^1)$, $\tilde{\sigma} = 1, 2, 3, 4$, where $\tilde{\sigma}$ denotes the index of the stage-1 candidate boxes. In case, any $B_{i,\tilde{\sigma}}^1(\mathcal{P}_{i,\tilde{\sigma}}^1)$ is degenerate, it was replaced with $B_i^0(\mathcal{P}_i)$. These five aggregate boxes are shown in Fig. 7.10.

The actual M-sum polytope, also shown Fig. 7.10, was obtained using the MPT toolbox [41]. Finally, the C-hull of these boxes was computed in MATLAB and is shown in Fig. 7.10. Since all vertices of the aggregate boxes lie inside the true M-sum, which is convex and compact, the C-hull of these aggregate boxes is also an inner approximation to the true M-sum polytope.

To assess the M-sum approximation accuracy, we computed the area ratios for our approximated boxes and the C-hull and compared these against the area of the true M-sum polytope. Using only B_0^{Aggr}, the M-sum approximation accuracy was 52%. Using both stage 0 and stage 1 candidate boxes, the accuracy increased to 71%, thus demonstrating the effectiveness of considering multiple homothets per device. Finally, with C-hull, the

Table 7.1: Area covered, as a fraction of the area of the original polytope, after every stage of HPD, for each inverter.

Stage, s	(A)	(B)	(C)	(D)
0	0.64	0.65	0.64	0.40
1	0.81	0.74	0.76	0.84
2	0.87	0.85	0.86	0.89
3	0.92	0.91	0.92	0.95
4	0.94	0.93	0.96	0.96

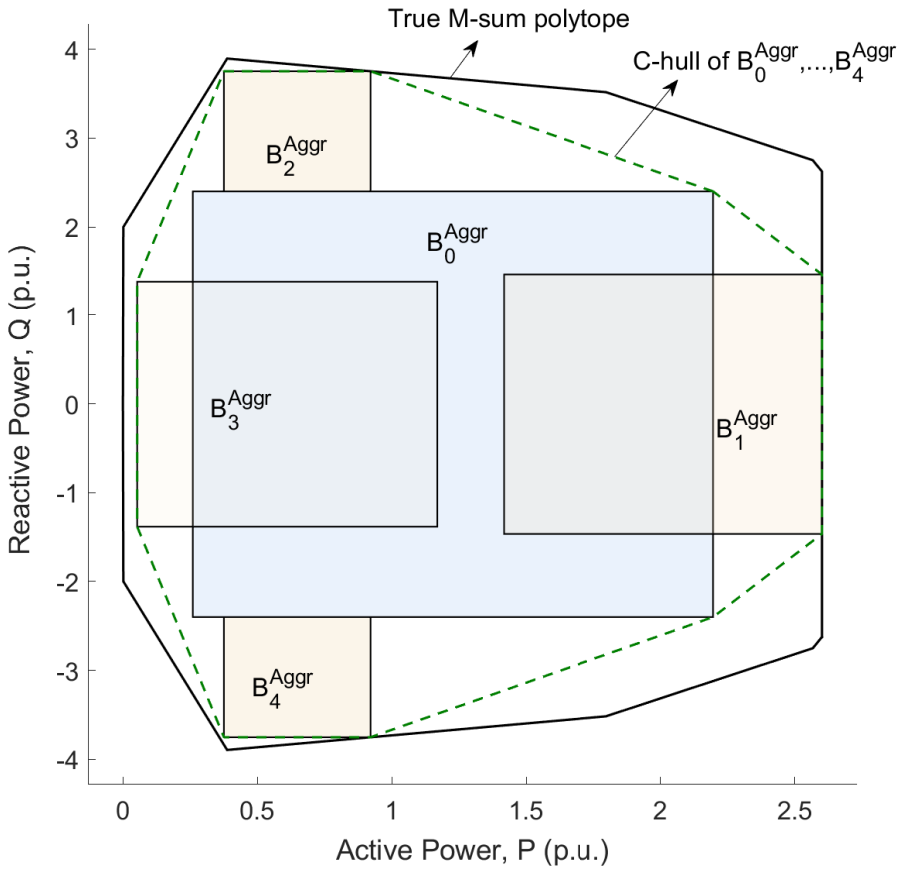

Figure 7.10: M-sum approximation using B_0^{Aggr} and $B_{\tilde{\sigma}}^{Aggr}$, $\tilde{\sigma} = 1, ..., 4$, the C-hull of the aggregate boxes, and the true M-sum polytope, $\mathcal{P}^{\mathbf{Aggr}}$.

accuracy was 85%.

As discussed before, considering all combinations of unions would cause exponential growth in complexity [101]. Instead, our policy used a fixed set of candidates. Once the results of the decomposition were available, computing the five aggregate boxes was a trivial task. Furthermore, given that the vertices of the aggregate boxes were readily available, the C-hull was also inexpensive to compute in $\mathbb{R}^2$, where C-hull computation has complexity $\mathrm{O}(n_v \log n_v)$ with n_v being the number of vertices [7]. .

7.5.2. Performance of Union-based M-sum for Controllable Loads and Storage Devices

To evaluate the performance of our proposed scheme in $\mathbb{R}^M$, consider storage-like loads and a 3 hour planning horizon with 30 minutes time steps. Hence, $M = 6$. Consider $n_d = 100$ devices with parameters, $\underline{P} = 0$, $\overline{P} \in [3, 4.5]$ kW, $a \in [0.9, 1]$, $e_0 \in [0.2, 0.6]$ (normalized), and $\gamma \in [0.035, 0.053]$. First, the polytope decomposition algorithm of section 7.4.2 was applied, with and without (7.20d) active. Imposing (7.20d) generally led to slower coverage of the feasible region of device polytopes in $\mathbb{R}^M$. Hence, we relaxed (7.20d) to improve volume coverage per stage. To evaluate accuracy of the decomposition phase, volume ratios were considered. Since volume computation in high dimension is an NP-hard problem, this is done here by finding a bounding box of the given polytope and generating random samples inside it. To obtain a Monte Carlo estimate of the volume of a polytope, we compute the fraction of samples that lie inside the polytope and then multiplying it by the volume of the bounding box [10]. For the decomposition phase, the volume ratios averaged 56% at the end of $s = 0$, 67% at the end of $s = 1$ and 82% at the end of $s = 2$. The decomposition up to $s = 1$, for each polytope, took on average 6.2s. Recall that the decompositions can be performed in parallel.

For computing the M-sum, we consider a limited set of candidate boxes from $s = 0$ and $s = 1$. Following the same procedure in section 7.5.1, we obtain $(2M + 1) = 13$ boxes in $\mathbb{R}^6$. In order to compute true M-sum polytope, we again used the MPT toolbox [41]. However, using MPT, it was only possible to obtain M-sums of 5 devices at a time, beyond which no solutions were reached in several hours. Hence, we sampled 5 devices from the population randomly and repeated the volume computation to obtain an estimate of the volume ratios. The accuracy of the M-sum approximation, compared to the actual volume, was 44% using only stage 0 boxes, and increased to 74% using the 13 candidate boxes. In our case, since the M-sum computation simply required an algebraic summation of the upper and lower bounds of intervals (or the scaling and translation coefficients of homothets in the previous example), the M-sum computation time was negligible and did not depend on the population size.

Given that at each decomposition stage (P2) must be solved, we can express time complexity in terms of the number of convex problems solved. During decomposition with axis-aligned boxes, two half-space inequalities are considered in each dimension. Given $x \in \mathbb{R}^M$, each box is defined by $2M$ constraints. Then, at stage s, for each box, $2M$

Table 7.2: Numbers of Convex Problems (P2) solved by end of stages, $s = 0, 1, 2, 3$, with increase in dimension, M

	M						
Stage, s	2	3	4	5	6	7	8
0	1	1	1	1	1	1	1
1	5	7	9	11	13	15	17
2	21	43	73	111	157	211	273
3	85	259	585	1111	1885	2955	4369

additional constraints are introduced for the subsequent stage. Thus, at stage, s, (P2) is solved at most $(2M)^s$ times.

Table 7.2 shows the maximum number of times (P2) must be solved by the end of stage s. While choosing a higher s assists in achieving convergence to the true M-sum, the complexity grows exponentially. Hence, in our approach, in $\mathbb{R}^4$ and above we suggest to choose $s = 1$, which ensures a polynomial time algorithm. Using our of candidate selection for computing a finite set of aggregate boxes, in $\mathbb{R}^M$, we obtain exactly $(2M+1)$ boxes. For the purpose of solving optimization problems, recall that (P1) can now be solved subject to each of these $(2M + 1)$ boxes, in parallel. Since only $2M$ half-space constraints are required to represent the boxes, the optimization problem is generally significantly simpler than (P0) subject to the true M-sum polytope.

7.6. Application: Voltage Control with Inverters

For an application, consider the following optimization problem (P3). The goal is to optimize the inverter set-points such that inverters provide reactive power support to ensure voltage stays within limits [23, 96, 98]. The objective of (P3) is to minimize the active power curtailment, P_{curt} from residential PV sources. The aggregate active power availability of all n_d PV inverters is $P_{\text{av}} = \sum_{i=1}^{n_d} s_i \overline{P_i}$. With inverter control, one can then obtain the aggregate feasible set $\mathcal{P}^{\text{Aggr}}$, which is the M-sum of the feasibility sets of all n_d inverters.

Assume the unit cost of curtailment is C_{curt}. Constraint (7.21b) ensures that the inverters' set-points must be within the aggregate flexibility polytope. Given, the aggregate real and reactive power demands of loads, P_{d} and Q_{d}, the net real and reactive power injections, P_{net} and Q_{net}, are given by (7.21d) and (7.21e). Constraint (7.21f) assumes an approximate linear relation between voltage and real and reactive power injections [107],

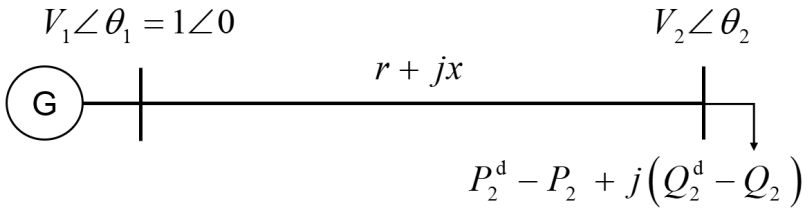

Figure 7.11: Two bus power flow.

which can be obtained from the standard power flow equations [103]. Finally, by (7.21g), the voltage, V, must be within minimum and maximum limits, $\underline{V}$ and $\overline{V}$, respectively.

$$
\begin{aligned}
\text{(P3)} \quad & \min_{P,Q} C_{\mathrm{curt}} P_{\mathrm{curt}} && \text{(7.21a)}\\
& s.t. \quad (P,Q) \in \mathcal{P}^{\mathrm{Aggr}}, && \text{(7.21b)}\\
& P + P_{\mathrm{curt}} = P_{\mathrm{av}}, && \text{(7.21c)}\\
& -P_{\mathrm{d}} + P = P_{\mathrm{net}}, && \text{(7.21d)}\\
& -Q_{\mathrm{d}} + Q = Q_{\mathrm{net}}, && \text{(7.21e)}\\
& V = R_1 P_{\mathrm{net}} + R_2 Q_{\mathrm{net}} + R_0, && \text{(7.21f)}\\
& \underline{V} \le V \le \overline{V}. && \text{(7.21g)}
\end{aligned}
$$

A simple two bus power system, shown in Fig. 7.11 is considered. Assume an aggregation of 32 houses at bus 2. Each house has 7 kW of installed PV capacity. Consider micro-inverters with rating of 255 W, hence, a total of 28 micro-inverters per PV system. The voltage limits are set to 0.92 and 1.04 p.u [23, 96]. Several studies have shown that power distribution systems can face voltage violations during afternoon hours when residential demand is low and solar PV outputs are high [96]. Therefore, in this example, consider hour 13 (1 pm in the afternoon). The real power demand per house is 2.2 kW and reactive power demand is 0.1 kVar. Consider 4 groups of 8 houses each. Each group's inverter ratings, availability and power factor angle are according to data given in section 7.5.1.

Recall that the standard AC power flow equations are nonlinear [103]. In order to obtain an approximate linear relation between voltage, P_{net} and Q_{net}, i.e. to estimate R_0, R_1, R_2 in (7.21f), we vary the real power and reactive power levels, solve the AC power flow equations and find the corresponding voltage levels at bus 2.

Finally, the result of inverter set-point control (P3) is applied. Three approaches were

considered:

- C1. Solve (P3) with $\mathcal{P}^{\text{Aggr}}$ = the true M-sum, computed by MPT toolbox [41] by vertex sums.
- C2. Solve (P3), in parallel, for the five boxes shown in Fig. 7.10 and choose the minimum from the candidate solutions, as outlined in (7.15).
- C3. Solve (P3) s.t.$\mathcal{P}^{\text{Aggr}}$ = C-hull shown in Fig. 7.10.

Table 7.3 shows the aggregate set-points for P and Q for inverters. P_{curt} and bus 2 voltage levels are also shown. For solving C2 and C3, we used MATLAB and YALMIP [65]. The optimal solutions using B_1^{Aggr} and the C-hull matched the solutions obtained using $\mathcal{P}^{\text{Aggr}}$, and no active power was curtailed. Using B_0^{Aggr}, only a sub-optimal solution was found, resulting in curtailment. B_4^{Aggr} resulted in infeasible solution since voltage limits would be violated at such reactive power level. Thus, the optimal solution by C2 is given by the solution corresponding to B_1^{Aggr}, and the minimum curtailment obtained by C2 agrees with the solutions from C1 and C3.

Table 7.3: Results with Inverter Set-point Control

Feature	$\mathcal{P}^{\text{Aggr}}$	B_0^{Aggr}	B_1^{Aggr}	B_2^{Aggr}	B_3^{Aggr}	C-hull
P	0.77	0.65	0.77	0.27	0.35	0.77
Q	-0.06	-0.09	-0.11	0.75	0	-0.14
P_{curt}	0	0.12	0	0.5	0.42	0
V	1.02	1	1.02	1.03	0.97	1.01

7.7. Chapter Conclusions

In this chapter, We developed and compared algorithms to compute inner approximations of the Minkowski sum of convex polytopes. As an application, we considered the computation of the feasibility set of aggregations of distributed energy resources (DERs), such as solar photovoltaic inverters, controllable loads, and storage devices. A convex polytopic representation for a feasible operating region of inverter interfaced DERs was developed first. We showed how homothets can be used to compute the M-sum and obtained analytical expressions in special cases. However, as heterogeneity increases, using a single homothet per device, may result in highly conservative inner approximation of the M-sum.

Hence, to fully account for the heterogeneity in the DERs while ensuring an acceptable approximation accuracy, we leveraged a union-based computation that advocates homothet or interval-based polytope decomposition. We show that the proposed algorithm can guarantee the inner approximation asymptotically converges to the true M-sum. However, union-based approached can in general lead to high-dimensionality concerns; to alleviate this issue, we showed how to define candidate sets to reduce the computational complexity. Accuracy and trade-offs have been analyzed through numerical examples. The flexibility polytopes of inverter-interfaced devices, controllable loads and storage can be integrated in power systems planning tools to provide various power system services.

Chapter 8.

Flexibility Aggregation in Distribution Systems

8.1. Chapter Introduction

In this case study, we plan to utilize the flexible polytopes of inverters and loads, developed in Chapter 7, to approximate the aggregate real and reactive power flexibility available at a distribution system substation. The net $P-Q$-flexibility set should be formed by capturing flexibility available from different nodes on the distribution feeder. This is challenging since the network constraints will affect the amount of flexbility that can be transferred from different nodes to the substation. A major benefit of being able to characterize the net $P-Q$-flexibility set at a substation is that transmission system operators can then better utilize the system resources without actually having to model any DERs at the distribution system. Standard IEEE radial test feeders will be studied to demonstrate the concept and the results.

Several recent works have therefore developed techniques to characterize the flexibility from DERs. In [75, 78, 106], polytopes containing feasible operating points of DERs are characterized. For performing aggregation, typically inner/outer approximations are sought by using *homothets* [76, 106], and *zonotopes* [75], which are convex sets with special structures. Considering these polytopes in any optimization problem (such as optimal power flow and economic dispatch) eliminates the need to consider DER state dynamic equations, their power and energy limits and hence can provide significant computational advantage. However, in previous work, network constraints are typically ignored, hence such techniques may overestimate the net flexibility of DERs (see Fig. 8.2 and Fig. 8.3).

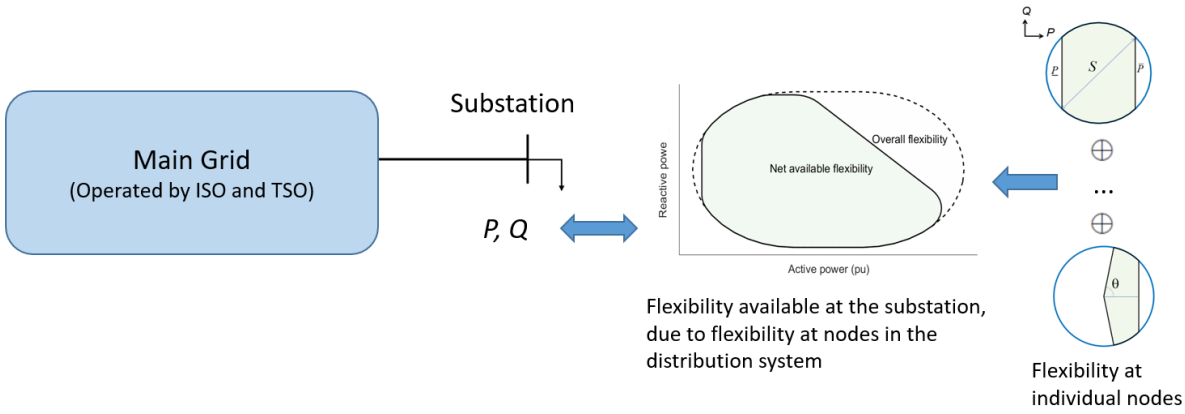

Figure 8.1: Aggregation of DER flexibility in distribution systems.

Considering a distribution system, specific parts of the network may significantly limit the flexibility that can be transferred from one node to its upstream or downstream nodes [67]. Hence, this chapter focuses on developing techniques to approximate the net aggregate real and reactive power flexibility available at a distribution system substation.

First, the nodal active and reactive power flexibility due to presence of DERs are approximated using convex sets and polytopes. Next, we show how network constraints and voltage limits act as constraints on such feasibility sets. The radial structure of distribution systems is exploited to efficiently perform the aggregation. Mainly, we propose three different techniques, geometric, optimization-based and Monte Carlo sampling-based, to compute the net active-reactive flexibility at the substation. The usefulness and shortcomings of each technique is illustrated. Instead of providing the TSOs or ISOs deterministic active and reactive power demand levels, each substation can then submit ranges of real and reactive powers, or P-Q flexibility polytopes (see Fig. 8.1). Optimal power flow formulations and economic dispatch problems can be solved efficiently considering these polytopes. The net P-Q polytopes allow TSO/ISOs to take into account the flexibility available in DERs in distribution feeders as well as the distribution level network constraints without actually modeling any of those.

Recently, in [92], an iterative technique has been presented to compute the net var capability curve at a substation, as a function of solar power curtailment. Linear DistFlow equations are used to model the power flow in 3-phase radial circuits. In our work, we consider both real and reactive power flexibility by considering DR resources, along with inverter-based flexibility. Instead of depending only on LinDistFlow, our optimization-

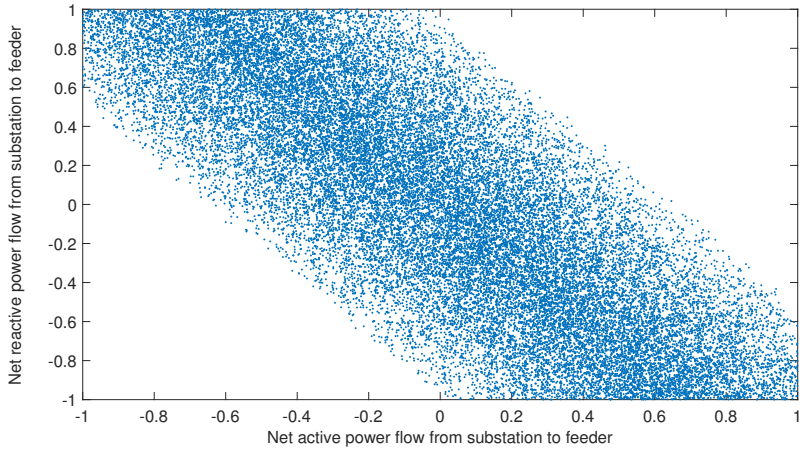

Figure 8.2: In a 2-node test system (see Fig. 7.11), Node-2 offers real and reactive power flexibility. At Node-2, active and reactive power levels can vary between -1 to 1 p.u. The resulting net active and reactive power-flows from the substation node (Node 1) towards Node-2 are shown. All samples satisfy voltage limits (0.95 to 1.05 p.u). Due to the network constraints and voltage limits, the net $P-Q$ flexibility set at the substation is more restricted than what is offered at Node-2.

based algorithm additionally includes a validation step using the full non-linear DistFlow equations. We also consider uncertainty from the DERs, hence propose adjustments to the algorithm to obtain a robust feasible set. Another major difference of our work compared to existing work on flexibility aggregation for distribution systems, is that we first obtain the M-sum representations of flexibility at each node with DERs and DR resources. With polytopes of nodal DER flexibility is included in the optimization problem, the dual variables show exactly which constraints are binding, hence in which direction flexibility sets get constrained. It also indicated when there is unused flexibility due to transfer limits due to network and voltage constraints. Finally, availability of the net P-Q polytope at a substation allows DSOs to follow simple yet effective DER control strategies, requiring minimal communication and local dispatch rules, for providing P-Q adjustments to provide grid support.

The rest of the chapter is organized as follows. Section 8.2 shows how nodal flexibility sets and their aggregate polytopic representations can be formed. Section 8.3 shows what the aggregate flexibility would be without network constraints. Section 8.4 described the network constraints and how to take these into account using three different techniques.

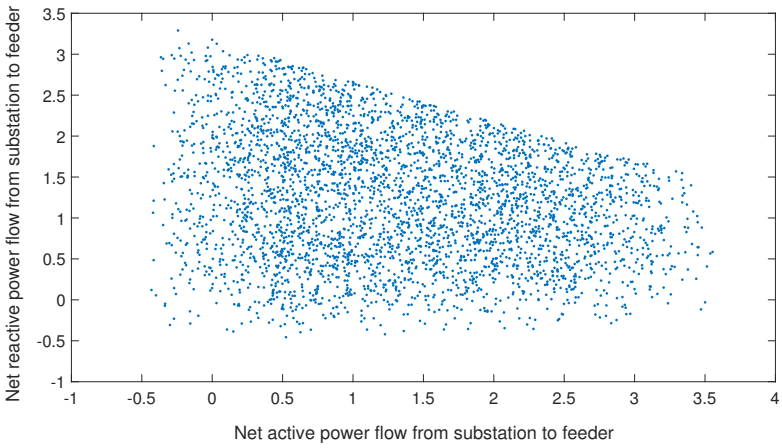

Figure 8.3: In a 4-node radial feeder, the terminal node is assumed to offer real and reactive power flexibility. Active and reactive power can vary between -1 to 3 p.u. The resulting net active and reactive power-flows from the substation node to the distribution feeder are shown. All samples satisfy voltage limits (0.95 to 1.05 p.u.).

Section 8.4.8 shows how DER dispatch rules can be designed. Then, section 8.5 shows results on a 33-bus radial feeder and finally, section 8.6 concludes.

8.2. Nodal Flexibility with Heterogeneous DERs

8.2.1. Inverter-interfaced DERs

Let $\mathcal{X} \subseteq \mathbb{R}^2$ be the set that contain an inverter-interfaced DER's real and reactive power operating points, $x = [p, q]^T, x \in \mathbb{R}^2$ [15, 52, 78]. Then, $\mathcal{X}$ can be written as ,

$$\mathcal{X}(s, \underline{p}, \overline{q}) = \{(p, q) : \underline{p} \leq p \leq \overline{p}, q^2 \leq s^2 - p^2\}. \tag{8.1}$$

Here, for a PV system, s is the apparent power rating of the inverter, $\underline{p} = 0$ and $\underline{p}$ is the available power injection (hence negative) based on solar irradiance. For a storage device interfaced with an inverter, p is the inverter's rating; $\underline{p}$ and $\overline{p}$ are the minimum and maximum real power available at a specific time. Note that $\mathcal{X}(s, \underline{p}, \overline{p})$ is a convex set. Additionally, a minimum power factor can be enforced, as shown in our recent work [78].

8.2.2. Controllable Loads

The feasible sets for controllable loads, such as variable speed drives, heating/ cooling loads, and pool pumps [13, 78], considering only real power flexibility, can be written as

$$\mathcal{X}(\underline{p}, \overline{p}) = \{p : \quad \underline{p} \leq p \leq \overline{p}\}. \tag{8.2}$$

For storage-like DERs, such as TCLs and EVs, the internal states (e.g. temperature, state of charge) also evolve with time. Hence, for M time-intervals, an M dimensional polytope can be formed [10, 74, 78, 106]. While our earlier work [78] dealt with such higher dimensional cases, here, since we focus on real and reactive power flexibility at a specific time instance, higher dimensional cases will not be discussed. However, it's possible to extend our proposed algorithms to such cases. for example, by computing inner-approximations to the M-dimensional polytopes [10, 74, 78, 106].

8.2.3. DER Flexibility Aggregation at a Single Node

With availability of inverter-interfaced DERs and controllable loads, the flexibility from DERs at a single node can be aggregated using Minkowski sum (M-sum), which is a set-theoretic operation. For two DERs, the M-sum, by definition, is the vector sum of all feasible points in first set, with all points in the second set, which is written as,

$$\mathcal{X}^{\text{node}} := \mathcal{X}_1 \oplus \mathcal{X}_2. \tag{8.3}$$

where $\oplus$ denotes the M-sum and $\mathcal{X}_i$ represents a DER's feasibility set of the form $\mathcal{X}(s, \underline{p}, \overline{p})$ and/or $\mathcal{X}(\underline{p}, \overline{p})$.

The above set-theoretic definition is however not directly used for computing the M-sum [101]. Instead various techniques from computational geometry are used to efficiently compute or approximate the M-sum.

Vertex Summation and Convex Hull

M-sum satisfies the property that the M-sum of convex sets is always convex [91]. Hence a general recipe for computing the M-sum of convex polytopes involves (i) performing the vertex summations, and (ii) taking the convex hull of the resulting points. For example, for $\mathcal{X}(s, \underline{p}, \overline{p})$ or $\mathcal{X}(s, \underline{p}, \theta)$ consider N_j points (vertices) on the boundary, as shown in our

earlier work [78], and for $\mathcal{X}(\underline{p}, \overline{p})$, simply $\underline{p}$ and $\underline{p}$. After performing vector sums of all vertex combinations, a convex hull gives the true M-sum. For problems in 2 dimensions and with limited number of DERs, the accurate M-sum can typically be computed very fast [41].

Inner/ Outer Approximating Homothets and/or Zonotopes

When the number of devices is large and in problems involving more than 2 dimensions, the above vertex summation-based method becomes highly inefficient. Hence, inner or outer approximations of the M-sum using special convex structures, such as homothets [52, 78, 106] and zonotopes [74] are typically sought.

8.2.4. Algebraic Techniques for Special Cases

While such generic recipes are valuable, in the case of DERs, we show that it is often possible to obtain analytic expressions for several commonly encountered cases, hence are presented next.

Case 1. Consider $\mathcal{X}_1(\underline{p}_1, \overline{p}_1)$ and $\mathcal{X}_2(\underline{p}_2, \overline{p}_2)$ as defined in (8.2). The aggregate flexibility is simply given by,

$$\mathcal{X}^{\text{node}} = \mathcal{X}_1(\underline{p}_1, \overline{p}_1) \oplus \mathcal{X}_2(\underline{p}_2, \overline{p}_2) = \mathcal{X}(\underline{p}_1 + \underline{p}_2, \overline{p}_1 + \overline{p}_2). \tag{8.4}$$

Case 2. Consider $\mathcal{X}_1(s_1, \underline{p}_1, \overline{p}_1)$ and $\mathcal{X}_2(s_2, \underline{p}_2, \overline{p}_2)$, as defined in (8.1). Assume only the s_1 and s_2 may vary, while $\underline{p_1}/s_1 = \underline{p_2}/s_2$ and $\overline{p_1}/s_1 = \overline{p_2}/s_2$. This situation can appear commonly when the two inverters have different ratings, but undergo similar solar irradiance conditions, which could be due to their geographic proximity (e.g. at the same node). Then, the aggregate flexibility set can be obtained by,

$$\begin{aligned} \mathcal{X}^{\text{node}} &= \mathcal{X}_1(s_1, \underline{p}_1, \overline{p}_1) \oplus \mathcal{X}_2(s_2, \underline{p}_2, \overline{p}_2) \\ &= \mathcal{X}\left(s_1 + s_2, \underline{p}_1 + \underline{p}_2, \overline{p}_1 + \overline{p}_2\right). \end{aligned} \tag{8.5}$$

Case 3. Consider $\mathcal{X}_1(s_1, \underline{p}_1, \overline{p}_1)$ and $\mathcal{X}_2(\underline{p}_2, \overline{p}_2)$, given by (8.1) and (8.2). The aggregate

flexibility set can be obtained by [15],

$$\begin{aligned}\mathcal{X}^{\text{node}} &= \mathcal{X}_1(s_1, \underline{p}_1, \overline{p}_1) \oplus \mathcal{X}(\underline{p}_2, \overline{p}_2) \\ &= \{(p,q) : \underline{p}_1 + \underline{p}_2 \leq p \leq \overline{p}_1 + \overline{p}_2, -f(p) \leq q \leq f(p)\},\end{aligned} \tag{8.6}$$

where $f(p)$ is given by,

$$f(p) = \begin{cases} s_1, & \text{if} \quad p \in [\underline{p}_2, \overline{p}_2] \\ \sqrt{s_1^2 - (p - \underline{p}_2)^2}, & \text{if} \quad p \in [\underline{p}_1 + \underline{p}_2, \underline{p}_2) \\ \sqrt{s^2 - (p - \overline{p}_2)^2}, & \text{if} \quad p \in (\overline{p}_2, \overline{p}_1 + \overline{p}_2] \end{cases} \tag{8.7}$$

Remark: For additional cases and detailed proofs of (8.4)-(8.7), please refer to authors' earlier work [78] and [14].

These nodal flexibility sets, $\mathcal{X}^{\text{node}}$, can be used in the following ways,

- Aggregation of net flexibility at an n-node distribution system, without considering the network constraints (see Section 8.3).
- Drawing samples from the each $\mathcal{X}_i^{\text{node}}$, $i = 1, 2, ..., n$, to perform simulations-based approximation of net flexibility, including the network constraints,
- Using the nodal flexibility polytopes in optimization formulations that consider the network and voltage constraints.

These three methods will be elaborated and compared in the subsequent sections.

8.3. Aggregate Flexibility without Network Constraints

Consider a distribution system with n nodes. The nodal flexibility are given by $\mathcal{X}_j^{node}$, $j = 1, 2, ..., n$. Then, the net real-reactive power flexibility at the substation (without considering impact of network constraints) is given by the M-sum,

$$\mathcal{X}^{\text{sub, wo}} := \oplus_{i=1}^{n} \mathcal{X}_j^{node}. \tag{8.8}$$

Techniques from 8.2.3 can be used. Additionally, for a few special cases, it is worth providing the analytic expressions.

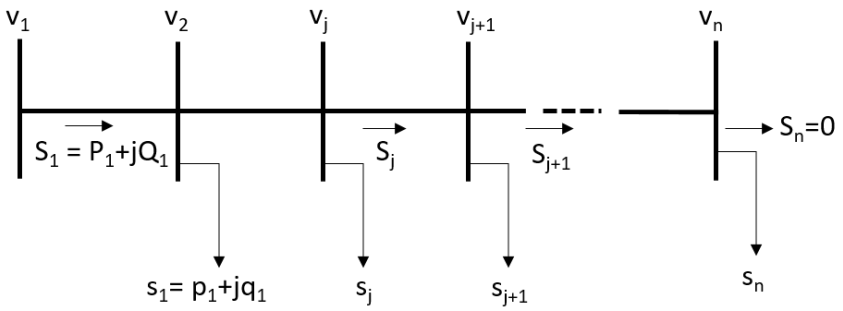

Figure 8.4: Branch flow in a radial distribution system.

Case 1. Consider the nodal flexibility in p_j and q_j, $\forall j$, are given or have been approximated using axis-aligned constraints, such as $\underline{p}_j \leq p_j \leq \overline{p}_j$ and $\underline{q}_j \leq q_j \leq \overline{q}_j$. Then, the real-reactive power flexibility at the substation is given by,

$$\begin{aligned}\mathcal{X}^{\text{sub, wo}} :=\{(p,q) : \sum_{j=1}^{n} \underline{p}_j \leq p \leq \sum_{j=1}^{n} \overline{p}_j, \\ \sum_{j=1}^{n} \underline{q}_j \leq q \leq \sum_{j=1}^{n} \overline{q}_j\}.\end{aligned} \tag{8.9}$$

Case 2. Consider $\mathcal{X}_{1,j}(s_j, \underline{p}_{1,j}, \overline{p}_{1,j})$ and $\mathcal{X}_{2,j}(\underline{p}_{2,j}, \overline{p}_{2,j})$ at nodes $j = 1, 2, ..., n$. In this case, using (8.6)-(8.7) to first obtain $\mathcal{X}_j^{node}$ will lead to difficulties in obtaining the M-sum overall nodes algebraically. However, notice that if conditions of case 2 from section 8.2.3 hold, then, $\mathcal{X}^{inv}(s^A, \underline{p}^A, \overline{p}^A)$, where $s^A = \sum_{j=1}^{n} s_j, \underline{p}^A = \sum_{j=1}^{n} \underline{p}_{1,j}, \overline{p}^A = \sum_{j=1}^{n} \overline{p}_{1,j}$ gives the M-sum of $\mathcal{X}_j(s_j, \underline{p}_{1,j}, \overline{p}_{1,j})$. The M-sum of $\mathcal{X}_{2,j}(\underline{p}_{2,j}, \overline{p}_{2,j})$ is simply $\mathcal{X}^L(\underline{p}^L, \overline{p}^L)$, where $\underline{p}^L = \sum_{j=1}^{n} \underline{p}_{2,j}$ and $\overline{p}^L = \sum_{j=1}^{n} \overline{p}_{2,j}$. Finally, (8.6)-(8.7) can be applied to obtain the M-sum of $\mathcal{X}^{inv}(s^A, \underline{p}^A, \overline{p}^A)$ and $\mathcal{X}^L(\underline{p}^L, \overline{p}^L)$.

Since network constraints are absent, the order of the M-sum operations does not matter, hence the most efficient order should be chosen. When network constraints and voltage limits are present, the above techniques may not give the true P-Q flexibility at the substation. However, $\mathcal{X}^{\text{sub, wo}}$ can still serve as an outer approximation of the net P-Q flexibility and represent the net capability available across all DERs.

8.4. Net Flexibility At Substation with Network Constraints

8.4.1. Constraints in a Distribution System

In a radial distribution feeder (e.g. Fig. 8.4), the real and reactive power flows and the nodal voltages are described by the classical DistFlow equations [9] as follows,

$$P_{j+1} = P_j - r_j \frac{P_j^2 + Q_j^2}{v_j^2} - p_{j+1} \tag{8.10a}$$

$$Q_{j+1} = Q_j - r_j \frac{P_j^2 + Q_j^2}{v_j^2} - q_{j+1} \tag{8.10b}$$

$$v_{j+1}^2 = v_j^2 - 2(r_j P_j + x_j Q_j) + (r_j^2 + x_j^2)\frac{P_j^2 + Q_j^2}{v_j^2} \tag{8.10c}$$

where v_j is the voltage at node $j = 1, 2, ..., n$, $\bar{S}_j = P_j + iQ_j$ is the apparent (complex) power flowing from node j to node $j+1$. $\bar{s}_j = p_j + iq_j$ is the net apparent power consumed at node j. Linear approximations to the DistFlow equations, also referred to as the LinDistFlow [9], are given by,

$$P_{j+1} = P_j - p_{j+1} \tag{8.11a}$$

$$Q_{j+1} = Q_j - q_{j+1} \tag{8.11b}$$

$$v_{j+1}^2 = v_j^2 - 2(r_j P_j + x_j Q_j) \tag{8.11c}$$

Additionally, the nodal voltages, v_j, at each j, respect the following limits,

$$\underline{v}_j \leq v_j \leq \overline{v}_j \tag{8.12}$$

Using squared quantities, (8.12) becomes,

$$\underline{v}_j^2 \leq v_j^2 \leq \overline{v}_j^2 \tag{8.13}$$

Finally, given $\mathcal{X}_j^{node}$, $j = 1, 2, ..., n$, we have the following constraints at each node,

$$(p_j, q_j) \in \mathcal{X}_j^{node},\ j = 1, 2, ..., n. \tag{8.14}$$

Using polytopic representation of $\mathcal{X}_j^{node}$, (8.14) becomes,

$$\mathcal{L}x \leq \mathcal{M} \tag{8.15}$$

The substation's aggregate flexibility can now be computed in several ways.

8.4.2. Simulation-based Approach to Approximate the Net $P-Q$ Flexibility

A sampling based algorithm, using nodal flexibility sets, is given below,

Algorithm-1.
S0. Initialize $k=0$ and $\mathcal{X}_j^{node}$, $j=1,2,...,n$.
S1. Sample $(p_j,q_j)\in\mathcal{X}_j^{node}$, with $j=1,2,...,n$.
S2. Compute DistFlow (8.10a)-(8.10c) solutions using forward-backward sweep.
S3. If condition (8.12) is satisfied, $k=k+1$ and $X_1[k]=P_1, X_2[k]=Q_2$.
S4. If $k\leq k^{\max}$ (a desired number of samples), repeat S1-S4.

$\mathbf{X_1}$ contains $k^{\max}$ number of feasible values for P_1, and $\mathbf{X_2}$ contains the correspondng Q_1 values. Hence, the set $\tilde{\mathcal{X}}([\mathbf{X_1},\mathbf{X_1}])$ shows the feasible $P-Q$ at the substation. Thus, the approach can be used to obtain intuitions of the net $P-Q$ feasibility, however, it lacks analytical guarantees and is cumbersome. In Section 8.4.4 an optimization-based algorithm will be presented.

8.4.3. The Impact of Voltage Constraints on Flexibility Sets

To analyze the impact of voltage constraints on net $P-Q$ feasibility, first resort to LinDist-Flow equations, where (8.11c) relates the nodal voltages with the active and reactive power flows over lines. Let $u_j=v_j^2,\forall j$. Then, (8.11c) can be written as,

$$u_{j+1}=u_j-2r_jP_j-2x_jQ_jP_j. \tag{8.16}$$

and Q_j can be expressed as,

$$Q_j=\frac{u_j-u_{j+1}}{2x_j}-\frac{r_j}{x_j}P_j. \tag{8.17}$$

Limits on node voltages (8.13) can be imposed by,

$$\underline{u}_j \leq u_j \leq \overline{u}_j \tag{8.18}$$

Consider $r_j \approx r$, $x_j \approx x$, $\forall j$. Then, the set of voltage constraints can be written as,
$u_2 = u_1 - 2rP_1 - 2xQ_1$,
$u_3 = u_1 - 2rP_1 - 2xQ_1 - 2rP_2 - 2xQ_2$,
...
Similarly, $u_n = u_1 - 2r\sum_{j=1}^{n-1} P_j - 2x\sum_{j=1}^{n-1} Q_j$.
Thus,

$$\sum_{j=1}^{n-1} Q_j = \frac{u_1 - u_k}{2x} - \frac{r}{x}\sum_{j=1}^{n-1} P_j.$$

At an arbitrary node, k, s.t. $2 \leq k \leq n$, we obtain,

$$\sum_{j=1}^{k-1} Q_j = \frac{u_1 - u_k}{2x} - \frac{r}{x}\sum_{j=1}^{k-1} P_j. \tag{8.19}$$

Similarly, expressing the nodal voltages in terms of upstream voltages, at an arbitrary node, k, s.t. $1 \leq k \leq n-1$, we obtain,

$$\sum_{j=k}^{n-1} Q_j = \frac{u_k - u_n}{2x} - \frac{r}{x}\sum_{j=k}^{n-1} P_j. \tag{8.20}$$

The relations (8.19)-(8.20) give hyperplanes which may or may not cut the nodal flexibility sets and limit the amount of flexibility that can be transferred from an node to its upstream or downstream nodes. This can be illustrated using a simple example.

Example 1. Consider a 2 bus network, where the nodal flexibility at bus 2, $\mathcal{X}_2^{node}$, is given by,

$$\mathcal{X}_2^{node} := \mathcal{X}_a(s_a, \underline{p}_a, \overline{p}_a) \oplus \mathcal{X}_b(\underline{p}_b, \overline{p}_b).$$

Since node 2 is terminal, $P_2 = Q_2 = 0$. Ignoring losses, $P_1 = p_1, Q_1 = q_1$. Thus,

$$(P_1, Q_1) \subseteq \mathcal{X}_2^{node}. \tag{8.21}$$

However, due to the voltage constraints, the following relation must also be satisfied,

$$Q_1 = \frac{u_1 - u_2}{2x} - \frac{r}{x} P_1. \tag{8.22}$$

along with nodal voltage limits (8.18).

From (8.22), we see that the values of x and $\frac{r}{x}$ play an important role in determining the how Q_1 might be restricted. For the lines given by (8.22), x along with $u_1 - u_2$ determine the y-intercept and $\frac{r}{x}$ determines the slope. The values of these would guide if the lines cut the feasible region. Three cases are considered. Values of r and x and voltage limits are provided in Table 8.1. The feasibility sets of P_1 and Q_1 are shown in Figs. 8.5-8.6. $u_1 - u_2$ is largest when $u_1 = \overline{u}_1$ and $u_2 = \underline{u}_2$, and smallest when the reverse occurs. Any realization of voltages between these would result in lines within these two.

Table 8.1: Parameters for a .

Test	r (p.u.)	x (p.u.)	$\underline{v}$	$\overline{v}$
(A)	0.15	0.15	0.9	1.1
(B)	0.6	0.5	0.9	1.1
(C)	0.15	0.15	0.95	1.05

In case (A), (8.18) and (8.22) do not cause any hyperplanes to cut the feasible region of the nodal flexibility set of node 2, i.e. $\mathcal{X}_2^{node}$. Therefore, the entire set is also the flexibility set at node 1. In case (B), due to high r and x values, we have hyperplanes that intersect the feasible region. Hence, the net flexibility available at node 1, i.e. the 'grey' shaded area in Fig. 8.6, is smaller than the flexibility available at node 2. Similarly, when voltage limits are reduced, in case (C), there is a cut in the feasible region.

To compare the results, we used Algorithm-1 by sampling from the feasible region and running the DistFlow equations. Both voltage feasible and voltage-infeasible solutions are shown in Fig. 8.7. The voltage-feasible solutions all lie within the 'grey' shaded region obtained for Case (B) shown in Fig. 8.6 (left sub-figure). Observe that the sampling approach, using DistFlow, also finds additional feasible solutions, beyond the 'grey' shaded region, which was approximated using LinDistFlow. This is not surprising because we know that LinDistFlow ignores the quadratic terms. Furthermore, observe that there are many voltage-infeasible solutions lying within the 'grey' shaded region. Again, since the solutions of Disflow, for the 2-bus case, are solutions to quadratic equations - all points surrounding a root may not be solutions to the DistFlow equations. Therefore, the 'grey' shaded region

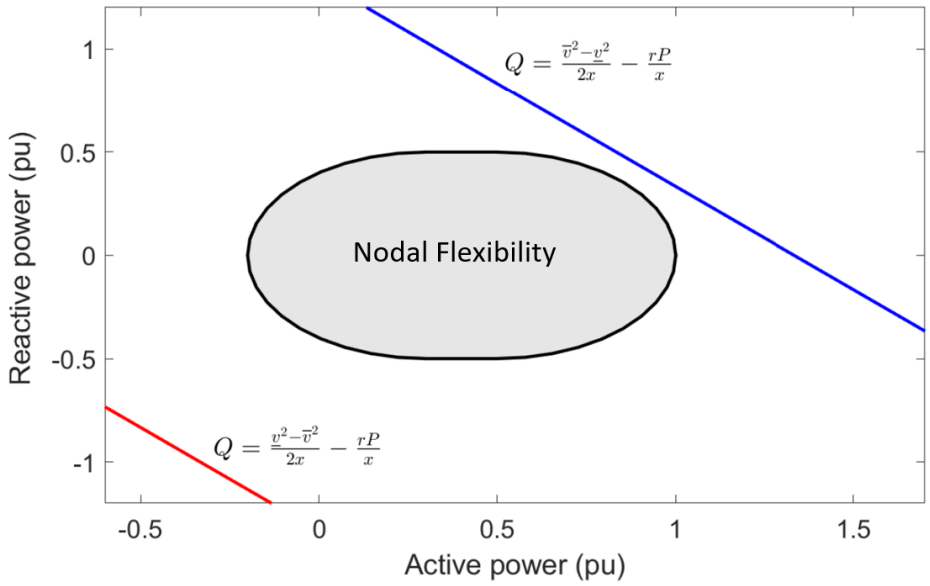

Figure 8.5: The net active and reactive power flexibility at node 1 in case (A).

approximated using the DistFlow equations can be interpreted such that if an arbitrary point, x^0, is picked from this region, a feasible solution of DistFlow is likely to be found at or near x^0. But it is not guaranteed that a voltage-infeasible solution will not lie near x^0. This observation has not been elaborated in existing work on flexibility aggregation using linear approximations, hence will be further elaborated in this chapter. □

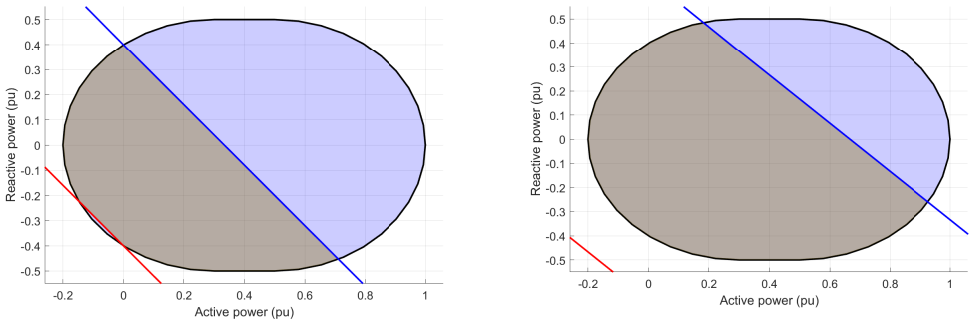

Figure 8.6: The net active and reactive power flexibility at node 1 in case (B) (left) and case (C) (right).

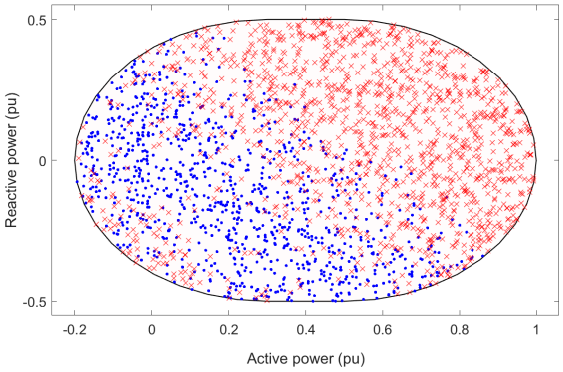

Figure 8.7: Limitations of a Monte carlo approach to approximate the net active and reactive power flexibility at node 1.

8.4.4. Optimization based Algorithm using LP-OPT

The constraints shown in section 8.4.1 can be incorporated in various optimization problems. In this section, we formulate linear optimization problem (LP-OPT) using DistFlow equations and nodal flexibility polytopes, and a nonlinear-counterpart (NL-OPT) using Distflow equations. LP-OPT will be used to approximate the net P-Q flexibility, which gives a convex polytope. The NL-OPT then performs validation as well as pruning to improve accuracy.

Using LinFistflow equations, consider the following optimization problem,

$$\begin{aligned} \text{(LP-OPT)} \quad \min \quad & \mathcal{F}(x) && \text{(8.23a)} \\ s.t. \quad & \mathcal{A}x \leq \mathcal{B} && \text{(8.23b)} \\ & \mathcal{H}x = \mathcal{G} && \text{(8.23c)} \end{aligned}$$

where $x = [P \ Q \ p \ q \ v^2]^\top$ and $\mathcal{F}(x)$ is the objective function. $\mathcal{H}x = \mathcal{G}$ contains the LinDistFlow equations (8.11a)-(8.11c) and $\mathcal{A}x \leq \mathcal{B}$ contains (8.13) and (8.15). Note that this formulation is an LP and a convex program, hence can be solved very efficiently.

Next, using the DistFlow equations, we can write the following optimization problem,

$$\begin{aligned} \text{(NL-OPT)} \quad \min \quad & \mathcal{F}(x) && \text{(8.24a)} \\ s.t. \quad & \mathcal{A}x \leq \mathcal{B} && \text{(8.24b)} \\ & \zeta(x) = \kappa && \text{(8.24c)} \end{aligned}$$

Here $\zeta(x) = \kappa$ contains full DistFlow (8.10a)-(8.10c), $\mathcal{F}(x)$ is an objective function, and $\mathcal{A}x \leq \mathcal{B}$ contains (8.13) and (8.15).

An LP-OPT-based iterative optimization based algorithm is proposed next to obtain the convex polytope of $P_1 - Q_1$ flexibility.

Algorithm-2.

S1. Construct the $\mathcal{A}, \mathcal{B}, \mathcal{H}, \mathcal{G}$ matrices.

S2. Solve LP-OPT with objectives $\min P_1$, $\max P_1$ to obtain $\underline{P_1}$ and $\overline{P_1}$.

S3. Vary P_1 from $\underline{P_1}$ to $\overline{P_1}$ with step size ΔP (total N_p levels).

S4. Solve LP-OPT for with objective $\min Q_1$ and $\max Q_1$ for each P_1 and record the optimal Q_1 values.

S5. Since the resulting set must be convex, take the convex hull of the $2N_p$ number of (P_1, Q_1) points to obtain the net flexibility polytope denoted by $\mathcal{X}^{\text{sub,w}}$.

Note that the step sizes, ΔP, can be adjusted based on the shape of $\mathcal{X}^{\text{sub,wo}}$.

8.4.5. Validation using NL-OPT

For validating the P-Q flexibility region obtained by LinDistFlow and LP-OPT, we can check if solution lies/ is feasible under a NL-DistFlow formulation (NL-OPT). For this, we can select N_r random points, $x^r = (\tilde{P}_1^r, \tilde{Q}_1^r)$, which are drawn inside $\mathcal{X}^{\text{sub,w}}$. For each point a feasibility problem is run using NL-OPT, which is also subject to the following constraints,

$$|P_1 - \tilde{P}_1^r| \leq \epsilon, \tag{8.25a}$$

$$|Q_1 - \tilde{Q}_1^r| \leq \epsilon. \tag{8.25b}$$

Here, $\epsilon \geq 0$ is a small constant (e.g. $10^{-}4$). Note that the absolute value constraints are first reformulated as linear constraints using standard techniques. If OPT-NL has a feasible solution, the test point produced by LP-OPT is an acceptable one. Thus, we compute the fraction of points to estimate accuracy. This serves as a Monte Carlo error estimate.

The inaccuracy appears mainly due to neglecting the loss terms and the non-linear terms $\beta = \frac{P_j^2 + Q_j^2}{V_j^2}$. However, in terms of computational speed, the benefits of using LP-OPT far out-weight the minor discrepancies. Furthermore, using a convex optimization guarantees solutions are connected, hence tracing the boundary is sufficient. However, if NL-OPT is

used, a similar iterative technique cannot be performed since the neighbouring region to a solution may contain holes. Besides validation, the NL-OPT solutions can be used to prune the net $P-Q$ polytope to improve accuracy, as will be shown in Section 8.5.

8.4.6. Applications of the Dual Variables from LP-OPT

The LP-OPT is further useful due to the availability of the dual variables, particularly for constraints (8.15) and (8.11c). Recall that the existence of a nonzero positive valued dual variable would suggest that the constraint is binding and would give the 'shadow price', hence suggest the value of adding additional resources to improve the objective function. The case study will demonstrate how the dual variables can be used to detect the nodes and branches that limit transfer of flexibility or have unusable flexible capacity.

8.4.7. Considering Uncertainty and Obtaining a Robust Set

Since solar power generation and response from controllable loads have uncertainties, a robust set, $\mathcal{X}^{\text{sub,r}}$ may be sought. While $\mathcal{X}^{\text{sub,r}}$ may be more conservative, it will provide feasibility guarantees. In this case, we apply the following modifications to the original LP-OPT.

Assume at an arbitrary node, for $\mathcal{X}_1(s_1, \underline{p}_1, \overline{p}_1)$, $\underline{p}_1$ is uniform within $[\underline{p}_1^-, \underline{p}_1^+]$ and $\overline{p}_1$ is uniform within $[\overline{p}_1^-, \overline{p}_2^+]$. Then, the robust set is given by, $\mathcal{X}_1^r(s_1, \underline{p}_2^+, \overline{p}_2^-)$. Similarly, for $\mathcal{X}_2(\underline{p}_2, \overline{p}_2)$, given $\underline{p}_2$ is uniform within $[\underline{p}_2^-, \underline{p}_2^+]$ and $\overline{p}_2$ is uniform within $[\overline{p}_2^-, \overline{p}_2^+]$, the robust set is given by, $\mathcal{X}_2^r(\underline{p}_2^+, \overline{p}_2^-)$. Similarly consider robust sets for all nodes. Then, following the procedures for nodal flexibility aggregation and LP-OPT, the robust net P_Q can be obtained. Examples are provided in Section 8.5.

8.4.8. DER Dispatch Scheme

Assume having access to the net $P-Q$ curves, the substation is willing to provide var support to the main grid. We explore if this can be achieved by simply by sending a common var adjustment signal to all DERs in the distribution feeder.

First, notice that linearization of (8.10c) w.r.t δQ_j gives [67],

$$\delta u_{j+1} = -2x\delta Q_j + \frac{P_j^2 + Q_j^2}{u_j}\delta Q_j. \tag{8.26}$$

For typical feeders, $Q_i \leq \frac{u_i}{2x_i}$ [67]. Hence, δu_{j+1} will decrease with increases in δQ_j. Next, see that (8.10b) are quadratic in Q_j, the solutions of which are of the form,

$$Q_j = \frac{v_j^2}{2x_j}\{1 \pm \zeta_2^{1/2}\}; \tag{8.27}$$

where $\zeta = \{1 - \frac{4x_j}{u_j}(Q_{j+1} + q_j + x_j P_j^2/u_j)\}$.

Then, linearization of (8.27) w.r.t. δ_{arg} ($= \delta Q_{j+1}$ or δq_j) gives,

$$\delta Q_j = \mp\{\zeta^{-1/2}\}\delta_{arg}; \tag{8.28}$$

Since $Q_j \leq \frac{u_j}{2x_j}$, (8.27) takes the negative root and hence (8.28) takes the positive. Therefore, changes in δQ_{j+1} or δq_j will result in change of same sign in Q_j.

This leads to the following local dispatch rule at node j at time t,

$$q_{j,t} = \begin{cases} q_{j,t-1}(1 + \bar{u}_t^q), \text{if (8.12) and } q_{j,k-1}(1 + \bar{u}_k^q) \in \mathcal{X}_j^{\text{node}} \\ q_{j,t-1}, \text{otherwise.} \end{cases} \tag{8.29}$$

where $\bar{u}_t^q$ is a small adjustment (in %) broadcast by the substation operator (DSO). A similar strategy can be derived for adjustment of active power set-points.

8.5. Numerical Results

8.5.1. Aggregation of DER Flexibility

To demonstrate the effectiveness of our proposed algorithm and to explain the impacts of various constraints, we formulate an NLP and LP using the IEEE-33 bus system data [9]. Fig. 8.8 shows the feeder. Every node of the system is populated with flexible DERs. Hence, instead of deterministic demands, we formulate nodal flexibility polytopes. For the base case (deterministic), consider $DR = 0.3$ and solar inverter rating, $S = 0.7$. For DR and S, as well as for solar availability forecast, we will also consider uncertainty.

Set, $V^{min} = 0.92$, $V^{max} = 1.04$ according to ANSI [23].

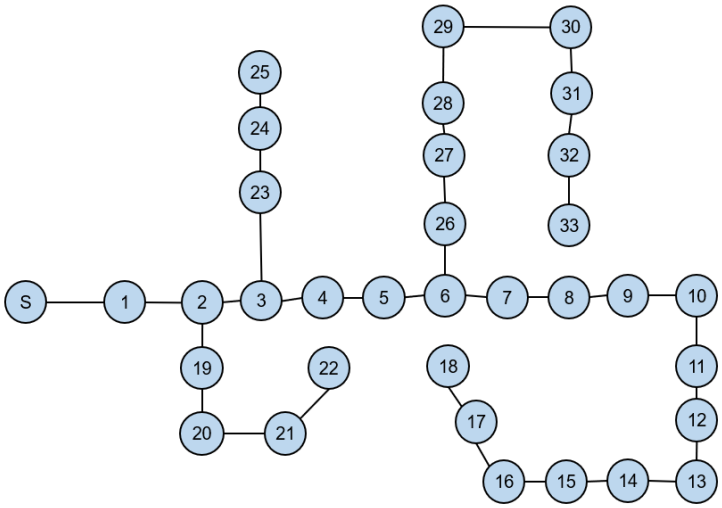

Figure 8.8: 33-bus feeder [9].

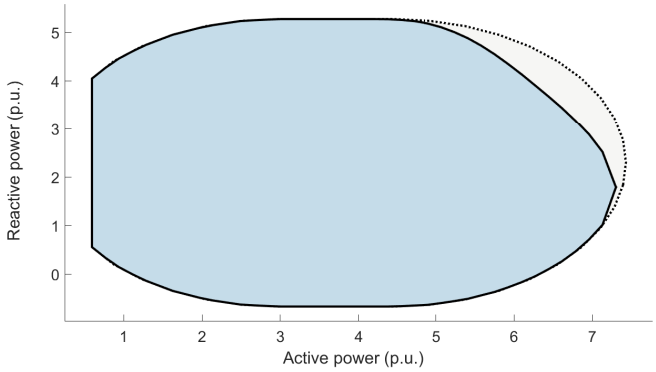

Figure 8.9: The net active and reactive power flexibility at node 1 using the IEEE 33-bus system.

8.5.2. Comparison with the NL-Distflow

The convex polytope $\mathcal{X}^{\text{sub,w}}$ approximated by the LinDisflow is fairly accurate in that almost 97.1 % of every point in that set contains is also a solution to the NL-OPT. From Fig. 8.10, we see that the points that do not satisfy the NL-OPT all lie near the boundaries, hence this suggests that the minor discrepancies are due to ignoring the loss terms. It intuitively makes sense that the infeasible points lie near the real and reactive power injection boudaries, and not near the absorption boundaries. This is because active and reactive losses over the lines are consumed, hence there is less amount of real and reactive

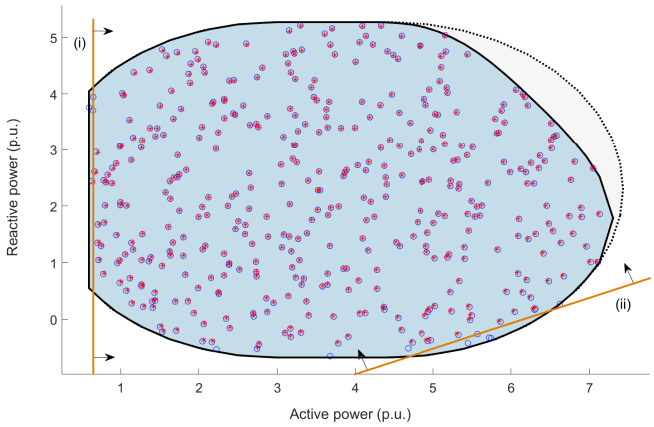

Figure 8.10: Validation of the net active and reactive power flexibility using NL-OPT and pruning to remove NL-OPT infeasible points using constraints (i) and (ii).

power injection (to grid) possible since a small percentage is lost. Using the identified infeasible samples, we can generate a more constrained convex set. For example, add pruning constraints that pass through an infeasible point and is parallel to the closest tangent hyperplane at boundary of $\mathcal{X}^{\text{sub,w}}$ (see (i) and (ii) shown in Fig. 8.10). The accuracy of this constrained flexibility polytope then improve even further, for example, reaches 99.2% when considering the further restricted set.

8.5.3. Transfer Limit and Usable Capacity from Dual Variable Values

Investigating the dual variables associated with (8.15) suggests that under all loading conditions, the constraints characterizing the upper right corners of $\mathcal{X}_i^{\text{node}}$ are almost never binding. Instead voltage constraints (8.11c) are binding and cut the feasible space. In all other regions of $\mathcal{X}_i^{\text{node}}$, the constraints can be binding depending on different objectives and loading conditions. Thus, having excessive reactive power consumption capability, when active power consumption is also high, is not useful. Such capability at nodes cannot propagate up to the substation due to network constraints. Given the network parameters, inverter sizing should therefore be done optimally, to avoid investing in unusable flexible capacity.

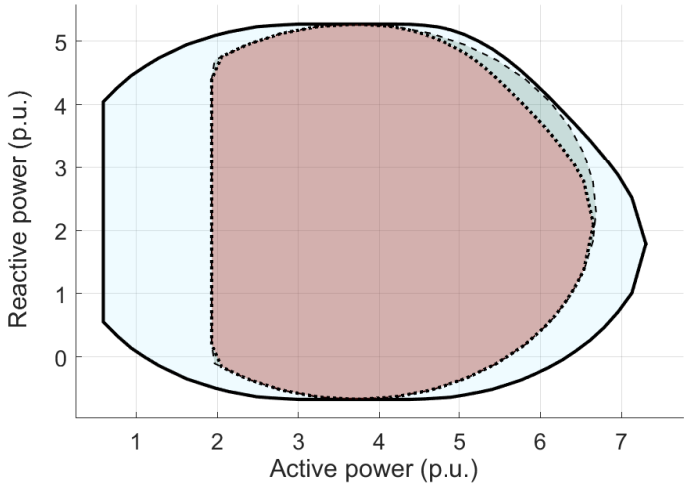

Figure 8.11: The net $P-Q$ flexibility at node 1 under three scenarios: (i) basecase (outer), (ii) moderate uncertainty in DR and solar injection (middle) and (iii) increased uncertainty in DR and solar injection (inner).

8.5.4. Robust Set Formation Considering Uncertainty

Three cases were considered, with increasing level of uncertainty in DR and solar power injection. The basecae (deterministic) and the robust sets are shown in Fig. 8.11. From the robust set we see that it is an inner approximation to the deterministic set, hence is safer and more conservative. Using sensitivity analysis one can identify how different uncertainty levels on different constraints may affect the substation's net $P-Q$ flexibility. A DSO can thus precompute and select a limited number of robust $P-Q$ polytopes for operations throughout different loading conditions, as well as DR and solar availability.

Finally, the dispatch scheme (8.29) has been tested on the 33-bus system for reactive power absorption from main grid, under three different loading (active power) conditions. The requested and provided amounts are shown in Fig. 8.12. Near nominal reactive power consumption levels, the slope is 1, but falls off as requested amount increases and different nodes' capabilities saturate. Eventually all nodes saturate, hence cannot absorb any more reactive power.

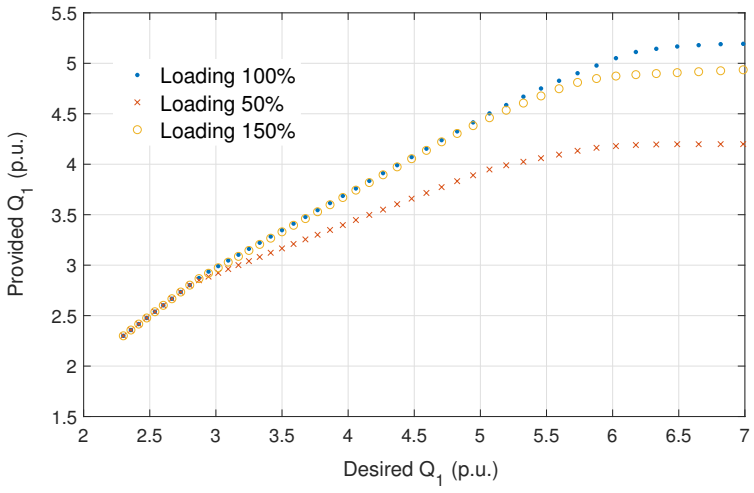

Figure 8.12: Increasing nodal reactive power consumption to absorb reactive power from the main grid (increase Q_1). Requested and provided amount using dispatch scheme (8.29)**, under three different loading (active power) conditions.**

8.5.5. Discussions

Regarding algorithm complexity, note that tracing the boundary requires solving $2N_p$ LPs. The convex hull computation is also simple in 2-D. To avoid frequent computation due to small changes in system conditions (e.g. solar irradiance and DR availability), DSOs can pre-compute and store a number of robust sets.

When the nodes have limited flexibility, the system is lightly loaded and voltage constraints are not active, the substation's net P-Q flexibility is usually the M-sum of the nodal flexibility sets, $\mathcal{X}^{\text{sub, wo}}$. Hence, all computation related to incorporating the network constraints can be safely avoided. Since the M-sum in several cases can be obtained algebraically, it is a fast and desirable technique. On the other hand, when system is heavily loaded and some network voltage constraints are active, the algebraic M-sum, $\mathcal{X}^{\text{sub, wo}}$, will not be the same as the true P-Q flexibility, $\mathcal{X}^{\text{sub, w}}$. $\mathcal{X}^{\text{sub, wo}}$ still serves as an outer-approximation to $\mathcal{X}^{\text{sub, w}}$.

Alternative approaches to compute the net substation P-Q flexibility often face practical challenges. For example, the central operators would need to know the user-equipment ratings and DR strategies to include those an the optimization problem. In the case of M-sum, the nodal flexibility polytopes are aggregate flexibility representations, hence users

might be more willing to share. Alternatively, these sets can be identified from data or using system identification techniques. Hence the our proposed framework would also fit data-driven models.

8.6. Chapter Conclusions

In this work, for radial feeders, we showed how nodal flexibility sets along with network constraints and voltage limits can be used to approximate the net P-Q flexibility curves at the substation. Due to a convex formulation, the dual variables indicate locations of unused flexible capacity in a network, hence can be used to restrict over-investment in unusable flexibility. Availability of net P-Q flexibility also allows applying DER dispatch rules, which are simple yet effective. A detailed case study is performed on a 33-bus test feeder.

Chapter 9.

Conclusions

To conclude this dissertation, a summary of the dissertation is presented, the main contributions of each chapter are summarized, and possible future research directions are discussed.

9.1. Dissertation Summary

The work on this dissertation focused on developing novel modeling, control, analysis and optimization techniques to facilitate large-scale integration of DERs in power systems.

Chapter 2 reviewed the mathematical models for capturing individual and aggregate TCL dynamics. Both coupled PDE model and Markov-chain-based state-bin transition model were presented. It was shown that, for homogeneous TCLs, the entries of the state-bin model can be obtained analytically using discretization of the PDE model. Since TCLs provide storage capabilities, a generalized battery-equivalent model has also been presented.

Chapter 3 developed analytically-based techniques to incorporate noise and heterogeneity in the Markov-chain model, i.e. the *state-bin model*. We found that the state-bin models inherently lack in their ability to accurately capture the dynamics of heterogeneous TCLs, which has been resolved by decomposing the population into multiple homogeneous groups. Using sensitivity analysis, we also captured the impacts of modeling error, noise and heterogeneity on aggregate dynamics of TCLs.

Chapter 4 presented simulations of market-based or Transactive energy coordination (TEC) and identified several cases showing undesirable behavior, such as emergence of temperature synchronization, exhibiting large oscillations in aggregate power and highly

fluctuating demand and prices. To address these challenges, we developed an extended state-bin model, which captures the TCL dynamics under TEC. An MPC was formulated, as a MIP/ a relaxed QP problem, to obtain the optimal price signals to coordinate the TCLs. A case study showed that a population of TCLs can be managed economically while avoiding congestion in a distribution grid. By considering additional costs/ constraints, synchronization of TCLs was also avoided.

Chapter 5 focused on modeling and analysis of load control algorithms using a generalized hybrid systems framework. Modal decomposition and study of the eigenvalues of the discretized hybrid system revealed when the aggregate TCL dynamics can undergo limit cycle oscillations or exhibit period-adding bifurcation. Synchronizing tendency and rate of converge to steady-state distributions were also studied for multiple control schemes. A candidate Lyapunov (energy) function has also been derived and used to examine if the trajectories of the controlled TCL ensembles remain stable.

Chapter 6 focused on quantifying and addressing performance limitations in TCL coordination when practical constraints are further imposed. First, we established performance limits for population of TCLs that are controlled via probabilistic dispatch for participating in fast energy balancing services. An analytical expression for expected RMS error due to probabilistic switching was also developed. Then we extended our analysis to aggregate bin-based models for which switching probabilities may be non-uniform. The close agreement between analytical and simulation-based results demonstrated the validity of the performance bounds and their dependence on target levels as well as population sizes. Alternative control techniques were also explored to reduce tracking error. Finally, we developed an 'explore-exploit'-controller, assuming TCL set-point information may not be directly available, hence must be estimated.

In **Chapter 7**, several techniques to compute and aggregate DER flexibility polytopes have been developed. To improve the accuracy of the inner-approximationing polytopes, we developed a novel technique mainly by considering unions of feasible intervals, homothets and/or zonotopes. The algorithm was shown to scale to higher dimensions (up to 8 dimensions were considered here), in which cases performing M-sums via other techniques is known to be extremely challenging. We analyzed the accuracy and performance tradeoffs compared to existing approaches and our algorithm showed superior performance. These flexibility polytopes can be efficiently integrated in optimal power flow formulations, economic dispatch problems or used for sampling feasible points for simulation based studies. A case study on voltage control using an inverter flexibility polytope was shown.

Chapter 8 then utilized the flexible polytopes of inverters and loads, developed in Chapter 7, to approximate the net aggregate real and reactive power flexibility available at a distribution system substation. For radial feeders, we showed how nodal flexibility sets along with network constraints and voltage limits can be used to approximate the net $P - Q$ flexibility curves. A convex optimization-based algorithm, was used to generate the polytope, whereas a non-linear counterpart was used to check feasibility and generate additional cuts to improve accuracy. The values of the constraints' dual variables were used to identify nodes with constrained or unused flexible capacity. Availability of net $P - Q$ flexibility polytope also allowed applying simple yet effective DER dispatch rules. Numerical results were shown on a 33-bus radial test feeder.

9.2. Summary of Contributions

This dissertation has made a number of contributions to the domains of ensemble TCL control, residential demand response, analysis of dynamical systems, computational geometry and optimization in power systems.

Chapter 2 contributed to the field of ensemble TCL control by summarizing the two fundamental aggregate modeling approaches, CPDE and Markov-chain based, and showing their equivalence for homogeneous TCLs. Chapter 3 contributed to the field of ensemble TCL control and residential demand response by capturing the effects of model error, uncertainty and heterogeneity in the widely used Markov-chain-based state-bin model of TCLs. Analytically-based numerical recipes are also provided for incorporating noise and heterogeneity in TCL bin-models. Chapter 4 contributed to the field of demand response by investigating and presenting unintended consequences of market-based or Transactive Energy Coordination (TEC). Chapter 4 then contributed towards the literature of ensemble load control by first showing how the price-response behavior of TCLs can be captured using an extended state-bin model. The model also took into account lock-out conditions. An accurate MIP formulation, and a fast relaxed QP formulation have been developed and tested. Chapter 5 contributed to the field of demand response by comparing several load control schemes using a common hybrid systems framework. Chapter 5 also contributed to analysis of dynamic systems, by showing how eigenvalues and eigen-modes of a discretized hybrid systems can be used to reveal nonlinear phenomenon such as limit cycle oscillations and period-adding bifurcation. With system matrices, which are state-constrained

and follow Markov-chain type behavior, a candidate Lyapunov function has been derived by using state reduction and transformation. Next, Chapter 6 contributed to the field of demand response by quantifying and addressing performance limitations when practical constraints are further imposed. Chapter 6 also contributed to the field of ensemble TCL control by developing a novel 'explore-exploit'-controller, assuming TCL set-point information is unknown/not readily available.

Next, Chapter 7 contributed to the field to computational geometry by overcoming the limitations in existing methods to compute inner-approximating polytopes using Minkowski-sum (M-sum), especially when dealing with heterogeneous shapes and problems in higher dimensions (greater than two or three). Chapter 7 also contributed to the field of optimization tools for power systems since the aggregate flexibility polytopes can be efficiently integrated in the optimal power flow (OPF) or economic dispatch (ED) problems. Recall that a single aggregate polytope replaces all individual DER dynamic constraints and limits on their state variables - hence OPF and ED can be solved very efficiently [15, 78]. Finally, Chapter 8 contributed to the field of optimization in power systems by providing numerical recipes to compute and validate net DER flexibility in distribution test systems while considering network constraints. Such flexibility will allow efficient coordination between the operators of distribution and transmission systems.

Overall, the work presented in this dissertation identified and filled important gaps in literature, and opened several avenues of future research related to ensemble load control and large scale DER integration. The numerical examples improved our understandings of the opportunities and limitations when using various schemes for controlling loads or techniques to aggregate DER flexibility. The work presented in this thesis can also have broader impact on our future energy systems, mainly by providing efficient and scalable tools to enable integration of large-scale distributed and renewable energy resources.

9.3. Future Directions

Several potential avenues of future research are described below,

1. Data-driven TCL modeling: In literature, the TCL parameters governing its dynamics are typically assumed to be fixed or are drawn from uniform or Gaussian distributions. However, in reality, the parameter distributions may vary considerably, hence should be estimated from data from real households. Since the parameter values may also change

with ambient conditions, one should identify how often or when to update the parameter values to ensure the models remain valid, i.e. can capture realistic behavior.

2. Impact of TCLs on Distribution Systems: As shown in Chapter 5, sycnhronized TCLs were shown to cause voltage violations in a 33-node distribution system. A similar, but more detailed investigation on prototypical distribution feeders should be performed to assess the probability of extreme events occurring due to synchronized TCLs, and capture other possible detrimental behavior.

3. Control schemes considering network models: The TCL control schemes discussed in Chapters 1- 5 did not consider network models. Hence, future research should investigate how network models and constraints can be taken into account without adding significant complexity. Techniques based on online convex optimization seem promising [57, 107], hence would be worth investigating.

4. Chance-constrained flexibility sets: In Chapters 7 and 8, we showed how to compute inner-approximating feasible sets or polytopes, and also showed how to obtain a conservative robust set when parameters follow a uniform distribution. A more general setting can be considered using chance-constraints. Instead of deterministic or robust sets, chance-constrained sets will allow system operators to balance tradeoffs between the flexibility that can be harnessed from the DERs versus the risk of choosing an infeasible operating point.

5. TSO-DSO Coordination techniques using flexibility polytopes: Transmission system operators (TSOs) typically treat active and reactive power demand at a distribution system substation to be a known quantity. The availability of real-reactive power polytopes at the distribution substations can cause a paradigm shift in how TSOs-DSOs would coordinate. Developing novel coordination mechanisms, designing new market rules and assessing the advantages/ disadvantages of different approaches, all would be important research directions and can have significant practical impacts.

Appendix A.

A Small-scale Example of a Bin-Model

Let $N_B = 2$ (Fig. A.1) and α_0, α_1 be the heating and cooling rates. Then, $\dot{x} = \mathcal{A}x$ with,

$$\mathcal{A} = \begin{bmatrix} -\alpha_0 & 0 & 0 & -\alpha_1 \\ \alpha_0 & -\alpha_0 & 0 & 0 \\ 0 & \alpha_0 & \alpha_1 & 0 \\ 0 & 0 & -\alpha_1 & \alpha_1 \end{bmatrix}. \tag{A.1}$$

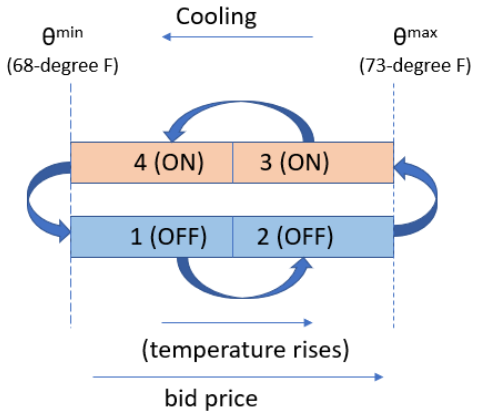

Figure A.1: Model with 4 bins.

The reset B-matrices with increasing clearing bin boundaries are given by,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{A.2}$$

Appendix B.

Counter-example on Heterogeneity

The following serves as a counter-example to prove that Markov property is not Satisfied when considering heterogeneity in a Bin-Model.

Consider total N_{TCL} TCLs are divided in two groups, each consisting half. In one time-step, TCLs in group 1 move exactly 1 bin, and TCLs in group 2 move exactly 2 bins. Hence, the transition matrices are given by,

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A_2 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{B.1}$$

Then, assume $X_1[0] = [1\ 0\ 0\ 0]^\top$ and $X_2[0] = [0\ 1\ 0\ 0]^\top$.
After one time-period, $X_1[1] = A_1 X_1[0] = [0\ 1\ 0\ 0]^\top$ and $X_2[1] = A_2 X_2[0] = [0\ 0\ 1\ 0]^\top$.
Overall, $\hat{X}[1] = X_1[1] + X_2[1] = [0\ 1\ 0\ 1]^\top$ Now, assume A_1 and A_2 are not known. Instead the following A, obtained by system identification and averaging (see Chapter 3), is used,

$$A = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \end{bmatrix}. \tag{B.2}$$

With $X[0] = [1\ 1\ 0\ 0]^\top$, after one time-period, $X[1] = AX[0] = [0\ 0.5\ 1\ 0.5]^\top$. The predicted $X[1]$ is not equal to actual $\hat{X}[1]$. Thus, the Markov property fails to hold unless groups of homogeneous TCLs and their associated system matrices are used.

Appendix C.

Candidate Lyapunov Function Derivation

Recall that the TCL population mode is given by $\dot{x} = \mathcal{A}x$. In this appendix, through state reduction and transformation, a candidate Lyapunov function associated with $\mathcal{A}$ is derived.

C.1. State Reduction and Transformations

Note that due to probability conservation, $\mathbf{1}^\top x = 1$, where $x \in \mathbb{R}^n$. Hence, $x_n = 1 - \mathbf{1}^\top \tilde{x}$, where $\tilde{x} = [x_1\ x_2\ ...\ x_{n-1}]^\top$, i.e. $\tilde{x} \in \mathbb{R}^{(n-1)}$ and $\dot{x_n} = -\mathbf{1}^\top \dot{\tilde{x}}$.
Recall that $x_{\rm s}$ is the steady-state eigen-vector [1] associated with $\mathcal{A}$. Let $z \in \mathbb{R}^n$ s.t. $z = x - x_{\rm s}$ and let the reduced state vector, $\tilde{z} = \tilde{x} - \tilde{x}_{\rm s}$, where $\tilde{x}_{\rm s} = [\tilde{x}_1\ \tilde{x}_2\ ...\ \tilde{x}_{n-1}]^\top$. Then, we obtain

$$\dot{z} = \mathcal{A}z, \tag{C.1a}$$

$$\dot{\tilde{z}} = \tilde{\mathcal{A}}\tilde{z}. \tag{C.1b}$$

where $\tilde{\mathcal{A}} \in \mathbb{R}^{(n-1)\times(n-1)}$. Assuming $\mathcal{A}$-matrix is given by (2.18)-(2.19), $\tilde{\mathcal{A}}$-matrix can be obtained in two steps: (i) setting $\tilde{\mathcal{A}}$ as the upper-left $(n-1)\times(n-1)$ sub-matrix of $\mathcal{A}$. (ii) since $x_n = 1 - \mathbf{1}^\top \tilde{x}$, setting the first row of $\tilde{\mathcal{A}}$ as $[\frac{-\alpha_0+\alpha_1}{\Delta^{\rm bw}} \quad \mathbf{1}^{1\times(n-1)}\frac{\alpha_1}{\Delta^{\rm bw}}]$. With (C.1), both z and $\tilde{z}$ now have equilibrium at zero.

[1] The eigen-vector associated with the single eigenvalue at zero of A is referred to as the steady-state eigen-vector.

Using $\tilde{x}$ instead of x, $\dot{x} = \mathcal{A}x$ can now be written as

$$\dot{\tilde{x}} = \tilde{\mathcal{A}}\tilde{x} + b, \tag{C.2}$$

where b is a vector of constant terms. Since $\tilde{x} = \tilde{z} + \tilde{x}_{\mathrm{s}}$, we have $\dot{\tilde{x}} = \dot{\tilde{z}} = \tilde{\mathcal{A}}\tilde{z} = \tilde{\mathcal{A}}\tilde{x} - \tilde{\mathcal{A}}\tilde{x}_{\mathrm{s}}$. Thus, $b = -\tilde{\mathcal{A}}\tilde{x}_{\mathrm{s}}$.

Since $x_n = 1 - \mathbf{1}^\top \tilde{x}$, an affine transformation of $\tilde{x}$ gives x in the following manner,

$$x = K\tilde{x} + K_0, \tag{C.3}$$

where

$$K = \begin{bmatrix} \mathbf{I}^{(n-1)\times(n-1)} \\ -\mathbf{1}^{1\times(n-1)} \end{bmatrix}, \quad K_0 = \begin{bmatrix} \mathbf{0}^{n-1} \\ 1 \end{bmatrix},$$

Additionally, since $\tilde{x} = \tilde{z} + \tilde{x}_{\mathrm{s}}$, we can write

$$x = K\tilde{z} + K\tilde{x}_{\mathrm{s}} + K_0. \tag{C.4}$$

z can also be obtained from $\tilde{z}$ using $z = K\tilde{z}$. K_0 vanishes since $\mathbf{1}^\top x_{\mathrm{s}} = 1$. Notice that K is not a square matrix. We can obtain a square matrix, K_1, by appending a zero column (has no effect on the dynamics) as

$$K_1 = \begin{bmatrix} K & 0 \\ -\mathbf{1}^{1\times(n-1)} & 0 \end{bmatrix}. \tag{C.5}$$

C.2. Obtaining a Candidate Lyapunov Function

Assume $V(\tilde{z}) = \tilde{z}^\top \tilde{P}\tilde{z}$, where $\tilde{P} = \tilde{P}^\top$ and $\tilde{P} > 0$. Here, $\tilde{P} \in \mathbb{R}^{(n-1)\times(n-1)}$. $V(\tilde{z}) = 0$ for $\tilde{z} = 0$ and $V(\tilde{z}) > 0$ for all $\tilde{z} \neq 0$. Then, $\dot{V}(\tilde{z}) = \tilde{z}^\top(\tilde{\mathcal{A}}^\top\tilde{P} + \tilde{P}\tilde{\mathcal{A}})\tilde{z}$ [48].
Solving $(\tilde{\mathcal{A}}^\top\tilde{P} + \tilde{P}\tilde{\mathcal{A}}) = -\mathbf{I}$ ensures that $\dot{V}(\tilde{z}) < 0$ for all $\tilde{z} \neq 0$. Here, $\mathbf{I}$ is an identity matrix of order $(n-1)\times(n-1)$.
Next, by applying the similar operation to convert $\tilde{x}$ to x, we can obtain a P matrix of dimension $(n \times n)$ as,

$$P = \frac{P_1 + P_1^\top}{2}, \tag{C.6}$$

where

$$P_1 = \begin{bmatrix} \tilde{P} & 0 \\ -\mathbf{1}^{1\times(n-1)} & 0 \end{bmatrix} \tag{C.7}$$

Finally, we obtain $V(x)$ using the P from (C.6) as,

$$V(x) = (x - x_{\mathrm{s}})^{\top} P (x - x_{\mathrm{s}}), \tag{C.8}$$

Remark 1: With (C.6), $P = P^{\top}$ and $P > 0$. The eigenvalues of P remain identical to those of $\tilde{P}$, except there is an additional eigenvalue at 0.

Remark 2: Since $z = x - x_{\mathrm{s}}$, $V(z) > 0$, $\dot{V}(z) < 0$, for all $z \neq 0$, and $\dot{V}_z = \dot{V}_x$ (by chain rule), $V(x)$ given by (C.8) is a valid Lyapunov function for (5.1a) (with $\mathcal{A}$-matrix time-invariant).

Appendix D.

Supporting Simulation Results for Chapter 6

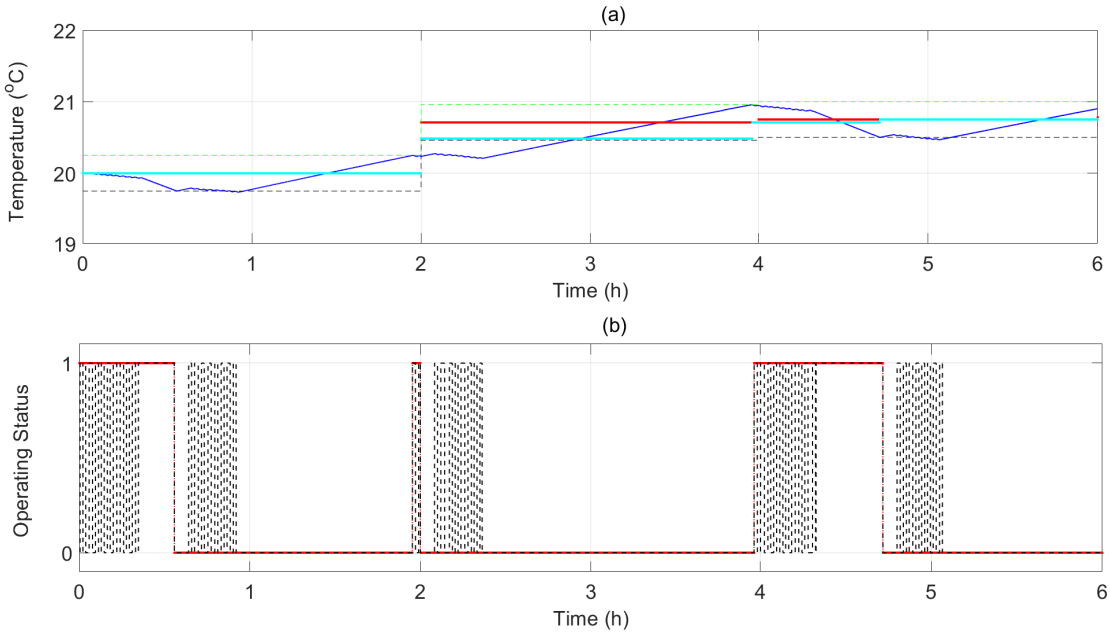

Figure D.1: (a) Temperature and set-point evolution, (b) Thermostat's internal cooling/ off mode and TCU's on/off/explore modes.

Table D.1: Simulation Results with the Explore-Exploit Controller

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Δ	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
τ^1	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	30.0	30.0	30.0	30.0
τ^0	300.0	300.0	300.0	300.0	300.0	300.0	300.0	300.0	300.0	300.0	120.0	300.0	120.0	120.0
Mean θ^{amb}	31.5	31.6	31.6	31.7	31.5	31.6	31.7	31.6	31.6	31.6	31.7	31.7	31.6	31.5
λ^{10}	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3	0.3	0.3
λ^{01}	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.3	0.3
$\tau^{explore,thresh}$ (mins)	15.8	15.8	22.6	22.6	29.4	29.4	38.4	38.4	45.1	45.1	22.6	38.4	45.1	45.1
Frac. of time in exploration	0.7	0.7	0.7	0.6	0.6	0.6	0.5	0.5	0.4	0.5	0.7	0.5	0.4	0.4
Mean set-point change (^{o}C)	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.0	0.2	0.1	0.2	0.1	0.2	0.0
Average switchings per hr	5.5	6.2	6.0	7.8	8.2	7.8	9.8	10.5	11.2	10.8	15.5	10.3	27.0	25.2

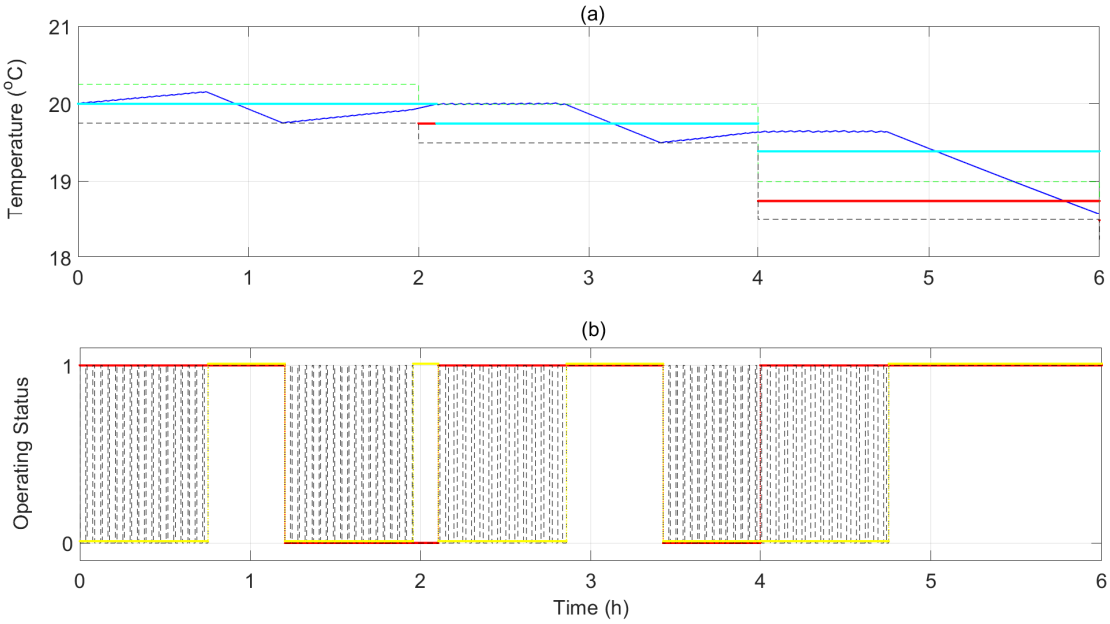

Figure D.2: (a) Temperature and set-point evolution, (b) Thermostat's internal cooling/ off mode and TCU's on/off/explore modes.

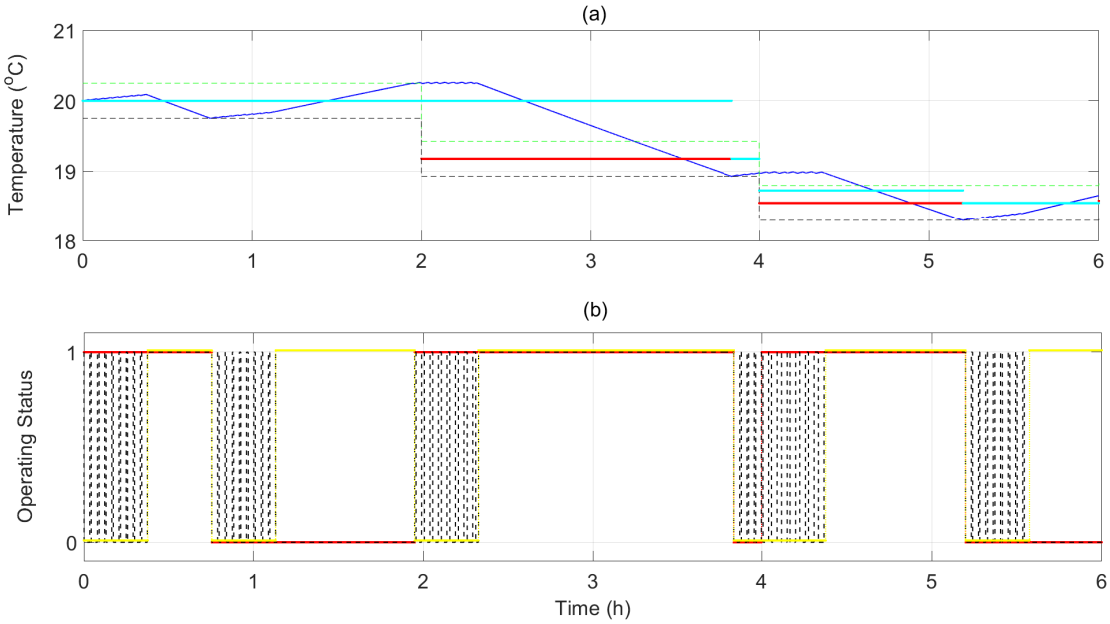

Figure D.3: (a) Temperature and set-point evolution, (b) Thermostat's internal cooling/ off mode and TCU's on/off/explore modes.

Bibliography

[1] M. Alizadeh, A. Scaglione, A. Applebaum, G. Kesidis, and K. Levitt. "Reduced-Order Load Models for Large Populations of Flexible Appliances". In: *IEEE Transactions on Power Systems* 30.4 (2015), pp. 1758–1774. arXiv: arXiv:1404.1958v1.

[2] M. Alizadeh, A. Scaglione, A. Goldsmith, and G. Kesidis. "Capturing aggregate flexibility in Demand Response". In: *Proceedings of the IEEE Conference on Decision and Control*. 2014.

[3] M. Althoff and B. H. Krogh. "Zonotope bundles for the efficient computation of reachable sets". In: *Proceedings of the IEEE Conference on Decision and Control*. 1. 2011, pp. 6814–6821.

[4] F. Alvarado. "The stability of power system markets". In: *IEEE Transactions on Power Systems* 14.2 (1999), pp. 505–511. ISSN: 08858950.

[5] and and and. "On the accuracy of the state-queuing model for the thermostatically controlled loads". In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. Oct. 2016, pp. 1990–1994.

[6] S. E. I. Association. *U.S. Solar Market Insight*. Mar. 2019. URL: https://www.seia.org/us-solar-market-insight.

[7] D. Avis. "On the complexity of finding the convex hull of a set of points". In: *Discrete Applied Mathematics* 4.2 (1982), pp. 81–86.

[8] L. Bai, J. Wang, C. Wang, C. Chen, and F. F. Li. "Distribution Locational Marginal Pricing (DLMP) for Congestion Management and Voltage Support". In: *IEEE Transactions on Power Systems* 99 (2017).

[9] M. E. Baran and F. F. Wu. "Network reconfiguration in distribution systems for loss reduction and load balancing". In: *IEEE Trans. on Power Delivery* 4.2 (Apr. 1989), pp. 1401–1407.

[10] S. Barot and J. A. Taylor. "A concise , approximate representation of a collection of loads described by polytopes". In: *International Journal of Electrical Power and Energy Systems* 84 (2017), pp. 55–63.

[11] S. Bashash and H. K. Fathy. "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads". In: *2011 American Control Conference*. 2011.

[12] S. Bashash and H. K. Fathy. “Modeling and control of aggregate air conditioning loads for robust renewable power management”. In: *IEEE Transactions on Control Systems Technology* 21.4 (2013), pp. 1318–1327.

[13] A. Bernstein, J. Y. L. Boudec, M. Paolone, L. Reyes-Chamorro, and W. Saab. “Aggregation of power capabilities of heterogeneous resources for real-time control of power grids”. In: *2016 Power Systems Computation Conference (PSCC)*. June 2016, pp. 1–7.

[14] A. Bernstein, N. J. Bouman, and J.-Y. Le Boudec. “Design of Resource Agents with Guaranteed Tracking Properties for Real-Time Control of Electrical Grids”. In: (2015). [Online] Available at: http://arxiv.org/abs/1511.08628.

[15] A. Bernstein and E. Dall’Anese. “Real-Time Feedback-Based Optimization of Distribution Grids: A Unified Approach”. In: *IEEE Transactions on Control of Network Systems* (2019).

[16] A. Bušić and S. Meyn. “Distributed Randomized Control for Demand Dispatch”. In: *55th Conference on Decision and Control (CDC)*. Las Vegas, 2016, pp. 6964–6971.

[17] D. S. Callaway. “Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy”. In: *Energy Conversion and Management* 50.5 (2009), pp. 1389–1400.

[18] D. S. Callaway and I. A. Hiskens. “Achieving Controllability of Electric Loads”. In: *Proceedings of the IEEE* (2011).

[19] G. Casella and R. L. Berger. *Statistical inference*. 2nd. 2002. arXiv: 9605103 [cs].

[20] C. Chong and A. Debs. “Statistical synthesis of power system functional load models”. In: *IEEE Conference on Decision and Control*. Vol. 18. 1979, pp. 264–269.

[21] S. J. Crocker and J. L. Mathieu. “Adaptive state estimation and control of thermostatic loads for real-time energy balancing”. In: *2016 American Control Conference (ACC)*. July 2016, pp. 3557–3563.

[22] E. Dall’Anese and A. Simonetto. “Optimal Power Flow Pursuit”. In: *IEEE Trans. on Smart Grid* (May 2016).

[23] E. Dall’Anese, S. V. Dhople, B. B. Johnson, and G. B. Giannakis. “Decentralized optimal dispatch of photovoltaic inverters in residential distribution systems”. In: *IEEE Transactions on Energy Conversion* 29.4 (2014), pp. 957–967.

[24] D. Docimo and H. Fathy. “Demand response using heterogeneous thermostatically controlled loads: Characterization of aggregate power dynamics”. In: *Journal of Dynamic Systems, Measurement and Control* 139.10 (2017), pp. 1–9. ISSN: 15289028.

[25] D. J. Docimo and H. K. Fathy. “Characterization of Damping and Beating Effects Within the Aggregate Power Demand of Heterogeneous Thermostatically Controlled Loads”. In: *ASME 2015 Dynamic Systems and Control Conference*. Columbus, OH, 2015.

[26] A. Dubey and S. Santoso. "Electric Vehicle Charging on Residential Distribution Systems: Impacts and Mitigations". In: *IEEE Access* 3 (2015), pp. 1871–1893.

[27] *Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States*. Tech. rep. NREL/TP-6A20-71500. Golden, CO: NREL, 2018.

[28] U. D. of Energy. *Wind Vision: A New Era for Wind Power in the United States*. Tech. rep. DOE/GO-102015-4557. Apr. 2015.

[29] *ERCOT Hourly Load Data*. URL: http://www.ercot.com/gridinfo/load/load%7B%5C_%7Dhist/.

[30] S. Esmaeil Zadeh Soudjani and A. Abate. "Aggregation and Control of Populations of Thermostatically Controlled Loads by Formal Abstractions". In: *IEEE Transactions on Control Systems Technology* 23.3 (2015), pp. 975–990. arXiv: 1307.6716.

[31] K. Fukuda. "From the zonotope construction to the Minkowski addition of convex polytopes". In: *Journal of Symbolic Computation* 38.4 (2004), pp. 1261–1272.

[32] J. C. Fuller, K. P. Schneider, and D. Chassin. "Analysis of residential demand response and double-auction markets". In: *IEEE Power and Energy Society General Meeting*. 2011, pp. 1–7.

[33] A. Ghaffari, S. Moura, and M. Krstic. "Analytics Modeling and Integral Control of Heterogeneous Thermostatically Controlled Load Populations". In: *Proceedings of ASME 2014 Dynamic Systems and Control Conference (DSCC 2014)*. 2014.

[34] A. Ghaffari, S. Moura, and M. Krstic. "Modeling, Control, and Stability Analysis of Heterogeneous Thermostatically Controlled Load Populations Using Partial Differential Equations". In: *Journal of Dynamic Systems, Measurement, and Control* 137.10 (2015), p. 101009.

[35] M. Grant and S. Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. http://cvxr.com/cvx. Mar. 2014.

[36] S. Hanif, H. B. Gooi, T. Massier, T. Hamacher, and T. Reindl. "Distributed Congestion Management of Distribution Grids under Robust Flexible Buildings Operations". In: *IEEE Transactions on Power Systems* 32.6 (2017), pp. 1–1.

[37] S. Hanif, T. Massier, H. Gooi, T. Hamacher, and T. Reindl. "Cost Optimal Integration of Flexible Buildings in Congested Distribution Grids". In: *IEEE Transactions on Power Systems* 32.3 (2017), pp. 2254–2266.

[38] H. Hao, C. D. Corbin, K. Kalsi, and R. G. Pratt. "Transactive Control of Commercial Buildings for Demand Response". In: *IEEE Transactions on Power Systems* 32.1 (Jan. 2017), pp. 774–783.

[39] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent. "Aggregate Flexibility of Thermostatically Controlled Loads". In: *IEEE Transactions on Power Systems* 30.1 (Jan. 2015), pp. 189–198.

[40] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent. "A generalized battery model of a collection of Thermostatically Controlled Loads for providing ancillary service". In: *2013 51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013*. 2013, pp. 551–558.

[41] M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari. "Multi-Parametric Toolbox 3.0". In: *European Control Conference (ECC)*. 2013.

[42] K. Hreinsson, A. Scaglione, and V. Vittal. "Aggregate load models for demand response: Exploring flexibility". In: *2016 IEEE Global Conference on Signal and Information Processing*.

[43] P. Huang, J. Kalagnanam, R. Natarajan, M. Sharma, R. Ambrosio, D. Hammerstrom, and R. Melton. "Analytics and Transactive Control Design for the Pacific Northwest Smart Grid Demonstration Project". In: *IEEE SmartGridComm*. 2010, pp. 449–454.

[44] S. Ihara and F. C. Schweppe. "Physically based modeling of cold load pickup". In: *IEEE Transactions on Power Apparatus and Systems* PAS-100.9 (1981), pp. 4142–4150.

[45] M. D. Ilić, L. Xie, and J. Y. Joo. "Efficient Coordination of Wind Power and Price-Responsive Demand—Part I: Theoretical Foundations". In: *IEEE Transactions on Power Systems* 26.4 (2011), pp. 1875–1884.

[46] C. ISO. *Fast Facts: What the duck curve tells us about managing a green grid*. 2016. URL: https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf.

[47] E. C. Kara, Z. Kolter, M. Berges, B. Krogh, G. Hug, and T. Yuksel. "A moving horizon state estimator in the control of thermostatically controlled loads for demand response". In: *2013 IEEE International Conference on Smart Grid Communications*. Vancouver, 2013, pp. 253–258.

[48] H. K. Khalil. *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice Hall, 2002.

[49] A. C. Kizilkale and R. P. Malhame. "Collective Target Tracking Mean Field Control for Markovian Jump-Driven Models of Electric Water Heating Loads". In: *Proceedings of the IFAC World Congress*. 2014.

[50] J. Knudsen, J. Hansen, and A. M. Annaswamy. "A Dynamic Market Mechanism for the Integration of Renewables and Demand Response". In: *IEEE Transactions on Control Systems Technology* 24.3 (2016), pp. 940–955. ISSN: 10636536.

[51] S. Koch, J. L. Mathieu, and D. S. Callaway. "Modeling and Control of Aggregated Heterogeneous Thermostatically Controlled Loads for Ancillary Services". In: *Proceedings of the 17th Power Systems Computation Conference*. 2011.

[52] S. Kundu, K. Kalsi, and S. Backhaus. "Approximating Flexibility in Distributed Energy Resources: A Geometric Approach". In: *2018 Power Systems Computation Conference (PSCC)*. 2018.

[53] S. Kundu and I. A. Hiskens. "Nonlinear Dynamics of Hysteresis-Based Load Controls". In: *IFAC Proceedings Volumes* 47.3 (2014). 19th IFAC World Congress, pp. 5419–5425.

[54] S. Kundu and N. Sinitsyn. "Safe Protocol for Controlling Power Consumption by a Heterogeneous Population of Loads". In: *2012 American Control Conference* July (2012), pp. 2947–2952. arXiv: arXiv:1207.1124v1.

[55] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens. "Modeling and Control of Thermostatically Controlled Loads". In: *Power Systems Computation Conference*. Stockholm: IEEE, 2011.

[56] G. S. Ledva, E. Vrettos, S. Mastellone, G. Andersson, and J. L. Mathieu. "Managing Communication Delays and Model Error in Demand Response for Frequency Regulation". In: *IEEE Transactions on Power Systems* 33.2 (Mar. 2018), pp. 1299–1308.

[57] A. Lesage-Landry and J. A. Taylor. "Setpoint Tracking With Partially Observed Loads". In: *IEEE Transactions on Power Systems* 33.5 (Sept. 2018), pp. 5615–5627.

[58] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. RI Providence: American Mathematical Society, 2009.

[59] R. Li, Q. Wu, and S. S. Oren. "Distribution locational marginal pricing for optimal electric vehicle charging management". In: *IEEE Transactions on Power Systems* 29.1 (2014).

[60] S. Li, W. Zhang, J. Lian, and K. Kalsi. "Market-Based Coordination of Thermostatically Controlled Loads—Part I: A Mechanism Design Formulation". In: *IEEE Transactions on Power Systems* 31.2 (2016), pp. 1170–1178.

[61] S. Li, W. Zhang, J. Lian, and K. Kalsi. "Market-Based Coordination of Thermostatically Controlled Loads—Part II: Unknown Parameters and Case Studies". In: *IEEE Transactions on Power Systems* 31.2 (2016), pp. 1–9.

[62] J.-M. Lien. "Point-Based Minkowski Sum Boundary". In: *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications, Washington, DC*. 2007.

[63] H. Lin and P. J. Antsaklis. "Stability and Stabilizability of Switched Linear Systems: A Survey of Recent Results". In: *IEEE Transactions on Automatic Control* 54.2 (Feb. 2009), pp. 308–322.

[64] M. Liu and Y. Shi. "Model Predictive Control of Aggregated Heterogeneous Second-Order Thermostatically Controlled Loads for Ancillary Services". In: *IEEE Transactions on Power Systems* 31.3 (2015), pp. 1–9.

[65] J. Lofberg. “YALMIP : a toolbox for modeling and optimization in MATLAB”. In: *2004 IEEE International Conference on Robotics and Automation*. 2004, pp. 284–289.

[66] R. Malhame and C.-Y. Chong. “Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system”. In: *IEEE Transactions on Automatic Control* 30.9 (Sept. 1985), pp. 854–860.

[67] J. A. Martin and I. A. Hiskens. “Reactive power limitation due to wind-farm collector networks”. In: *2015 IEEE Eindhoven PowerTech*. June 2015, pp. 1–6.

[68] J. Mathieu, M. Kamgarpour, J. Lygeros, G. Andersson, and D. Callaway. “Arbitraging intraday wholesale energy market prices with aggregations of thermostatic loads”. In: *IEEE Transactions on Power Systems* 30.2 (Mar. 2015), pp. 763–772.

[69] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway. “Energy Arbitrage with Thermostatically Controlled Loads”. In: *European Control Conference*. Zurich, 2013, pp. 2519–2526.

[70] J. L. Mathieu, S. Koch, and D. S. Callaway. “State estimation and control of electric loads to manage real-time energy imbalance”. In: *IEEE Transactions on Power Systems* 28.1 (2013), pp. 430–440.

[71] D. Métivier, I. Luchnikov, and M. Chertkov. “Power of Ensemble Diversity and Randomization for Energy Aggregation”. In: *Scientific Reports* 9.1 (2019), p. 5910. ISSN: 2045-2322.

[72] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. 2009.

[73] R. E. Mortensen and K. P. Haggerty. “Stochastic computer model for heating and cooling loads”. In: *IEEE Transactions on Power Systems* 3.3 (1988), pp. 1213–1219.

[74] F. L. Muller, O. Sundstrom, J. Szabo, and J. Lygeros. “Aggregation of Energetic Flexibility Using Zonotopes”. In: *Proceedings of the IEEE Conference on Decision and Control*. Osaka, 2015.

[75] F. L. Muller, J. Szabo, O. Sundstrom, and J. Lygeros. “Aggregation and Disaggregation of Energetic Flexibility from Distributed Energy Resources”. In: *IEEE Transactions on Smart Grid* (). to appear.

[76] M. S. Nazir and I. A. Hiskens. “A Dynamical Systems Approach to Modeling and Analysis of Transactive Energy Coordination”. In: *IEEE Transactions on Power Systems* 34.5 (Sept. 2019), pp. 4060–4070.

[77] M. S. Nazir and I. A. Hiskens. “Analysis of Synchronization in Load Ensembles”. In: *Electric Power Systems Research, S.I.: 21st Power Systems Computation Conference, 2020* (To appear).

[78] M. S. Nazir, I. A. Hiskens, A. Bernstein, and E. Dall'Anese. "Inner Approximation of Minkowski Sums: A Union-Based Approach and Applications to Aggregated Energy Resources". In: *2018 IEEE Conference on Decision and Control (CDC)*. Dec. 2018, pp. 5708–5715.

[79] M. S. Nazir, F. D. Galiana, and A. Prieur. "Unit Commitment Incorporating Histogram Control of Electric Loads With Energy Storage". In: *IEEE Transactions on Power Systems* 31.4 (2016), pp. 2857–2866.

[80] M. S. Nazir and I. A. Hiskens. "Load Synchronization and Sustained Oscillations Induced by Transactive Control". In: *IEEE Power and Energy Society General Meeting*. 2017.

[81] M. S. Nazir and I. A. Hiskens. "Noise and Parameter Heterogeneity in Aggregate Models of Thermostatically Controlled Loads". In: *Proceedings of the 20th IFAC World Congress*. 2017, pp. 8888–8894.

[82] S. Nazir, S. C. Ross, J. L. Mathieu, and I. A. Hiskens. "Performance Limits of Thermostatically Controlled Loads under Probabilistic Switching". In: *Proceedings of the 20th IFAC World Congress*. 2017, pp. 8873–8880.

[83] *New record-breaking year for Danish wind power*. Jan. 2016. URL: https://en.wikipedia.org/wiki/Wind_power_in_Denmark.

[84] D. Paccagnan, M. Kamgarpour, and J. Lygeros. "On the Range of Feasible Power Trajectories for a Population of Thermostatically Controlled Loads". In: *IEEE 54th Annual Conference on Decision and Control*. 2015.

[85] K. X. Perez, W. J. Cole, J. D. Rhodes, A. Ondeck, M. Webber, M. Baldea, and T. F. Edgar. "Nonintrusive disaggregation of residential air-conditioning loads from sub-hourly smart meter data". In: *Energy and Buildings* 81 (2014), pp. 316–325. arXiv: arXiv:1502.03908v1.

[86] C. N. Perfumo. "Dynamic modelling and control of heterogeneous populations of thermostatically controlled loads". Ph.D. University of Newcastle, 2013.

[87] C. Perfumo, J. Braslavsky, J. K. Ward, and E. Kofman. "An analytical characterisation of cold-load pickup oscillations in thermostatically controlled loads". In: *2013 3rd Australian Control Conference*. 2013, pp. 195–200.

[88] *Renewables 2018 Global Status Report*. Tech. rep. Renewable Energy Policy Network for the 21st Century, 2018.

[89] M. Roozbehani, M. A. Dahleh, and S. K. Mitter. "Volatility of power grids under real-time pricing". In: *IEEE Transactions on Power Systems* 27.4 (2012), pp. 1926–1940. arXiv: 1106.1401.

[90] B. M. Sanandaji, H. Hao, and K. Poolla. "Fast regulation service provision via aggregation of thermostatically controlled loads". In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. 2014, pp. 2388–2397.

[91] R. Schneider. *Convex Bodies : The Brunn – Minkowski Theory.* 1993.

[92] A. Singhal and V. Ajjarapu. "A Framework to Utilize DERs' VAR Resources to Support the Grid in an Integrated T-D System". In: *2018 IEEE Power Energy Society General Meeting (PESGM).* Aug. 2018, pp. 1–5.

[93] N. A. Sinitsyn, S. Kundu, and S. Backhaus. "Safe protocols for generating power pulses with heterogeneous populations of thermostatically controlled loads". In: *Energy Conversion and Management* 67 (2013), pp. 297–308. ISSN: 0196-8904.

[94] A. G. Thomas and L. Tesfatsion. "Braided Cobwebs: Cautionary Tales for Dynamic Pricing in Retail Electric Power Markets". In: *IEEE Transactions on Power Systems* 33.6 (Nov. 2018), pp. 6870–6882.

[95] H. R. Tiwary. "On the hardness of computing intersection, union and minkowski sum of polytopes". In: *Discrete and Computational Geometry* 40.3 (2008), pp. 469–479.

[96] R. Tonkoski, D. Turcotte, and T. H. El-Fouly. "Impact of high PV penetration on voltage profiles in residential neighborhoods". In: *IEEE Transactions on Sustainable Energy* 3.3 (2012), pp. 518–527.

[97] L. C. Totu, R. Wisniewski, and J. Leth. "Demand Response of a TCL Population Using Switching-Rate Actuation". In: *IEEE Transactions on Control Systems Technology* 25.5 (Sept. 2017), pp. 1537–1551.

[98] K. Turitsyn, P. Sulc, S. Backhaus, and M. Chertkov. "Options for Control of Reactive Power by Distributed Photovoltaic Generators". In: *Proc. of the IEEE* 99.6 (2011), pp. 1063–1073.

[99] L. Vandenberghe, S. Boyd, and S.-P. Wu. "Determinant Maximization with Linear Matrix Inequality Constraints". In: *SIAM J. Matrix Anal. Appl.* 19 (1998), p. 499.

[100] R. A. Verzijlbergh, L. J. De Vries, and Z. Lukszo. "Renewable Energy Sources and Responsive Demand. Do We Need Congestion Management in the Distribution Grid?" In: *IEEE Transactions on Power Systems* 29.5 (2014), pp. 2119–2128.

[101] C. Weibel. "Minkowski Sums of Polytopes: Combinatorics and Computation". PhD thesis. 2007.

[102] S. Wong, W. Muneer, S. Nazir, and A. Prieur. *Designing, Operating, and Simulating Electric Water Heater Populations for the Smart Grid.* Tech. rep. 2013-136 (RP-TEC). Varennes: CanmetENERGY, Natural Resources Canada, 2013, p. 88.

[103] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation, and Control.* New York, NY: John Wiley & Sons, 1984.

[104] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi. "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response". In: *IEEE Transactions on Power Systems* 28.4 (2013), pp. 4655–4664.

[105] L. Zhao and W. Zhang. "A Unified Stochastic Hybrid System Approach to Aggregate Modeling of Responsive Loads". In: *IEEE Transactions on Automatic Control* 63.12 (Dec. 2018), pp. 4250–4263.

[106] L. Zhao, W. Zhang, H. Hao, and K. Kalsi. "A Geometric Approach to Aggregate Flexibility Modeling of Thermostatically Controlled Loads". In: *IEEE Transactions on Power Systems* 32.6 (2017), pp. 4721–4731.

[107] X. Zhou, E. Dall'Anese, L. Chen, and A. Simonetto. "An Incentive-Based Online Optimization Framework for Distribution Grids". In: *IEEE Transactions on Automatic Control* (2018).