Data-driven Modeling and Analytics for Greening the Energy Ecosystem

John Wamburu
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Computer Sciences Commons

Recommended Citation
https://doi.org/10.7275/32675285 https://scholarworks.umass.edu/dissertations_2/2788

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
DATA-DRIVEN MODELING AND ANALYTICS FOR GREENING THE ENERGY ECOSYSTEM

A Dissertation Presented
by
JOHN WAMBURU

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2023

Manning College of Information & Computer Sciences
DATA-DRIVEN MODELING AND ANALYTICS FOR GREENING THE ENERGY ECOSYSTEM

A Dissertation Presented
by
JOHN WAMBURU

Approved as to style and content by:

________________________________________
Prashant Shenoy, Chair

________________________________________
David Irwin, Member

________________________________________
Mohammad Hajiesmaili, Member

________________________________________
Jay Taneja, Member

Ramesh K. Sitaraman, Associate Dean for Educational Programs and Teaching
Manning College of Information & Computer Sciences
ACKNOWLEDGMENTS

This thesis is the culmination of an immense learning experience that I will forever cherish. I am deeply grateful to everyone who has played a role in the successful completion of my Ph.D. journey.

First and foremost, I would like to thank my advisor, Prof. Prashant Shenoy. Throughout the Ph.D. journey, he has guided me through the process of doing research, from working on the right research questions, to rigorously looking at possible solutions. He has also helped me overcome multiple challenges that I faced during the program, all the way to helping me evaluate career options towards the end. I have learned so much from him, and will keep these lessons with me forever. I consider myself incredibly fortunate to have had Prof. Shenoy as my advisor, and I couldn’t have asked for a better mentor.

I would also like to acknowledge the support, guidance and feedback I have received from my collaborators. I am especially grateful to Prof. David Irwin who has always been there to discuss my research, and with whom I have worked on multiple projects during my Ph.D. I am also immensely grateful to Prof. Mohammad Hajiensmaili for providing me with a broader perspective of optimization, which formed a significant part of this thesis. I would also like to thank Prof. Jay Taneja, whose guidance and mentorship started all the way before my Ph.D., and has continued up to the end of the program.

During the second half of my Ph.D., I was fortunate to become part of the ELEVATE program. I would like to thank the ELEVATE program for providing a home, funding and a diverse set of friends and colleagues for my research. I am especially
grateful to Prof. Erin Baker and Prof. Matthew Lackner for their research discussions and feedback. I would also like to thank Emma Grazer and Prof. Christine Crago with whom I worked on multiple projects.

My research journey began at IBM, and it’s from here that I received the encouragement and support to start my Ph.D. journey. I am especially grateful to Dr. Aisha Walcott-Bryant who was pivotal in developing my interest in research, and whose guidance as a supervisor prepared me for the Ph.D. I am also thankful to Dr. Solomon Assefa for his mentorship and advice on a wide variety of topics such as research, career after Ph.D. and many more.

My Ph.D. journey wouldn’t have been complete without the three summers I spent interning at IBM and LinkedIn. I am particularly grateful to Dr. Levente Klein for guiding me through two internships at IBM. I am also grateful to Dr. Hendrik Hamann for giving me the opportunity to intern at IBM for the two summers. I am also immensely grateful to Qi Guo, Muchen Wu and Sriram Vasudevan for making my time at LinkedIn enjoyable and full of learning.

My time in Amherst would not have been so enjoyable were it not for the friends I made. I would particularly like to thank Phuthipong Bovornkeeratiroy (Nikko) for being a great friend and collaborator. I thank him for his encouragement, support and advice even in issues outside of the Ph.D. I am also immensely grateful to Dr. Stephen Lee, who was instrumental in shaping the way I think about research, and for his guidance and advice during the program. I would also like to thank Hia Gosh for being a great friend, and for all the time we’ve spent on activities outside the lab. I am also grateful to Albert Williams, who helped me settle down when I first moved to the US. I would also like to thank Purity Mugambi for being a great friend, and for always reminding me about home. I am also immensely grateful to my colleagues Bin, Priyanka, Akanksha, Amee, Srin, Bhawana, Qianlin, Walid, Mehmat, Jorge, Adam,
Prateek, Lucas, Camellia, Abel and Noman — for making the lab a great place to work.

Finally, I would like to thank my family — my father, my dear mother, my two brothers and sisters — for their love and support. They have always been there for me, and were it not for them, I probably wouldn’t be here. Lastly, I want to thank Nelly Wanjiku, who has supported me in every way, and whose presence has carried me through the highs and lows of graduate school.
ABSTRACT

DATA-DRIVEN MODELING AND ANALYTICS FOR GREENING THE ENERGY ECOSYSTEM

FEBRUARY 2023

JOHN WAMBURU
B.Sc., KENYATTA UNIVERSITY, KENYA
M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Prashant Shenoy

The energy ecosystem is undergoing a major transition from primarily using carbon-intensive energy sources to greener and renewable sources of energy. For instance, electric vehicles (EVs) are rapidly increasing in popularity thereby eliminating gas-based carbon emissions. Similarly, the increased adoption of solar is injecting greener energy into the grid, thus reducing the grid’s overall carbon footprint. At the same time, the proliferation of networked devices and sensors in the grid is enabling energy usage analysis at fine granularity.

In this thesis, I argue that data-driven modeling and analytics applied to energy usage data can facilitate optimal carbon reduction in the energy transition. I present four studies that use principles from machine learning, optimization, and statistical time series analysis to study, analyze and understand the carbon footprint of various energy sectors and devise carbon reduction strategies.
First, I study the impact of residential EVs on the demand experienced by a city-wide distribution grid. Since the residential distribution grid was built in a pre-EVA era, it was not designed to account for EV loads, and challenges such as transformer overloading can arise with increased EV energy demand. I quantify and show how grid energy storage and smart charging technologies can mitigate this increased demand and increase transformer lifetime.

Second, I examine the feasibility, costs, and carbon benefits of using electric bike sharing - a low carbon form of ride sharing - as a potential substitute for shorter ride sharing trips with the overall goal of greening the ride sharing ecosystem. I present a linear optimization framework that employs a hybrid mix of regular and electric bikes to perform substitution and quantify the carbon reduction achieved from such substitution.

Third, I discuss the inequity that exists in the energy transition. I show that data driven approaches for building energy efficiency analysis may have inherent biases that prevent them from producing equitable results. I argue for design of equitable and fair analytic approaches to ensure that benefits of energy improvements and decarbonization brought about by the energy transition are distributed equitably across the whole society.

Finally, I study the potential of electric heat pumps to reduce CO$_2$ emission by replacing gas heating in buildings. I present a flexible multi-objective optimization framework that optimizes CO$_2$ reduction while also maximizing other aspects of the energy transition such as carbon efficiency and minimizing energy inefficiency in buildings. I quantify the tradeoffs that exist in such multifaceted transition and present results that shed light into the expected load exerted on the electric grid by transitioning from natural gas to electric heat pumps.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Thesis Contribution</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Mitigating the Impact of EVs on the Grid</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Carbon Reduction Through Ride Substitution</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Incorporating Equity in the Energy Transition</td>
<td>5</td>
</tr>
<tr>
<td>1.2.4 Optimal Decarbonization of Residential Heating Systems</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Thesis Outline</td>
<td>6</td>
</tr>
<tr>
<td>2. BACKGROUND AND RELATED WORK</td>
<td>8</td>
</tr>
<tr>
<td>2.1 The Energy Transition</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1 Electric Heat Pumps</td>
<td>9</td>
</tr>
<tr>
<td>2.1.1.1 Challenges</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Electric Vehicles</td>
<td>11</td>
</tr>
<tr>
<td>2.1.2.1 Challenges</td>
<td>12</td>
</tr>
<tr>
<td>2.1.3 Grid-based Energy Storage</td>
<td>12</td>
</tr>
</tbody>
</table>
5.4.3 Carbon-first approach ........................................ 90
5.4.4 Equitable transition ........................................ 91
5.4.5 Targeted transition ........................................ 93

5.5 Related Work .................................................. 95
5.6 Conclusions ..................................................... 96

6. MULTI-OBJECTIVE DECARBONIZATION OF RESIDENTIAL HEATING SYSTEMS .................. 97

6.1 Motivation ..................................................... 97
6.2 Problem and Methodology ................................... 100
  6.2.1 Problem Statement ....................................... 100
  6.2.2 Description of Datasets .................................. 101

6.3 Energy Usage and Carbon Analysis .......................... 102
  6.3.1 Energy Demand Analysis ................................. 102
  6.3.2 Carbon Emission Analysis ............................... 104

6.4 Multi-objective Decarbonization ............................ 105
  6.4.1 Optimizing for Carbon Emissions Reduction ............ 105
  6.4.2 Optimizing for Carbon-efficiency ....................... 107
  6.4.3 Targeting Energy Inefficient Buildings ................. 110

6.5 Evaluation .................................................... 112
  6.5.1 Experimental Setup ..................................... 112
  6.5.2 Optimizing Carbon Emissions Reduction ............... 113
  6.5.3 Maximizing Carbon-efficiency ......................... 114
  6.5.4 Targeting Energy Inefficient Buildings ............... 116
  6.5.5 Impact on Energy Consumption Reduction .............. 119
  6.5.6 Impact on the Electric Grid ............................ 120

6.6 Related Work ................................................ 121
6.7 Conclusions .................................................. 124

7. SUMMARY AND FUTURE WORK ................................. 126

7.1 Thesis Summary .............................................. 126
7.2 Future Work .................................................. 127

BIBLIOGRAPHY ....................................................... 129
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Examples of popular electric vehicles with different battery</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>characteristics.</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Key characteristics of the dataset</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Charging Summary of electric vehicle models.</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of the peak utilization of transformers</td>
<td>29</td>
</tr>
<tr>
<td>4.1</td>
<td>Ride Sharing Datasets</td>
<td>46</td>
</tr>
<tr>
<td>6.1</td>
<td>Probable building faults alongside their underlying</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>characteristics.</td>
<td></td>
</tr>
</tbody>
</table>


**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Distribution of transformer capacities.</td>
<td>23</td>
</tr>
<tr>
<td>3.2</td>
<td>Distribution of smart meters connected to transformers of varying capacities in the distribution grid.</td>
<td>24</td>
</tr>
<tr>
<td>3.3</td>
<td>This graph illustrates the seasonal variation in the load profile of a representative transformer.</td>
<td>26</td>
</tr>
<tr>
<td>3.4</td>
<td>Demand profile clusters across transformers. The number and percentage of transformers in each cluster is listed in the caption. The clusters are qualitatively different, with some exhibiting daytime peaks and others exhibiting evening peaks.</td>
<td>26</td>
</tr>
<tr>
<td>3.5</td>
<td>Distribution of transformer overloads over a year, based on the peak transformer utilization (a), the number of total hours the transformers experienced overloads (b), and the maximum sustained period of overloading (on a log scale) (c).</td>
<td>30</td>
</tr>
<tr>
<td>3.6</td>
<td>Peak utilization distribution of the transformers for varying EV penetration levels (a). Additional transformers that are at risk of overloading or are overloaded due to EVs (b). As EV penetration increases, the maximum sustained overloading in transformers increase depicted by the distribution shifting to the right (c).</td>
<td>31</td>
</tr>
<tr>
<td>3.7</td>
<td>Best case scenario, where EV adoption is skewed to low-to-moderate transformers, causing fewer transformers to become overloaded or critically overloaded.</td>
<td>35</td>
</tr>
<tr>
<td>3.8</td>
<td>Worst case scenario, where EV adoption is skewed to highly utilized and above transformers, causing the number of transformers at risk to increase.</td>
<td>36</td>
</tr>
<tr>
<td>3.9</td>
<td>Energy storage capacity required to limit utilization to no more than 125% across all transformers (a), and the distribution of energy storage capacity needed to limit overloading based on transformer capacities (b).</td>
<td>37</td>
</tr>
</tbody>
</table>
3.10 Increase in the median battery size necessary with different EV penetration levels to limit overloaded transformers (above 125%), utilization exceeding 90% .......................... 38

3.11 Reduction in number of overloaded and heavily utilized transformers with increasing battery scale factor without EVs. ...................... 39

3.12 Reduction in number of overloaded and heavily utilized transformers with increasing battery scale factor at 10% EV penetration. .......... 40

3.13 Reduction in the number of transformers that are overloaded when using smart charging policy. ................................. 41

3.14 Reduction in the necessary battery capacity when smart charging is combined with energy storage. ......................... 42

4.1 Distribution of trip distances for bikes, taxis, and for-hire (ride-share) vehicles. The median ride is short, with median trip distances ranging from 2.1-3.4km ........................................ 47

4.2 Cumulative distribution functions of trips duration. Short bike rides of 2km or less are faster than FHV rides. FHV trips have a greater variance due to variable traffic conditions ...................... 49

4.3 Distribution of pickup/dropoff locations to the closest bike stations ................................................................. 52

4.4 Distribution of number of passengers in each taxi and FHV car trip. A quarter of the trips involve more than one passenger .......................... 52

4.5 (4.5a) Distribution of bike trip demand by time of day during rainy, snowy and sunny (other), and (4.5b) distribution of taxi trip demand during rainy, sunny and snow days, and (4.5c) distribution of bike trip demand by time of day during working days and holiday days, and (4.5d) distribution of taxi trip demand during working days and holiday days. ...................... 54

4.6 Trip demand by (a) month of the year and (b) hour of the day. ........ 55

4.7 (a) Distribution of e-bikes and regular bikes at varying distance thresholds. (b) Bike capacity enhancement to service varying taxi trip substitution .............................................. 63
4.8 (4.8a) Monthly emissions from taxis, FHV\textemdash s and electric bikes under various ride substitution scenarios. (4.8b) Overall carbon emission reduction under different ride substitution scenarios. (4.8c) Overall carbon emission in the current bike share system under varying distance threshold. ........................................ 67

5.1 Average house value (left y-axis) and average annual gas energy usage intensity (right y-axis) across homes in different income groups. .... 76

5.2 Distributions of household income by race (a), distribution of EUI by house value (b), and distribution of house value by race (c). ......... 76

5.3 Distribution of EUI by age of building i.e. year the building was built. ......................................................... 81

5.4 Distribution of number of homes per census block (a), number of homes by age (b), and values of homes (c). ......................... 85

5.5 Aggregate gas demand during the whole period for all homes. ........ 87

5.6 Carbon emission reduction achieved from varying levels of transition. ................................................................. 87

5.7 Median house value for homes selected at varying levels of transition. ................................................................. 88

5.8 Carbon emissions achieved for decarbonization strategy that targets equity based on income group (a) and census block (b).......... 88

5.9 House values for homes selected based on approaches that target equity based on income group (a) and census block (b).......... 89

5.10 Distribution of EUI by age group at 5\% transition. ................. 92

5.11 Distribution of total winter emissions by census block. ............ 92

6.1 Aggregate gas demand during the year 2020. ....................... 102

6.2 Probability distributions for aggregate daily gas demand (a), average gas demand over the course of the day (b), and CO\textsubscript{2} emissions from buildings throughout the year. ......................... 103

6.3 Relationship between the energy generated from gas consumption and temperature. .................................................. 107
6.4 Relationship between energy demand of an electric heat pump and the temperature. ........................................ 108

6.5 Reduction in carbon emission at varying levels of transition to electric heat pumps. ........................................ 114

6.6 Distribution of normalized carbon emissions reduction across all buildings.................................................. 115

6.7 Carbon emissions at varying levels of electric heat pump transition while jointly optimizing for carbon emissions reduction and carbon-efficiency. ........................................ 116

6.8 Distribution of energy inefficiencies across buildings by age group......................................................... 118

6.9 Carbon emissions at varying levels of electric heat pump transition while jointly optimizing for carbon emissions reduction and targeting energy inefficient buildings......................................................... 119

6.10 (6.10a) Pareto curve for jointly maximizing CO\textsubscript{2} efficiency and reduction, and (6.10b), Pareto curve for jointly maximizing CO\textsubscript{2} emission reduction and the number of inefficient buildings transitioned. ........................................ 120

6.11 Distribution of average gas load demand by time of day after transitioning to electric heat pumps. ....................... 121

6.12 Distribution of energy consumption reduction across buildings after transition to electric heat pumps. ....................... 122

6.13 Cumulative distribution functions of additional electric demand exerted on the system after transitioning from gas to electric heat pumps. ........................................ 123
CHAPTER 1
INTRODUCTION

The energy ecosystem is undergoing a major transition from primarily using carbon-intensive energy sources to greener and renewable sources of energy. This transition is characterized by fundamental changes in key energy domains such as the electric grid, buildings, transportation, and others. For instance, electric vehicles (EVs) are rapidly increasing in popularity thereby eliminating gas-based carbon emissions. Similarly, the increased adoption of solar is injecting greener energy into the grid, thus reducing the grid’s overall carbon footprint.

As the transition progresses, the electric grid will need to evolve and adapt to the changing energy needs. For example, EVs are powered by batteries that must be charged frequently, e.g., often daily, using electricity drawn from the grid. As EVs become commonplace, their impact on the grid will be profound, as major aspects of the grid such as distribution transformers were installed during the pre-EV era, and thus were not sized to handle large loads from EVs. Similarly, in the building sector, there’s a strong push towards cleaner and more efficient sources of heat energy, such as electric heat pumps. Since natural gas is the most-used heating energy source, this transition will involve load migration from one system i.e. the gas distribution network, to another i.e. the electric grid, and further underscores the need for the electric grid to adapt to changing energy patterns.

At the center of the transition are key societal issues of equity and energy justice. Traditionally, disadvantaged parts of the society have borne the higher burden of pollution, lack of access to clean and renewable energy and higher energy costs. As
the world moves towards a carbon-free future, it is important to not only maximize carbon reduction and facilitate evolution of the future grid, but also consider how to do so *equitably*.

This thesis discusses challenges and opportunities in designing data-driven techniques that lead to optimal carbon reduction in the context of the energy transition. The implications of this work are both immediate e.g. optimal transition in the face of the current migration to electric vehicles, while others are more enduring e.g. ensuring equity in the transition, and will lead the world into a just and equitable energy future.

1.1 Motivation

To reduce carbon emission, fundamental changes are required in sectors such as the electric grid, building and residential energy usage and transportation. In most countries in the world, these sectors still rely on non-renewable, fossil fuel-based energy sources — such as coal, natural gas — for majority of their energy needs. For example, fossil fuel-based energy resources fulfilled more than 79% of U.S. energy consumption in 2021 [7]. To mitigate the effects of climate change, there is a push towards cleaner sources of energy. These changes have already begun, with the grid currently undergoing a major transition from primarily using carbon-intensive energy sources to greener and renewable sources of energy. Similarly, newer forms of energy in the transportation sector such as electric vehicles are also gaining momentum.

The transition towards cleaner sources of energy can be achieved individually for each of the major sectors of energy consumption, such as transportation, buildings, and agriculture, by systematically transitioning the major sources of energy in those sectors. However, prior studies have suggested that a more effective pathway is to transition our energy needs to electricity while intensifying the efforts to cleaning the sources of electricity production [73, 22]. This hypothesis is supported by re-
cent estimates that suggest electricity’s carbon-intensity (in g·CO₂/kWh) in the U.S.
decreased 30% between 2001 and 2017, largely due to the replacement of coal-fired
power plants with natural gas and wind generation [110, 57]. This trend is expected
to continue as the use of renewable energy resources for electricity production in-
creases. In particular, the electrification of buildings and the transportation sector
has received significant attention to accelerate progress in the energy transition.

At the same time, the rise in networked devices has enabled data analytic tech-
niques in facilitating these changes. For example, smart meters installed in the grid
have led to the widespread study of energy patterns at fine granularity, while sensors
installed in electric vehicles have been used to study energy usage patterns and inform
optimization techniques. This has brought about new opportunities for optimization,
as well new challenges such as equity that were not present before the transition.
In this thesis, I explore the opportunities presented by these networked devices to
facilitate optimal carbon reduction in the energy transition. Further, I study other
tangential aspects of the transition such as carbon and energy efficicny, as well as
address problems of equity that arise from such transition. In particular, I seek to
address the following questions in various thematic areas.

1. **EVs and the Electric Grid.** What is the impact of increased penetration of
   EVs on the load distribution seen by edge transformers? At what penetration
   levels does the distribution grid see significant overload problems? How does
   the skewed the penetration of EVs to particular areas (e.g., affluent) neighbor-
   hoods affect edge transformers? Can smart charging techniques that defer (or
   rate limits) EV charging loads during peak periods help alleviate transformer
   overloading?

2. **Ride Sharing.** What is the feasibility of substituting ride sharing trips with
greener bike sharing rides? What is the optimal mix of regular and electric bikes
needed to accommodate such substitution, and what are the carbon benefits of such an approach?

3. **Equity in the Energy Transition.** What biases exist in data-driven and machine learning energy efficiency analysis techniques, and to which degree are such biases present? What approaches can be adopted to alleviate such biases?

4. **Buildings and the Energy Transition.** What is the impact of replacing gas heating with electric heat pumps on energy consumption and CO₂ emissions? What is the optimal order in which buildings should be transitioned from gas to electric heat pumps in order to minimize CO₂ emission? How is this ordering impacted when additional goals such as carbon/energy inefficiency of buildings are introduced? How is CO₂ reduction impacted, and what are the tradeoffs?

By addressing these questions, I seek to draw upon principles from machine learning, time series analysis and optimization.

### 1.2 Thesis Contribution

This thesis proposes novel data-driven techniques that combine principles from machine learning, optimization, and statistical time series analysis to devise carbon reduction strategies, as well as study their effects on the grid and society as a whole.

**1.2.1 Mitigating the Impact of EVs on the Grid**

Electric cars, which are the most common type of EV, are becoming increasingly popular. This comes with the promise of reduced emissions since electric energy such as hydro has a lower carbon footprint compared to gas fuel. Residential EVs which are charged at home can cause a higher peak charging load since they are plugged in during evening hours, which coincides with higher residential electric usage. This can impose a significant load on edge transformers in the distribution grid. To understand
and quantify the extent of this impact, I perform a data-driven study that quantifies
the impact of EVs on the distribution grid of an entire city, and the impact on
transformer overloading. To mitigate the adverse effects of increased loads, I show
how grid-level energy storage and smart charging algorithms can be used to curtail
the peak load imposed by EVs on the grid.

1.2.2 Carbon Reduction Through Ride Substitution

While ride-sharing has emerged as a popular form of transportation in urban areas
due to its on-demand convenience, it has become a major contributor to carbon emis-
sions, with recent studies suggesting it is 47% more carbon-intensive than personal
car trips. This is mainly because of “dead miles” travelled to pick passengers up.
It has further driven people away from public transit which has inn turn increased
emissions from the transportation sector. To mitigate these problems, I examine the
feasibility, costs, and carbon benefits of using electric bike-sharinga low carbon form
of ride-sharingas a potential substitute for shorter ride-sharing trips, with the overall
goal of greening the ride-sharing ecosystem. I propose a linear optimization frame-
work that employs a hybrid mix of electric and regular bikes to reduce emissions while
maintaining ride convenience for users.

1.2.3 Incorporating Equity in the Energy Transition

Buildings consume nearly 40% of the total energy consumption and 70% of the
total electricity in many countries. This comes at a high carbon cost e.g., buildings
contributed over 1850 million metric tons of greenhouse gases in 2019 [?]. As the
grid makes a profound energy transition towards a carbon-free future, improving the
energy efficiency and carbon footprint of the buildings sector will play an important
role in meeting our society’s sustainability goals.

Data-driven energy analytic techniques play a key role in this transition. For
instance, they are used to identify inefficient buildings and target them for energy
improvements with the overall goal of reducing carbon emission. However, such techniques are not always equitable. In this thesis, I show that data driven approaches for building energy efficiency may have inherent biases that prevent them from producing equitable results. I further argue for design of equitable and fair energy analytic approaches to ensure that benefits of energy improvements and decarbonization schemes are seen equitably across the whole society.

1.2.4 Optimal Decarbonization of Residential Heating Systems

Residential energy usage contributes nearly 20% of all greenhouse gas emissions in the United States [47]. Heating and cooling account for roughly 38% of these emissions [76]. Because of this, there is a strong interest in decarbonizing residential heating systems using new technologies such as electric heat pumps. In this thesis, I study the potential of replacing gas heating with electric heat pumps to reduce CO$_2$ emission in buildings. I present a flexible multi-objective optimization (MOO) framework that optimizes carbon emission reduction while also maximizing other aspects of the energy transition such as carbon-efficiency and minimizing energy inefficiency in buildings.

1.3 Thesis Outline

The remainder of this thesis is structured as follows. Chapter 2 provides background on shared mobility, electric vehicles, equity in the energy transition and electric heat pumps. Chapter 3 describes the impact of electric vehicles on the distribution grid and proposes grid-level storage as well as smart charging as mitigation strategies. Chapter 4 describes an optimization approach to reduce carbon emission in ride sharing through ride substitution. Chapter 5 discusses the inequity that exists in energy efficiency analysis techniques and proposes equity as a key design goal in such techniques. Chapter 6, examines the potential of electric heat pumps to replace fossil fuel based heating and presents a multi-objective optimization (MOO) frame-
work that enables the flexible selection of a subset of homes for heat pump retrofits with the aim of achieving decarbonization goals. Finally, Chapter 7 concludes with the thesis summary and future work.
CHAPTER 2
BACKGROUND AND RELATED WORK

This chapter presents background and related work on CO₂ reduction in various sectors of the energy domain. Specifically, we discuss the energy transition in residential and building energy usage, and shared mobility.

2.1 The Energy Transition

Most countries in the world still rely on non-renewable, fossil fuel-based energy sources — such as coal, natural gas — for majority of their energy needs. For example, fossil fuel-based energy resources fulfilled more than 79% of U.S. energy consumption in 2021 [7]. To curtail the effects of climate change, there is a push towards more cleaner sources of energy. The energy transition can be achieved individually for each of the major sectors of energy consumption, such as transportation, buildings, and agriculture. However, prior studies have suggested that a more effective pathway is to transition our energy needs to electricity while intensifying the efforts to cleaning the sources of electricity production [73, 22]. This hypothesis is supported by recent estimates that suggest electricity’s carbon-intensity (in g·CO₂/kWh) in the U.S. decreased 30% between 2001 and 2017, largely due to the replacement of coal-fired power plants with natural gas and wind generation [110, 57]. This trend is expected to continue as the use of renewable energy resources for electricity production increases. The electrification of buildings in particular has received significant attention to accelerate the energy transitioning progress. In this section, we discuss the various modes in which buildings are decarbonized by focusing on shifting major
loads such as heating and cooling towards greener sources of energy. Further, we
discuss emerging societal issues arising as part the energy transition such as equity
and fairness among different demographics in society.

2.1.1 Electric Heat Pumps

We begin by examining the state of carbon reduction in buildings sector by transi-
tioning heating from fossil-fuel sources of energy to greener energy sources. We do so
by studying the potential of electric heat pumps to replace gas heating in buildings.

Electric heat pumps are a new and energy-efficient alternative to gas furnace
heating during cold seasons, as well as space cooling during summer seasons. During
winter seasons, heat pumps pull warm air from outside and concentrate it into your
home space, making the inside warm. Conversely, during summer seasons, a heat
pump moves heat from within a building to the outside atmosphere which cools the
inside of the building. Since the main principle behind heat pump operation is heat
transfer instead of heat generation, heat pumps are more energy efficient than fossil
fuel based burners.

The most popular type of heat pump available in the market today is an air-
source heat pump [89], which transfers heat between the inside of a building and
the outside air. Because these heat pumps rely on air heat transfer, as the outside
temperature decreases, their heating capacity degrades. In the past, such heat pumps
required a backup energy source to be used during extremely low temperatures, such
as a gas furnace or electric heating [51]. However, recent advances in heat pump
technology have made them efficient even at low temperatures, which makes them
an ideal replacement for gas heating even in cold climates [38, 111]. In addition to
increased energy efficiency, heat pumps also have other advantages over natural gas.
Since they use electricity, as more electricity is sourced from renewable sources, their
carbon footprint is lower than that of natural gas. Moreover, due to their reduced
energy usage, heat pumps can reduce the cost of heating a building by up to 60%. This makes them an attractive source of heating from a carbon, energy efficiency, and cost perspective.

Because of their carbon and cost benefits, electric heat pumps have received significant attention as a viable alternative to replace gas heating in buildings [23, 124, 58, 20, 68]. However, multiple challenges still exist on the path to widespread usage of electric heat pumps.

2.1.1.1 Challenges

Because heat pumps rely on heat transfer, as the outside temperature decreases, their heating capacity degrades. Traditionally, heat pumps have required a backup energy source to be used during extremely low temperatures, such as a gas furnace or electric heating [51], which lessens the gains made towards carbon reduction. Therefore, it is important to understand their ability to generate the required amount of heat energy even in very cold climates before widespread deployment.

Further, building selection for heat pump retrofits still remains non-trivial. This is mainly because the transition to heat pumps incurs significant capital costs, and may be unreachable for many homeowners. Additionally, since the transition will not occur instantaneously across the whole building stock, it is important to determine the order in which buildings should be transitioned. Determining the order of transition is non-trivial, as multiple factors must be taken into consideration before transitioning. For instance, building selection strategies that aim to achieve maximum possible CO₂ reductions fail to take advantage of other aspects of the energy transition such as improving energy and carbon efficiency in buildings. This thesis aims to provide solutions to these selection problems in the energy transition.
<table>
<thead>
<tr>
<th>EV model</th>
<th>Range (mi)</th>
<th>Size (kWh)</th>
<th>Rate (kW)</th>
<th>Charge Time (hour) at 220V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nissan Leaf (electric)</td>
<td>150</td>
<td>40</td>
<td>6.6</td>
<td>8h</td>
</tr>
<tr>
<td>Tesla Model S (electric)</td>
<td>315</td>
<td>100</td>
<td>10</td>
<td>10.7h</td>
</tr>
<tr>
<td>Chevrolet Bolt (electric)</td>
<td>238</td>
<td>60</td>
<td>7.2</td>
<td>9.3h</td>
</tr>
<tr>
<td>Chevrolet Volt (hybrid)</td>
<td>53</td>
<td>18.4</td>
<td>3.6</td>
<td>4.5h</td>
</tr>
<tr>
<td>Prius Prime (hybrid)</td>
<td>25</td>
<td>8.8</td>
<td>3.3</td>
<td>2.1h</td>
</tr>
</tbody>
</table>

Table 2.1: Examples of popular electric vehicles with different battery characteristics.

2.1.2 Electric Vehicles

In transportation, another mode of transition that has gained significant momentum in the recent past is electric vehicles (EVs). Electric cars, which are the most common type of EV, are becoming increasingly popular. They come with the promise of reduced emissions due to electric vehicles using cleaner energy compared to gas-based cars. Many manufacturers now include one or more types of EV in their product line up (see Table 2.1). Two types of electric vehicles are particularly common — pure EVs, which are solely powered using batteries, and plugin hybrid EVs (PHEV), which are powered using a combination of a gas-powered and electric motor. Examples of pure EVs include the Tesla Model S, Nissan Leaf, and Chevy Bolt, while examples of plugin hybrids include the Chevy Volt and the Toyota Prius Prime.

The larger battery sizes of pure EVs imply a larger load, which results in a higher peak charging load and a longer charging time to fully charge the battery. Regardless of the type, EVs can impose a significant load on edge transformers in the distribution grid. For instance, electric Type 2 chargers, a standard charger used for charging electric cars, draws roughly 7kW of power, while a central air conditioner, which is typically the “largest” load, draws 3.5kW. In other words, an electric car charger imposes twice the load of largest load, the central AC, in many homes on the distribution grid. Since distribution edge transformers were sized based on pre-EV era peak loads, increasing penetrations of EVs can have a significant negative impact. Thus, in the summer, a home with a central AC and an EV may exhibit a peak load
3× the previous peak load based on central AC alone (e.g. 10.5kW versus 3.5kW). As these electric cars become more commonplace, the demand of electric energy at the residential level will increase as charging power is drawn from the grid. This brings about a number of challenges in ensuring constant power supply from the grid.

2.1.2.1 Challenges

The architecture of the electric grid has three key components: generation, transmission, and distribution. While electricity is transmitted at high voltages through transmission lines, the distribution grid network uses a series of transformers to progressively step down the voltage and supply end-consumers. However, since EVs are connected directly to home power, they can impose a significant load on edge transformers in the distribution grid. This has the potential to cause power blackouts and reduce transformer lifetime, and needs to be considered in the face of the global transition towards electric cars.

Further, due to the uncoordinated nature of charging events in residential EV charging, peak prediction and control becomes challenging with the introduction of EVs into the grid. For example, many people plug their cars after arrival in the evening. This exacerbate the typical evening peak experienced by the electric grid, and necessitates further power generation from backup sources to mitigate the increased peak. This has led to the need for deployment of other mitigation strategies in the grid such as energy storage, which we discuss in the next section.

2.1.3 Grid-based Energy Storage

Grid-level energy storage, in the form of batteries, has emerged as a promising approach for various grid optimizations. Battery-based grid energy storage can be deployed at various points in the grid’s hierarchy — generation, transmission, or the distribution part of the grid network. Prior studies have shown that battery-based storage is especially appealing to handle the intermittency exhibited by renewable
energy sources, such as solar and wind, by using storage to smooth out the fluctuations [67, 18]. Similarly, battery-based storage has been used for peak load shaving [92, 91].

Although the cost of battery-based energy storage remains high, prices are dropping more rapidly than expected even a few years ago, and commercial products and deployments are beginning to ramp up. For instance, Tesla sells PowerWall battery packs to both residential users and to utility companies. The largest deployment of grid batteries, a capacity of 100 MWh, was recently installed by Tesla in Australia [6]. In this work, we consider the deployment of energy storage batteries alongside distribution edge transformers to mitigate overloads caused by EVs and enhance transfer for lifetimes — a use of batteries that has not seen much attention in the distribution network. Utilities are especially interested in using such application in the future as prices continue to fall.

2.1.3.1 Challenges

The deployment of energy storage at the grid level comes with unique challenges and considerations that must be made to ensure successful operation. First, storage sizing, which heavily depends on the ability to predict demand in advance. This thesis proposes a peak-shaving algorithm that clips the maximum contiguous peak above a given threshold. Further, with the increased penetration of renewable energy sources in the grid, an estimation of the aggregate energy generation expected from the distributed renewable sources is required. This thesis focuses on storage at the transformer level to mitigate increased loads from residential EV penetration.

2.1.4 Equity in the Energy Transition

Equity in energy usage is measured in three main concepts [64, 84]. First, distributional equity is concerned with ensuring that the burdens and benefits of the energy transition are accrued equitably across populations, i.e., some sections of the popu-
lation do not receive a disproportionate share of the burdens, while the benefits are experienced by a small percentage of the population. Second, *procedural equity* is concerned with ensuring that the public engagement processes for planning and implementing the energy transition are conducted in a diverse and inclusive manner. Third, *recognition equity* is concerned with ensuring that historical injustices against certain demographics are acknowledged, and conscious efforts to remedy such inequalities are made. While decarbonization is achieved through replacement of natural gas heating with electric heat pumps, a key goal is to ensure that the benefits of such transition are distributed equitably across the whole population.

2.1.4.1 Challenges

To accurately measure energy equity in the context of decarbonization strategies, suitable equity metrics are required. *Energy use intensity (EUI)*, which measures the energy consumption per unit area is one such metric. EUI has been used widely to measure disparity in energy usage across different demographic profiles, e.g., high versus low income households [118, 19, 102]. However, analyzing high versus low EUI has intrinsic challenges. For instance, larger buildings might be found to be more “efficient” having lower EUI, while in fact, they might have a higher energy consumption compared to smaller buildings. Further, identifying bias and inequity that exists in various data-driven energy analytic techniques is a key challenge that requires special focus.

2.2 Shared Mobility

The sharing economy has had a significant impact on personal transportation and has led to ride sharing becoming common, or even the preferred, form of transport in urban areas [70]. The ability to get a ride on demand using a smartphone “anywhere” has made them a popular alternative to traditional taxi rides and private car
ownership. A side effect of this convenience has also caused them to supplant more carbon-efficient public transit options. Researchers have analyzed the climate impact of ride sharing and found ride sharing trips to generate 47% more carbon emissions than a private car trip, mainly due to the “wasted” driving between two hailed trips [13, 4]. This observation motivates our study on whether other types of ride sharing, such as bike sharing, which have a low carbon footprint could be used for short rides, while providing similar convenience such as on-demand pick-up/drop-off and wide availability.

There has been significant research attention on the design of bike sharing systems over the years. A range of problems such as optimizing station placement, capacity planning, rebalancing algorithms, and bike placement methods have been studied. However, the focus of this thesis is on sustainable carbon-aware design and on using bike sharing as a viable alternative for shorter car rides from a carbon standpoint.

Since this work aims to substitute car rides with bikes, electric bikes make this an even more viable option. An electric bike provides pedal assist to its rider using an inbuilt motor and battery, which makes biking nearly effortless and attractive for longer rides or rides on uphill roads. Their attractiveness for handling more challenging rides and reducing the biking effort makes them a key design element for encouraging substitution of short car rides with bikes. Electric bikes have long been a popular form of transportation in countries like China [81]. Some bike share program such as in Riverside [105] and Raleigh [99] in the USA and Guildford [52] in UK use only electric bikes. Moreover, New York’s Citibike program has plans to add up to 4000 electric bikes to its fleet [31].

2.2.1 Challenges

While an all-electric bike system requires higher capital and operational cost, e.g., electric bikes are more expensive to acquire, and a charging infrastructure must
be put in place for operation, a hybrid system is cheaper to build and maintain as only a subset of bikes and infrastructure will require electric capability. Given the higher deployment and maintenance costs of electric bikes, a key goal of this work is to highlight the benefits of hybrid systems that use a combination of regular and electric bikes as a cheaper alternative to an all electric-bike system while retaining the key advantage of electric bikes for enabling ride substitution.

Given this background, this thesis analyzes the feasibility of substituting shorter but more carbon-intensive car rides using bike sharing and to understand the carbon benefits of such ride substitution as well as the costs of handling a higher bike sharing demand resulting from such substitution.
CHAPTER 3
IMPACT OF EVS ON THE GRID

In this chapter, I analyze the impact of residential electric vehicles (EVs) on the demand experienced by a city-wide distribution grid. I show how EV loads lead to transformer overloading in the grid. To mitigate this problem, I show how grid-level energy storage and a smart energy aware charging strategy can be used to mitigate the impact of EV loads on distribution transformers.

3.1 Motivation

Advancements in battery and electric vehicle (EV) technology, combined with public policy initiatives, is rapidly accelerating the electrification of transportation. Major car and truck manufacturers have all announced new EV products, making it likely that EVs will become mainstream in the coming years. Nearly 200,000 EVs were sold in 2017 in the U.S. alone—a 25% increase in sales over 2016 [5]. Reports from Norway indicate that 70% of all new cars being sold are now EVs. Of course, EVs are powered by batteries that must be charged frequently, e.g., often daily, using electricity from the grid. Consequently, as EVs become commonplace, their impact on the electric grid will be profound. At a macro scale, all of the energy used to power automobiles, currently supplied by gasoline, will need to be provided by the electric grid, resulting in a manifold increase in electricity usage. At a micro scale, the residential distribution grid was built in a pre-EVA era and was not designed to account for EV loads. For example, a typical home in the U.S. has an average load 1.2kW, while an electric car such as Nissan Leaf adds an additional load of 6.6kW,
effectively doubling or tripling the peak electric demand of the home. As a result, distribution grid transformers that were sized before EVs may become overloaded and not be able to reliably support high EV penetrations.

At the same time, the emergence of the smart electric grid has resulted in new technologies for more flexible demand-side load management and load mitigation in the grid. In particular, grid-level energy storage is emerging as a key technology for supporting future smart grids, since it can smooth out fluctuations from intermittent renewable energy sources, such as solar and wind, as well as enable grid optimizations, such as shaving peak loads and serving as backup power to reduce outage durations [86, 92, 91]. Interestingly, grid-level energy storage can also be used to mitigate the impact of EV loads on distribution transformers. If judiciously deployed adjacent to distribution transformers, energy storage batteries can reduce or eliminate transformer overloads due to EV charging and increase transformer lifetimes. A complementary smart grid technology is intelligent load management via load shifting [94, 34]. In the context of electric vehicles, this technique translates to smart charging where the EV intelligently coordinates its charging with the distribution grid often by deferring its charging from peak to off-peak periods whenever necessary [114, 116]. Together, energy storage and smart charging have the potential to mitigate the impact of EV loads on the distribution grid, but how much and to what extent is unclear based on actual transformer capacities, projected EV loads, and current demand profiles.

In this chapter, we study the impact of residential EVs on the demand experienced by a city-wide distribution grid in the New England region of United States and then analyze whether and how much grid energy storage and smart charging technologies can mitigate this increased demand. Our study is empirical in nature and is based on analyzing real load data from i) 13,523 residential homes and 1,353 distribution transformers gathered at 5 minute granularity over a 2-year period and
ii) real charging data from over 91 EVs in use over a one year period. While there has been prior work on analyzing the impact of EV loads [32, 119, 100], our study differs from prior work in several key aspects. For example, Clement-Nyns et al. [32] largely focuses on characterizing the aggregate load impact from EVs, and does not consider the issue of mitigating the load impact using grid storage, while Verzijlbergh et al. [119] focuses on peak load analysis and thus only considers mitigating the one day that experiences the peak annual load.

In contrast, we analyze the impact of EVs on transformer loads throughout the grid over a 2-year period and specifically study how the distribution of loads changes as the penetration of EVs increases. As we show later, understanding the impact on the probability distribution of loads is as important as analyzing the peak load alone. While Ramanujam et al. [100] examines a similar problem, it drives its simulations using synthetic estimates of existing loads, rather than real-world empirical data, and is thus not an accurate characterization of real-world conditions. We analyze long-term fine-grained transformer load data across an entire city to characterize the real-world implications of increasing EV penetration, and examine ways to mitigate problems using grid-scale energy storage. In conducting our empirical analysis, this chapter makes the following contributions:

**Transformer Distribution Analysis.** We use a city-scale dataset to conduct an in-depth analysis of the existing transformers and quantify their different load profiles. Surprisingly, we observe that most transformers are not over provisioned in the network and all transformers are already designed to gracefully handle temporary overloads. Moreover, we find that 19.2% of transformers are heavily overloaded, having a utilization of over 100%.

**Impact of Electric Vehicles.** We analyze the effect of increasing penetrations of EVs and the effect on the load experienced by transformers in the grid and their lifetime under multiple different scenarios, e.g., uniform and skewed distributions of
EVs. Our results indicate that the percentage of critically overloaded transformers is low for small levels of EV penetration (1-5% of homes), but increases significantly at higher penetrations levels (20-40% of homes).

Mitigation Strategies. Since our results demonstrate that the current distribution system is not provisioned for high levels of EV penetration, we examine the effect of two mitigation strategies—the use of energy storage and smart EV charging—to reduce transform overloads, extend their lifetime, and improve grid reliability. Our results show that even deployed a small amount of energy storage capacity, e.g., 24kWh, can dramatically reduce the risk of failures in transformers. We also show that smart charging is highly effective at reducing the number of critically overloaded transformers at high EV penetrations levels. In addition, when used in conjunction with energy storage, we show that smart charging can reduce the battery capacity necessary (by 41.3%-69.6%) to prevent transformers from exceeding their capacity.

3.2 Problem and Methodology

In this section, we present the problem and key research questions we address, and then describe the datasets and experimental methodology that we use to answer those questions.

3.2.1 Problem

The primary goal is to understand the impact of varying levels of EV penetration on the loads experienced by distributed edge transformers, so as to understand how much slack capacity is currently present and to identify when grid transformers become overloaded. An additional goal is to understand when emerging technologies, such as smart EV charging or battery-based grid storage, can alleviate the overloads or what extra upgrades will be necessary to accommodate the growing number of EVs. Specifically, we seek to answer the following research questions.
1. What is the distribution of load experienced by edge transformers? What are the daily and seasonal variations in this load, specifically the peak load, seen by edge transformers? What does this load analysis reveal about the current slack present in the distribution grid? For those transformer with little or no slack, how loaded or overloaded are they?

2. How does progressively increasing the penetration of EVs impact the load distribution seen by edge transformers? How does the resulting load increase change the fraction of highly utilized and overloaded transformers? At what penetration levels does the distribution grid see significant overload problems? How does skewing the deployment of EVs to particular (e.g., affluent) neighborhoods change these results?

3. Can smart EV charging that defers (or rate limits) charging loads during peak periods help alleviate transformer overloads? How much energy storage is necessary for overload shaving of edge transformers at different EV penetration levels? How much additional benefits can be obtained by combining smart charging and grid-level energy storage? What do these results reveal about the relative size and feasibility of energy storage as a mitigation strategy and how much more penetration can be accommodated?

3.2.2 Datasets and Experimental Setup

The answers to these questions will vary from region to region, and clearly depend on the current state of the distribution grid in terms of its load over time, transformer capacities, and the resulting slack. In this work, we use a small city in the New England region of United States and attempt to answer these questions for this city by conducting a city-wide data analysis. Since the distribution grid design in this city is typical of many regions in North America, we believe that our high level insights are broadly applicable.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. of transformers</td>
<td>1353</td>
</tr>
<tr>
<td>Num. of commercial meters</td>
<td>1566</td>
</tr>
<tr>
<td>Num. of residential meters</td>
<td>13523</td>
</tr>
<tr>
<td>Transformer sizes</td>
<td>5kVA to 1500kVA</td>
</tr>
<tr>
<td>Electric meter granularity</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Duration</td>
<td>2015 to Sep 2017</td>
</tr>
</tbody>
</table>

(a) Grid Distribution Dataset

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total num. of electric cars</td>
<td>91</td>
</tr>
<tr>
<td>Num. of Tesla S</td>
<td>12</td>
</tr>
<tr>
<td>Num. of Nissan Leaf</td>
<td>18</td>
</tr>
<tr>
<td>Num. of Chevrolet Volt</td>
<td>61</td>
</tr>
<tr>
<td>Granularity</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Duration</td>
<td>2016</td>
</tr>
</tbody>
</table>

(b) Electric Vehicles Dataset

Table 3.1: Key characteristics of the dataset.

**Distribution Grid Dataset.** Our dataset consists of electricity usage (load) data recorded by 15,089 smart meters that serve every residential and commercial user in the city. These 15,089 meters are served by 1,353 distribution edge transformers. Our dataset includes a mapping of each meter to its edge transformer, and also includes a detailed specification of each transformer, including its rated capacity. The load data is recorded at a five minute granularity and spans from 2015 to 2017. Since data from late 2017 was not yet available when performing our analysis, we limit our analysis to two full calendar years—2015 and 2016—for which data is available.

Since these edge transformers are low voltage transformers that are directly connected to end-customers, the load on each transformer can be computed by summing the load recorded by each meter connected to that transformer. Doing so yields highly detailed load information experienced by each distribution edge transformer over the two year period of the study. The availability of detailed load information for all 1,353 edge transformer in a city is a distinguishing feature of our study. Prior work
has only considered the total grid load across a city rather than transformer-level loads [100]. In contrast, we study probability distribution of loads as well as the time of day/seasonal impacts that other studies did not consider.

Table 3.1 summarizes the key characteristics of our dataset discussed above. Figure 3.1 then depicts the diversity of transformer capacities in the distribution grids and the distribution of transformers across varying sizes. Note that, since the rated capacity is in apparent power as kVA, in our later analysis, we use the average power factor to convert it into kilowatts (kW) to make our results more intuitive. We use the equation below for the conversion.

\[
kW = kVA \cdot PF
\]  

Here, \(0 \leq PF \leq 1\) is the power factor. For our analysis, we use power factor of 0.9 and 0.95 for summer and winter, respectively, which represents the average power factor in these seasons.

Figure 3.1 shows that transformer capacities can vary from 5 kVA all the way to 1500 kVA. Most of the deployed transformers are ”small” and have a rated capacity
of less than 150 kVA — a few transformer are large with a capacity of 500 kVA to 1500 kVA. Generally, the small transformers serve a small number of residential customers (e.g., 2-4 homes). The larger transformers serve apartment complexes, office buildings, other light commercial customers.

Figure 3.2 shows the distribution of meters connected to transformers of various sizes. We observe that the number of connected residential meters increases with the increase in transformer capacity. In contrast, fewer meters are connected to transformers that provide electricity to commercial buildings as they tend to consume higher energy. The median number of meters connected to these transformers ranges from 2 to 28.

**Electric Vehicle Dataset.** Since our study seeks to understand the impact of electric vehicles, we use the Dataport dataset from Pecan St.\(^1\) — a real-world trace that consists of power consumption from 91 electric vehicles gathered at five minute resolution in 2016. The 91 EVs in the dataset represents a mix of 3 popular electric

\(^1\)Dataport dataset. [http://dataport.pecanstreet.org](http://dataport.pecanstreet.org)
<table>
<thead>
<tr>
<th>Car Model</th>
<th>Summary (kWh)</th>
<th>Max</th>
<th>Median</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tesla S</td>
<td>Daily Energy Demand</td>
<td>28.07</td>
<td>13.69</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>#Charging Session</td>
<td>340</td>
<td>235</td>
<td>112.8</td>
</tr>
<tr>
<td>Nissan Leaf</td>
<td>Daily Energy Demand</td>
<td>10.41</td>
<td>4.98</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>#Charging Session</td>
<td>337</td>
<td>174.5</td>
<td>122.7</td>
</tr>
<tr>
<td>Chevrolet Volt</td>
<td>Daily Energy Demand</td>
<td>7.19</td>
<td>5.01</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>#Charging Session</td>
<td>351</td>
<td>266</td>
<td>123.3</td>
</tr>
</tbody>
</table>

Table 3.2: Charging Summary of electric vehicle models.

car models — Tesla Model S, Nissan Leaf, and Chevy Volt. Table 3.2 depicts the different types of EVs in the dataset. The dataset includes detailed information, such as the power drawn and the time and duration the car was connected to the power outlet. Table 3.2 also shows the statistics of the charging profiles for each car model in the dataset.

Since our dataset only includes 91 EV traces, we supplement our dataset by constructing additional synthetic EV traces as follows. First, we randomly choose a particular car from the existing dataset. We then take the charging data for the entire year and permute the weekdays and weekends over the year for that car. That is, each weekday trace is mapped to a random other weekday and each weekend is randomly mapped to a different weekend. Doing so yields a synthetic trace that is based on permutations of the initially chosen trace. We repeat this process to construct 24,000 synthetic EV traces, 8000 for each model to supplement our real dataset in our analysis of increasing EV penetrations.

To simulate the effect of increased EV penetration on the distribution transformers, unless otherwise specified, we randomly assign EVs to residential homes. We then calculate the net load in each transformer after the addition of the electric vehicles. We repeat the above simulation 50 times and show the results for the average case over multiple runs. We also study skewed EV deployment, where we concentrate a greater fraction of EVs to specific neighborhoods, such as affluent neighborhoods that
Figure 3.3: This graph illustrates the seasonal variation in the load profile of a representative transformer.

(a) 134 (9.9%)  (b) 65 (4.8%)  (c) 22 (1.6%)  (d) 508 (37.5%)  (e) 624 (46.2%)

Figure 3.4: Demand profile clusters across transformers. The number and percentage of transformers in each cluster is listed in the caption. The clusters are qualitatively different, with some exhibiting daytime peaks and others exhibiting evening peaks.

are more likely to experience a higher fraction of EV adopters, rather than uniformly distributing them across the whole city.

3.3 Analysis of Edge Transformer Loads

In order to understand the impact of EV penetration on transformer loads, we must begin with an analysis of the current ("as-is") loads on edge transformers before the introduction of any EVs. Such an analysis reveals the slack available at various
transformers, as well as the transformers that are already heavily utilized and have little available slack.

3.3.1 Demand Profiles of Edge Transformers

We begin with an analysis of the monthly and daily loads seen by the 1,353 edge transformers across the city. Figure 3.3 depicts the monthly load experienced by a representative transformer over 2016. The figure illustrates the seasonal variation in the load, and is characterized by two peak demand periods — winter and summer. The winter peak occurs due to increased use of electric heaters during the winter, while the summer load coincides with the increased use of air conditioning on hot summer days. Although the winter peak is slightly higher than the summer one, the summer peak has a greater impact on transformer efficiency and lifetimes. Prior studies have shown that a high ambient temperature can have an adverse impact on transformer lifetimes [117, 43], as a high ambient temperature contributes to the effect of overloading by further heating up (and evaporating) the insulation oil, which protects transformers from overheating. With increased energy demand from EVs, summers are likely to have a greater adverse impact on transformers than other seasons. Since the spring and fall seasons see lower peak loads, there is more slack and cooler temperatures, which makes the transformers less vulnerable during these periods.

Next, we analyze the daily load profile of edge transformers to identify the most common types of transformers based on their load profile. For this analysis, we clustered the average daily profile of all transformers using \( k \)-Means clustering. Since the transformers are of different sizes, we normalize the daily load profile of each transformer to a range between 0 and 1 (e.g., using MinMaxScaler in scikit-learn), and then perform clustering. Figure 3.4 depicts the five clusters that emerge when using \( k \)-means with \( k=5 \); the figure shows the result for 2016 (the other years yield
qualitatively similar results and are omitted). We selected k=5, since 5 was the highest value of k that yielded clusters that were qualitatively different, and also did not yield an outlier cluster with few transformers. The red line depicts the centroid of the clusters, while the grey line shows the energy profiles of all the transformers in the cluster.

The five clusters reveal interesting patterns. For example, Figures 3.4(a) and (b) depict transformers that exhibit daytime peaks, while Figures 3.4(c), (d) and (e) depict transformers that exhibit evening peaks. The captions depict the number and percentage of transformers in each cluster. We hypothesize that the transformers exhibiting daytime peaks, in Figures 3.4(a) and (b), serve office buildings that have a 9 am to 5 pm workday or businesses, such as retail stores, that have 9 am to 9 pm work hours. These transformers have a low load during the late evening and nighttime hours.

The clusters shown in Figures 3.4(c), (d), and (e) all exhibit evening peaks and also exhibit a nontrivial amount of nighttime usage — we hypothesize that these are large residential customers with different daily routines. The cluster in Figure 3.4(c) shows transformers that see a low load during the day — these are likely users who are away from home (i.e., working) during the day and at home in the evening and night. Figures 3.4(d) and (e) show residential customers with evening peaks, but also a non-trivial amount of daytime and nighttime usage. These are likely to be families where someone is at home during the day, where the increased evening activities result in an evening peak — these two clusters (d) and (e) also account for a large fraction of the transformer, 37.5%, and 46.2%, respectively.

From the perspective of EV loads, these load profiles have interesting implications. Transformers with daytime peaks, which are offices and businesses, may deploy EV chargers in their parking lots that will service users charging their EVs while at work, and thus causing the already-high daytime peaks to rise further. Transformers, such
<table>
<thead>
<tr>
<th>Group Name</th>
<th>Utilization</th>
<th>#Transformers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low to Moderate</td>
<td>&lt; 90%</td>
<td>976 (72%)</td>
</tr>
<tr>
<td>Heavy utilization</td>
<td>≥90% to &lt;125%</td>
<td>283 (21%)</td>
</tr>
<tr>
<td>Overloaded</td>
<td>≥125% to &lt;150%</td>
<td>63 (5%)</td>
</tr>
<tr>
<td>Critically overloaded</td>
<td>≥150%</td>
<td>31 (2%)</td>
</tr>
</tbody>
</table>

Table 3.3: Summary of the peak utilization of transformers

as those in Figure 3.4(c), where users are away during the day, will likely see evening charging of EVs when users return home, causing evening peaks to increase further. In both cases, EVs may exacerbate the already-high peaks. Transformers in clusters (d) and (e) have the most flexibility, since users may be home during the day and may charge their vehicles at day or night. Of course, charging during the peak load periods, when feasible, exploits more slack in the transformer than the other way around.

3.3.2 Analyzing Peak Loads

Next, we analyze the peak loads (defined as the 99.9th percentile of the load serviced by a transformer over the year) experienced by the edge transformers. Using raw meter readings as they are leads to erroneous estimates of transformer peaks brought about by spurious reads, often way higher than the normal peaks. We use the 99.9th percentile to eliminate these values. Using (3.1), we compute the rated capacity of transformers in kW and then compute the utilization by normalizing the load observed at the transformer with its rated capacity.

We then group the transformers into four categories, explained in Section 3.2.2 and depicted in Table 3.3: low-to-moderate, highly utilized, overloaded, and critically overloaded. Figure 3.5 depicts the peak load distribution of the transformer across the whole city, while Table 3.3 shows the number and percentage of transformers that fall in each category. Since transformers have a typical lifetime of 20-30 years, one would expect careful sizing, such that the peak load is well below the rated capacity.
Figure 3.5: Distribution of transformer overloads over a year, based on the peak transformer utilization (a), the number of total hours the transformers experienced overloads (b), and the maximum sustained period of overloading (on a log scale) (c).

However, as shown in Table 3.3 and Figure 3.5(a), only 72% of the transformers service a peak load of less than 90% utilization over the course of the year. Around 21% of the transformers are heavily utilized and service a peak load of up to 125% of capacity. Note that this implies that the transformer operated at or above its rated capacity for at least part of the time over the year. Around 3.8% of the transformers are overloaded and see a peak load that exceeds 125% utilization, while an additional 2% of the transformers are critically overloaded with peak load exceeding 150%. As explained earlier, it is not “abnormal” for a transformer to exceed 100% utilization for short periods, since they have mineral oils to insulate them from overheating, although sustained overloads for long periods are dangerous. Therefore, we next analyze the duration of the overloads experienced by transformers.

We consider only the transformers that are in the overloaded and the critically overloaded groups and compute the number hours over the year for which they service a load exceeding 125% of their rated capacity, and also compute the maximum ”session duration” over which the transformer is continuously overloaded. Figure 3.5(b) plots the total number of hours for which transformers are overloaded or critically overloaded over a year. The figure shows that the overload distribution is long-tailed — the majority are overloaded for 162 hours over a year, while a few see overloads
Figure 3.6: Peak utilization distribution of the transformers for varying EV penetration levels (a). Additional transformers that are at risk of overloading or are overloaded due to EVs (b). As EV penetration increases, the maximum sustained overloading in transformers increase depicted by the distribution shifting to the right (c).

of as many as 1000 - 3000 hours. Figure 3.5(c) analyzes each continuous period that experiences an overload, and plots the longest continuous duration for which a transformer was overloaded. The figure, plotted on a log scale, shows the median duration of overload was 45 min, while some transformers see a sustained overload of 143 hours.

**Implications.** Our analysis shows that roughly two-thirds of the transformers have slack due to low-to-moderate peak loads. However, our temporal analysis reveals that the amount of slack has high seasonal variations — i.e., there may be less slack during the summer or winter peaks and less slack during peak hours of the day, which vary based on the transformer’s load profile. Conversely, around 21% of the transformers are heavily utilized and have almost no slack to accommodate EVs, while around 6% are already overloaded or critically overloaded. Further, energy storage may be beneficial for these 6% of the transformers to absorb the overloads, even without any EV. Finally, one surprising aspect of our analysis is our finding that shows roughly 19% of the transformers routinely operate over capacity at least for a portion of time each year, with some experiencing long sustained overloads of many days.
3.4 Impact of Electric Vehicles

In this section, we analyze the impact of increasing EV penetrations on the peak loads experienced by edge transformers. We first assume a uniform distribution of EVs across households (and transformers) in the city, and analyze the impact of varying levels of EV penetration on transformer peak loads. We also examine the impact of a skewed distribution, where EVs are disproportionately concentrated in specific (e.g., affluent) neighborhoods, and study the effects of different penetrations for such skewed distribution.

We first introduce different levels of EV penetration into the grid, namely 1, 2, 5, 10, 20 and 40% — where penetration represents the percentage of smart meters that service an EV load. To do so, we randomly select an EV trace (from our synthetic trace of Tesla, Chevy and Nissan EV) and map it to a randomly chosen smart meter (selected from a uniform distribution). The EV charging trace is overlaid on the smart meter trace, and the transformer load is recomputed accordingly. We repeat each experiment for 50 runs with a different random mapping of EVs to transformers to ensure our results have tight confidence intervals.

Figure 3.6 shows the impact of varying levels of EV penetration on the peak loads. Figure 3.6(a) depicts the distribution of transformers seeing different peak loads for the no EV case (current grid) and at 10% and 40% penetration. As expected, the peak loads experienced by a transformer increases due to EV loads, such that the mass and tail of the distribution shifts to the right. Furthermore, while the median peak load is 0.7 in the no EV case, the median peak load increases to 0.76 and 1.03 at 10% and 40% penetration respectively.

Next, we analyze the increase in the number (and percentage) of highly utilized, overloaded, and critically overloaded transformers at different penetration levels. We assume the current state (from Table 3.3) as the baseline such that Figure 3.6(b) reports the additional percentage of transformers in each group (over the baseline)
at each penetration level. The figure shows that for low penetration levels of 1%, 2%, and 5%, the additional transformers that become heavily utilized or overloaded are relatively small (1-2% in each group). In these cases, since the number of EVs is relatively small, there is sufficient slack in the transformer load to accommodate them. Generally, we see that as penetration levels rise, so does the percentage of transformers in each category. At 10% penetration, an additional 4% transformers become heavily utilized, while an additional 3% transformers see overloads or critical overloads. The peak loads rise quickly at 20% and 40% penetration, with up to 18% of transformers becoming overloaded or critically overloaded. Since many transformers become overloaded, rather than highly utilized, this case yields a slight drop in the heavily utilized transformers.

Figure 3.6(c) shows the maximum sustained duration of overloads seen by all transformers in 2016 for the no EV case, and for 10% and 40% penetration levels. The figure shows that with increasing EVs, not only do the peak loads rise, the duration for which these peak loads persist also rises. The median overload duration rises from 0.75 hours in the no EV case to 1.1 and 2.8 hours for 10% and 40% penetration respectively.

Implications. Overall, the results show that the distribution grid can easily accommodate up to 5% EV penetration, and potentially up to 10% penetration. The impact of 5% penetration is relatively small, while a 10% penetration level causes an increase in highly utilized transformers (which are still considered within normal operating range) and a moderate 3% increase in the overloaded transformers. Higher penetration levels above 10% cause an increasing problem with overloading, and indicate that mitigation strategies are necessary to accommodate these higher levels of EVs.
3.4.1 Impact of Skewed EV Penetration

While the above analysis assumes that EVs are uniformly distributed across meters and transformers, it is entirely likely that “early adopters” of EVs may be concentrated in certain neighborhoods (e.g., affluent households can pay the higher price for electric cars). In this scenario, EVs will be concentrated in a certain neighborhood and not uniformly distributed. To understand the impact of a skewed distribution, we repeat the above analysis for a skewed mapping of EVs to transformers. We perform two types of analysis representing both an optimistic best case and a pessimistic worst case scenario. For the optimistic case, we disproportionately skew the assignment of EV to low and moderately utilized transformers — by assigning 75% of the EVs to the group and the remaining 25% to the remainder of the transformers. For the pessimistic case, we do the opposite, and disproportionately skew the assignment of EVs to highly utilized and overloaded transformers. Like before, we conduct at least 50 runs for each penetration level. The two scenarios study the impact of EVs in neighborhoods with transformers with the greatest and the least slack, respectively. Figures 3.7 and 3.8 depict our results for the optimistic and the pessimistic scenario, respectively.

Figure 3.7 shows that skewing EV adoption to neighborhoods that have transformers with low-to-moderate loads (and the greatest slack) allows a higher penetration level compared to the uniform distribution (Figure 3.6). Specifically, at low penetration levels of up to 5%, there is minimal impact on overloaded transformers and a small increase in the heavily utilized transformer. Even at 20% penetration level, the increase in heavily-utilized and critically overloaded transformers is minimal, although the number of overloaded transformers nearly doubles. The 40% penetration level sees a dramatic increase in the percentage of critically overloaded transformers, increasing from 9.1% (in Figure 3.6(b)) to 17.8%
Figure 3.7: Best case scenario, where EV adoption is skewed to low-to-moderate transformers, causing fewer transformers to become overloaded or critically overloaded.

Figure 3.8 shows that skewing EV penetrations to neighborhoods with highly utilized transformers permits a lower penetration level. Interestingly, even in this pessimistic worst case scenario, there is only a modest rise of 4.6% of overloaded transformers at a 5% penetration level, indicating that there is adequate slack to accommodate up to 5% EVs even in the worst case. However, the percentage of overloaded transformer rises quickly at 10% and higher penetration levels, indicating that additional mitigation strategies are necessary to handle the worst-case scenario.

### 3.5 Mitigation Strategies

Having examined the effect of EV-based loads on distribution edge transformers, we now evaluate mitigation strategies to help alleviate transformer overloads. Specifically, we explore two emerging technologies in the smart grid — energy storage and smart charging to understand how using them in isolation or in combination can help in reducing the number of overloaded and critically overloaded transformers in the
Figure 3.8: Worst case scenario, where EV adoption is skewed to *highly utilized and above* transformers, causing the number of transformers at risk to increase.

grid. We also evaluate how many additional EVs can be accommodated if utilities introduce these technologies.

### 3.5.1 Energy Storage

In Section 3.3, we show that even in the absence of EV-based loads, a small fraction of the grid consists of overloaded or critically overloaded transformers. Since we are only concerned with how EVs impact the grid, we first examine how much energy storage capacity is required to eliminate the overloaded and critically overloaded transformers. We then examine how much additional energy storage is required when EVs are serviced by the edge transformers.

To calculate the energy storage capacity required per transformer, we propose a simple peak-shaving algorithm that clips the maximum contiguous peak above a given threshold. Specifically, for each transformer, our algorithm scans over its load and computes the contiguous period when the load exceeds the threshold. Our algorithm then computes energy storage capacity by computing the energy above the threshold across the periods, and selects the maximum. Our premise is that the energy storage
Figure 3.9: Energy storage capacity required to limit utilization to no more than 125% across all transformers (a), and the distribution of energy storage capacity needed to limit overloading based on transformer capacities (b).

that can flatten the maximum contiguous peak can also provide energy to flatten the smaller peaks experienced at other periods.

We begin by analyzing the distribution of storage capacity required to limit the maximum transformer utilization to 125%. Figure 3.9(a) shows that energy storage capacity can vary between 1kWh and 915kWh. We note that the 90th percentile of energy storage is 24kWh, which indicates that even a small battery size can dramatically reduce the risk of failures in transformers. In particular, 85% of overloaded and critically overloaded transformers can benefit from an energy storage capacity of 24kWh or less.

Since, the battery capacity is a function of the size of the transformer capacity, we plot the distribution of battery size against transformer size. Figure 3.9(b) shows the median energy storage capacity increases with increases in transformer capacity. The larger energy storage capacity can be attributed to the higher number of homes that larger transformers serve.

Figure 3.10 shows that the median battery capacity required to eliminate overloads in the overloaded and heavily utilized transformers increases from 1.9kWh prior to introduction of EVs to 15.6kWh at 40% penetration. The difference in energy stor-
Figure 3.10: Increase in the median battery size necessary with different EV penetration levels to limit overloaded transformers (above 125%), utilization exceeding 90%

The difference in capacity between heavily utilized and overloaded transformers is also small, not exceeding 2kWh at all penetration levels, showing that most of the transformers are in the overloaded region.

Next, we analyze the effect of adding storage as a function of transformer sizing. We define battery scale factor as a 1:1 ratio to a transformer’s kVA rating i.e. for each kVA, what effect would adding 1kWh of storage to the transformer have on the transformer’s overload status. Figure 3.11 shows the result of adding storage using this factor in the no EV case. By adding a 0.1 factor of storage, we are able to reduce the number of heavily utilized and overloaded transformers by up to 42% and 55% respectively. Figure 3.12 shows that at 10% EV penetration, the number of heavily utilized and overloaded transformers can be reduced by up to 70% and 76% respectively using a 0.5 factor of storage.
Figure 3.11: Reduction in number of overloaded and heavily utilized transformers with increasing battery scale factor without EVs.

3.5.2 Smart Charging

We now evaluate the reduction in peak utilization due to smart charging of electric vehicles. The goal of our analysis is to understand how flexibility in EV charging can reduce the number of overloaded transformers. Our hypothesis is that the demand profiles of transformers have sufficient low usage periods, especially during the night, such that EVs can be charged without significantly increasing the peak utilization of the transformers.

For the purpose of our analysis, we assume an ideal EV charging algorithm that has full knowledge of future transformer loads and EV charging profiles. We also assume that transformers and EV chargers are able to communicate over a network. We then compute an optimal threshold that minimizes the transformer’s utilization while ensuring that all EV requirements are met ahead of time. In the event that transformers do not have enough future slack for EV charging, our smart charging algorithm allows the threshold to be exceeded in order to meet EV demands. We then allocate EV charging schedules on a first come first serve basis. Whenever the
Figure 3.12: Reduction in number of overloaded and heavily utilized transformers with increasing battery scale factor at 10% EV penetration.

threshold is reached, additional chargers are not allowed to start charging immediately. As additional slack becomes available either due to connected EVs reaching full capacity or general household power usage reducing, the remaining EV chargers are scheduled.

Figure 3.13 shows the reduction in the number of transformers that are heavily utilized, overloaded, and critically overloaded at different EV penetration levels. The graph demonstrates that smart charging becomes more important as the EV penetration increases. Smart charging has little to no effect on reducing over-capacity transformers at small EV penetrations between 1% and 5%. However, the reductions in over-capacity transformers increases for EV penetrations between 10% and 20%. Once EV penetration reaches 40%, smart charging becomes critical, as it is able to reduce the number of critically overloaded transformers by nearly 20% and the number overloaded transformers by 7.4%. These results demonstrate that smart charging is an important tool for maintaining grid reliability as EV penetration ramps up.
3.5.3 Combing Storage and Smart Charging

Finally, we examine the effect of combining grid-level battery-based energy storage with EV smart charging. Since batteries are still expensive to deploy and maintain, we examine how much battery capacity is necessary to limit utilization to no more than 125% across all transformers when used in conjunction with our smart charging algorithm. Figure 3.14 shows the results at different levels of EV penetration. The graph shows that smart charging can reduce the battery capacity substantially, ranging from 41.3% to 69.6%. At low penetration, smart charging is able to take advantage of the available slack. As penetration increases, the threshold boundary is crossed, because EVs still need to be charged within the time period, and the reduction in overall battery size reduces. Since smart charging would incur very little capital expenses on the utility side, this reduction would significantly decrease the capital expenses related to deploying and maintaining batteries.
Figure 3.14: Reduction in the necessary battery capacity when smart charging is combined with energy storage.

3.6 Related Work

**Distribution Grid Network.** There have been numerous studies on the distribution network [11, 77, 115]. For instance, [11] studied the grid’s resilience to disruptions in the distribution network. Others have studied the feasibility, or have examined the cost-benefit analysis, of integrating renewables in the distribution network [115, 77]. However, these studies do not analyze the load on distribution edge transformers or examine the effects of EVs on edge transformers. Prior work has also studied the impact of load on transformer lifetimes [21, 53, 117, 108, 56]. These approaches provide thermal modeling of transformers, and examine how load and external factors affect transformer lifetime. Our work is complementary to this work, as we provide a broader analysis of the current state of distribution edge transformers in a city over a 2-year period. In addition, these studies do not characterize the impact of increased penetration of large EV loads. Prior work has also studied demand patterns at both the household and grid level [71, 10, 101, 62]. These include studies to understand the types of demand profiles for setting power tariffs or enabling demand-response
programs [88, 123]. Again, our work differs, as we focus on classifying load profiles across edge transformers, and characterize the current state of the grid to study the effect of emerging technologies, such as EVs and energy storage.

**Electric Vehicles.** There has also been a significant amount of prior research on EVs [48, 103, 97, 56, 50, 14, 119, 33, 74]. While some studies have focused on the effects of EVs on power quality [48, 103, 97, 50], other work has focused on controlled EV charging [14]. In contrast, our work focuses on characterizing the load impact from EVs on edge transformers, and approaches to mitigate these impacts. Prior work has also studied flexible charging or co-ordinated EV charging in the grid [114, 116]. Our work is complementary to this work, as these smart charging methods can be employed reduce the overloading of distribution edge transformers. Prior work has also analyzed the impact of EVs on the distribution grid [100, 32, 119]. However, as discussed in Section ??, our work differs from this work, as the dataset used is limited or synthetically generated. Instead, we use fine-grained load data and provide empirical analysis on potential strategies that can be used to mitigate impact of EVs.

**Energy storage.** Prior studies have explored the benefits of using energy storage in conjunction with renewable energy [60, 98, 44, 67]. These studies focus on control policies to meet certain cost objectives. In addition, the use of energy storage has been studied in the context of load shifting, where energy storage charges itself during periods of excess generation (or off-peak pricing periods) and discharges when the demand is high [37, 94, 34, 61]. Similarly, prior work has proposed algorithms to shave peaks at both the individual home or the grid level [92, 91, 86]. Again, our work is complementary, as this work does not study the use of peak-shaving techniques to mitigate the impact of EVs on distribution edge transformers at city-scale.
3.7 Conclusions

This chapter analyzes both the current load on transformers from a small city, and the expected load as EV penetrations increase. We find that many transformers are over-provisioned in today’s grid, but a significant fraction of them will become overloaded once EV penetrations reach 20% and above. We then examine mitigation strategies for reducing transformer overloads using grid-level energy storage and smart EV charging strategies. Our results indicate that both mitigation strategies can reduce over-capacity transformers at high EV penetration levels, and can also be used in combination to achieve significant reductions. At 40% EV penetration, we can reduce the number of critically overloaded transformers by 18% using smart charging, and up to 90% by deploying energy storage of up to 15kWh per transformer. We expect our work to spur further work on the impact of the changing electric grid, with higher penetrations of EVs and renewable energy sources, on the grid’s distribution system and its edge transformers.
CHAPTER 4

GREENING THE RIDE SHARING ECOSYSTEM

In this chapter, we analyze the feasibility, costs, and carbon benefits of using electric bike sharing - a low carbon form of ride sharing - as a potential substitute for shorter ride sharing trips with the overall goal of greening the ride sharing ecosystem. We propose a linear optimization framework that employs a hybrid mix of regular and electric bikes to perform substitution and quantify the carbon reduction achieved from such substitution.

4.1 Motivation

While ride-sharing has emerged as a popular form of transportation in urban areas due to its on-demand convenience, it has become a major contributor to carbon emissions, with recent studies suggesting it is 47% more carbon-intensive than personal car trips. In this chapter, we examine the feasibility, costs, and carbon benefits of using electric bike-sharing—a low carbon form of ride-sharing—as a potential substitute for shorter ride-sharing trips, with the overall goal of greening the ride-sharing ecosystem. Using public datasets from New York City, our analysis shows that nearly half of the taxi and rideshare trips in New York are short trips of less than 3.5km, and that biking is actually faster than using a car for ultra-short trips of 2km or less. We analyze the cost and carbon benefits of different levels of ride substitution under various scenarios. We find that the additional bikes required to satisfy increased demand from ride substitution increases sub-linearly and results in 6.6% carbon emission reduction for 10% taxi ride substitution. Moreover, this reduction can be achieved
through a hybrid mix that requires only a quarter of the bikes to be electric bikes, which reduces system costs. We also find that expanding bike-share systems to new areas that lack bike-share coverage requires additional investments due to the need for new bike stations and bike capacity to satisfy demand but also provides substantial carbon emission reductions. Finally, frequent station repositioning can reduce the number of bikes needed in the system by up to a third for a minimal increase in carbon emissions of 2% from the trucks required to perform repositioning, providing an interesting tradeoff between capital costs and carbon emissions.

4.2 Feasibility Analysis

We begin by describing the datasets we leverage in the study and conducting a feasibility analysis that examines the degree to which ride sharing trips are substitutable by bike trips.

4.2.1 Overview of Datasets

Our analysis is based on two public datasets from New York City.

CitiBike Dataset. CitiBike is the official bike share system for New York City and has released extensive data about its operations. For the purposes of our work, we use a 12 month period (July 2018 - June 2019) that comprises of 18.98 million trips. During this period, the CitiBike system had 17,954 bikes and 941 bike stations. The data contains a trip’s duration, its start and end time, the ID of start and end

<table>
<thead>
<tr>
<th></th>
<th>CitiBike 2019</th>
<th>TLC 2019</th>
<th>TLC 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Type</td>
<td>Bikes</td>
<td>Taxi and FHV</td>
<td>Taxi</td>
</tr>
<tr>
<td>No. of vehicles</td>
<td>17,954 bikes/941 station</td>
<td>14K taxi, 78K FHV</td>
<td>22K taxi</td>
</tr>
<tr>
<td>No. of trips</td>
<td>18.98 million</td>
<td>354 million</td>
<td>156.77 million</td>
</tr>
<tr>
<td>Coordinates</td>
<td>GPS</td>
<td>Geo-fenced Zones</td>
<td>GPS</td>
</tr>
</tbody>
</table>
stations, the station names, the start and end GPS locations (latitude and longitude),
the type of user, age and gender, and the bike ID.

**New York TLC Dataset.** The New York City Taxi and Limousine Commission (TLC) [1] provides comprehensive statistics for taxi trips as well as ride sharing vehicles (referred to as For-Hire-Vehicles (FHV)) such as Uber, Lyft, Juno, Via and Limousine series. There are a total of 354 million rides, of which 101 million come from taxis and 253 million come from FHVs, with Uber and Lyft accounting for a large majority (> 90%) of the FHV rides. Trip-level records are available for all of these 354 million rides, which include the pickup and drop-off date and time, pickup and drop-off location, trip distance, number of passengers, and itemized fare.

### 4.2.2 Feasibility Based on Distance

We begin our feasibility analysis by computing distributions of trip distances for bike, taxi, and FHV car sharing for the above datasets.

We analyze trip distances either by using the actual trip distance whenever reported in the trip records, or by computing the shortest road distance between the pickup and dropoff locations. For taxi rides, the actual trip distance is available in trip records. For bike rides, we use the GPS coordinates of the pickup and dropoff...
bike stations to estimate the bike trip distance. For the FHV rides, the trip distance is not available in trip-level records. Further, the pickup and drop-off locations are reported in terms of coarse-grain geo-fenced zones. To estimate the trip distance from coarse-grain pickup and dropoff zones, we compute the centroid of each pickup and dropoff zone and compute the shortest road distance between the two centroids. For trips that originate and end within the same zone, we assume the trip started at the zone’s centroid and ended at the zone boundary and use half the zone length as the distance.

Figure 4.1 depicts our results. As shown in Figure 4.1a, most bike rides are short, with the median distance of a bike trip at 2.1km (for our analysis, we use median, instead of mean, since all our trip distributions have a long tail, which can skew the mean). The figure also shows that 75% of the trips, represented by the 75-th percentile of the distribution, are 3.2km or shorter. At the same time, we find that the distribution has a long tail with a small number of trips as long as 12.5km.

Figure 4.1b depicts the distribution of taxi trips from the TLC dataset. Like our bike results, we find that most taxi trips are short—the median taxi trip in 2019 was 2.7km long; this is virtually unchanged from the 2016 TLC data (not shown here for brevity) where the median taxi trip was 2.8km long. The 75-th percentile of the distribution shows that three-quarters of the trips are 5km or shorter, with the distribution exhibiting a long tail.

Figure 4.1c depicts the distribution of FHV ride distances. The results show that the median FHV ride is 3.4km, which is slightly longer than the median taxi ride. Note that many FHV rides start and stop within the same zone, and due to lack of exact pickup and dropoff coordinates for FHV rides, we approximated all of these rides to be half the zone distance, which may bias our result somewhat. Nevertheless, the majority of FHV rides are short (<3.5km), which is similar to our observations for taxi trips; the 75-th percentile of FHV trip distance is 6.1km.
(a) Bike short trips (b) FHV short trips (c) Bike medium trips (d) FHV medium trips

Figure 4.2: Cumulative distribution functions of trips duration. Short bike rides of 2km or less are faster than FHV rides. FHV trips have a greater variance due to variable traffic conditions

4.2.2.1 Key Takeaways

- The median car trip is quite short (2.7km for taxis and 3.4km for FHVs), and this is comparable to the median bike trip (2.1km). Further, three quarters of all car trips, represented by the 75-th percentile of the distribution, are less than 5 and 6.1km, for taxis and FHVs, which represent bikeable distances.

- Around 43% of the bike rides are medium-distance trips (between 2 and 5km) long, which indicates that current bike riders are amenable to biking longer distances even without electric bikes. This implies that some of medium-distance car rides are also be amenable to substitution, in addition to shorter rides. Electric bikes have the potential to increase the feasibility of such substitutions.

4.2.3 Feasibility Based on Convenience

Next, we examine the convenience of ride substitution by examining two key user convenience metrics: end-to-end ride duration and distance to the nearest bike station.

4.2.3.1 Convenience Based on Trip Duration.

The ride duration is a key convenience metric since users will have less motivation to substitute a bike ride by a car ride if bike trips take significantly longer than car trips. To analyze the feasibility of ride substitution based on trip durations, we
compute the CDFs of trip durations of short distance (less than 2km) bike and FHV rides. We also compute the CDFs of trip durations for medium distance (between 2 and 5km) rides. Figure 4.2 depicts our results.

As shown, the median short bike trip takes 6 minutes (Figure 4.2a), while the median short FHV trip is 9.9 minutes long (Figure 4.2b). This reveals an interesting result—although bicycles are slower than cars, for short rides, *biking is faster than a car ride*—likely since dedicated bike lanes in cities such as New York are less congested than regular roads, so the bikes do not get caught in congestion.

Importantly, Figure 4.2b shows that distribution of short FHV rides has a long tail and high variance, which means that many short car rides can have longer durations during periods of traffic congestion (the 90-th percentile is 26.8min and the standard deviation is 11, even when traveling less than 2km). In contrast, Figure 4.2a shows that the bike trip duration distribution has a relatively short tail, with the 90-th percentile at 12.5min and a standard deviation of 7.1. Similarly, Figures 4.2c and 4.2d plot the CDF of the duration of medium distance (between 2 and 5km) trips. The result shows, that similar to short trips, the median duration of medium trips is shorter for bikes (13.8 minutes) as compared to FHV (14.2 minutes), though the difference is not as substantial as the short trips.

*The key takeaway is that for very short rides of 2km or less, biking is 40% faster than an equivalent FHV ride due traffic congestion seen by cars on city roads and the availability of dedicated bike lanes.*

### 4.2.3.2 Accessibility of Bike Stations.

While the above analysis considered trip durations in terms of travel times, a bike trip has two components that impact trip durations: the time to walk to a bike station and the actual travel time on the bike. Users will have less motivation to use
a bike if the nearest pickup or dropoff bike stations are inconveniently located, which increases the walking distance.

To analyze accessibility of bike stations, for each car ride, we consider their pickup and dropoff coordinates and estimate the distance to the nearest bike station from those locations. Note that this analysis requires precise (GPS) pickup and dropoff coordinates, and hence, we use the 2016 TLC dataset for this analysis. For each trip, we compute the distance to the nearest bike station from the pickup and dropoff GPS coordinates and consider the greater of the two values as the nearest station for that trip.

Figure 4.3 shows the distribution of the distance to the nearest bike station for each trip. As can be seen, 18% of rides are within 100m of a bike station, 45% of the rides are within 150m and 69% are within 200m of a bike station, where 200m distance takes less than 5 minutes to walk. *This implies that about 70% of riders are within a short walking distance of a bike station, making ride substitution convenient.*

Together, Figures 4.2 and 4.3 indicate that even if we were to add a 5 minute pickup and dropoff overhead to each bike trip (equivalent to a 200m walk to/from a station), the end-to-end duration of bike trips for short rides is still comparable to the FHV trips. If we were to add pickup wait times to FHV trip durations, end-to-end duration for FHV rides would increase as well, making short bike trips more attractive from a time standpoint.

### 4.2.4 Challenges in Ride Substitution

Despite the results above, there are many scenarios where ride substitution may not be feasible even for short trips. First, if a taxi or FHV trip has multiple passengers, we assume that ride substitution is infeasible, i.e., only single passenger trips can be substituted using bike trips. The TLC data includes the number of passengers for each trip, hence in Figure 4.4, we plot the distribution of passengers per ride for taxi
Figure 4.3: Distribution of pickup/dropoff locations to the closest bike stations and FHV. The results show that for both taxi and FHV trips, more than 70% of trips are single passenger, which implies that 30% are multi-passenger trips and are not substitutable.

Second, the weather plays an important role in bike usage. Figure 4.6a depicts the demand for bike trips by the month of the year. As can be seen, bike usage drops by half in winter months in New York and with an even greater drop during January, which is the coldest month of winter. In addition, an analysis of weather data for New York shows that it rained for 170 days of the year, and there were 158 days with an average temperature of less than 10°C. To quantify the effect of weather on trip demand, we analyze the number of trips during snow, rain, and other weather conditions. We group days as either snow, rain, or ‘other’ as follows; for example, if snow (or rain) falls during a day, we label the day as snow (or rain) day. However, if both rain and snowfall exist on the same day, we label the day as a snow day. All other days are labeled ‘other’ and capture other weather conditions during which biking is more convenient than during rain or snow. We then compute the cumulative number of trips every hour for each day and then compute the hourly average across all days of the year.
Figure 4.5a shows the distribution of the number of bike trips for different weather conditions. Not surprisingly, rainy or snowy days see lower bike share usage than sunny days. Figure 4.5b depicts a similar distribution for taxi trips. The morning peak for taxi trips shows higher demand for taxi rides during sunny days than rainy and snowy days. However, we observe a slightly higher demand for taxi trips in the evening hours during rain and snow than sunny days, which indicates that people generally prefer to avoid unfavorable weather after work hours. Since our optimization focuses on the upper bound of trip demand (the number of bikes is as many as is required to satisfy peak demand), the reduction in demand during rain and snow conditions is not likely to increase the number of bikes required to satisfy demand. We, therefore, do not consider these conditions for the rest of the chapter.

Third, social factors such as holidays also play a role in bike usage. To quantify their effect on trip demand, we compute the average hourly number of trips during holidays and regular working days. We consider holiday days to be Federal Holidays only and all other days as working days. Figure 4.5c shows the distribution of bike trip demand during holidays and working days, while Figure 4.5d shows the same distribution for taxi trip demand. In both cases, we find that the average hourly demand for both bike and taxi trips is lower during holidays than during working days. Our analysis in the following sections considers all days; those with favorable weather and those without, and days with holidays. However, since the above analysis shows that demand drops during holidays and bad weather days, such factors do not need to be considered separately since lower demand will not increase bike capacity needed to handle peak demand.

4.2.4.1 Key Takeaways

In summary, our data-driven analysis shows that the median taxi and FHV ride is quite short (2.7 km and 3.4km) and around three quarters of these trips are less than
Figure 4.5: (4.5a) Distribution of bike trip demand by time of day during rainy, snowy and sunny (other), and (4.5b) distribution of taxi trip demand during rainy, sunny and snow days, and (4.5c) distribution of bike trip demand by time of day during working days and holiday days, and (4.5d) distribution of taxi trip demand during working days and holiday days.

5km, which represent distances that could be traveled on a bike. Importantly, for very short rides of 2km or less, biking is actually faster than taking a ride share trip. Further, nearly 70% of the car rides start and stop within 200m of a bike station. Even some of the medium distance car trips between 2 and 5km appear to be substitutable by bikes, since 43% of bike rides involve commuting these longer distances already, and future availability of electric bikes can further reduce the biking effort involved. The above data-driven observations show that a large fraction of taxi and FHV rides are feasible for substitution with bike rides.

In practice, however, a much smaller fraction of the rides may be amenable to such substitution due to several challenges involved in bicycling. For our subsequent analysis, we focus on the costs and benefits of substituting only a modest fraction of car rides (e.g., 1% to 10% of the total rides) using bike rides—especially since such a modest amount of ride substitution may actually be feasible with the introduction of electric bikes into a bike sharing system and appropriate user incentives.

Lastly, while our analysis has focused on New York City, our insights are broadly applicable to other dense urban areas. For instance, other studies have shown that a non-trivial fraction of the FHV and Uber rides in downtown urban districts of many cities are short [55] and share similar characteristics with our New York City
Figure 4.6: Trip demand by (a) month of the year and (b) hour of the day.

analysis. Traffic congestion in cities is a global problem, and our observation that biking is faster than a car ride for ultra-short rides will likely hold for many other cities as well.

4.3 Analysis of Cost and Carbon Benefits of Substitution

Having analyzed the feasibility of substituting shorter car rides with bike rides, we next analyze the cost and carbon benefits of modest levels of ride substitution. Specifically, we analyze the number of additional bikes needed to handle the higher demand due to ride substitution, the costs of using a hybrid mix of electric and regular bikes to meet this higher demand, and the carbon emission reductions resulting from such ride substitutions. We first describe our research methodology and then our results.

4.3.1 Estimating Ride Substitution Demand.

We begin with the CitiBike dataset described in Table 4.1 since it represents the baseline (existing) demand for bikes in the bike share system. We then consider various scenarios, each representing a different fraction of ride substitution of car
rides by bike rides. For each scenario, we augment the bike dataset with car rides that have been substituted by bike rides—by sampling the TLC dataset and adding those sampled trips to the bike dataset. We use three sampling strategies to estimate ride substitution demand. Our baseline strategy is described here. To illustrate, given a certain fraction of ride substitution, say 1%, we sample the TLC dataset to extract 1% of the total trips (here, 1% of the 354 million rides) and add them to the bike dataset. This augmented bike dataset represents the enhanced demand for bike rides as a result of that level of ride substitution.

Note that sampling of the TLC dataset must be done carefully to ensure only feasible rides are extracted from the overall dataset. Consequently, our sampling method only considers rides with the following constraints when sampling.

1. Only rides less than a threshold distance are considered, which eliminates the long tail of the distribution representing long taxi and FHV rides. That is, we exclude all trips longer than 10km from the sampling and strongly bias the sampling towards short and medium distance car rides (<5km) in the TLC dataset, since these trips are the most likely candidates for substitution.

2. Only rides that start and stop within a certain threshold walking distance (e.g. 200m) from a bike station are considered, which ensures only “convenient” rides are considered for sampling and substitution.

3. Rather than uniformly sampling the TLC dataset across time, we bias the sampling to follow the temporal distribution of bike rides. As shown in Figure 4.6b, car rides occur at all times of the day—even when it is dark—while bike rides are more common during daylight hours and exhibit peaks that are correlated with the morning and evening commute hours. We assume that ride substitution is more likely when bikers prefer to bike, and hence we sample the TLC dataset based on the temporal distribution of bike rides (i.e. sampling of car
rides from the TLC data follows the temporal distribution of bike rides shown in Figure 4.6b). As a result, the enhanced bike data set follows the same temporal distribution as the original Citibike dataset.

We use the process above to create multiple enhanced bike datasets, each representing a different level of ride substitution (e.g. 1%, 2%, 3%, 5%, 10%). This yields enhanced datasets with the total number of trips in the enhanced datasets ranging from 18 to 54 million (see Table ?? in appendix for details about the enhanced datasets).

Finally, since the above sampling process needs to ensure substituted car rides start and stop within a threshold walking distance of a bike station, we need precise GPS coordinates of pickup and dropoff locations for trips in the TLC data. Recall that the 2019 TLC data only provides coarse-grain zones for location coordinates. Hence, we use a combination of the 2019 and 2016 TLC data for constructing the enhanced bike datasets. In Section 4.2, we showed the median distance of the 2016 and 2019 taxi trips was similar (2.8 and 2.7km). Further, our analysis (omitted for brevity) shows that both 2019 and 2016 TLC datasets exhibit similar spatial and temporal distributions. Hence, we use the 2019 TLC data to determine the volume of trips that need to be sampled for a certain level of ride substitution and then sample rides from the 2016 data to construct an enhanced bike dataset with complete trip level records having the desired substitution volume.

4.3.2 Optimization Problem

We now cast an optimization problem with the objective of minimizing the number of regular and electric bikes needed to satisfy the demand represented by the above enhanced bike dataset. We divide the trips in the dataset into short, long, and medium categories by using two distance thresholds: $T_{short}$ and $T_{long}$. Short trips are assumed to have distances less than $T_{short}$, long trips are assumed to have distances
greater than $T_{\text{long}}$, and medium trips lie between the two. We assume that the demand for short and long trips is handled by regular and electric bikes, respectively. Demand for medium trips can be handled by either regular or electric bikes, based on the preference of the rider. Our optimization formulation finds the allocation of medium trips between regular and electric bikes such that the overall number of bikes is minimized (in addition to computing the optimal number of regular and electric bikes needed to handle the demand from short and long trips, respectively).

We define our optimization model as follows. Let $\mathcal{S} = \{1, \ldots, n\}$ denote the set of bike stations in a bike sharing system, each indexed by $i$. We assume a time-slotted model where in each slot $t$ and station $i$, we have incoming trips denoted by $I_i(t)$. We use mean bike trip duration as the value of $t$. Trips that start and end in a particular time slot are considered independently, i.e., trips that start in one time slice and end in another are considered as starting in the first time slot and ending in the other, respectively. We further divide the incoming trips $I_i(t)$ into $I^l_i(t)$, $I^m_i(t)$, and $I^s_i(t)$ to indicate the number of incoming long, medium, and short incoming trips to station $i$ at time $t$, respectively, and we have

$$I_i(t) = I^l_i(t) + I^m_i(t) + I^s_i(t), \quad \forall i, \forall t. \quad (4.1)$$

Next, let $O_i(t)$ denote the number of outgoing trips from station $i$ at time $t$. Similarly, outgoing trips can be further divided into long $O^l_i(t)$, medium $O^m_i(t)$, and short $O^s_i(t)$ outgoing trips, so, we have

$$O_i(t) = O^l_i(t) + O^s_i(t) + O^m_i(t), \quad \forall i, \forall t. \quad (4.2)$$

As the optimization variables, let $x_i(t)$ and $y_i(t)$ denote the number of e-bikes and regular bikes available at station $i$ and time $t$. Then, the flow conservation constraint
indicates that the outgoing flows from station \(i\) at time \(t\) should be less than or equal to the incoming flow and available bikes at the station, i.e.,

\[ I_i(t) + x_i(t) + y_i(t) \geq O_i(t), \quad \forall i, \forall t. \quad (4.3) \]

Note that bikes from incoming trips can also be used to satisfy outgoing trips. To optimize the bike usage for the medium trips, for station \(i\) at time \(t\), let us define \(I_{i}^{M,\ast}(t)\) and \(O_{i}^{M,\ast}(t)\) as additional optimization variables that determine the e-bike incoming and outgoing medium trips respectively, and \(I_{i}^{M,r}(t)\) and \(O_{i}^{M,r}(t)\) denote medium trips satisfied by regular bikes. Now, the flow conservation constraint could be further divided to separately enforce the dedicated e-bike and regular flows, i.e.,

\[
I_i^r(t) + I_i^{M,\ast}(t) + x_i(t) \geq O_i^r(t) + O_i^{M,\ast}(t), \quad \forall i, \forall t, \quad (4.4)
\]

\[
I_i^s(t) + I_i^{M,r}(t) + y_i(t) \geq O_i^s(t) + O_i^{M,r}(t), \quad \forall i, \forall t. \quad (4.5)
\]

Note that with the formulation of Equations (4.4) and (4.5) formulated, Equation (4.3) becomes redundant, and is stated for the sake of better illustration of the flow conservation constraints. We further emphasize while \(I_i^M(t)\) and \(O_i^M(t)\) are the inputs to the optimization problem, the allocation of medium trips to regular and e-bikes is performed by the optimization problem, and we have

\[
I_i^M(t) = I_i^{M,\ast}(t) + I_i^{M,r}(t), \quad \forall i, \forall t, \quad (4.6)
\]

\[
O_i^M(t) = O_i^{M,\ast}(t) + O_i^{M,r}(t), \quad \forall i, \forall t. \quad (4.7)
\]

The next constraint determines the evolution of available bikes across time slots, i.e.,

\[
x_i(t + 1) + y_i(t + 1) = x_i(t) + y_i(t) + I_i(t) - O_i(t), \quad \forall i, \forall t. \quad (4.8)
\]
That is, the overall bikes available of station $i$ at the next slot $t+1$ is equal to the aggregation of incoming and available bikes subtracted by the outgoing bikes at $t$. Similarly, the bike evolution constraints could be further partitioned to determine the number of e-bikes and regular bikes available, i.e.,

\begin{align}
  x_i(t + 1) &= I_i^L(t) + I_i^{M,r}(t) + x_i(t) - O_i^L(t) - O_i^{M,e}(t), \quad \forall i, \forall t, \\
  y_i(t + 1) &= I_i^S(t) + I_i^{M,r}(t) + y_i(t) - O_i^S(t) - O_i^{M,e}(t), \quad \forall i, \forall t.
\end{align}

Equation (4.9) (resp. Equation (4.10)) states that the e-bikes (resp. regular bikes) available at $t+1$ is equal to the aggregation of the incoming long (resp. short) trips, medium trips dedicated to e-bikes (resp. regular bikes), and the available e-bikes (resp. regular), subtracted by the outgoing long (resp. short) and medium trips handled by e-bikes (resp. regular). The last set of constraints simply enforce the availability of bikes in each slot.

\begin{align}
  x_i(t) &\geq 0 \quad \forall i, \forall t, \\
  y_i(t) &\geq 0 \quad \forall i, \forall t.
\end{align}

Finally, let $x_i(1)$ and $y_i(1)$ denote the number of bikes available at the beginning in each station, then the optimization objective is to minimize the number of regular and electric bikes at the first slot, while respecting the constraints. Hence, the optimization problem that determines the optimal mix between the regular and electric bikes (called \textbf{OptMix}, hereafter) could be formally formulated as

$$
[\text{OptMix}] \quad \min \sum_{i=1}^{n} x_i(1) + y_i(1) \\
\text{s.t.,} \quad \text{Equations (4.1) – (4.12),}
\text{vars.,} \quad x_i(t), y_i(t), I_i^{M,r}(t), I_i^{M,e}(t), O_i^{M,r}(t), O_i^{M,e}(t), \quad \forall i, \forall t.
$$

60
We use our enhanced bike datasets, each representing a certain percentage of ride substitution demand, as the input to this optimization problem; since each dataset represents one year of data, our approach assumes that demand estimates for the entire year are available in advance to solve this optimization problem. Our optimization approach yields the minimum possible number of bikes to fully cover the trips in the dataset.

Note that the actual number of bikes deployed in the system needs to be greater than this minimum (lower bound) solution to reduce the so-called blocking probability—where a user arrives at a bike station and finds the station empty. Bike share systems usually overprovision the number of bikes well above the minimum levels needed to match estimated demand in order to reduce the blocking probability of turning away bikers, especially since real-world demand will not exactly follow this estimated demand and will exhibit real-time stochastic variations.

In addition, bike-sharing systems often rebalance (or reposition, used interchangeably) bikes twice a day to cater to demand surge or depleted stations [90]. Hence with repositioning, the number of available bikes in each station changes and the bike evolution constraints (Equations (4.9) and (4.10)) will need to be modified to account for arrivals, departures, and repositioning within the system.

To capture the effects of repositioning, we can run the OptMix problem in each repositioning period separately. Specifically, let $\mathcal{P} = \{1, \ldots, P\}$, be the set of repositioning periods, each indexed by $p$. To find the overall optimal mixture of regular and electric bikes, we solve $P$ instances of OptMix, separately. In addition, let $x_i^p(1)$ and $y_i^p(1), i \in \mathcal{S}$ be the optimal number of e-bikes and regular bikes for the $p$-th instance of OptMix that takes inputs from the $p$-th repositioning period. In addition, $\text{OptMix}(p) = x_i^p(1) + y_i^p(1)$ is the optimal heterogeneous mix of bikes that will satisfy all bike trip demand within repositioning period $p$. Given the optimal solution of all instances, and to compute the global optimal mix of bikes, we take the maximum
across all instances, i.e.,

\[ x^*_i(1) = \max_{p \in P} x^p_i(1), \quad y^*_i(1) = \max_{p \in P} y^p_i(1), \quad \text{OptMix}' = x^*_i(1) + y^*_i(1), \quad (4.13) \]

where \( x^*_i(1) \) and \( y^*_i(1) \) are the maximum number of e-bikes and regular bikes required across all repositioning periods, and \( \text{OptMix}' \) is the global optimum of the \( \text{OptMix} \) problem across all repositioning periods, and we have \( \text{OptMix}' \geq \max_{p \in P} \text{OptMix}(p) \).

### 4.3.3 Cost Analysis

Given the above optimization approach, we use our enhanced bike dataset to compute the total minimum number of bikes needed to handle a certain level of ride substitution, as well as the mix of regular and electric bikes. Unless specified otherwise, we assume \( T_{\text{short}} = 2 \text{km} \) and \( T_{\text{long}} = 5 \text{km} \), which implies that very short rides of 2km or less are satisfied using regular bikes, rides longer than 5km are satisfied using electric bikes and all medium rides of 2 to 5km can be satisfied using either type of bike based on user preference. Our analysis below assumes a rebalancing frequency of 8 hours (twice a day). Since the CitiBike system sees peak demand during the morning and evening rush hours (see Figure 4.6b), we assume that bikes are repositioned once after the morning rush to fill in depleted stations, and once more after the evening rush hours [90]. We run our optimization algorithm for various levels of ride substitution ranging from 1% to 10% using the enhanced datasets, and compute the total number of bikes needed to satisfy the corresponding demand as well as the relative proportion of electric and regular bikes. Figure 4.7b depicts our results.

Note that the original CitiBike dataset has around 18 million trips. Ride substitution of 5% (i.e. 5% of 354 million TLC rides) adds around 17.7 million additional bike trips and doubles the aggregate bike trip demand. Ride substitution of 10% adds 35.4 million additional bike trips to the dataset, and triples the demand. As shown
in Figure 4.7b, the total number of bikes grows sub-linearly with increasing demand from increased ride substitution. The graph yields the following observations.

First, we observe that for 5% ride substitution—which doubles the total number of bike trips—the total number of bikes increases to 13,348 (from 9,199), roughly a 52% increase. For 10% ride substitution—which triples the total number of trips—the total number of bikes increases to 19,065, a 117% increase.

This sub-linear increase in bike share capacity to handle a linear increase in demand is an interesting result and attractive from a greening perspective. The primary reason for this sub-linear increase is due to the skewed nature of demand that causes some bike stations to see high demand while others have slack. As overall demand increases, heavily utilized bike stations need additional bikes to handle the extra demand, but less utilized stations can absorb some of the extra demand using available slack rather than requiring extra bikes. Similarly, any increase in demand during off-peak hours typically does not require a proportionate increase in bikes, since there is surplus slack available in the system during off-peak periods. As a result of these spatial and temporal skews in usage, a linear increase in demand requires a sub-linear
increase in bike capacity. Note that our optimization computes the minimum (lower bound) on the number of bikes and yet benefits from slack in the system. In a real deployment with over-provisioning over this lower bound, the actual slack is even higher. Note that while this sub-linear increase may not be the main driver for expanding an existing bike fleet, our main focus is on ride substitution and each bike trip added results in a net positive i.e. there are lower emissions, better space management due to reduced parking space and less traffic on roads.

Second, we observe that the minimum number of bikes needed to handle the baseline demand (depicted as 0% ride substitution) is 9,199. The current CitiBike system (Table 4.1) has 17,954 bikes, which implies that the system is already over-provisioned by a factor of 2 above the minimum required capacity computed by our optimization. This is in order to minimize blocking probability of turning away users due to unavailable bikes.

A final observation about Figure 4.7b is that around 26% of the total bikes are electric bikes - a significant saving to the system over using all electric bikes which tend to be more expensive to buy and maintain.

Next, we vary the distance threshold $T_{\text{long}}$ that represents the cutoff distance above which all rides are satisfied using electric bikes. To simplify the analysis, we set $T_{\text{short}} = T_{\text{long}}$ for each experiment run, which yields a single threshold above which electric bikes are used. We vary this threshold from 0 to 10km in steps and run our optimization algorithm to compute the bike capacity in each case.

Figure 4.7a depicts our results. As shown, a distance threshold of 0km implies an all-electric bike system, while that of 10km implies using mostly regular bikes (with a small number of electric bikes for satisfying ultra long bike rides that exceed 10km). The black curve depicts the sum of the regular and electric bikes needed for each distance threshold value. As can be seen, the total number of bikes needed is a bit higher for intermediate values of the threshold $T_{\text{long}}$ than at the extremes. This is
because the bike share system uses a mix of bikes for intermediate $T_{long}$ values. In this case, it is not sufficient to have a non-empty bike station when a user arrives to check out a bike - the correct type of bike needs to be available as well (e.g. a non-empty bike station that only has non-electric bikes available still causes blocking if the user wants to use an electric bike for a longer ride). Since we cannot substitute one type of bike with another, the total bike capacity needed for intermediate values is around 30% higher. At the two extremes, most of the demand is serviced by a single type of bike reducing the need for such "over-provisioning".

The figure shows that for $T_{long} = 3$km (roughly equal to the median trip distance), the number of regular and electric bikes needed is equal. For $T_{long} = 5$km, the percentage of electric bikes needed drops to 26%. This result indicates that a hybrid system can yield significant cost savings over an all-electric bike system by requiring between a quarter and a half of the total number of bikes to be electric bikes.

4.3.3.1 Key Takeaways.

First, a linear increase in ride substitution results in a sub-linear increase in bike capacity, due to available slack in the bike system. When the number of bike trips is doubled, it results in a 52% increase in bike capacity. Second, although the overall number of bikes may be higher in a hybrid bike system compared to an all-electric bike system, it can still yield cost savings as the overall number of e-bikes drops to 26% of total capacity.

4.3.4 Carbon Analysis

Next, we analyze the carbon emission reductions from substituting car rides with bike rides. Our methodology for estimating the carbon emissions from cars and bikes is as follows.

To compute the emissions from a car trip, we assume the gasoline consumption rate of taxis and FHVs is equivalent to the USA average consumption rate for passenger
vehicles, which is 22.3 miles per gallon (9.48 km/litre) [3]. The amount of gasoline consumed for a specific trip can be converted into equivalent CO₂ emissions as follows.

\[
\text{CO}_2 \text{ emissions per trip} = \frac{\text{CO}_2 \text{ emissions per litre} \times \text{total trip length}}{\text{km per litre}}. \tag{4.14}
\]

When a car trip is substituted with a non-electric bike, we assume that CO₂ emissions are zero for that bike trip. When a car trip is substituted with an electric bike, we need to estimate the small amount of emission from charging the battery. To compute the carbon emission, we assume the average e-bike energy consumption of 7.9 kWh/km [45], which captures a broad range of energy consumption scenarios from when a user starts an e-bike trip, to reaching maximum speed and coasting such that the motor is no longer turning. We then convert the electricity consumed during a bike ride into equivalent CO₂ emission as follows.

\[
\text{CO}_2 \text{ emissions per trip} = \frac{\text{CO}_2 \text{ emissions per kWh} \times \text{total trip length}}{\text{km per kWh}}. \tag{4.15}
\]

Finally, we use the Greenhouse Gas Equivalencies Calculator CO₂ emission factors of electricity and gasoline [2]: (1) Electricity: 0.486 MT/MWh, and (2) Gasoline: 0.00234 MT/litre.

Figure 4.8a depicts the monthly carbon emissions from taxis and FHVs for all trips during that month. The blue curve labelled "Current" depicts the existing CO₂ emission footprint with no ride substitution. Since the trips see a seasonal drop during winter months, the emission footprint is also lower in the winter than in other seasons. Overall, the CO₂ emission ranges between 32,000 and 35,000 MT per month over a 12 month period. The figure shows that there is a noticeable drop in CO₂ emission under a 5% and 10% ride substitution scenario.
Figure 4.8: (4.8a) Monthly emissions from taxis, FHV's and electric bikes under various ride substitution scenarios. (4.8b) Overall carbon emission reduction under different ride substitution scenarios. (4.8c) Overall carbon emission in the current bike share system under varying distance threshold.

Figure 4.8b depicts the percentage reduction in CO₂ emission under different ride substitution scenarios. As shown, 5% ride substitution yields a 3.3% reduction in CO₂ emission, while a 10% ride substitution yields 6.6% reduction in CO₂ emission. Even a small 2% ride substitution yields 1.3% savings in CO₂ emission. These results show the greening potential of ride substitution.

Interestingly, the reduction, while good at an absolute level, is sub-linear with respect to the percentage of ride substitution. This is due to the nature of sampling used to determine trips for ride substitution. If we had uniformly sampled the TLC dataset, a certain percentage of trips would yield a corresponding reduction in miles driven and an equivalent reduction in CO₂ emission. However, since the sampling is biased towards short and medium distance car trips less than 5km, these shorter rides yield a lower reduction in total miles driven (by virtual of being short), and a sub-linear reduction in CO₂ emission.

There are two reasons why these estimates are likely to be conservative. First, a reduction in trips reduces traffic congestion and brings additional benefits to other vehicles (e.g. faster moving traffic, less idling at traffic lights), which reduces carbon emissions. Second, taxis and FHV's consume dead miles driving between hailed trips,
which increases CO$_2$ emission by 47% as per a study [13]. Our analysis does not account for these dead miles, which will substantially increase the carbon emissions reported in Figure 4.8b.

Finally, Figure 4.8c shows CO$_2$ emissions from the bike share system. The distance threshold results in different number of electric bikes, and hence different levels of emission. As shown, even for a threshold of zero which yields an all-electric bike system, the total emissions for a year are 188MT, which is negligible compared to the 32,000MT of taxi and FHVs. The result highlights the benefit of using a zero or low carbon form of ride sharing for greening the overall ride share ecosystem.

4.3.4.1 Key Takeaways.

The overall emissions represented as the sum of taxi and e-bike emissions see a reduction of 7% under 10% ride substitution. Under this scenario, e-bikes account for only 0.05% of the emissions, indicating substantial carbon benefits from greening ride shares.

4.4 Related Work

The growth of bike-sharing systems has motivated significant research interest in analyzing usage patterns [104, 80, 78, 12], optimizing infrastructure [28, 29], and improving service availability [42]. Studies analyzing bike infrastructure at a city-scale provide insights into reducing congestion, pollution, improved public health, and land use [85]. This, in turn, enables policymakers to invest and transition into sustainable transportation. Separately, there have been multiple surveys that indicate e-bikes displace other modes of transport and are often used for long-distance commute [104, 30, 96]. The analysis in [96] indicates that users find e-bikes more fun to use, thereby increasing the trips made on bikes. While these studies have looked at other aspects of bike-sharing systems, our work is the first to quantify the sustainability of bike-
share systems by analyzing the conversion of taxi rides to bike rides and factoring in the emissions brought about by electricity production.

Analysis of usage patterns and trip history is especially useful in urban planning that enables the placement of bike stations [125, 28, 29]. For example, prior work uses demand predictions to recommend locations for planning future stations with high demand potential [28]. Extending bike-sharing programs at city-scale requires extensive planning and thus propose the use of data-driven approaches to keep cost low. Data-driven approaches typically rely on characterizing factors (e.g., neighborhood density, point of interest popularity) that impact bike demand mobility. Locations that show high traffic demand indicate the potential for future bike stations [29]. However, our work’s main goal is analyzing carbon reduction by substitution for-pay car rides and therefore uses FHV car trips to predict demand. Our work has also looked at different strategies of trip substitution, e.g., weather and traffic-aware substitution. Our work has also explored the expansion of existing bike-share systems in a multimodal hybrid manner, i.e., by using different distributions to drive trip substitution and using trip characteristics to inform the split of electric and regular bikes.

Rebalancing bikes has also been extensively studied [79, 90, 83, 46, 113, 27, 41, 122, 121]. Bike deployment models focus on improving the availability of bikes in stations while reducing the bikes shuffled across stations. Modeling demand surge periods, depleted bike stations, hotspots for bike pick up and drop off can play a key role in rebalancing [113, 90]. Since there is an operating cost associated with rebalancing bikes, most approaches optimize for truck routes with the number of trucks and time of day as additional constraints. For example, in [90], the authors limit the number of trucks for rebalancing and consider different rebalancing strategies during rush hour and night time [90] since traffic during rush hour is vastly different during this period. Differently, [83, 46] propose online re-positioning strategies that consider truck routes and future expected demand. In [122], a data-driven model is
proposed to predict the safe rebalancing range for bike stations. In comparison, while prior work focuses on reducing the cost of repositioning bikes by optimizing routes, our work focuses on quantifying the carbon footprint of repositioning activity and analyzing the tradeoff between capital cost and frequency of repositioning. Our work also presents an optimal way of bike-share expansion by finding clusters of high FHV activity with high potential for trip substitution, thereby identifying optimal carbon reduction opportunities by converting car trips to bike trips.

Mobility patterns from taxi cabs [42, 59, 82], and extrinsic sources such as mobile, social networks, demographics, have also been used for modeling and analyzing bike systems [66, 39, 72]. These studies mainly focus on designing and optimizing bike infrastructure systems. However, our approach provides a novel analysis of the carbon footprint in designing hybrid bike-sharing systems.

4.5 Conclusions

To green the ride sharing ecosystem, we conducted a data-driven approach to study the feasibility, costs, and carbon benefits of using hybrid electric bike share system as an alternative to ride sharing systems. From the feasibility perspective, our analysis using publicly available datasets showed that a large fraction of car rides is feasible for substitution with bike rides. We also showed that various sampling strategies yield similar results (within 11.8% of one another on average) for bike capacity needed to handle ride substitution. Our analysis showed that only 10% ride substitution reduces the carbon emissions of the entire ride share system by up to 7%, while e-bikes account for only 0.05% of the emissions. We also proposed a data-driven clustering approach to expand the bike coverage area and analyzed the benefits of such expansion from the emissions’ perspective. Finally, we studied the cost vs. emissions tradeoff of frequent repositioning of the system to prevent empty or full stations, and our results showed that the annual carbon emissions from repositioning
vehicles can be twice as much as emissions from riding e-bikes. However, they are still negligible compared to the emissions from cars.

While our results were specific to New York City, many of our observations apply more broadly to other cities. Biking using hybrid electric bike share systems can be a viable form of transport for short rides in many congested cities, and it can provide a reduction in carbon emissions from cars. As part of future work, we plan to study how the use of electric cars in ride sharing schemes can be used in conjunction with electric bike sharing to further green the ride sharing ecosystem.
CHAPTER 5

EQUITY AWARE DECARBONIZATION

In this chapter, I study the inequity that exists in the energy transition. I show that data driven approaches for building energy efficiency analysis may have inherent biases that prevent them from producing equitable results. Further, I conduct an in-depth analysis into the energy patterns of buildings revealing a huge disparity between lower and higher income households, with the former incurring up to 24% higher energy cost than high income households. I argue for design of equitable and fair approaches to the energy transition to ensure that benefits of energy improvements and decarbonization brought about by the energy transition are distributed equitably across the whole society.

5.1 Motivation

Residential sector accounts for 21% of all energy consumption and contributes nearly 20% of all greenhouse gas emissions in the United States [47]. Home cooling and heating contribute to around 38% of residential sector emissions [76]. The decarbonization of homes will be necessary as the society transitions towards a carbon-free future in an attempt to avert the disastrous effects of climate change.

Even today, a significant fraction of homes in colder climates, such as regions of North America and Europe, depend on natural gas, propane, or oil for heating homes in the winter. To reduce the carbon footprint of residential heating, electric heat pumps have been proposed as an alternative to fossil fuel-based HVAC systems. Recent advances in heat pump technology have made them efficient even at very low
temperatures of -15° C [38, 111], which makes them suitable for the cold climates of North America and Northern Europe. Electric heat pumps have two main decarbonization advantages over gas-based heating. First, they use less energy than gas furnaces to generate the same amount of heat energy [89]. Second, since they run on electricity, as more electricity gets sourced from greener and renewable sources, their carbon footprint will reduce progressively. Therefore, replacing gas furnaces with energy-efficient electric heat pumps has great potential to not only reduce a building’s energy usage, but also reduce its overall carbon footprint.

From a utility’s perspective, a decarbonization strategy involves a planned gradual shift of its customers from gas and oil-based heating to electric heat pump-based heating and cooling. This involves determining which customers to choose for heat pump retrofits in order to meet targets for reducing carbon emissions, in line with commitments made at the UN’s Paris Climate Agreement [9]. One strategy is for a utility to identify its largest emitting customers — which are homes with the highest heating bills in the winter — and prioritize them for heat pump retrofits. While such a strategy will yield the greatest initial reduction in carbon emissions from residential heating, it will not be equitable from a societal perspective. The homes with higher heating bills are likely to be large-sized homes housing affluent residents, resulting in inadvertently benefiting high-income groups. Hence, we should take an equity perspective while devising a strategy to reduce carbon emissions of heating systems.

Decarbonization strategies that target bigger homes, due to their high carbon footprint, might perpetuate social inequity against lower income households in multiple ways. First, low income households will not benefit from the high energy efficiencies of newer heating technologies. Second, since gas customers pay for the cost of maintaining the utility’s gas network, customers who cannot transition will pay higher costs as the number of gas customers dwindles over time. Finally, high income households are better equipped to bear the capital cost of replacing a heating system without
any subsidies. Therefore, the decarbonization studies should ensure that low income households are also able to benefit from and participate in decarbonization efforts. Therefore, an equitable decarbonization framework must not only consider carbon reduction potential but also quantify how socially equitable decarbonization strategies are. Recently, multiple cities have begun to factor in the social equity in their decarbonization policies by informing the distribution of decarbonization investments such as financial rebates [24, 107, 112].

In this chapter, we conduct a data-driven optimization-based study to analyze the decarbonization potential of replacing gas heating with electric heat pumps in a city-wide distribution grid. We analyze real-world gas-based heating data from 4,729 residential buildings gathered at hourly granularity over 1-year period. We analyze the heating demand of buildings and quantify their carbon footprint. Our analysis of the Energy Usage Intensity (EUI) of buildings underscores the disparity between higher and lower value homes. We then show how a transition strategy that prioritizes carbon emissions perpetuates such disparity. We then propose an equity-aware carbon-reduction approach that incorporates both the carbon reduction and social equity goals into the transition strategy. In conducting our empirical analysis and designing our equity-aware decarbonization algorithms, this chapter makes the following contributions.

**Energy demand and usage patterns.** We conduct a data-driven analysis to demonstrate the need for an equitable decarbonization strategy for transitioning homes away from gas-based heating. We analyze the heating energy demand of buildings to quantify Energy Usage Intensity (EUI) and carbon footprint of buildings. We observe that the lower income homes have a higher EUI and incur a higher energy cost per unit area than higher income homes. This disparity extends along racial lines and the neighborhoods occupied by predominately non-white races are
disproportionately affected. Finally, we show that EUI increases as a buildings age, indicating an opportunity to target such homes in decarbonization process.

**Decarbonization benefits of electric heat pumps.** We analyze the energy usage reduction and decarbonization potential of transitioning gas-based heating systems to electric heat pumps. Our results indicate that such transition allows an average home to cut energy usage by up to 60% and reduce carbon emissions by up to 80%. At city-scale, carbon emissions can be reduced by ≈55% by transitioning 40% of homes from gas to heat pumps. We also demonstrates that a carbon-first approach perpetuates social inequity by preferring higher value homes first in gas-to-electric transition.

**Equity-aware optimization for decarbonization** We present an equity-aware optimization algorithm for selecting a subset of customers for heat pump retrofits. Our algorithm embeds equity metrics into a carbon-first optimization technique and enables a flexible approach that balances the need to maximize carbon reductions while also ensuring equitable outcomes for residential households. We present both equity-aware and targeted policies to determine an overall decarbonization strategy. We will release the source code for our transition strategy as an open source tool with sample datasets.

**Equitable carbon reduction.** We conduct an empirical evaluation of our equity-aware approach. Our results show that this strategy achieves significant carbon emission reduction in transition to electric heat pumps while reducing the disparity in the value of selected homes by 5× compared to a carbon-first approach.

### 5.2 The Case for Equitable Energy Transition

Data-driven techniques that leverage building and energy data to facilitate the energy transition have recently become an active research area. For instance, researchers have proposed several techniques that use data-driven analysis and machine learn-
Figure 5.1: Average house value (left y-axis) and average annual gas energy usage intensity (right y-axis) across homes in different income groups.

Figure 5.2: Distributions of household income by race (a), distribution of EUI by house value (b), and distribution of house value by race (c).

... that facilitate the transition towards a carbon-free future. Therefore, as we strive for the decarbonization of the world, it is crucial to consider how different demographic groups are affected by these changes and make sure that everyone has access to affordable and sustainable energy solutions. This requires not only identifying energy inefficiencies in buildings and recommend areas of improvement [15, 63]. However, most of the studies focus only on identifying energy usage patterns without contextualizing the underlying demographic and societal causes of such observations. As a result, these studies lack an equity perspective.

Traditionally, disadvantaged parts of the society have borne the higher burden of pollution, lack of access to clean and renewable energy, and higher energy costs. This marginalization has continued even in the energy transition happening today, where less privileged communities are left out of opportunities that facilitate the transition towards a carbon-free future. Therefore, as we strive for the decarbonization of the...
grid, it is important to not only focus on carbon reduction, but also consider how to do so equitably. A prerequisite for designing equitable techniques is that we must understand the underlying inequity that exists in energy usage.

To quantify the social inequity in energy usage patterns across different demographic profiles, we leverage two main datasets which we describe in more detail in Section 6.2.2. First, we use fine grained energy usage data to discover patterns in energy usage at the household level. Second, we combine the observed patterns with demographic and tax data at the community level to discover how energy usage correlates with social and demographic constructs. We show how different demographic profiles, including income level, value of a home, age of a home, and race of the residents, impact energy usage patterns. We also show how such insights can help devise equitable energy transition and decarbonization strategies.

We begin by analyzing how the level of income impacts energy usage at the household level. We focus on gas energy usage and compute the energy use intensity (EUI) for each home in our dataset. Since our end goal involves replacing gas-based heating with electric heat pumps, we convert gas usage data (measured in volume consumed per unit time in CCF) to the equivalent energy usage in kWh. To do so, we first compute the amount of heat energy generated from gas heating, and then compute the equivalent electric energy required to generate the same amount of heat energy. We then use census data (as reported by the US Census Bureau) to group homes according to household income relative to the whole population. We divide homes into three groups — low, medium, and high income homes. We classify homes with income below the 20th percentile as low income homes and homes with income above the 80th percentile as high income homes. All the other homes are categorized as medium income homes. Our subsequent analysis of EUI and house values uses the resulting income groups.
Figure 5.1 depicts the average EUI and house value for each income group. The average EUI for high income homes is 162.9, while low income homes have a 24% higher average EUI of 202.3. This highlights a large disparity in EUI by income and reveals that low income households pay more per unit area for energy than high income homes. This is despite that high income homes have higher purchasing power than lower income homes, representing an inequitable distribution of the energy cost burden across the two demographics. We hypothesize that this disparity is because lower income homes tend to have poor house insulation, which results in less energy-efficient heating and high energy usage per unit area. Finally, our analysis also indicates that homes in the high income group are 12% higher in value than homes in the low income group, but are less expensive to heat than their low-income counterparts.

We next examine the racial distribution of population in each income group. To do so, we use the racial distribution for each census block — as reported by the Census Bureau — and compute the ratio of white to non-white population. Figure 6.2a, demonstrates that the lower income census blocks are predominantly made up of non-white populations, while the wealthiest blocks are predominantly white. This further indicates the racial disparities present in society. The lower income group, which is largely non-white population, experiences the highest energy cost burden. We also examine the impact of house value on the resulting EUI of the home. To do so, we group homes into deciles based on the individual home value. Figure 6.2b depicts the distribution of EUI for homes in each decile. It indicates that lower value homes have higher EUI than higher values homes. The lowest 10% of homes have an average EUI of 90.8, compared to an average EUI of 66.8 in the top 10% of homes. An EUI disparity of 36% and means that lower value homes have a higher energy cost per unit area further exacerbating the inequity between the two groups. Further examining the racial profiles based on home values reveals racial inequity
in energy usage. Figure 6.2c depicts the relationship between the value of a home and the racial distribution of residents. The figure shows that lower value homes are predominantly located in areas with a high percentage of non-white residents, while the most expensive homes are primarily located in areas with a high percentage of white residents.

Finally, we examine the impact of age on the EUI of a home. To do so, we first group homes based on the year the building was built (buckets of 20 years each). Figure 5.3 depicts distribution of EUI for all homes in each age group. The figure shows that the older homes have a higher EUI compared to newer homes. This is because the building envelope degrades over time, the building becomes less energy efficient, and the EUI increases. This reveals an important insight for designing decarbonization strategies. Newer buildings are already less carbon-intensive and older homes should be prioritized in energy transition and decarbonization efforts.

**Summary and Key Takeaways.** Our data-driven analysis yields the following key observations.

1. Low income homes have higher EUI than high income homes. They pay a higher energy cost per unit area despite having lower purchasing power than high income homes and share a disproportionate energy cost burden.

2. Income based inequity disproportionately affects non-white populations as they are more likely to be low-income earners than white populations. This also means that energy inequity also affects non-white population more than white population.

3. Lower value homes have higher EUI than high value homes. They pay a higher energy cost per unit area than their high value counterparts. Since non-white populations are more likely to live in lower value homes, house value inequity also affects non-white populations more than white populations.
4. Older homes have a higher EUI than newer homes. This presents an opportunity for targeted transition based on the home age.

Given these observations, the primary goal of this chapter is to devise a decarbonization strategy to transition homes from gas-based heating to electric heat pumps that achieves highest carbon reductions while satisfying equity constraints. Specifically, we seek to answer the following research questions.

1. How can we design an optimization framework that maximizes carbon reduction by transitioning a subset of homes from a group from gas-based heating to electric heat pumps?

2. How can we embed equity metrics into a framework for maximizing carbon reductions to ensure equitable transition?

3. How do the carbon-first and equity-aware approaches for transitioning impact carbon emissions? What is the impact of varying levels of transition on the carbon footprint as well as the energy usage intensity of homes in our dataset?

5.3 Data-Driven Decarbonization

In this section, we present our data-driven decarbonization approach. Our primary goal is twofold — first, to maximize the amount of carbon reduction while transitioning a subset of homes from gas-based heating to electric heat pumps, and second, to enable selection criteria to either ensure equity or targeted selection of homes for transition. An equitable transition means reducing disparity in EUI across different demographic profiles. A targeted transition may refer to selecting a group of homes built in a certain time period to retiring old gas lines. Our data-driven approach is therefore a two-step process. First, we develop an optimization problem with an objective to maximize the carbon reduction by selecting the highest emitting homes.
Figure 5.3: Distribution of EUI by age of building i.e. year the building was built.

Second, we embed additional constraints to ensure equity in home selection, as well as target specific homes that meet the specified criteria such as age and location.

5.3.1 Carbon-first approach

We develop a linear optimization model whose goal is to maximize the amount of carbon reduction achieved while transitioning homes from gas-based heating to electric heat pumps.

Let \( H = \{h_1, h_2\ldots h_n\} \) denote the set of buildings, each indexed by \( i \). Let \( X_i^g \) denote the total carbon emissions from the cumulative annual gas consumption for heating for building \( i \). Let \( X_i^e \) denote the total carbon emissions from the cumulative annual electricity consumption of heat pumps for building \( i \). Let \( \tau_i \) be a binary variable that represents the status of transition for the building \( i \) and \( \Gamma \) denotes the target number of buildings for transition. Our objective is to select \( \Gamma \) buildings from the set \( H \) that result in the highest carbon emissions reductions after transitioning from gas-based heating to electric heat pumps. This objective can be described as follows.
\[
\min \sum_{i=1}^{n} \left(1 - \alpha_i\right) \cdot X_{gi} + \tau_i \cdot X_{ei}
\]
\[
s.t., \quad \text{Equations (6.2) - (6.4)}
\]
\[
\text{vars., } X_{gi}, X_{ei}, \tau_i, \Gamma \quad \forall i
\]

Our first constraint relates to the level of transition. To ensure that only \(\Gamma\) buildings are transitioned, the sum of all values of \(\tau_i\) must equal \(\Gamma\), as stated below.
\[
\sum_{i=1}^{n} \tau_i = \Gamma \quad (5.2)
\]

Our second set of constraints simply ensure that a building cannot have negative carbon emissions from either the gas consumption or the electric demand.
\[
X_{gi} \geq 0 \quad \forall i \quad (5.3)
\]
\[
X_{ei} \geq 0 \quad \forall i \quad (5.4)
\]

The annual carbon emissions from gas consumption, \(X_{gi}\), for a building \(i\) is a multiple of the total heating gas demand \(Y_{gi}\) and the carbon intensity of gas \(I_g\).
\[
X_{gi} = Y_{gi} \times I_g \quad (5.5)
\]

The annual carbon emissions from electric demand, \(X_{ei}\), for a building \(i\) is a multiple of the total electricity demand \(Y_{ei}\) and the carbon intensity of the electric grid \(I_\epsilon\).
\[
X_{ei} = Y_{ei} \times I_\epsilon \quad (5.6)
\]

### 5.3.2 Equitable transition

To ensure equitable distribution of selected homes, we analyze two strategies that ensure an equitable distribution of homes across both the income groups and the census blocks.
5.3.2.1 Equitable distribution across income groups

The goal of this strategy is to select an equitable number of homes from each income group as the homes selected for transition. We begin by applying our carbon-first approach to select a subset of homes to transition which is ordered by the level of carbon emissions reduction. We set $\Gamma = \Gamma'$ as the target number of homes, where $\Gamma'$ is higher than the actual target $\Gamma$. In our experiment, we set $\Gamma' = \Gamma \times 3$ — therefore, the optimization model selects the top $\Gamma'$ emitting homes. We then apply our equitable sampling strategy to this output to select $\Gamma' \div 3$ homes, which is the actual expected number of homes.

To select the subset $\Gamma' \div 3$ from $\Gamma'$, we perform the following steps. We first compute $\Gamma_{low}$, $\Gamma_{medium}$ and $\Gamma_{high}$, where $\Gamma_{low} = \frac{\text{no. of low income homes}}{n} \times \Gamma$, $\Gamma_{medium} = \frac{\text{no. of medium income homes}}{n} \times \Gamma$, and $\Gamma_{high} = \frac{\text{no. of high income homes}}{n} \times \Gamma$. We then select $\Gamma_{low}$ low income homes from $\Gamma'$, $\Gamma_{medium}$ medium income homes from $\Gamma'$, and $\Gamma_{high}$ high income homes from $\Gamma'$. The final number of homes $\Gamma = \Gamma_{low} + \Gamma_{medium} + \Gamma_{high}$, and is made up of an equal proportion of homes from each income group in the subset.

5.3.2.2 Equitable distribution across census blocks

The goal of this strategy is to select an equitable number of homes for transition from each census block, i.e., we select homes from each block as a proportion of the total homes present in that block. Similar to the previous strategy, we first apply our carbon-first approach to select a subset homes $\Gamma'$, where $\Gamma' = \Gamma \times 3$. From this subset, we then select the target number of homes $\Gamma$ as follows. We first compute $\Gamma_b$ for each block, where $\Gamma_b = \frac{\text{no. of homes in block}}{n} \times \Gamma$ for all blocks. We then select $\Gamma_b$ homes from $\Gamma'$ for each block. The final number of homes $\Gamma = \Gamma_{b1} + \Gamma_{b2} + \ldots + \Gamma_{bk}$, where $k$ is the total number of blocks, and is made up of an equal proportion of homes from each block as a fraction of the total number of homes in that block.
5.3.3 Targeted transition

In addition to equitable transition, we also analyze strategies that provide targeted transition for homes that meet specific criteria. Targeted transition from gas is crucial as it can improve the safety of the infrastructure and reduce the cost of transition. For example, older homes can be prioritized in transition because they are typically serviced by older gas lines that may pose risks of leakage, etc. To transition such homes, targeted strategies can be used to select homes built within a particular decade as the first homes for transition while also maximizing carbon reduction. Similarly, since transitioning homes from gas may involve retiring an entire gas line, a targeted strategy can be used to select homes in a certain geographic location that are serviced by the same gas line. We analyze two targeted transition strategies that maximize carbon reduction while prioritizing homes based on age and geographic location.

5.3.3.1 Age-based selection

In this strategy, we skew the selection towards homes within a particular age group. This strategy allows targeting homes that have higher carbon footprint due to their age. Our analysis in Section ?? showed that homes built between 1920-1960 years have a disproportionately high EUI compared to homes built later. We can prioritize homes in these age groups in our transition criteria. To do so, we first select $\Gamma'$ homes using our carbon-first approach, where $\Gamma' = \Gamma \times 3$. From the resulting $\Gamma'$ homes, we then select a subset $\Gamma$ as follows. First, we prioritize homes within $\Gamma'$ whose age group falls in the selected targets. If the number of homes in the targeted age group is less than $\Gamma$, we select the remainder of homes from $\Gamma'$ in order of carbon emissions. The result is a subset of homes that maximizes carbon emissions reduction and falls within the specified age group.
Figure 5.4: Distribution of number of homes per census block (a), number of homes by age (b), and values of homes (c).

5.3.3.2 Block-based selection

Our next strategy targets homes based on census blocks and offers two main advantages. First, since the transition from gas will involve migrating customers from existing gas lines, targeting a group of homes that get served by the same gas line lowers the cost of transition from a gas utility’s perspective. If all homes on a certain line are migrated, maintenance costs on that line will be eliminated, and the line can be shut off from supply. This strategy can also be used to target old lines which would otherwise need to be replaced or upgraded. Second, neighborhoods with high aggregate carbon footprint can be targeted to maximize carbon reduction by transitioning all homes within such blocks.

To perform targeted transition by blocks, we skew selection towards certain blocks by selecting all homes within that block for transition. To do so, we first use our carbon-first approach to select $\Gamma'$ homes where $\Gamma' = \Gamma \times 3$. We then compute the aggregate carbon footprint for each block that exists in the selected $\Gamma'$ homes. We then compute the total number of homes in each block $\Gamma_b$. If $\Gamma_b$ is less than the required target $\Gamma$, we select all homes in that block, and eliminate homes in $\Gamma'$ that do not fall in the specified blocks.
5.4 Evaluation

In this section, we evaluate the performance of our carbon-first approach, as well equitable and targeted strategies on carbon reduction. We evaluate different levels of transition i.e. from 5-40%. At each transition level, we compute the overall carbon reduction, as well as analyze the EUI and value of homes in the selected subset. We then conduct equity-aware analysis to evaluate the transition from an equity and a targeted perspective. We will release the source code with sample data as an open source tool for reproducibility.

5.4.1 Experimental setup

The gas and electricity consumption data from the buildings (described in Section 6.2.2) provides building-level metering of the gas and electricity demand. We first disaggregate the gas and electric demand data into two components: first used for heating purposes, second used by all the other appliances such as stoves. For gas demand data, we compute the average gas usage during the summer months, which corresponds to non-heating uses of utility gas. We subtract this from the year-round data to get the heating components of the gas demand. We also account for energy loss due to inefficiencies of the gas furnace. We use an efficiency value of 87.5%, which lies between the typical efficiency for a standard and a high efficiency furnace. To compute total carbon emissions from gas usage, we use a gas carbon intensity value of 0.0551 MT/MCF [8].

To compute the corresponding electric heat pump emissions, we compute the total heat energy generated by the volume of gas consumed and use the Heating Seasonal Performance Factor (HSPF) of electric heat pumps to compute the electric energy required to generate the equivalent amount of heat energy. For all experiments, we assume a HSPF rating of 8.5, which is typical of many efficient heat pump models. Finally, to compute the carbon emissions from electric heat pump usage, we use
a carbon intensity value of 0.000386 MT CO₂/kWh, corresponding to the average carbon intensity value for the United States electric grid [110, 57].

5.4.2 Key characteristics of the data

5.4.2.1 Gas distribution and usage dataset

Our energy usage dataset consists of electric and gas usage data from 14,094 buildings in a small city. The data reports electricity usage in kWh and gas usage in Ccf. 56% of homes in the data (7,846) are both electric and gas customers and we limit our experiments to the 7,846 homes only. The dataset also contains real estate information that includes a building’s size, type of home e.g. single vs multi-family,
Figure 5.7: Median house value for homes selected at varying levels of transition.

![Median house value](image)

Figure 5.8: Carbon emissions achieved for decarbonization strategy that targets equity based on income group (a) and census block (b).

![Emission by Income Group](image)

![Emission by Census Block](image)

type of building e.g. apartment, school etc, and the year the building was built. We filter out apartments, factories, schools, etc — whose reported size may be inaccurate. We perform our data-driven analysis on the remaining 4,729 homes. The usage data spans 2014 to 2019 but our study focuses on one year period between Jan-Dec 2019.

Figure 5.4a depicts the distribution of the number of homes in each census block. The figure shows a long tail, with most blocks having between 1-15 homes. The median block has 27 homes, with the most populated blocks having more than 150 homes. These characteristics present interesting opportunities for equitable as well as targeted transition. For example, a targeted transition strategy may focus on homes
in the same block to minimize disruptions in the gas distribution lines. Similarly, an equitable transition strategy may aim to select an equal number of homes from each block to ensure equity across the whole population. Figure 5.4b depicts the number of homes by the year built. The figure shows that most homes are old and built in the 19\textsuperscript{th} century. This presents an opportunity for targeted transition where older buildings are prioritized. Such transition can improve safety by retiring older gas lines that were installed in the 19\textsuperscript{th} century. Figure 5.4c depicts the distribution of house value in the city. The figure shows the median house value is \( \approx \$165,000 \), with a few homes having a value above \( \$600,000 \).

Figure 5.5 depicts the aggregate gas energy demand for the whole city during 2019. The figure shows two peaks — between Jan-Feb and Nov-Dec months, which coincides with the winter season in the region of study. The average daily gas demand during winter months is 94.4MMCF, which is 6\( \times \) the daily average demand during summer (15.6MMCF). The figure also shows strong negative correlation (-0.9) between temperature and gas demand — as the temperature falls, the gas demand rises due to increased space heating in homes.
5.4.2.2 Demographic profile data

We used Geocode API\(^1\) to collect demographic data such as race, population and median income. We use census data from the year 2020, which represents the most recent year whose census data is available. The data is provided by the U.S. Census Bureau per census block which is the smallest geographical unit for which the Bureau provides statistical data. We use geocoding to map each home in the dataset to the parent census block and use the block data to compute our equity metrics.

5.4.3 Carbon-first approach

We begin by analyzing the effect of transitioning homes from gas to electric heat pumps using the carbon-first approach. To do so, we simulate electric heat pump transition by converting emission from transitioned homes from gas to electric heat pump emissions. We first run our carbon-first framework on the data while varying the target number of homes from 5-40%. At each transition level, we compute the total amount of carbon emissions reduced as well analyze the effect on EUI after such transition.

Figure 5.6 depicts the total carbon reduction achieved by transitioning a varying number of homes using the carbon first approach. The figure shows a linear relationship between the amount of carbon reduction achieved and the number of transitioned homes. For example, transitioning 5% of homes from gas to electric heat pump results in a 14.2% reduction in total emissions. At the other end, replacing 40% of gas-based heating results in a 54.8% reduction in carbon emission. This is because the carbon intensity of electricity in our area of study is lower than the carbon intensity of the natural gas, and the amount of energy required by electric heat pumps to generate heat is also lower than what is required by gas.

\(^{1}\)https://www.geocod.io/docs/
We next examine the value of homes selected by the carbon-first approach. Figure 5.7 depicts the median house value for homes selected at the various levels of transition. The figure shows an inverse relationship between the number of homes transitioned and the median value of homes in the subset. At low transition levels, the median house value is high, indicating this approach prefers higher value houses to lower value ones. This indicates bias against lower value homes in selection and perpetuates inequity.

5.4.4 Equitable transition

We next analyze the impact of transition strategies that take equity into consideration. To do so, we apply the equitable transition approaches described in Section 5.2 and analyze the reduction in carbon emissions, as well as the demographic properties of selected homes. Since these approaches focus on optimizing the trade-off between carbon reduction and equity, the overall amount of emission reduction is lower than the carbon-first approach. However, disparity in energy usage across different demographics is minimized. We next analyze the performance of equitable transition by income groups and geographical location i.e. census blocks.

Figure 5.8a depicts the results for a strategy that targets equity based on income groups. The figure shows a slight reduction in the total amount of carbon reduction achieved compared to the carbon-first approach. Carbon emissions reduction decrease by 0.2% from 14.2 to 14% at 5% transition, and by 1.1% from 45.6 to 44.5% at 30% transition. This is because some homes that have lower emissions are being added to the subset while some high emitting homes are removed to ensure equity in the selection process. However, the figure still indicates a super-linear relationship between the number of homes converted and the amount of carbon reduction achieved.

Since census blocks share similar demographic characteristics, as discussed in in Section ??, equalizing the number of selected homes from each blocks ensures an eq-
Figure 5.10: Distribution of EUI by age group at 5% transition.

Figure 5.11: Distribution of total winter emissions by census block.
uitable selection across different demographic profiles. Figure 5.8b depicts the carbon emission reductions achieved for a strategy that targets equity across census blocks. As expected, the total amount of carbon emissions reduction decrease compared to the carbon-first approach. At 5% transition, carbon emissions reduction decrease by 2.3% from 14.2% to 11.9%. This is because this strategy considers lower emitting homes that are in under-represented census blocks in carbon-first approach, which ensures equity in representation across census blocks.

Next, we quantify disparity by analyzing the value of homes for equity-aware technique. Figure 5.9a depicts the median house value at various transition levels for the equitable income group strategy. The figure shows an equitable distribution of house value as the median house value in each subset is closer to the overall median. Similarly, Figure 5.9b depicts the median house value for a strategy that targets equitable distribution by census block. The figure also shows an equitable selection from a house value perspective with the median value in each group being closer to the overall median compared to the carbon-first approach. To quantify the reduction in disparity, we compute the RMSE between the median value of selected homes and the overall median. The RMSE of the carbon-first approach in median house value is 25.78, compared to 5.077 in the equitable selection, indicating a reduction of 5× in disparity.

5.4.5 Targeted transition

We next analyze the carbon emissions and impacts on EUI for targeted transition using age-based and block-based approaches.

Our analysis in Section 5.2 showed that buildings built between the years 1920-1960 have higher energy cost per unit area than homes built during the other decades. Therefore, we perform targeted transition which skews selected buildings towards homes built during these decades and then analyze the carbon and EUI after tran-
sition. Figure 5.10 depicts the distribution of EUI for the carbon-first and targeted strategies at 5% transition. The targeted approach prioritizes two age groups i.e. between 1920-1960. The figure shows that the mean and median EUI after running both the carbon-first and targeted approaches are lower than the current EUI distribution. The biggest gain in EUI reduction occurs in the third bucket (1920-1940) where the average EUI in the targeted approach (19.6) is 30% lower than the current average EUI in the same bucket (28). This indicates that the targeted approach selects most homes from this bucket, suggesting that these homes are the most energy inefficient. The figure also shows that the targeted approach prioritizes the targeted age groups at the expense of other age groups e.g. 1900-1920, where the carbon-first EUI is significantly lower than the targeted approach EUI. Finally, in both approaches, at 5% transition, none of the newer homes (≥2000) are selected, suggesting that newer homes have higher energy efficiency than older ones.

To examine the impact of targeting buildings by geographical location, we perform targeted transition by census blocks on the dataset. Figure 5.11 depicts the distribution of aggregate carbon emissions for census blocks in the data. The figure shows that the median block emits 743.6 MT CO₂ during winter, with most blocks emitting between 5-500 MT CO₂. The figure also shows a long tail, with some blocks emitting more than 4000 MT CO₂ (≥5× the median emission). This shows a huge emission disparity in the highest emitting blocks, with the top 10% block accounting for more than 33% of all emission in the city, which shows the importance of performing targeted transition by census blocks. Similar to targeting homes by age, our analysis here shows that carbon reduction can be achieved by transitioning a small number of sections of the entire grid.
5.5 Related Work

In this section, we discuss prior work on decarbonizing heating in buildings, electric heat pumps, and equity in energy transition.

**Decarbonizing heating.** There is a significant prior work on decarbonizing heating in the building sector [93, 120, 75, 58, 16]. Padovani et al [93] quantify the decarbonization and economic impacts of replacing propane heating with electric heating such as solar heat pumps in isolated, rural residential buildings. Waite & Modi [120] propose a hybrid transition approach that combines existing fossil fuel systems with electric heat pumps with a plan to gradually phases out fossil fuels over time. Leibowicz et al develop an optimization model for decarbonizing residential buildings that considers switching to cleaner energy sources, upgrading to energy-efficient appliances, and improving the thermal properties of buildings. Hopkins et al propose transitioning to electric heat pumps. Finally, Baldino et al [16] analyze the cost and decarbonization benefits of using hydrogen and renewable-based electricity for heating. While all these studies focus on the decarbonization benefits of transitioning heating, they ignore equity. Our work is complementary to this work, as we introduce equity into the decarbonization process. We not only show how to decarbonize heating, but also how to do so equitably.

**Electric heat pumps.** There has been numerous studies on the viability of electric heat pumps as a replacement for gas heating in residential buildings [124, 95, 120, 69, 68, 87]. These studies either focus on evaluating the performance of heat pumps in extreme temperatures or analyze their decarbonization potential at various geographical scales. For example, Johnshon et al [65] analyze the cost of transiting to electric heat pumps and how it varies across different regions across United States. Zhang et al [124] study the decarbonization benefits of electric heat pumps using a simulated energy system of an entire city. Our work is complementary to these studies as it
introduces an equity perspective to the deployment of such technology, and analyzes how such transition efforts affect the community across various demographic profiles.

**Equity in the energy transition.** Prior studies have evaluated social inequality in energy use and how it can be used to inform more equitable distribution of energy resources in the energy transition. For instance, Tong et al [118] analyzed the disparities in energy usage across different demographic profiles such as race and income. Other studies have analyzed the inequity in the energy transition and shown that people with lower incomes and those who belong traditionally marginalized races are negatively impacted by the emerging energy transition [17, 106]. Other studies have also proposed incorporating equity into the energy transition policies [35, 36]. Our work complements the prior work as we develop an optimization framework that fuses together deployment of cleaner sources of energy for heating as a decarbonization strategy, and incorporates equity into the process such that decarbonization is more equitable across different demographic profiles.

### 5.6 Conclusions

In this chapter, a data-driven analysis to quantify the decarbonization potential of replacing gas heating with electric heat pumps in a city-wide distribution grid was conducted. Our goal was to reduce the overall carbon footprint in an equitable manner. We conduct an in-depth analysis into the energy patterns of buildings revealing a huge disparity (24%) between low and high income households. We first analyzed the decarbonization potential of a strategy that prioritizes reducing carbon emissions. We then presented equity-aware approaches for selecting buildings to transition that balance the need to maximize carbon reduction with ensuring equitable outcomes for residential households. We showed that the equitable strategies achieve significant levels of carbon emissions reduction while reducing the disparity in value of the selected buildings by 5× compared to a carbon-first approach.
CHAPTER 6
MULTI-OBJECTIVE DECARBONIZATION OF RESIDENTIAL HEATING SYSTEMS

In this chapter, I conduct a data-driven optimization study to analyze the potential of replacing gas heating with electric heat pumps to reduce CO₂ emission in a city-wide distribution grid. I present a flexible multi-objective optimization (MOO) framework that optimizes carbon emission reduction while also maximizing other aspects of the energy transition such as carbon-efficiency, and minimizing energy inefficiency in buildings.

6.1 Motivation

To date, a significant fraction of buildings in colder climates, such as regions of North America and Europe, depend on natural gas, propane, or oil for residential heating in the winter. For example, 82% of Massachusetts households use non-electric sources of energy — such as utility gas, heating oil, or propane — for heating [26]. On the other hand, only 16% of households use electric heating. The low adoption of electric heating is attributed to the historical inefficiency of electric heat pumps in extremely cold climates. However, recent technological advancements have made it possible to operate electric heat pumps efficiently even at very low temperatures of -15° C [38, 111]. This has made modern heat pumps viable candidates for replacing fossil fuel based heating even in the extreme climates of North America or Northern Europe.

Electric heat pumps offer two key decarbonization advantages over fossil fuel based heating, such as utility gas. First, they are more energy-efficient, which means they
use less energy than gas furnaces to generate the same amount of heat energy [89]. Second, their reliance on electricity means that as the electric grid transitions towards greener and renewable sources for energy production, the carbon footprint of electric heat pumps will also decrease. In contrast, the carbon footprint of fossil fuel based heating will remain constant as the energy efficiency of gas furnaces is reaching its limits [40]. As a result, replacing gas furnaces with energy-efficient electric heat pumps has great potential to not only reduce a building’s energy usage, but also reduce its overall carbon footprint.

A push for transition to electric heat pump can come from either the utility or the end consumers. Although consumers do not have a direct incentive to reduce their carbon emissions, they do have a strong financial incentive to reduce their energy consumption, and ultimately their utility bills. Such transition requires identifying a set of buildings to be retrofitted with electric heat pumps. The selection of buildings is non-trivial and traditionally depends on various factors, such as the total energy consumption, energy-efficiency, and insulation levels. However, for decarbonization, the two of the most important factors are the total carbon footprint and the carbon-efficiency of a building. The total carbon footprint quantifies the total amount of emissions for heating purposes, irrespective of how much heat was generated and the efficiency of the process. Carbon efficiency, on the other hand, quantifies the amount of heat generated per unit of carbon emitted. The two metrics are related but distinct. For example, a building may have a large total carbon footprint, but be highly carbon-efficient. Therefore, while a carbon reduction strategy that targets the highest emitting buildings yields the greatest initial reduction in CO₂ emissions, it does not fully exploit additional opportunities for improvements, such as increasing building energy efficiency. In addition, there are additional questions that need to be answered. How does the choice of one metric impact the other? How does carbon-efficiency differ from energy-efficiency from a decarbonization standpoint? How do
carbon emissions reduction the energy consumption (also a proxy for heating cost)
for the end consumers? The answers to these questions are non-trivial and require
an in-depth analysis of real energy consumption data.

In this chapter, we conduct a data-driven optimization study to analyze the po-
tential carbon emission reductions from replacing gas-based heating with electric heat
pumps in a city-wide distribution grid. Our empirical study is based on analyzing
real natural gas and electric data from 13,800 and 6,445 smart electric and gas meters
respectively collected over a one year period. We conduct an in-depth analysis of the
heating demand of buildings and quantify their carbon footprint. We quantify CO₂
emission reductions obtained when a carbon-optimal transition strategy is applied
to the conversion from gas to electric heating. We then introduce additional goals
such as CO₂ efficiency and improving building efficiency to take advantage of fur-
ther energy improvements in addition to CO₂ reduction. In conducting our empirical
analysis, this chapter makes the following contributions.

**Energy Consumption and Emission Analysis.** We use a city-scale dataset to
conduct an in-depth analysis of its gas consumption and the resulting CO₂ emissions.
One of our key findings is that the median building produces ≈ 32 MT of CO₂
annually, with some buildings emitting ≥250MT CO₂, which is 7.8× the median. This
analysis motivates transition strategies that target buildings with higher emissions to
meet aggressive decarbonization goals.

**Optimal CO₂ Reduction.** We present a multi-objective optimization (MOO)
framework that enables the flexible selection of a subset of homes for heat pump
retrofits to achieve decarbonization goals. Our analysis of a transition and build-
ing selection strategy that achieves maximum possible initial CO₂ reductions sug-
gests that it fails to take advantage of other aspects of energy transition such as
improving energy and carbon efficiency in buildings. Consequently, we update our
multi-objective optimization framework to consider additional objectives of energy efficiency and carbon efficiency.

**Joint CO₂ and energy-efficiency optimization.** In addition to a carbon emissions analysis, we analyze the energy inefficiency of buildings and its causes. We show energy efficiency can be improved by transitioning buildings from gas to electric heat pumps, to reduce emissions, while simultaneously improving energy efficiency via renovations, such as adding insulation to the building. We show the effect of prioritizing energy efficiency on energy demand and CO₂ emissions. Our analysis finds that older buildings are generally less efficient and should be prioritized in transition.

### 6.2 Problem and Methodology

In this section, we present the problem statement and key research questions we address in this chapter. We also describe the datasets and experimental methodology we use to answer these questions.

#### 6.2.1 Problem Statement

Given a set of residential buildings in a city or town, the primary goal of our work is to quantify the impact of replacing gas heating with electric heat pumps on carbon emission reductions, and the optimal order in which homes should be transitioned. Another goal is to understand the impact of introducing additional goals such as carbon-efficiency and energy inefficiency in buildings as priorities for such a transition, and the tradeoffs such goals have on emissions reduction. Specifically, we seek to answer the following questions.

1. What is the distribution of heating energy consumption, and how much gas is required to meet these heating requirements? What are the daily and seasonal variations in gas consumption? How much CO₂ is emitted from this gas consumption, both for individual residential buildings and in the aggregate?
2. What is the impact of replacing gas heating with electric heat pumps on energy consumption and CO₂ emissions? What is the optimal order in which buildings should be transitioned from gas to electric heat pumps in order to minimize CO₂ emission?

3. How is this ordering impacted when additional goals such as carbon/energy inefficiency of buildings are introduced? How is CO₂ reduction impacted, and what are the tradeoffs?

6.2.2 Description of Datasets

The answers to these questions vary based on regions and largely depend on seasonal factors such as the severity of winter weather, which in turn influences gas demand for heating. In this chapter, we use data collected from a small city in the Northern region of United States to answer these questions. Since the gas and electric system design in this city is typical of many regions with similar seasonal patterns, we believe that our insights are widely applicable.

Gas and Electric Usage Data. Our dataset consists of electric and gas consumption data recorded by 13,800 electric and 6,445 gas meters. The data also includes a mapping of electric and gas meters installed at each building. To compute the aggregate load profile of a building, we sum up the load from the electric and gas meters installed in the building. Electricity demand data is recorded at 5 minute granularity and spans >5 years. Gas consumption data is recorded at hourly granularity, and spans the same duration. For the purpose of our study, we limit our analysis to the full calendar year 2020, which is the latest year whose complete data was available.

Building Property Data. In addition to load data, we collect property data for all buildings present in our dataset using public real-estate records. This includes the size of the building, type of building, e.g., single vs multi-family, etc. We use this data to augment our analysis, e.g., to generate a building’s energy profile, we normalize the
load by the building’s size to enable comparative analysis across different buildings. We gather and parse this data from publicly available property information recorded as part of tax records.

**Weather Data.** Since our analysis involves measuring the impact of weather on energy usage, we gather weather data for the city from the Dark Sky API \(^1\). We collect multiple data points such as temperature, humidity from the API. We gather this data at hourly granularity to match our hourly gas load data.

### 6.3 Energy Usage and Carbon Analysis

To understand the impact of transitioning buildings from gas to electric heat pump heating, we begin with an analysis of the current load on the gas system and the resulting CO\(_2\) emission. Specifically, we study the daily, seasonal and annual variations in gas energy usage across the whole system.

#### 6.3.1 Energy Demand Analysis

Figure 6.1 depicts the aggregate gas demand for the city under-consideration over the course of an year. There are two peak periods — between Jan-Feb and Nov-Dec

---

\(^1\)https://darksky.net/dev
Figure 6.2: Probability distributions for aggregate daily gas demand (a), average gas demand over the course of the day (b), and CO$_2$ emissions from buildings throughout the year.

months. These peaks coincide with the most severe winter months. The average daily gas demand during winter months is 89.3 MMCF, which is 6× the daily average during summer months (14.5 MMCF). The data also demonstrates a strong negative correlation (-0.9) between temperature and gas demand — as the temperature falls, gas demand rises due to increased residential heating in buildings. Figure 6.2(a) depicts the daily aggregate demand for gas across the system. The figure shows that on most days, aggregate demand is < 25 MMCF. This is mainly due to the use of non-heating appliance such as stoves. The figure also shows a spread of high usage days during which demand is highest. For instance, the peak day consumes > 150 MMCF, which is 3.5× the average usage. Since these high usage days are predominantly made up of heating consumption, replacing gas heating with electric heat pumps has great potential to curtail CO$_2$ emission.

In addition to analyzing the aggregate daily demand, we study the variation in gas demand by time of day, and the periods during which the daily peak demand occurs. Figure 6.2(b) depicts the average gas demand by time of day during winter and summer months. During winter, gas demand exhibits a bi-modal peak — a sharp peak between 8-9am, and a moderate peak between 5-8pm. This coincides with the morning and evening routines during which occupancy and activity in homes is highest. The peak hourly demand is 5.08 MMCF, while the average demand is 3.72,
indicating a 1.4 peak-to-average ratio. Lastly, gas demand during summer months
does not show significant variation over the course of the day. This is because gas
usage during summer is predominantly made up of appliance usage which is fairly
constant throughout the year.

6.3.2 Carbon Emission Analysis

The combustion of natural gas produces carbon dioxide as a byproduct which
is released into the atmosphere. When gas is used for heating, the amount of CO$_2$
emitted is driven by the amount of gas required to generate enough heat for a building.
This is in turn driven by the temperature e.g. as the temperature decreases, more
heat is required to raise the indoor temperature, as well as the building size i.e.
larger spaces require more energy to heat. Further, building characteristics such as
insulation affect how much gas is consumed e.g. buildings with poor insulation lose
heat to the atmosphere faster than those with better building envelope, and therefore
have higher gas demand.

To examine the CO$_2$ emission generated directly from gas heating, we compute
the emission for each building by multiplying the total gas consumption for the year
with the emission factor of gas. About 0.0551 MT of CO$_2$ is produced for each MCF
of natural gas burned [8]. To estimate heating gas consumption, we subtract the
recorded gas usage from the summer average i.e. we compute emission for the heating
load only instead of the overall consumption. Figure 6.2(c) depicts the distribution
of CO$_2$ emitted by each building from heating gas combustion. The figure shows that
the median building emits 32.4 MT of CO$_2$ every year. The figure also shows a long
tail, with a small number of buildings emitting a lot more CO$_2$ compared to others.
These buildings in particular form good candidates for transition to in order to reduce
CO$_2$ emission from the heating. The highest emitting buildings contribute >250 MT
CO₂ during the year which is $8.1 \times$ the median emission and $7 \times$ the average CO₂ emission.

6.4 Multi-objective Decarbonization

In this section, we present a data-driven multi-objective optimization (MOO) framework that enables flexible selection of a subset of homes for heat pump retrofits to achieve decarbonization goals. We start the optimization with an initial goal of maximizing CO₂ reductions and iteratively add additional objectives of maximizing carbon efficiency and targeting energy inefficient buildings.

6.4.1 Optimizing for Carbon Emissions Reduction

Let $H = \{h_1, h_2, \ldots, h_n\}$ denote the set of buildings, each indexed by $i$. Let $C_i^g$ denote the total CO₂ emission from the cumulative gas consumption for building $i$ required for heating during the year. Let $C_i^e$ denote the total CO₂ emission from the cumulative electric consumption for building $i$ required for heating when using an electric heat pump. Let $\alpha_i$ represent the transition-to-electricity status for the building $i$ and $S$ represent the target number of buildings to transition to electric heat pump heating.

Given that, our objective is to select $S$ buildings from the set $H$ which when transitioned to electric heat pumps result in the lowest aggregate CO₂ emission possible across buildings. This objective can formally be described as follows.

$$\min \sum_{i=1}^{n} (1 - \alpha_i) \cdot C_i^g + \alpha_i \cdot C_i^e$$

s.t., Equations (6.2) - (6.4)

vars., $C_i^g, C_i^e, \alpha_i, S \forall i$
Our first constraint relates to the level of transition. Let $\alpha_i$ denote a binary variable which indicates the state of transition for each building $i$ such that $\alpha_i \in \{0, 1\}$. When set, the building is transitioned to electric heat pump heating, and when not set, the building remains on gas. Further, let $S$ denote the target number of buildings to transition to electric heat pump heating. To ensure that only $S$ buildings are transitioned, the sum of all values of $\alpha_i$ must equal $S$, as stated below.

$$\sum_{i=1}^{n} \alpha_i = S \quad (6.2)$$

Our final set of constraints simply ensure that a building cannot have negative carbon emissions from either the gas consumption or the electric demand.

$$C_i^g \geq 0 \quad \forall i \quad (6.3)$$

$$C_i^e \geq 0 \quad \forall i \quad (6.4)$$

The total CO$_2$ emission from gas consumption, $C_i^g$, over the course of a year for a building $i$ is a multiple of the total heating gas demand $D_i^g$ and the carbon intensity of gas $I_g$.

$$C_i^g = D_i^g \times I_g \quad (6.5)$$

The total CO$_2$ emission from electric demand, $C_i^e$, over the course of a year for a building $i$ is a multiple of the total electricity demand $D_i^e$ and the carbon intensity of the electric grid $I_e$.

$$C_i^e = D_i^e \times I_e \quad (6.6)$$

It should be noted that a simple ordering of homes based on their total carbon emissions can achieve the singular goal of selecting a set of buildings that maximizes
carbon emission reductions after transition. However, we present this as a flexible multi-objective optimization framework so that additional objectives, discussed in subsequent sections, can be integrated into the same framework.

6.4.2 Optimizing for Carbon-efficiency

Optimizing for total carbon emission reduction targets buildings with highest carbon footprint. However, the large footprint may be a result of large residential area or large number of residents and the building itself may be making a highly efficient use of its gas demand. To capture this effect, we define a notion of carbon-efficiency. We define carbon-efficiency as the amount of CO₂ emitted when one unit area of a building is raised by one unit of temperature. We further elaborate the notion of carbon-efficiency next.

The notion of carbon-efficiency is based on our observation that electric heat pumps consume lower energy compared to gas to heat the same building from a lower temperature $T_{low}$ to a higher temperature $T_{high}$. Figure 6.3 depicts the relationship between energy generated from gas consumption and temperature for a building dur-
Figure 6.4: Relationship between energy demand of an electric heat pump and the temperature.

In winter months. The figure shows an inverse relationship between energy and temperature. As the temperature decreases, the energy required to heat the building increases. The rate of change (captured in the slope of the fit line) indicates the amount of energy required to raise the building’s temperature by one unit of temperature. This is directly proportional to the CO₂ produced for each unit temperature.

Figure 6.4 depicts the relationship between the energy demand of an electric heat pump and the temperature for the same building. Similar to gas energy consumption, there is an inverse relationship between electric energy and temperature. However, the slope is significantly less steep than that of gas. This is because electric heat pumps consume lower electrical energy to generate the same amount of heat energy. Since carbon emissions are directly proportional to energy consumption, the CO₂ produced for each unit temperature is lower for electric heat pumps. Since buildings have different sizes, energy consumption alone is not enough to compare usage between buildings of varying size. Before computing the energy slope, we normalize both gas and electric heat pump energy by size. We then compute carbon emissions per unit size.
Note that maximizing carbon efficiency introduces a tradeoff between reduction and efficiency. Since the most carbon efficient buildings are not necessarily the highest emitters in absolute scale, a portion of CO₂ reduction must be foregone to maximize efficiency. However, since gas furnaces are inherently inefficient, maximizing carbon efficiency places a tighter bound on wasted CO₂ emission, and leads to better energy utilization. We extend the multi-objective optimization framework defined in Equation 6.1 to jointly maximize the carbon-efficiency and minimize the total carbon emissions. In doing so, we introduce a new set of variables that are defined next.

First of all, we use the absolute value of the carbon emission slopes for both gas and electricity as a substitute for carbon-efficiency. This formulation of the problem allows us to keep the overall objective as minimization of carbon emission reductions and slopes of the emission curves (representing carbon-efficiency). Given that, let λₗᵣ be the absolute slope of gas CO₂ emissions for the building i. Let λₑᵣ be the absolute slope of electric CO₂ emission for the building i. Our joint optimization of minimizing carbon emissions and maximizing carbon-efficiency can be stated as follows.

\[
\begin{align*}
\min \ & \sum_{i=1}^{n} (1 - \alpha_i) \cdot C_i^g + \alpha_i \cdot C_i^e \\
\min \ & \left[ \sum_{i=1}^{n} (1 - \alpha_i) \cdot \lambda_i^g + \alpha_i \cdot \lambda_i^e \right] \cdot \frac{1}{n} \\
\text{s.t.,} \ & \text{Equations (6.2) - (6.4)} \\
\text{vars.,} \ & C_i^g, C_i^e, \alpha_i, \lambda_i^g, \lambda_i^e, S \ \forall i
\end{align*}
\] (6.7)

As stated before, to maximize CO₂ efficiency, we minimize the average absolute slope of CO₂ emissions curve across all buildings. To solve this multi objective formulation, we use a lexicographic approach. Specifically, we optimize for CO₂ efficiency, and then solve for optimal CO₂ reduction along with the optimal efficiency. We do
so because using this approach enables us to solve the optimization within a short period of time.

6.4.3 Targeting Energy Inefficient Buildings

In addition to carbon-efficiency, building decarbonization strategies may also want to target energy inefficient buildings. The sources of energy inefficiencies include poor insulation, high temperature set points for heating and cooling, and inefficient appliances. In this section, we extend our optimization formulation to target buildings that have one or more energy inefficiencies. To do so, we extend our analysis to not only consider gas energy usage only, but also electric usage. We learn a building energy model and use it to identify energy inefficiencies which we target in decarbonization.

Let $U = \{h_1, h_2...h_p\}$ denote the set of buildings with the heating inefficiency i.e. high heating slope, each indexed by $k$. Let $V = \{h_1, h_2...h_q\}$ denote the set of buildings with cooling inefficiency i.e. high cooling slope, each indexed by $l$. Further, let $W = \{h_1, h_2...h_r\}$ denote the set of all other buildings i.e. all buildings with any other inefficiency except heating and cooling, as well as those without any inefficiency, each indexed by $m$.

Let $C_{k}^{q,u}$, $C_{l}^{q,v}$ and $C_{m}^{q,w}$ be the total carbon emissions from gas consumption in heating inefficient, cooling inefficient, and the remaining buildings, respectively. Further, let $C_{k}^{e,u}$, $C_{l}^{e,v}$ and $C_{m}^{e,w}$ be the total carbon emissions from electricity usage in heating inefficient, cooling inefficient and remaining buildings, respectively. Let $\zeta_k$, $\beta_l$ and $\gamma_m$ be the binary variables that indicate the transition status of heating inefficient, cooling inefficient, the remaining buildings, respectively. All of the binary variables can only take a value of either 0 or 1, which means that $\zeta_k, \beta_l, \gamma_m \in \{0, 1\}$ for all $k$, $l$, and $m$. To transition only $S$ buildings, the sum of all set variables from all building groups must be equal to $S$. 

110
\[ \sum_{k=1}^{p} \zeta_k + \sum_{l=1}^{q} \beta_l + \sum_{m=1}^{r} \gamma_m = S \]  

(6.8)

Lastly, since buildings cannot have negative energy usage and therefore negative emission, we ensure that emission from buildings in all groups is greater than or equal to zero.

\[ C_{k}^{g,u}, C_{l}^{g,v}, C_{m}^{g,w}, C_{k}^{e,u}, C_{l}^{e,v}, C_{m}^{e,w} \geq 0 \quad \forall k, l, m \]  

(6.9)

With these constraints in place, our objective is to select \( S \) buildings from the sets \( U \), \( V \) and \( W \) such that when the \( S \) buildings are transitioned to electric heat pumps, carbon emissions are minimized, while the portion of \( S \) buildings selected from the heating and cooling inefficient groups is maximized. This multi-objective optimization problem can be formally stated as follows.

\[
\min \quad f_u(u) + f_v(v) + f_w(w) \\
\min \quad \sum_{k=1}^{p} (-1 \cdot \zeta_k) + \sum_{l=1}^{q} (-1 \cdot \beta_l) \\
s.t., \quad \text{Equations (6.8) - (6.9)} \\
\text{vars.,} \quad C_{k}^{g,u}, C_{l}^{g,v}, C_{m}^{g,w}, C_{k}^{e,u}, C_{l}^{e,v}, C_{m}^{e,w}, \zeta_k, \beta_l, \gamma_m, S \quad \forall k, l, m
\]

The composite functions \( f_u, f_v, \) and \( f_w \) are defined as follows.

\[ f_u(u) = \sum_{k=1}^{p} (1 - \zeta_k) \cdot C_{k}^{g,u} + \zeta_k \cdot C_{k}^{e,u} \]  

(6.11)

\[ f_v(v) = \sum_{l=1}^{q} (1 - \beta_l) \cdot C_{l}^{g,v} + \beta_l \cdot C_{l}^{e,v} \]  

(6.12)

\[ f_w(w) = \sum_{m=1}^{r} (1 - \gamma_m) \cdot C_{m}^{g,w} + \gamma_m \cdot C_{m}^{e,w} \]  

(6.13)
Note that to maximize the number of buildings selected from the heating and cooling inefficient groups, we minimize the negation of all set binary variables from the two sets.

Similar to jointly optimizing CO₂ efficiency and reduction, we use a lexicographic approach to solve this formulation. Specifically, we optimize for maximizing the number of inefficient buildings transitioned, and then solve for optimal CO₂ reduction along with the optimal number of buildings. This approach enables us to solve the optimization within a short period of time.

6.5 Evaluation

In this section, we present the results for various decarbonization strategies presented in Section 6.4 and evaluate their efficacy in reducing carbon emissions and increasing energy efficiency. To do so, we introduce varying levels of transition across the system — where the transition rate represents the percentage of buildings converted from gas to electric heat pumps.

6.5.1 Experimental Setup

The gas and electricity consumption data from the buildings (described in Section 6.2.2) provides building-level metering of the gas and electricity demand. We first disaggregate the gas and electric demand data into two components: first used for heating purposes, second used by all the other appliances such as stoves. For gas demand data, we compute the average gas usage during the summer months, which corresponds to non-heating uses of utility gas. We subtract this from the year-round data to get the heating components of the gas demand. We also account for energy loss due to inefficiencies of the gas furnace. We use an efficiency value of 87.5%, which lies between the typical efficiency for a standard and a high efficiency furnace. To
compute total carbon emissions from gas usage, we use a gas carbon intensity value of 0.0551 MT/MCF [8].

To compute the corresponding electric heat pump emissions, we compute the total heat energy generated by the volume of gas consumed and use the Heating Seasonal Performance Factor (HSPF) of electric heat pumps to compute the electric energy required to generate the equivalent amount of heat energy. For all experiments, we assume a HSPF rating of 8.5, which is typical of many efficient heat pump models. Finally, to compute the carbon emissions from electric heat pump usage, we use a carbon intensity value of 0.000386 MT CO₂/kWh, corresponding to the average carbon intensity value for the United States electric grid [110, 57].

6.5.2 Optimizing Carbon Emissions Reduction

We begin by analyzing the impact on carbon emissions after transitioning buildings from gas to electric heat pumps using a strategy that optimizes carbon emission reductions. Our main goal here is to reduce carbon emissions — therefore, at each transition level, the buildings that lead to the highest reduction in carbon emission are selected for transition. We run this optimization on our gas consumption data and compute the resulting carbon emission at each transition level. Figure 6.5 presents the results for this analysis and there are two interesting observations. First, carbon emissions reduce at an exponential rate at the lower levels of transition i.e. 1-10%. This is because since we are only optimizing for carbon emissions reduction, the biggest emitters are selected first, which leads to a disproportionately high carbon emissions reduction at the start. As transition rate increases, carbon emission reductions enter another phase characterized by a linear growth (with low slope) from 10-100%, where the rest of buildings with moderate emissions are transitioned. Second, results also shows that at 100% transition, electric heat pumps have the potential to cut carbon emissions by up to 81%. This is a noteworthy observation, and demonstrates the
viability of heat pumps to replace natural gas for heating and at the same time, help-
ing make significant strides towards decarbonizing the building sector and achieving climate goals.

We next analyze the carbon emission reductions per unit area. We compute the total carbon emission reductions for each building, and normalize the difference with the size of the building. Figure 6.6 depicts the distribution of carbon emission reductions per unit area across all buildings. The figure shows that normalized CO\textsubscript{2} reduction is normally distributed with the average building seeing an annual reduction of 0.018 MT/ft\textsuperscript{2}. Given that the median house size of single family home in United States is 2273ft\textsuperscript{2} [25], each home has a potential to reduce 40.9 MT each year.

### 6.5.3 Maximizing Carbon-efficiency

Next, we analyze the impact of optimizing for carbon-efficiency on carbon emissions reduction. The goal here is to quantify the tradeoff between carbon emissions reduction and efficiency, i.e. how much carbon emissions can be eliminated while also ensuring that carbon emissions per unit area is minimized. We solve the optimization
problem described in Section 6.4.2 and compare the aggregate carbon emissions after the transition with the carbon-optimal strategy results presented in Section 6.5.2. Figure 6.7 depicts the results for this analysis. The figure shows that carbon emissions reduction is lower than the optimal case for up to \( \approx 85\% \) transition, after which carbon emissions are similar to the optimal scenario and converge towards the optimal case near full transition (100\%). The initial growth of carbon emissions reductions is also at a lower magnitude compared to the optimal case. This is because some of the highest emitting buildings are also highly carbon-efficient. In joint optimization, they are foregone in favor of less-efficient buildings which have a lower absolute carbon footprint. The largest deviation occurs at 15\% transition, where 71 GT of carbon emissions reduction is foregone in favor of maximizing efficiency. However, carbon-efficiency increases by 1.9\times. Since there is not a significant deviation in carbon emissions reduction, utility companies can maximize carbon-efficiency while sacrificing only a small amount of carbon emissions reduction compared to the optimal case.
Figure 6.7: Carbon emissions at varying levels of electric heat pump transition while jointly optimizing for carbon emissions reduction and carbon-efficiency.

6.5.4 Targeting Energy Inefficient Buildings

We next examine the tradeoff in carbon emissions reduction introduced by prioritizing inefficiencies in buildings. We begin by performing building segmentation based on their unique energy inefficiencies and the underlying faults that cause such inefficiencies. Our fault analysis is based on the technique proposed in [63]. We apply the proposed technique to our data. Table 6.1 shows the indicator characteristics identified for each building along with the possible faults that underlay such inefficiencies. The third column shows the optimization group that we place each building in based on the identified characteristics. Specifically, we target homes that have heating and cooling inefficiencies since these would benefit most from transitioning from gas to electric heat pumps.

Figure 6.8 depicts the distribution of energy inefficiencies identified in buildings in our dataset. The figure shows that poor building envelope is the leading cause of energy inefficiency. This is true across buildings of all age groups. It also reveals that inefficient HVAC and heating units are the second and third most prevalent causes of energy inefficiencies of buildings, respectively. Since electric heat pumps are capable
Table 6.1: Probable building faults alongside their underlying characteristics.

<table>
<thead>
<tr>
<th>Indicator Characteristics</th>
<th>Probable Faults</th>
<th>Optimization Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>High heating slope</td>
<td>Inefficient heater, Poor building envelope</td>
<td>Inefficient heating</td>
</tr>
<tr>
<td>High cooling slope</td>
<td>Inefficient HVAC, Poor building envelope</td>
<td>Inefficient cooling</td>
</tr>
<tr>
<td>High heating temperature</td>
<td>High set point, Poor building envelope</td>
<td>Other</td>
</tr>
<tr>
<td>Low cooling temperature</td>
<td>Low set point, Poor building envelope</td>
<td>Other</td>
</tr>
<tr>
<td>High base load</td>
<td>Inefficient appliances</td>
<td>Other</td>
</tr>
</tbody>
</table>

of operating as both heating and cooling units based on the season, this distribution of faults underpins the importance of targeting energy inefficient buildings in transition. The figure also shows that older buildings are more prone to being energy inefficient, while newer buildings show less prevalence probably due to improved building standards. This segmentation of buildings based on underlying energy inefficiency informed the basis of our targeted optimization, presented in Section 6.4.3. Targeting inefficient buildings offers multiple advantages over optimizing for carbon emissions alone. For example, transitioning to electric heat pumps typically comes with additional benefits such as building retrofits. This enables buildings to take advantage of these additional benefits during transition. Moreover, the amortized cost of transition may be reduced by performing multiple upgrades at once.

To quantify the tradeoff in carbon emissions reduction and targeting inefficiency buildings, we run the optimization described in Section 6.4.3 on our datasets and compare the resulting carbon emissions reduction with the carbon optimal scenario. Figure 6.9 depicts the results of this experiment, and presents some interesting observations. First, carbon emissions reduction show a gradual linear decrease from start to finish compared to the optimal case, and only converges at near full transition ($\approx 98\%$). Since older buildings are more prone to energy inefficiency, this figure also indicates that this strategy has the effect of selecting older buildings first. The figure also indicates that the highest emitters are not necessarily the most inefficient, which
Figure 6.8: Distribution of energy inefficiencies across buildings by age group.

presents an important tradeoff between carbon emissions reduction and improving building energy efficiency.

Next, we examine the tradeoff between different sets of solutions to our multi-objective optimizations. To do so, we generate the Pareto frontier (i.e. the set of all Pareto efficient solutions) for each formulation. Figure 6.10a depicts the Pareto curve for the multi-objective optimization that jointly maximizes CO₂ efficiency and reduction at 40% transition. The figure shows that there exists a tradeoff in CO₂ reduction from 3,837-4,556 MT (i.e. a 19% increase). The figure also shows that CO₂ efficiency ranges from 0.00019-0.0002, indicating a 5% increase. On the other hand, Figure 6.10b depicts the Pareto curve for the multi-objective optimization that jointly maximizes CO₂ reduction and the number of energy inefficient buildings that are transitioned at 40% transition. The figure shows that CO₂ reduction ranges from 3818-5088 MT, a 33.26% increase. The figure also shows that the number of inefficient buildings ranges from 160-800, which indicates a multiplicative increase. This presents policy makers with a wide range of optimal solutions for transition.

Finally, we evaluate the impact of transitioning to electric heat pumps on the daily gas demand. Figure 6.11 depicts the average hourly demand of gas during
winter and summer months after 100% transition to electric heat pumps. Similar to the observations made in Figure 6.2 (b), we find that gas demand exhibits a bi-modal peak — between 8-9am and 5-8pm. The figure also makes two interesting observations. First, the average peak demand reduces by 78% compared to the case before transition. Second, the extremity of the peak is also reduced significantly. Before transition, the peak demand was $1.4 \times$ the average hourly demand. Post transition, the peak-to-average ratio is 1.2, indicating a 14.3% reduction compared to the value before transition. Lastly, the figure shows no significant change in daily usage pattern of gas during summer months since consumption is mainly made up of appliance usage which does not change with the introduction of heat pumps.

6.5.5 Impact on Energy Consumption Reduction

Figure 6.12 depicts the distribution of potential energy reduction for buildings in our dataset. It shows that electric heat pumps can reduce annual energy usage by 1,193 GWh, with a median reduction of 107.5 MWh. Most buildings reduce energy
Figure 6.10: (6.10a) Pareto curve for jointly maximizing CO₂ efficiency and reduction, and (6.10b), Pareto curve for jointly maximizing CO₂ emission reduction and the number of inefficient buildings transitioned.

usage by up to 200 MWh, with a few large energy consumers seeing annual reductions of up to 800 MWh, which is 7.5× the median. Reducing energy consumption makes buildings more energy efficient, and this makes electric heat pumps an attractive replacement for gas heating.

6.5.6 Impact on the Electric Grid

After converting heating from gas to electric heat pumps, it is important to understand the expected load exerted on the electric grid. To compute the expected electric demand, we estimate the amount of heat energy generated from a building’s gas consumption, and compute the electric energy required to generate the equivalent amount of heat energy. This is a multi-step process which we describe in more detail in Section 6.4. Figure 6.13 depicts the CDF of electric demand required by heat pumps to generate the equivalent amount of heat energy generated by gas across the entire system. The figure shows that the median buildings increases electric demand by ≈ 72 MWh annually. The figure also indicates that most buildings increase electric demand by up to ≈ 200 MWh and only a few buildings having an additional annual
demand of $> 200$ MWh. Finally, the median annual gas energy is 179.1 MWh, which indicates a 60\% reduction in absolute energy consumption. An additional study is needed to study the impact of the extra load on the electric grid in depth. However, these preliminary results show how the grid is expected to change as the penetration of heat pumps increases in buildings in the future.

### 6.6 Related Work

In this section, we discuss prior work on the energy transition, decarbonizing heating in buildings and electric heat pumps.

**The energy transition.** Multiple studies on the pathways for transitioning to a clean energy future have been carried out. Majority of these studies examine the economic, environmental and societal benefits of a successful energy transition. For instance, Santamarta et al [109] evaluated the potential of transitioning to geothermal energy showing that in addition to CO$_2$ emission reduction, up to 66\% energy savings and up to 13\% ROI can be realized. Heinisch et al [54] propose an optimization
Figure 6.12: Distribution of energy consumption reduction across buildings after transition to electric heat pumps.

model that interconnects various sectors of the energy ecosystem i.e. the electric grid, heating requirements and transportation, showing that availability of renewable sources such as solar further increases the benefits of electric heating technologies such as heat pumps. Gonzalez-Salazar et al [49] explore pathways to phasing out coal-fired heating stations in favor CO₂-free energy sources. These studies are performed at macro scale i.e. energy generation and CO₂ mitigation are performed at from centralized point of view. Our work is complementary to this work as it evaluates the potential of transition at high granularity.

**Decarbonizing heating.** There have been numerous studies on decarbonizing space heating in the building sector [93, 120, 75, 58, 16]. For instance, Padovani et al [93] quantified the economic and decarbonization implications of replacing propane heating with cleaner electric energy sources such as solar heat pumps in isolated, rural residential buildings. Waite & Modi [120] propose and analyze a dual transition approach that maintains existing fossil fuel systems alongside electric heat pumps. Instead of replacing all existing fossil fuel heating with electric heat pumps, they
Figure 6.13: Cumulative distribution functions of additional electric demand exerted on the system after transitioning from gas to electric heat pumps.

propose a mix of both energy sources that gradually phases out fossil fuels over time. Leibowicz et al develop an energy system optimization model for decarbonizing residential buildings that incorporates transitioning to greener energy sources, migrating to more energy-efficiency appliances and improving the thermal properties of buildings e.g. through insulation retrofits. Hopkins et al propose transitioning to electric heat pumps for heating building. Finally, Baldino et al [16] analyze the cost and decarbonization benefits of hydrogen and renewable electricity as a replacement for heating. Our work is complementary to this work, as we evaluate the impact of multiple building selection strategies on CO₂ reduction. Since transitioning involves shifting energy load from one system to another i.e. from gas to electric, our work also quantifies the impact of such transition on the electric grid.

**Electric heat pumps.** The viability of electric heat pumps as a replacement for gas heating in residential buildings has also been widely studied [124, 95, 120, 69, 68, 87]. While some studies have focussed on the evaluation of the applicability and performance of heat pumps in extreme temperatures [87, 38, 111], others have analyzed
their potential to decarbonize heating at various geographical scales. For instance, Johnshon & Krishnamoorthy [65] analyzed the cost and economic implications of transitioning to electric heat pumps, and how it varies across different regions in the entire United States. Zhang et al [124] studied the decarbonization benefits of electric heat pumps using a simulated energy system of an entire city. Other studies [87, 38, 111] have analyzed on the applicability of heat pumps especially in extremely low temperatures, and shown that as heat pump technology continues to evolve, modern heat pumps can be deployed even in very cold regions. Our work is complementary to this work as we evaluate the viability of heat pumps at high granularity using real world data.

6.7 Conclusions

In this chapter, we conducted a data-driven optimization study to analyze the potential of replacing gas heating with electric heat pumps to reduce carbon emissions in a city-wide distribution grid. We performed an in-depth analysis of gas consumption in the city and showed that $\approx 17$ BCF of gas is consumed directly resulting in $\approx 360$ GT of CO$_2$ emission annually. We presented a flexible multi-objective optimization (MOO) framework that optimizes carbon emissions reduction while also maximizing other aspects of the energy transition such as carbon-efficiency and energy inefficiency in buildings. We showed that transitioning to electric heat pumps can cut carbon emissions by up to 81% and energy required for heating by up to 60%. We also showed that optimizing for other aspects such as carbon-efficiency and energy inefficiency introduces tradeoffs with carbon emissions reduction that must be considered in a transition strategy. For example, carbon emissions can be foregone in favor of maximizing other aspects of transition such as energy inefficiency. Finally, we presented preliminary results that examine the expected additional load on the
electric grid by transitioning gas to electric heat pumps. We showed that a median building will add an annual load of 71.6 MWh to the electric grid.
7.1 Thesis Summary

This thesis has explored the opportunities and challenges presented by the energy transition, and designed data-driven techniques to facilitate CO\(_2\) reduction. In this dissertation, I have demonstrated how machine learning, time-series analysis and optimization principles can used to build optimal and fair strategies that facilitate the transition towards greener sources of energy across various sectors. I performed several experiments and proposed new optimization formulations for emerging challenges in the energy transition. I so doing, I have made the following contributions.

1. **Impact of EVs on the Grid:** First, I discussed how the current electric system is not designed for EVs, and how the increased EV loads can potentially lead to problems such as blackouts and reduction in transformer lifetime. To address these challenges, I proposed smart a EV charging algorithm that defers (or rate limits) charging loads during peak periods to help alleviate transformer overloads. Further, I proposed transformer level energy storage to shave transformer peaks thereby increasing transformer lifetime. I evaluate these techniques on real world energy usage data providing first of a kind insights into transformer load demand.

2. **Decarbonizing Transportation via Hybrid Bike Share Systems:** Second, I examined the feasibility, costs, and carbon benefits of using electric bike sharing — a low carbon form of ride sharing — as a potential substitute for
shorter ride sharing trips with the overall goal of greening the ride sharing ecosystem. I presented a linear optimization framework that employs a hybrid mix of regular and electric bikes to perform substitution and quantified the carbon reduction achieved from such substitution. I evaluate the approach using real world ride sharing data from the city of New York, showing that significant CO₂ reduction can be achieved while impacting travel times negligibly.

3. **Equity in the Energy Transition**: Third, I discussed the emerging inequity in the energy transition. I showed that lower income households pay more per unit area of energy consumed than their wealthier counterparts. Further, I showed that such inequity is perpetuated across demographic, economic, racial and census block lines. I proposed fair analytic approaches for selecting homes for energy retrofits such as electric heat pumps. I evaluate these techniques using real world economic and energy usage data, showing significant improvement in equity outcomes.

4. **Decarbonizing Residential Heating Systems**: Finally, I evaluated the potential of electric heat pumps to reduce CO₂ emission by replacing gas heating in buildings. I present a flexible multi-objective optimization framework that optimizes CO₂ reduction while also maximizing other aspects of the energy transition such as CO₂ efficiency and minimizing energy inefficiency in buildings. Further, I quantified the tradeoffs that exist in such multifaceted transition and presented results that shed light into the expected load exerted on the electric grid by transitioning from natural gas to electric heat pumps.

### 7.2 Future Work

This thesis covers a broad range of areas of the energy transition, particularly, CO₂ reduction in the energy and transportation domains. It gives rise to several
promising directions to enable an optimal and fair energy transition. In this section, I outline possible directions for future work emerging from this thesis.

1. **Renewable energy in the grid**: The smart charging and grid storage techniques presented in Chapter 3 can be extended to model increased renewable energy sources in the grid. Due to their intermittent and unpredictable nature, when combined with increased EV loads, these new energy sources will lead to further challenges that require non-trivial solutions such as charge scheduling, injecting excess energy back into the grid, and others. This combination of varying and renewable sources of energy at different penetration levels forms the basis of the future smart grid.

2. **Incorporating EVs in ride sharing**: The linear optimization based formulation presented in Chapter 4 shows that a hybrid mix of electric and regular bikes can be systematically deployed to reduce CO₂ emission in shared rides. In the same light, similar formulations can be extended to electric cars in ride sharing schemes in conjunction with bike sharing to further green the ride sharing ecosystem.

3. **Optimal pruning of the gas network**: This thesis presented the tradeoffs that exist in transitioning buildings from gas to electric heating, particularly with carbon and energy efficiency. Since the transition will not occur instantaneously, it is important to quantify the costs associated with the transition. For instance, it is more beneficial to transition a whole street compared to a single building within a street. This ensures that the remaining maintenance cost of the gas network is minimized, and gives rise to a class of optimization problems geared towards to remainder of the underlying gas infrastructure.
BIBLIOGRAPHY


[27] Cavagnini, Rossana, Bertazzi, Luca, Maggioni, Francesca, and Hewitt, Mike. A two-stage stochastic optimization model for the bike sharing allocation and rebalancing problem, 2018.


[31] Citi bike electric fleet will grow to 4,000 with $2 fee. https://nyc.streetsblog.org/2019/02/28/exciting-but-also-expensive-citi-bike-electric-fleet-will-grow-to-4000-with-fee-2019-


[35] Drehobl, Ariel, and Ross, Lauren. Lifting the high energy burden in americas largest cities: How energy efficiency can improve low income and underserved communities.


[52] You can now rent a ‘boris bike’ in guildford from the university of surrey. [https://www.getsurrey.co.uk/news/surrey-news/you-can-now-rent-boris-14982984](https://www.getsurrey.co.uk/news/surrey-news/you-can-now-rent-boris-14982984). (Accessed on 05/05/2020).


[78] Li, Yexin, and Zheng, Yu. Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering* (2019).


[96] Popovich, Natalie, Gordon, Elizabeth, Shao, Zhenying, Xing, Yan, Wang, Yunshee, and Handy, Susan. Experiences of electric bicycle users in the sacramento, california area. Travel Behaviour and Society 1, 2 (2014), 37–44.


[125] Zhang, Jiawei, Pan, Xiao, Li, Moyin, and Yu, Philip S. Bicycle-sharing systems expansion: station re-deployment through crowd planning. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2016), pp. 1–10.