Improving Public Transit Systems at Intersection and City-wide Scales

by

Haoyu Chen

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Engineering - Civil and Environmental Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Michael Cassidy, Co-Chair
Professor Carlos Daganzo, Co-Chair
Professor Alexander Skabardonis
Assistant Professor Weihua Gu
Professor Zuo-Jun(Max) Shen

Spring 2016
Improving Public Transit Systems at Intersection and City-wide Scales

Copyright 2016
by
Haoyu Chen
Abstract

Improving Public Transit Systems at Intersection and City-wide Scales

by

Haoyu Chen

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Michael Cassidy, Co-Chair

Professor Carlos Daganzo, Co-Chair

Designing a highly efficient public transit system can help reduce the congestion and pollution in large cities. This dissertation tries to improve public transit systems at both the micro and macro scales.

At the microscale, a Transit Signal Priority strategy is developed to integrate a pre-signal scheme. Buses often compete with cars for limited road space and signal times at intersections. Providing signal priority together with exclusive lanes for buses will often diminish car discharge flows into the intersection. This can create huge car delays, which will undermine the benefit brought by the priority measures to buses. To solve this problem, we propose a midblock pre-signal to recover the loss in the cars’ discharge capacity caused by the bus priority measures. We show how the pre-signal can be coordinated with the intersection’s main signal under a simple signal priority scheme. Analytical models are formulated to examine the effects of our strategy on two performance metrics: the expected bus delay and the car capacity. Pareto frontiers between the two performance metrics are developed. Numerical case studies show that huge benefits for both buses and cars can be achieved by the proposed strategy, as compared to baseline scenarios where a bus signal priority scheme or pre-signal is absent.

At the macroscale, two continuum approximation (CA) optimization models are formulated to design city-wide transit systems at minimum cost. Transit routes are assumed to lie atop a city’s street network. Model 1 assumes that the city streets are laid out in ring-radial fashion. Model 2 assumes that the city streets form a grid. Both models can furnish hybrid designs, which exhibit intersecting routes in a city’s central (downtown) district and only radial branching routes in the periphery. Model 1 allows the service frequency and the route spacing at a location to vary arbitrarily with the location’s distance from the center. Model 2 also allows such variation but in the periphery only.

We show how to solve these CA optimization problems numerically, and how the numerical results can be used to design actual systems. A wide range of scenarios is analyzed in this way. It is found among other things that in all cases and for both models: (i) the optimal...
headways and spacings in the periphery increase with the distance from the center; and (ii) at the boundary between the central district and the periphery both, the optimal service frequency and the line spacing for radial lines decrease abruptly in the outbound direction. On the other hand, Model 1 is distinguished from Model 2 in that the former produces in all cases: (i) a much smaller central district, and (ii) a high frequency circular line on the outer edge of the central district. Parametric tests with all the scenarios further show that Model 1 is consistently more favorable to transit than Model 2. Cost differences between the two designs are typically between 9% and 13%, but can top 21.5%. This is attributed to the manner in which ring-radial networks naturally concentrate passenger’s shortest paths, and to the economies of demand concentration that transit exhibits. Thus, it appears that ring-radial street networks are better for transit than grids.

To illustrate the robustness of the CA design procedure to irregularities in real street networks, the results for all the test problems were then used to design and evaluate transit systems on networks of the “wrong” type. Grid networks were outfitted with transit systems designed with Model 1 and ring-radial networks were designed with Model 2. Cost increased on average by a little: only 2.7%. The magnitude of these deviations suggests that the proposed CA procedures can be used to design transit systems over real street networks when they are not too different from the ideal, and that the resulting costs should usually be very close to those predicted.
To my parents who bring me to this world and always like to make wrong decisions for me.

To all THU guys who help me along the way, especially let me recognize how stupid I am.

To my CAL undergraduate friends, who play with me for five years.
# Contents

## Contents

List of Figures  
List of Tables  

1. **Introduction**  
   1.1 Motivation  
   1.2 Literature Review  
   1.3 Summary of Research Contributions  
   1.4 Organization of the Dissertation  

2. **Improving efficiency for signalized intersections by pre-signal and TSP**  
   2.1 Proposed Strategy and Signal Plan  
   2.2 Analytical Model  
   2.3 Numerical Case Studies  
   2.4 Summary  

3. **Analytical Models to Design Transit Network for Circular and Grid Cities**  
   3.1 City Forms and the Supply of Transit  
   3.2 Demand: Patron Choices of Routes and Transfers  
   3.3 Model Formulations  
   3.4 Solution Procedures  
   3.5 Summary  

4. **Parametric Analysis for Transit Networks**  
   4.1 System Structure Insights  
   4.2 Comparisons between Ring-radial and Grid Cities  
   4.3 Model robustness  
   4.4 Comparisons between CA and Non-CA Models  
   4.5 Summary  

5. **Conclusions**
5.1 Contributions ................................................................. 40
5.2 Future Work ................................................................. 41

A A List of Notations for Chapter 2 42
B Derivation of the Minimum Length of Sorting Area 44
C Derivation of Equation (2.12) 45
D Calculation of $E[t_L]$ 47
E Model 1 for Ring-radial Networks 49
F Grid Transit Network Model 58

Bibliography 62
List of Figures

1.1 A pre-signal controlled intersection approach .............................................. 7
2.1 An intersection approach .................................................................................. 9
2.2 Phase plan of the main signal .......................................................................... 10
2.3 Pre-signal timing coordination ........................................................................ 14
2.4 Comparison of scenarios ($\lambda_b = 30$ buses/hr): (a) $N = n = 3, l = 0.2$; (b) $N = n = 4, l = 0.2$; (c) $N = n = 3, l = 0.3$; (d) $N = n = 4, l = 0.3$; (e) $N = n = 3, l = 0.4$; (f) $N = n = 4, l = 0.4$ ..................................................... 19
2.5 Comparison of scenarios ($\lambda_b = 90$ buses/hr): (a) $N = n = 3, l = 0.2$; (b) $N = n = 4, l = 0.2$; (c) $N = n = 3, l = 0.3$; (d) $N = n = 4, l = 0.3$; (e) $N = n = 3, l = 0.4$; (f) $N = n = 4, l = 0.4$ ..................................................... 21
3.1 (a) circular city and its ring-radial transit network; (b) square city and its grid transit network ................................................................. 23
3.2 (a) route choice for ring-radial network; (b) route choice for grid network ........ 25
4.1 Fraction of total vehicle kms flowing along a circular (square) crown of unit width and radius $x/R$ ................................................................. 33
4.2 Ring-radial network example for $S = 225$ km$^2$ .............................................. 34
4.3 Grid network example for $S = 225$ km$^2$ ......................................................... 35
4.4 Parametric comparisons .................................................................................... 36
4.5 Tests of model robustness ................................................................................. 38
C.1 Queueing diagram of through-moving vehicles in Baseline Scenario 2 ......... 46
E.1 cases for average waiting time per trip ............................................................. 51
E.2 transit stations in the central district ................................................................. 52
E.3 Cases for average access time per trip ............................................................. 54
E.4 In-vehicle travel time ....................................................................................... 55
List of Tables

2.1 Comparison of modeling scenarios .............................................. 13
4.1 Technology parameters for the considered scenarios. ...................... 31
A.1 Notations for Chapter 2 (part (a)) ............................................ 42
A.2 Notations for Chapter 2 (part (b)) ............................................ 43
E.1 Notations for ring-radial networks ............................................ 49
F.1 Notations for grid networks .................................................. 58
Acknowledgments

I would like to express my very great appreciation to Professor Daganzo and Professor Cassidy, my research advisors, for their patient guidance, enthusiastic encouragement and useful critiques of my research work. I really enjoy the research meeting with them, and 290T group discussion. Their help is not only limited on my research but also on my life at Berkeley. It is very enjoyable to share my life with them. I am very honored and grateful to have Professor Daganzo and Professor Cassidy as my advisors.

I feel so lucky to have Professor Gu, Professor Skabardonis, and Professor Shen on my dissertation committee. I want to thank Professor Gu, who helps me transform from an undergraduate student to a PhD student and helps me get used to the live at America. I want to thank Professor Skabardonis, who teaches me a lot of knowledge on highway traffic, and Professor Shen, who teaches me a lot on logistics system and supply chain.

My thanks also go to Professor Canny, who teaches me a lot on computer system and machine learning algorithm. The knowledge on computer science helps me a lot on doing simulations of transit system, and make my view broader.

I want to give special thanks to my roommates Biye Jiang and Yang Gao. They not only provide a good living circumstance for me, but also provide a lot of good suggestions on my research.

This dissertation marks the end of my 21 years of school life, from primary school, middle school, highschool, to Tsinghua, and to Berkeley. I would like to thank all the teachers and students who taught me knowledge and lesson of life. I also want to thank Tsinghua for providing the second home for my heart, and Berkeley for opening the door of research. I will be a THUer and a CAL forever.

Finally, I thank my parents to support me along the way. They provide deep love and understanding to me.
Chapter 1

Introduction

1.1 Motivation

Transportation is one of the most important components of a city. Yet, many cities, especially large ones, suffer from long travel times, heavy congestion, and severe engine emissions. Public transit is often considered the key to solving these problems. Regrettably, many city dwellers find travel by public transit to be unduly time consuming and uncomfortable, and are therefore hesitant to choose that option. Thus, designing a highly efficient and affordable public transit system is an important task for cities.

Developing this sort of transit system is not an easy task. Problems exist at different geographic scales that must be solved. At the micro scale, passengers need to board/alight transit vehicles at stations. Because of space constraints, stations may be congested. Similar constraints at intersections mean that decisions must be made regarding how to allocate space between transit vehicles and automobiles. It is difficult to provide good public transit services without negatively impacting travel by private car. At the macro scale, a public transit system needs to be designed to accommodate the city’s passenger demand and the geometry of its street network. In addition, the transit system needs to strike a balance between passenger expectations and budgetary constraints that can limit a transit agency’s capacity to meet those expectations.

Given these multi-scale challenges, no single method can be used to design all aspects of a transit system. Some researchers have focused on the macro-scale by designing transit networks with fewer transfers. Others have sought to optimize vehicle scheduling, say to minimize fleet size without degrading service levels. Still others focus their work on micro-scale solutions, such as plans to combat congestion at bus stops or to keep vehicles operating on schedule so as to provide more reliable service.

This thesis focuses on ways to improve public transit at both the macro and micro scales. At the micro scale, we seek to develop better strategies for prioritizing transit vehicle movements at signalized intersections. At the macro scale, we seek better ways to design a network that uses the continuous approximation method to minimize generalized cost.
CHAPTER 1. INTRODUCTION

A review and critique of the relevant literature is furnished in Section 1.2. The contributions of this research are summarized in Section 1.3. The structure of the remaining chapters of this dissertation is presented in Section 1.4.

1.2 Literature Review

The literature review in this section addresses two issues: 1) transit signal priority (TSP); and 2) transit network structure and design.

Transit Signal Priority

TSP is an operational strategy that facilitates the movement of transit vehicles (usually those in service), either buses or streetcars, through traffic-signal controlled intersections [21]. The main task of TSP is to improve the reliability and speed of buses. There are three types of TSP strategies: 1) passive priority; 2) active priority; and 3) real-time control. A passive priority system is based on historical traffic data to calculate signal settings offline to favor transit [39]. Passive priority does not need special hardware to detect buses or adjust signals according to the real-time locations of buses. Passive priority systems also do not require exclusive bus lanes to let buses bypass cars queueing at signalized intersections. However, uncertainty in bus arrival times at intersections may reduce the benefits of passive priority systems [45]. Thus, passive priority systems are only useful for moderate to heavy uniform transit service in light overall traffic conditions [34]. Strategies for active priority systems include green extension, early green (or red truncation), actuated transit phases, phase insertion, and phase rotation [21]. Although bus exclusive lanes are not a requirement for these TSP strategies, TSP has been found ineffective under congested or near congested conditions [14] without exclusive bus lane. To overcome this problem, exclusive bus lanes [55] or intermittent bus exclusive lanes ([47], [27], [9]) are applied. Real-time control applies more complex strategies that involve communicating with approaching buses. In order to accurately estimate the arrival times of buses or bypass car queues at intersections, an exclusive bus lane is required as well.

Many cities have implemented various bus signal priority schemes that grant extra green time to buses approaching intersections [15]. Studies of real-world cases have demonstrated the benefits of these strategies in reducing bus delays [8]. However, reduced delays for buses often come at the cost of damaging the car traffic ([8], [31]). Car traffic problems are exacerbated when bus signal priority is implemented alongside an exclusive bus lane [56], which is usually the case in practice as we noted in the above paragraph. Bus signal priority and exclusive bus lanes together can significantly reduce car capacity resulting in oversaturation at intersections and even gridlock in a congested urban area [5]. Therefore, bus signal priority is generally considered inappropriate when car traffic is heavy and the number of travel lanes in the intersection approach is relatively few [31].
CHAPTER 1. INTRODUCTION

Alternative strategies that offset the negative impacts on car traffic have been proposed in the literature. For example, some researchers have proposed so-called "intermittent bus lanes" or "bus lanes with intermittent priority" ([47], [9] [44]). In these schemes, bus lanes are open to cars when no bus is present. Similar strategies have been developed by using a mid-block pre-signal (i.e., a signal added upstream of the intersection [54]) to let buses and cars alternatively use a shared lane [43]. To their credit, these strategies can increase the car discharge flow at an intersection without compromising bus priority. Still, cars suffer a capacity loss (albeit a smaller one) due to the need to minimize the interference of cars with buses. For example, cars have to be cleared from the approach lane when an arriving bus is detected. This capacity loss grows significantly as bus flow increases [52].

In this dissertation, we propose a pre-signal strategy that can work with TSP to decrease the bus delay at intersection and compensate for the capacity loss of cars. We build analytical models to reveal the benefits of our strategy.

Transit Network Structure and Design

Public transit network design involves determining where transit lines go, the frequency of service, and the spacing between stops. The objective of design is to find a balance between the needs of users and the limited resources of the service provider (or, as we call it here, the agency). From the user’s perspective, a good public transit system should cover a large area with dense transit lines, making it easier for users to access the system. A good system should provide convenient scheduling and fast, reliable service. From the agency’s perspective, a public transit system should have relatively low construction and maintenance costs but satisfy basic passenger capacity requirements. There are three equivalent ways to formulate a design problem that combines user and agency objectives: 1) optimize the user cost by constraining the agency cost; 2) optimize the agency cost by constraining user cost; and 3) optimize the generalized cost by using time cost to combine user cost and agency cost. One of the key tasks of transit network design is to formulate the cost of the whole network and, based on this cost formulation, find the optimal design solution.

There are two approaches to formulate the cost of the transit network: 1) discrete optimization and 2) continuous approximations. Discrete optimization treats transit line and stop locations, and frequency of service as discrete variables. This approach usually takes the OD table as the passenger demand information. Since the objective function is non-convex and complex, the optimization problem usually is NP-hard. However, there are simplification methods: sequential approaches and bi-level approaches [37]. Sequential approaches decompose the design process into line generation and transit assignment stages [12]. [25][24][23] applied similar decompositions to maximize the number of direct trips and minimize transfer times. Bi-level approaches minimize the total travel time and the corresponding operation cost, and solve a transit assignment problem at a lower-level ([20], [49], and [50]). Decomposition of the objective problem can greatly simplify the optimization problem. However, decomposition is not intrinsic to the system, and the solution may not converge with the global optima. Some researchers have tried to optimize the un-decomposed non-convex ob-
jective function of transit network design by metaheuristic algorithms. For instance, [51] [35] embedded the transit network design problem in genetic algorithms (GA).

The discrete method can provide numerical results for transit network design, but is complex and problem formulation varies greatly from city to city. The method requires information from an OD table.

In order to obtain an analytical solution and the general structure of the transit network, continuous approximation (CA) approaches can be used instead. The CA method was first proposed by [18] [19]. In the context of transit system design, the method models station and route spacings as continuous functions of location. Similarly, headways are treated as a continuous function of location and time. Example applications of the CA method include: [18], which modeled a single transit station with time-dependent demand; [42], which modeled a corridor; [13], which modeled a many-to-one system; and [38], which modeled a many-to-many system. These studies illustrate how the design problem can be greatly simplified by modeling a systems numerous decision variables with just a few decision functions. Such simplifications give CA models a certain idealized quality. Yet, the output from these models can serve as guidelines for designing actual transit systems in real-world settings. The design objective is a system that fits a citys street layout and achieves actual costs close to those predicted by adhering to the guidelines of the model as closely as possible. Some design examples with cost comparisons for many-to-one problems can be found in [13] and [53].

Idealized CA models have long been used to determine how transit networks should be organized. [10] appears to have been the first study to examine a city-wide system. It explored how a bus grid should be organized in terms of its vehicle headways and station spacings so as to provide optimum service for a given cost. Other works subsequently analyzed other network structures, including those that are purely radial and purely ring-and-radial in form ([3], [1], and [38]).

More recently, [4] examined a hybrid structure that serves a city by means of a square grid of transit routes within a central (e.g. downtown) district, in combination with radial routes that branch out to all areas of the periphery. This hybrid structure generalizes those previously studied. The structure is fully described by just three parameters: the station spacing, which is assumed to be even over the entire city; vehicle headway in the central district; and the physical size of that district relative to the size of the entire city. The model unveiled how distinct transit modes (bus, BRT and rail) can serve distinct city forms. The hybrid concept was more recently adapted to ring-radial transit networks in [26].

Despite their favorable attributes, [38] appears to be the only previous attempt to use CA methods to design a many-to-many system with spatially varying vehicle headways and line spacings. That model minimizes passenger trip time subject to a fleet size constraint. It looks for an optimum ring-radial transit system atop a dense ring-radial street network, but the model does not allow for a hybrid structure and lacks realism by allowing for arbitrary headway functions, which require numerous transfers, and then ignores passenger waiting times while transferring. Non-monotonic headway functions require that each radial line be served by multiple transit routes with distinct start and end points. As a result, many
patrons would have to transfer between those routes when traveling radially. The constraint of passenger transfer times needs to be properly captured in a model.

1.3 Summary of Research Contributions

This research makes two primary contributions: 1) We propose a pre-signal strategy that can compensate for the capacity loss to cars at intersection approaches caused by bus signal priority and the dedication of a bus lane; and 2) we develop CA models to compare two typical city network patterns (i.e. ring-radial and grid network).

Our pre-signal strategy is built upon one introduced in a previous study that used a mid-block pre-signal to reorganize queues of different vehicle flows at the intersection in a tandem fashion [54]. The design proposed in the earlier study is illustrated in Figure. 1.1. A pre-signal allows batches of left-turning and through-moving vehicles to alternately enter the sorting area between the pre-signal and the intersection. When the main signal is green, vehicles of each (left-turning and through-moving type) can discharge simultaneously using all (three) approach lanes. The capacity gained from this design is tremendous, as compared to conventional designs where each vehicle flow only uses a portion of the lanes to discharge (e.g., one lane for left-turning vehicles and two for through-moving ones). The feasibility and benefits of this design have been demonstrated by real-world applications at busy intersections in Shanghai, China (e.g., [22]).

Figure 1.1: A pre-signal controlled intersection approach

To prioritize buses, we now propose converting the shoulder lane to bus use only, and use the pre-signal to control the remaining car lanes; see Figures 2.1a and 2.1b as examples. Using the simplified kinematic wave theory [16], a strategy is developed to integrate the pre-signal scheme proposed in [54] with a simple bus signal priority scheme.

In order to compare the ring-radial and grid network patterns and provide a general methodology for transit network design, we developed two CA models. Like [38], Model 1 assumes a circular city with a dense ring-radial street network. It generalizes this reference by allowing for a hybrid transit structure that features both ring and radial lines in the central district, and radial lines alone in the city’s periphery. Like [4], Model 2 assumes a
square city with a dense square grid of streets. Model 2 generalizes that reference by allowing headways and line spacings to vary within the periphery.

1.4 Organization of the Dissertation

This dissertation is organized as follows.

Chapter 2 first presents the strategy we proposed to integrate the previously-published pre-signal scheme with a proposed bus signal priority scheme. Secondly, analytical models are formulated for estimating the car discharge capacity and bus delays that our strategy entails, and three baseline scenarios are presented for comparison. Thirdly, we present quantitative case studies showing that our strategy can significantly reduce bus delays without compromising car discharge capacities.

Chapter 3 presents the CA models to design the transit network for both grid and circular cities. It includes a procedure to solve the optimization problem for these models (i.e. to minimize the cost of transit networks with constraints).

Chapter 4 presents the numerical analysis of these CA models. Firstly, the general transit network patterns revealed by our models are summarized. Secondly, the optimal costs of two kinds of transit networks (i.e. ring-radial and grid) are compared. Thirdly, the robustness of the CA design procedure in dealing with irregularities in real-world street networks is demonstrated. Fourthly, the optimal costs of our CA models are compared with non-CA models, that fix the transit line spaces.

Finally, Chapter 5 concludes the dissertation by summarizing its contributions, and discussing potential research opportunities that build upon the present work.

Appendices A-F presents notations and math formulations of our proposed models.
Chapter 2

Improving efficiency for signalized intersections by pre-signal and TSP

This chapter is an abridgment of [6]. In this Chapter, we describe the proposed strategy which combines pre-signal and TSP together to provide priority for transit with minor effects on cars. Analytical models are built to examine the benefits of the strategy. Numerical studies are performed in parametric fashion.

2.1 Proposed Strategy and Signal Plan

The intersection approach’s geometric layout is presented in Section Geometric Layout. The main signal’s phase plan together with the bus signal priority scheme is furnished in Section The Main Signal and Bus Signal Priority. The phase plan of the pre-signal is described in Section Pre-signal Plan. For the readers’ reference, a list of notation is furnished in Appendix A.

Geometric Layout

We consider an approach to a four-way intersection controlled by a pre-timed (main) signal (see Figure 2.1a). The approach consists of three or more travel lanes, the left-most of which is for left-turning vehicles only. Those vehicles enjoy a dedicated signal phase.

To implement our strategy, the right-most lane will be converted to a curbside bus lane. For simplicity, we assume that buses do not make turns at the intersection; and that right turning cars are ignored\(^1\). A pre-signal will then be installed upstream of the intersection to control the remaining car lanes (see Figure 2.1b).

\(^1\) In reality, right-turning cars are often allowed to use a curbside bus lane temporarily if their maneuvers do not impede the buses; e.g., see the policy enacted by the City of New York Government [33].
Figure 2.1: An intersection approach

(a) before the bus lane and pre-signal are installed

(b) after the bus lane and pre-signal are installed
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

The Main Signal and Bus Signal Priority

The main signal’s phase plan is described in Figure 2.2, where $G_L$ and $G_T$ denote the duration of green phases for the left-turning and through-moving traffic in the subject direction (denoted as EW left-turning and EW through-moving in Figure 2.2), respectively, and $R_L$ and $R_T$ the duration of phases for the cross-street traffic \(^2\) (denoted as NS left-turning and NS through-moving in Figure 2.2). For phase transition, a yellow period of 4 sec denoted by $t_y$ is inserted between any two consecutive phases.

![Figure 2.2: Phase plan of the main signal](image)

To prioritize buses, we implement a green extension scheme at the main signal, as what has been commonly used in the real world ([15], [40]). Specifically, for a through-moving bus that is predicted to arrive to the intersection within $t_m (t_m \leq G_T)$ \(^3\) after the end of the last EW through-moving phase \(^4\), that EW through-moving phase will be extended to let the bus pass without delay. For the other bus arrivals, no change will be made to the main signal. This signal priority scheme would benefit those buses that would otherwise experience the longest signal delays; i.e., those who would just miss the green signal.

Following the extension of EW through-moving phase, three other phases (NS left-turning, EW left-turning, and NS through-moving) will be postponed but not shortened. Only the next EW through-moving phase will be shortened to offset the green extension

\(^2\)We choose this signal phase plan because it fits well with the operating requirement of pre-signals [54]. We note that another signal plan, i.e., NS left-turning - NS through-moving - EW left-turning - EW through-moving, seems to be commonly used in reality. However, this latter signal plan requires a much longer sorting area when a pre-signal is installed. This requirement is difficult to satisfy in many real cities with short blocks [54].

\(^3\)Here we set $t_m \leq G_T$, because we want to limit the effects of extension $G_T$ in one cycle, i.e. we can always decrease the duration of $G_T$ in the next cycle to avoid changing the duration of other phases.

\(^4\)We assume that the bus arrival time to the intersection can be predicted with high accuracy using today’s on-board GPS technology, and that the prediction can be made early enough to alter the green phase if needed.
so that the following signal cycles are not affected. In so doing, the discharge capacity for cross-street traffic will not be compromised.\(^5\)

**Pre-signal Plan**

We first present the pre-signal timing plan when no signal priority is applied. This fixed-cycle pre-signal plan is determined in a similar fashion to what was proposed by [54]. It is notwithstanding furnished below to form a basis for the discussion that follows, which explains how the pre-signal can be coordinated with the bus signal priority scheme presented in Section The Main Signal and Bus Signal Priority.

A pre-signal cycle consists of two green phases for left-turning and through-moving vehicles, whose durations are denoted by \(g_L\) and \(g_T\), respectively. They are separated by a \(t_y\), plus possibly a red phase. Without TSP, the green phases are timed so as not to oversaturate the main signal where vehicles discharge in all the lanes simultaneously (see again Figure 1.1). Thus we have:

\[
\begin{align*}
g_L n_L & \leq G_L N, \\
g_T n_T & \leq G_T N, 
\end{align*}
\]

where \(n_L\) and \(n_T\) are the numbers of lanes assigned to left-turning and through-moving vehicles upstream of the pre-signal, respectively; and \(N\) the total number of car lanes at the main signal. The equalities hold when the saturation flows of the two signals are equal.

An offset of \(d\) is introduced in between the ends of pre-signal left-turning and EW left-turning phases, and those of pre-signal through-moving and EW through-moving phases, where \(d\) denotes the length of the sorting area between the pre-signal and the main one, and \(v_f\) the vehicles’ free-flow travel speed. This offset ensures that the last vehicle passing the pre-signal in a green phase will discharge into the intersection without delay. An example of this pre-signal plan under saturation vehicle flows is illustrated by a time-space diagram in Figure 2.3a. The diagram describes the states of traffic in the sorting area, where \(O\) denotes the no-traffic state, \(J\) the jam state, \(S\) the state of saturation flow, \(A_L\) and \(A_T\) the states of left-turning and through-moving vehicle flows discharging from the pre-signal, respectively.

These traffic states are modeled using a triangular fundamental diagram ([30], [36], [16]) as shown in Figure 2.3d, where \(q_s\) denotes the car capacity of a travel lane; \(q_T\) and \(q_L\) the saturation discharge flows from the pre-signal for the left-turning and through-moving vehicles, respectively; \(w\) the backward wave speed in a car queue; \(w_T\) and \(w_L\) the speeds of

\(^5\)Real-world transit signal priority schemes often create lost green times to cross-street traffic and those lost times will not be compensated. The overall performance of the priority scheme will thus be undermined by the long cross-street queues that might be created when those approaches are nearly saturated. This negative impact on cross-street traffic will not be an issue for our priority scheme.
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

the shock waves that separate states \(A_T\) and \(A_L\) from \(J\), respectively. We have:

\[
q_L n_L \leq q_s n_T, \quad (2.2a)
\]
\[
q_T n_T \leq q_s n_L, \quad (2.2b)
\]

\[
w_T = \frac{q_T}{N q_s \left(\frac{1}{v_f} + \frac{1}{w}\right)} - \frac{q_T}{v_f} = \frac{n_T}{N - n_T} + \frac{N}{w}, \quad (2.3a)
\]
\[
w_L = \frac{n_L}{N - n_L} + \frac{N}{w} \quad (2.3b)
\]

Note that the pre-signal plan illustrated in Figure 2.3a is valid when the following two inequalities hold: \(g_L \leq R_L + G_L + t_y\) and \(g_T \leq R_T + G_T + t_y\). If one of them is not satisfied, the pre-signal’s left-turning and through-moving phases would overlap. In that case the leading one of the two overlapping green phases has to be shifted to an earlier time to avoid overlapping. Figure 2.3b shows an example when \(g_L > R_L + G_L + t_y\). Note in the figure that pre-signal through-moving phase is shifted leftward by \(g_L - (R_L + G_L + t_y)\) and additional delay is incurred to the through-moving vehicles.

The figures also show that the pre-signal plan requires a minimum length of sorting area to hold the vehicle queues. This length can be calculated as shown below (the derivation is relegated to Appendix B):

\[
d \geq d_{min} = \max \left\{ \frac{g_T n_T}{N} + \frac{n_L}{N - n_L} \max \left\{ \frac{g_T n_T}{N} + g_L - (G_T + R_L + G_L + 2t_y), 0 \right\}, \frac{g_L n_L}{N} + \frac{n_T}{N - n_T} \max \left\{ \frac{g_L n_L}{N} + g_T - (G_L + R_T + G_T + 2t_y), 0 \right\} \right\} / \left( \frac{1}{w} + \frac{1}{v_f} \right) \quad (2.4)
\]

Now we explain how the pre-signal will respond to a green extension of the main signal, as may occur when a bus is approaching the intersection. Figure 2.3c illustrates this case, where the double-line arrow represents the bus trajectory. We recognize that accurate predictions of the bus arrival time to the main signal can be made only when the bus is very near. Thus, there is no time to alter the pre-signal through-moving phase that corresponds to the extended main signal EW through-moving phase (i.e., only the prioritized bus will pass the main signal during the added green time). The following pre-signal left-turning phase will remain unchanged too. The only pre-signal phase to be altered would be the next through-moving, in response to the truncation of its corresponding EW through-moving phase (see Figure 2.3c). This pre-signal through-moving phase will be truncated from the start to ensure that inequality (1b) still holds after the truncation.

Our bus signal priority scheme and the pre-signal timing are simple and conservative. Yet they can produce huge benefits for both buses and cars. To see this, we next formulate analytical models to assess the benefits.
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

(a) An example of separated queues

(b) An example of superimposed queues

(c) When bus signal priority is applied

(d) Fundamental diagram of the sorting area

Figure 2.3: Pre-signal timing coordination
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

<table>
<thead>
<tr>
<th>No pre-signal</th>
<th>Baseline Scenario 1</th>
<th>Signal priority and bus lane</th>
<th>Baseline Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-signal</td>
<td>Baseline Scenario 2</td>
<td></td>
<td>Scenario 4</td>
</tr>
</tbody>
</table>

Table 2.1: Comparison of modeling scenarios

2.2 Analytical Model

To examine the benefits of the strategy proposed in Section 2.1, we consider four scenarios: three baseline ones for comparison and a fourth one where the proposed strategy is applied. The baseline scenarios are: 1) no pre-signal is used, and buses receive no priority; 2) a pre-signal is used to sort left-turning and through-moving vehicles (including both cars and buses) but buses receive no priority; and 3) no pre-signal is used, but buses enjoy an exclusive lane and signal priority. The differences between the four scenarios are illustrated in Table 2.1.

Analytical models are presented in two batches. The first batch of models produce the signal phases and lane assignments for the two signals on an intersection approach that maximize its total vehicle discharge capacity. These models are similar to those furnished in [54], and are described in Section Optimizing Signal Timing and Lane Assignment. Built upon the optimal solutions to the first batch of models, we formulate models for estimating the Pareto frontiers between expected bus delays and car capacities for each of the four scenarios. Those latter models are furnished in Section Estimating Expected Bus Delays and Car Capacities.

Optimizing Signal Timing and Lane Assignment

Given are: the signal’s cycle length as $T$; the total green time assigned to the subject approach in a main signal cycle as $G$ (i.e., $G = G_L + G_T$); the left-turning ratio as $l$; and the numbers of travel lanes available at the main signal $N$.

For the conventional design where no pre-signal is used, the decision variables are $N_L$ and $N_T$, the numbers of lanes assigned to left-turning and through-moving traffic. The optimal lane assignment ($N_L^*, N_T^*$) can be obtained by solving the following integer program (see [54]):

$$ Q_M(N) = \max \left\{ Q = \frac{q_s G}{T \left( \frac{1}{N_L} + \frac{l \cdot l}{N_T} \right)} \right\}, $$

subject to:

$$ N_L + N_T = N, $$

$$ N_L, N_T \geq 1, $$

$$ N_L, N_T \text{ are integers.} $$
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

The optimal objective $Q^*_M(N)$ denotes the maximum vehicle capacity that can be achieved at an intersection approach with $N$ travel lanes; and the corresponding optimal $G^*_L$ and $G^*_T$ are given by:

$$G^*_L = \frac{Tl}{q_s N_L} Q_M(N),$$  \hspace{1cm} (2.6a)  
$$G^*_T = \frac{T(1-l)}{q_s N_T} Q_M(N).$$  \hspace{1cm} (2.6b)

When a pre-signal is installed upstream, we are also given the number of lanes, $n$, at the pre-signal. The decision variables are now $n_L$ and $n_T$, the number of lanes reserved for left-turning and through-moving at the pre-signal. The optimal lane assignment turns out to be the solution to the following integer program (see again [54]):

$$Q_P(N,n) = \max \left\{ Q = q_s \cdot \min \left\{ \frac{NG}{T}, \frac{1 - \frac{2n_T}{T}}{\frac{1}{n_L} + \frac{1}{n_T}} \right\} \right\},$$  \hspace{1cm} (2.7a)  
subject to:  
$$n_L + n_T = n,$$  \hspace{1cm} (2.7b)  
$$n_L, n_T \geq 1,$$  \hspace{1cm} (2.7c)  
n_L, n_T \text{ are integers.}$$  \hspace{1cm} (2.7d)

The optimal objective $Q^*_P(N,n)$ denotes the maximum vehicle capacity of an intersection with $N$ lanes at the main signal and $n$ lanes at the pre-signal. The optimal green periods of the two signals are:

$$g^*_T = \frac{T(1-l)}{q_s n_T} Q_P(N,n),$$  \hspace{1cm} (2.8a)  
$$g^*_L = \frac{Tl}{q_s n_L} Q_P(N,n),$$  \hspace{1cm} (2.8b)  
$$G^*_L = \frac{g_L n_L}{N},$$  \hspace{1cm} (2.8c)  
$$G^*_T = G - G^*_L. \quad ^6$$  \hspace{1cm} (2.8d)

Since $N$ and $n$ usually are small integers, the two integer programs (2.5a-d) and (2.7a-d) can be easily solved by examining the objective function values for all the feasible combinations of the decision variables.

---

\(^6\)When the main signal green time is more than needed, i.e., $G > \frac{(g_L n_L + g_T n_T)}{N}$, the redundant green time is all allocated to the EW through phase to compensate the potential loss imposed by signal priority.
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

Estimating Expected Bus Delays and Car Capacities

Given the optimal signal settings and lane assignments, we now estimate the expected bus delay, $E[W_{B,i}]$, and through-moving car capacity, $E[Q_{T,i}]$, for each scenario $i (i = 1, 2, 3, 4)$ in Table 2.1. We do this as a function of the bus arrival rate to the intersection. For simplicity, we assume that virtual bus arrivals (i.e. unaffected by traffic) follow a stationary Poisson process.

Baseline Scenario 1

Since virtual bus arrivals are uniformly distributed over time, the expected bus delay is equal to the average delay per through-moving car. The latter can be easily obtained from examining a queueing diagram of traffic flows that discharge into the main signal (e.g., [17]):

$$E[W_{B,1}] = \frac{qs}{2T(q_s - q_A)}(T - G_T)^2, \quad (2.9)$$

where $q_A$ denotes the stationary vehicle inflow per through-moving lane. We have:

$$q_A = \frac{Q_{T,1} + \lambda_b \delta}{N_T} \quad (2.10)$$

where $Q_{T,1}$ denotes the inflow of through-moving cars; $\delta$ the Passenger Car Equivalent (PCE) of a bus, which usually takes a value of 3.5 ([32]); and $\lambda_b$ the Poisson bus arrival rate. We use the $G_T^*$ and $N_T^*$ in (2.9) and (2.10) obtained by solving (2.5a-d) and (2.6b). Now combining (2.9) and (2.10), we have the following equation:

$$E[W_{B,1}] = \frac{qs}{2T\left(q_s - \frac{Q_{T,1} + \lambda_b \delta}{N_T}\right)}(T - G_T^*)^2 \quad (2.11)$$

This expectation gives the bus delay that arises for different car flows, $q_A$. It is an increasing function that jumps to infinity when $q_A$ exceeds capacity. As such it describes the Pareto frontier between car flow and bus delay for Baseline Scenario 1.

Baseline Scenario 2

In this case, $E[W_{B,2}]$ is again equal to the average delay per through-moving car, which can be calculated again using a queueing diagram. The resulting Pareto frontier between car flow and bus delay for Baseline Scenario 2 is shown below. The derivation is relegated to Appendix C.

\[\text{Note } E[Q_{T,i}] \text{ is the car capacity. It is the maximum outbound through-moving car flow in the intersection with given signal plan and intersection geometry.}\]
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

\[ W_{B,3} = \left\{ \begin{array}{ll}
\frac{(Q_{T,2} + \lambda_0 \delta) T}{2q_s N} + T - G_T + \max\{0, g_L - (R_L + G_L + t_y)\}, & \\
q_s N \left( T - G_T + \max\{0, g_L - (R_L + G_L + t_y)\} \right)^2, & \\
q_s N \left( G_T - \max\{0, g_L - (R_L + G_L + t_y)\} \right) \geq (Q_{T,2} + \lambda_0 \delta) T, & \\
q_s N \left( G_T - \max\{0, g_L - (R_L + G_L + t_y)\} \right) < (Q_{T,2} + \lambda_0 \delta) T, & \\
\end{array} \right. \] (2.12)

where \( G_T, G_L, \) and \( g_L \) come from solving (2.7a-d) and (2.8b-d). This again describes the programming formula for values of car inflow below capacity.

**Baseline Scenario 3**

With the exclusive bus lane, we have: \( W_{B,3} = 0 \) for \( 0 < t_a \leq t_m \) and \( T - G_T < t_a \leq T \); and \( W_{B,3} = T - G_T - t_a \) otherwise. Since \( t_a \) is uniformly distributed in \([0, T]\), we have:

\[ E[W_{B,3}] = \frac{(T - G_T - t_m)^2}{2T} \] (2.13)

Since \( E[W_{B,3}] \) does not depend on \( Q_{T,3} \), the Pareto frontier of Baseline Scenarios 3 (i.e. (2.13)) will be a horizontal line. This is true even if the car inflow exceeds the capacity of Scenario 3: \( q_s(N - 1)G_T \).

**Baseline Scenario 4**

Similar to Scenario 3, we have:

\[ E[W_{B,4}] = \frac{(T - G_T - t_m)^2}{2T} \] (2.14)

However, in this case the signal priority scheme will create green time loss for the through-moving traffic. Let \( t_L \) denote the green time loss for through-moving vehicles at the main signal in a cycle. We have:

\[ E[Q_{T,4}] = Q_P (N - 1) (1 - l) \left( 1 - \frac{E[t_L]}{G_T} \right) \]

\[ = Q_P (N - 1) (1 - l) \left( 1 - \max\{0, t_m - G_T + \frac{g_T T}{N - 1}\} + \frac{1}{\lambda_0} \left( e^{-\lambda_0 \max\{0, t_m - G_T + \frac{g_T T}{N - 1}\}} - 1 \right) \right) \] (2.15)

where the calculation of \( E[t_L] \) is furnished in Appendix D. The Pareto frontier can be obtained by eliminating \( t_m \) from (2.14) and (2.15). The calculated frontier only spaces values of \( Q_{T,4} \) that are below saturated flow. For greater values, \( E[W_{B,4}] \) is the value at capacity since the system is separated as if the flow matched capacity.
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

2.3 Numerical Case Studies

The models in this section are programmed in Matlab to examine the proposed strategy under various numerical cases. For the first batch of numerical cases, we assume: \( N = n = 3, l = 0.2, T = 120 \text{ sec}, G = 60 \text{ sec}, R_L = 20 \text{ sec}, R_T = 40 \text{ sec}, q = 1800 \text{ vehicle/hr}, \lambda_b \) buses/hr, and \( t_m \) takes value from 0 to \( G_T \). The Pareto frontiers of Scenarios 1-4 are shown in Figure 5a with square, triangle, cross, and circle markers, respectively.

Comparison between Scenarios 1 and 4 in Figure 2.4a unveils that, by deploying bus priority schemes and the pre-signal concurrently, not only buses enjoy the reduced delay, but cars also benefit from the increased capacity; see that the circle-marked curve is entirely in the top-left corner of any square marker. Further comparisons show that our strategy furnishes much smaller bus delays than Scenario 2, and higher car capacities than Scenario 3.

The above benefits hold in general. To show this, we compare the Pareto frontiers of the four scenarios for other combinations of \( N = n = 3 \) or 4, and \( l = 0.2, 0.3, \) or 0.4, with other parameters the same as above. These Pareto frontiers are shown in Figures 5b-f. The figures again show that our strategy furnishes higher car-carrying capacities than Scenarios 1 and 3, and much smaller bus delays than Scenarios 1 and 2.

To examine the effectiveness of the strategy under a relatively high bus flow, we also plot the Pareto frontiers in Figures 6a-f for the cases where \( \lambda_b = 90 \text{ buses/hr}, N = n = 3 \) or 4, and \( l = 0.2, 0.3, \) or 0.4. These figures unveil similar patterns as Figures 2.4a-f. The figures also shows that, with a higher bus flow, the car capacity diminishes more rapidly as bus delay reduces; see the steeper left part of circle-marked curves in Figures 2.5a-f. In the worst case (Figure 2.5a), the lowest car capacity from our strategy is only slightly below the maximum capacity that Scenario 1 could achieve. In all the other cases, our strategy still outperforms both Scenarios 1 and 3 in terms of car-carrying capacity.

On a related note, the pre-signal has to be located at least \( d_{\text{min}} \) from the intersection to accommodate the maximum inflow of cars. Suppose \( w = 6.26 \text{m/s} \) and \( v_f = 15.64 \text{m/s} \) ([2]), the \( d_{\text{min}} \) for cases (a-f) are respectively: 200m, 175m, 150m, 224m, 206m, and 179m.

2.4 Summary

We proposed a strategy that integrates a mid-block pre-signal and a simple bus signal priority scheme with a dedicated bus lane in an intersection approach. The combined effect includes increase in car discharge capacity and significant reduction in bus delays. Numerical results from our analytical models show that the benefit of this strategy is huge, as compared against the mixed-traffic scenario, and the scenarios where only the pre-signal or only the bus priority scheme is implemented.

Our findings have significant practical implications. For example, they demonstrated that bus priority schemes can be deployed without losing car-carrying capacity even in a busy arterial with only three lanes per travel direction. Note how this would make conventional
bus priority strategies applicable in a much wider range of operating conditions. Note further that the benefits are achieved under a simple signal priority scheme that involves only green extension. Even better results are expected when more realistic and comprehensive priority schemes are taken into account. For example, by integrating red truncation and green insertion schemes (e.g., [46]), bus delays can be further reduced. On the other hand, by making the bus lane intermittently available to cars (e.g., [43]), the car discharge capacity can be further improved. Both of the above can be done with modest modifications to our models.

Other limitations of our work include that: i) the analysis based upon kinematic waves ignored the bounded acceleration of vehicles; ii) only through-moving buses are considered; and iii) the effects of neighboring bus stops are ignored. However, the former can be settled by simply increasing the $t_y$ to account for the lost time due to vehicle acceleration. When some buses need to make left-turns at the intersection, the pre-signal can be reprogrammed to hold off all the vehicles upstream, so that a left-turning bus can change to the left-most lane without being impeded by cars. Finally, work is under way in regard to integrating our strategy with previous studies on the effects of near-side/far-side bus stops on the intersection discharging traffic ([48], [29]).
CHAPTER 2. IMPROVING EFFICIENCY FOR SIGNALIZED INTERSECTIONS BY PRE-SIGNAL AND TSP

Figure 2.4: Comparison of scenarios ($\lambda_b = 30$ buses/hr): (a) $N = n = 3, l = 0.2$; (b) $N = n = 4, l = 0.2$; (c) $N = n = 3, l = 0.3$; (d) $N = n = 4, l = 0.3$; (e) $N = n = 3, l = 0.4$; (f) $N = n = 4, l = 0.4$
Figure 2.5: Comparison of scenarios ($\lambda_0 = 90$ buses/hr): (a) $N = n = 3, l = 0.2$; (b) $N = n = 4, l = 0.2$; (c) $N = n = 3, l = 0.3$; (d) $N = n = 4, l = 0.3$; (e) $N = n = 3, l = 0.4$; (f) $N = n = 4, l = 0.4$
Chapter 3

Analytical Models to Design Transit Network for Circular and Grid Cities

This chapter and the following chapter is an abridgment of [28]. The chapter presents the models and the analysis methods to design city-wide transit networks. We build two CA models for circular (Model 1) and grid city (Model 2). Sections 3.1 and 3.2 introduce the basic assumptions: section 3.1 for supply and section 3.2 for demand. Then, section 3.3 formulates the problems and section 3.4 discusses how to solve them.

3.1 City Forms and the Supply of Transit

This subsection introduces notation and explains the assumptions regarding the structure of the transit systems. This is done first for Model 1 and then for Model 2.

It is assumed for Model 1 that the city is circular with radius \( R \), and that its streets form a dense and rotationally-symmetric network; see Fig. 3.1a. It is also assumed that the city has a central district of radius \( r \) (to be determined) which enjoys double-coverage transit service; i.e. such that every point is near two transit lines – one circular and one radial. The headways on the circular ring lines and the spacings between these lines shall be denoted \((H_c, s_c)\). Similarly, the headways and spacings pertaining to the radial routes shall be denoted \((H_r, s_r)\).

The four headway and line spacing values at a location \((H_c, s_c, H_r, s_r)\) are allowed to vary with the location’s distance from the center, \( x \), as in [38]. Now, however, these functions are constrained to ensure that all patrons can travel in the radial direction without any transfers. Transfers are avoided by stipulating that all transit vehicles that travel along radial routes: (i) depart from a city’s boundary; (ii) travel through the central district; and (iii) eventually arrive at the city boundary on the opposite side of town. This is illustrated by the route shown in bold in Fig. 3.1a. Since vehicles are dispatched at the city’s boundary and conserved, the radial flow of buses across any circular ring, \( \frac{2\pi x}{s_r(x)} \left( \frac{1}{H_r(x)} \right) \), must then be
the same for all $x$; i.e. $s_r(x)$ and $H_r(x)$ must satisfy the following conservation constraint:

$$\frac{2\pi R}{s_r(R)H_r(R)} = \frac{2\pi x}{s_r(x)H_r(x)}, \forall x \quad (3.1)$$

Lines with spacings that approximately conform to $s_r(x)$ are deployed by allowing outbound radial lines to bifurcate and inbound lines to merge; e.g., as in Fig. 3.1a.

Consider now Model 2 and Fig. 3.1b. As in the figure, cities are assumed to be square with ”radius” $R$ (side $2R$) and to have a dense, square grid street system. As with Model 1, a central district of ”radius” $r$ (side $2r$) enjoys double coverage. In this case this coverage is achieved by lines forming a homogenous square grid with spacing $s_g$ and sharing a common headway $H_g$. The radial routes in the periphery are prolongations of the central routes that extend all the way to the edge of the city; see the figure. Their headways and spacings are denoted $(H_r, s_r)$. The latter are allowed to be arbitrary functions of the distance from the city center, where such distance is measured by the ”radius” $x$ of the square ring at each location. As with Model 1, $(H_r, s_r)$ are related by the vehicle conservation constraint (3.1); and as in that model, line spacings that approximately conform to $s_r(x)$ are achieved by bifurcating routes.

1The factor $\frac{2\pi x}{s_r(x)}$ is the number of radial lines across the ring at $x$, and $\frac{1}{H_r(x)}$ is the bus flow on each radial line at $x$. Thus, $\frac{2\pi x}{s_r(x)H_r(x)}$ is the radial flow of buses across the ring at $x$.

2Route bifurcation/convergence is shown in the figure in abrupt, piecewise fashion, as might occur in real settings when transit lines are fit to underlying streets. Because we assume a continuum (as if a city’s streets were infinitely dense), our numerical analyses will produce approximations in which bifurcations and convergences occur gradually over space; see Figs. 3.3 and 3.4.
3.2 Demand: Patron Choices of Routes and Transfers

It is assumed in this paper that the demand is temporally variable but spatially uniform; i.e., with origins and destinations uniformly and independently distributed in the interior of the city.\(^3\) As explained in [4], temporal variations can be described for modeling purposes with only two parameters: the average daily demand and the rush-period rate. Furthermore, because the latter plays only a secondary role (it is just used to determine fleet size and the vehicles’ passenger-carrying capacity) and to avoid the proliferation of parameters, this paper will fix the rush period demand at 250% of the average, just as in [4]. Thus, the demand is fully characterized in the present paper by a single parameter: the trip generation rate per unit area, \(\rho\) (trips/hr-km\(^2\)).

Regarding travel behavior, it is assumed that a patron walks from her origin to the nearest transit station at speed, \(w\) (km/h), and ultimately departs the system via the station nearest her destination. It is further assumed that the patron chooses the transit route that offers the shortest trip distance, and breaks ties by first minimizing transfers and then choosing randomly among all the routes that remain. This is now fleshed out in more detail for each model.

Consider Model 1 and let \(\theta\) be the angle (in radians) between the patron’s origin and destination radii; e.g. see the trip labeled 1 in Fig. 3.2a. Consideration shows that if \(\theta > \frac{\pi}{2}\), a patron on a ring-radial network will travel by radial line to the city center and transfer there to another radial line to reach her destination. If instead \(\theta < \frac{\pi}{2}\), a patron will travel via both circular and radial lines, using the circle that is closest to the trip terminus that is itself closest to the city center. Proofs can be found in [10]. This travel behavior is exemplified by trips 2 and 3 in the figure. Note in particular trip 3 in which a patron’s origin and destination both lie in the periphery, such that two transfers are required. All other trip types require only one transfer.\(^4\)

On hybrid grid networks, the routing pattern and transfers are as described in [4]. Travelers use two intersecting routes if the trip can be achieved with only one transfer, in which case distance is minimized; or else using the three routes that minimize travel distance, and transferring twice in the process. As illustrated by Fig. 3.2b, the latter situation arises when a patron’s origin and destination lie in the same quarter of the periphery. Consideration shows that it also arises if the origin and destination lie on opposite quarters of the periphery, and that all other origin-destination locations require only one transfer.

3.3 Model Formulations

We concern ourselves with the system’s generalized cost in the city per unit time, denoted \(Z\) ($/hour). This cost is the sum of the costs experienced by the transit agency, the public

---

\(^3\)Though the proposed models can be extended to accommodate spatially inhomogeneous demands patterns with rotational symmetry, this is not done for the sake of simplicity.

\(^4\)We conservatively ignore the small likelihood that a trip can be made without requiring a single transfer.
CHAPTER 3. ANALYTICAL MODELS TO DESIGN TRANSIT NETWORK FOR CIRCULAR AND GRID CITIES

Figure 3.2: (a) route choice for ring-radial network; (b) route choice for grid network.

at large (externalities) and the system’s users. These cost types are further subdivided into categories, $i$, that depend in a similar way on the decision variables. Thus, we write: $Z = \sum_i Z_i$, where the subscript $i$ denotes a specific cost category.

For the transit agency three categories are considered: infrastructure, rolling stock operation and rolling stock capital. These costs depend on specific features of the system configuration, i.e., our design. The infrastructure cost is assumed to depend linearly on the total length of the transit routes, $L$; the rolling stock operating cost to depend linearly on the vehicle kilometers traveled per unit time, $V$; and the rolling stock capital cost on the fleet size $M$. For convenience these variables will be used to label the categories, so for the agency $i = L, V$ or $M$.

It seems reasonable to assume that the system’s externalities are linked to what the agency does for this reason can be approximately expressed as a linear combination of $L, V$ and $M$. As such, the external costs can be grouped together with the agency costs without involving any extra system configuration features if appropriate coefficients are chosen in the cost function. In other words, it is assumed that the two costs together are: $L_L + V_V + M_M$.

For the patrons, we recognize four cost categories: the access cost of walking to and from the origin and destination stations; the cost of waiting to board at the origin station and at transfer stations; the in-vehicle cost of riding from origin station to destination station; and a transfer penalty. These costs can also be related to the system design through some key variables. As the access cost is assumed to be proportional to the total distance walked per hour $A$, the waiting cost to the total time waited per hour, $W$, the in-vehicle travel cost to the total in-vehicle time per hour, $T$, and the transfer penalty to the total number of transfers done per hour, $e_T$. So as in the case of the agency cost and the externalities, the total user cost can be expressed as a linear combination of these variables: $A_A + W_V + T_T + e_T e_T$. It will be assumed in the numerical examples that $A_A = V_W = T_T = \mu$, where $\mu$ is a value of
time that would be reasonable for an application.\textsuperscript{5}

In summary, if we use \( Y_i \) for the design variables \((L, V, M, A, W, T, e_T)\) and \( \$_i \) for the cost coefficients, the systems’ total cost is:

\[
Z = \sum_i Z_i = \sum_i \$_i Y_i.
\]

In agreement with the CA modeling framework we further break-down each of the above categories into: a set of localized cost densities, \( z_i(x) = \$_i y_i(x) \) \($/time-distance\), that are associated with the circular crowns defined by radii \((x, x + dx)\), and that can be summed across all crowns from \( x = 0 \) to \( x = R \); and a set of global fixed costs, \( F_i \) \($/time\). Thus we write:

\[
Y_i = F_i + \int_0^R y_i(x)dx.
\]

The generalized cost to be minimized is therefore:

\[
Z = \sum_i Z_i = \sum_i \$_i \left[ F_i + \int_0^R y_i(x)dx \right] = \$_i \left[ \sum_i F_i \right] + \$_i \left[ \int_0^R \sum_i y_i(x)dx \right]
\]

Appendices A and B explain how the \( F_i \) and \( y_i \) depend on the parameters of the problem and on the decision variables. The problem parameters are: \( R \) (or the area of the city, \( S \)); the demand density for trips, \( \rho \) (patrons/h/km\(^2\)); the transit vehicle’s passenger-carrying capacity, \( C_{pax} \) and cruise speed, \( v \) (km/h);\textsuperscript{6} the fixed lost time spent at stations due to acceleration, deceleration and doors opening and closing, \( \tau \) (h/station), and the added time for each patron’s boarding time, \( \tau' \) (h/patron). (Alighting times are not considered, since these tend to be small and because boarding and alighting movements often occur simultaneously through separate doors.) Recall that the decision variables are the spacings, headways and the central district radius. In what follows only the dependence on the decision variables is important so this dependence is now explicitly expressed.

It will be convenient to separate the decision variables into a set of global variables, \( G \), and a set of location-dependent decision functions of \( x, D(x) \). The local decision functions are \( D(x) = \{s_r(x), H_r(x), s_c(x), H_c(x)\} \) for ring-radial Model 1, and \( D(x) = \{s_r(x), H_r(x)\} \) for grid Model 2. These functions define the system in those parts of the city where spacings and headways are allowed to vary.

The global decision variables define the general characteristics of the system. These include the radius of the central district, \( r \), and an auxiliary variable, \( Q \), which shall denote the flow of radial-line buses across every ring. The global variables also include the service headway, \( H_B \), of the outermost ring route, i.e. the route on the boundary between the central district and the periphery, which is allowed to be smaller than \( H_c(r) \). The headway for this ring route is allowed to be smaller than \( H_c(r) \) because this is the only route that serves as a circulator for trips starting and ending in the periphery and therefore carries considerably more traffic than other ring roads nearby. We shall also define an auxiliary variable for this route, \( s_B \), representing the spacing between the route’s stops that reside at the intersections of the radial lines and the boundary route. We will require that \( s_B = s_r(r) \),\textsuperscript{5}

\textsuperscript{5}It would be possible to value the different components of travel time differently but this is not done to avoid the proliferation of parameters.

\textsuperscript{6}This speed accounts for disruptions from traffic signals and other interferences along a route, but not for the stops made to serve demand. It depends on the type of technology and infrastructure used.
however. It will be convenient to jointly refer to the two parameters for the boundary route as \( f_B = (s_B, H_B) \). Finally, two additional variables are required for Model 2 (because this model has fewer local decision functions): the uniform spacing in the central district, \( s_g \), and the common headway, \( H_g \). It will be convenient to refer to these variables as \( f_G = (s_g, H_g) \) and to define \( f_G = \emptyset \) for ring-radial Model 1. In summary, the global decision variables are:

\[
G = \{Q, r, f_B, f_G\} \text{ where } f_B = (s_B, H_B) \text{ and } f_G = \emptyset \text{ for Model 1 or } f_G = (s_g, H_g) \text{ for Model 2.}
\]

Appendices A and B show that for both Models the elements of the objective function depend on the decision variables as indicated in the problem formulation below:

\[
\min_{D(x), G} \{ Z = \sum_i s_i[F_i(r, f_B, f_G)] + \int_0^R \sum_i s_i[y_i(D(x), r, x)]dx \} \quad (3.2)
\]

The appendices also show that the decision variables are constrained by the following relations, where \( C, C_G \) and \( K \) are functions of the indicated variables:

- \( s_B = s_r(r) \) (definitional) \quad (3.3a)
- \( H_B \leq H_c(r) \) (definitional for ring-radial network) or \( H_B \leq H_g \) (definitional for grid network) \quad (3.3b)
- \( C(D(x), x, Q) = 0, \forall x \geq r \) (vehicle conservation constraint) \quad (3.3c)
- \( C_G(f_G, Q) = 0, \) (vehicle conservation constraint in central district, Model 2) \quad (3.3d)
- \( C(D(x), x, Q) = 0, \forall x < r \) (vehicle conservation constraint in central district, Model 1) \quad (3.3e)
- \( K(D(x), x) \leq 0, \forall x \) (vehicle’s passenger-carrying capacity constraint) \quad (3.3f)
- \( K(f_B, r) \leq 0. \) (vehicle’s passenger-carrying capacity constraint on boundary route) \quad (3.3g)

### 3.4 Solution Procedures

For both models, we find a feasible (near-optimum) solution, and then compare the near optimum cost with a lower bound to confirm the quality of the feasible solution. We will consider Model 1 first and then Model 2.

To construct a feasible solution for Model 1, fix the global variables \( G = \{r; Q; f_B; f_G = \emptyset\} \) and consider the subproblem for the remaining decision function, \( D(x) \). This problem only involves location-dependent constraints \((3.3c, 3.3e, \text{ and } 3.3f)\) unless \( x = r \). The objective function is still \((3.2)\) but the first term is fixed. The optimal decision function for the
CHAPTER 3. ANALYTICAL MODELS TO DESIGN TRANSIT NETWORK FOR CIRCULAR AND GRID CITIES

subproblem only depends on $Q$ and for this reason will be denoted $D^o(x|Q)$. This subproblem can be seen to be a calculus of variations problem that further decomposes by $x$. In other words, for every $x$, $D^o(x|Q)$ is simply the value of $D$ that minimizes the integrand portion of (3.2) subject to constraints (3.3c, 3.3e and 3.3f):

$$D^o(x|Q) = \arg\min_D \left\{ \sum_i z_i(D, r, x) : C(D, x, Q) = 0; K(D, x) \leq 0 \right\}, \forall x. \quad (3.4)$$

Because the optimization problem in (3.4) is not convex, we solve it numerically. For each instance of the problem, a gradient search method was used ten times with ten randomly selected initial points to identify ten local optima. The least cost optimum was then identified as a candidate solution. This process was then repeated 15 times with different 10-point random samples of initial points to identify fifteen candidate solutions. In every instance of the problem we solved, the fifteen candidates coincided. This strongly suggests that these candidates are the globally optimal solutions (except when $x = r$) and that a single ten-point sample suffices to identify each globally optimal solution.

To finish the process of obtaining a near-optimum feasible solution, now solve problem (3.2) and (3.3) with $D(x)$ replaced by $D^o(x|Q)$ (even when $x = r$) in order to identify $G^o = \{Q^o, r^o; f_B^o, f_G^o = \emptyset\}$. This is slightly suboptimal because $D^o(r|Q)$ is suboptimal. The numerical solution is simple because the problem is now an ordinary minimization problem with only four variables and a simple structure.\(^8\) The final result, $\{G^o, D^o(x|Q^o)\}$, is the sought-after, near-optimum feasible solution. Since the solution is feasible, it is a upper bound to the exact optimum.

For the lower bound, we find the optimum of problem (3.2), (3.3) after relaxing constraints (3.3a) and (3.3b). Conditional on $r$ and $Q$ the relaxed problem now decomposes perfectly into two subproblems for the two sets of variables: $G$ and $D(x)$. The subproblem for the second set $D(x)$ involves, as before, the second term of the objective function and location-dependent constraints (3.3c, 3.3e, and 3.3f), and its solution is the same as before. The subproblem for the first set $\{f_B; f_G = \emptyset\}$, involves the first term of the objective function and the remaining constraint (3.3g). This problem is simple as it only involves two variables. To find the unconditional optimum across all $r$ and $Q$, which identify the lower bound, we look for the $r$ and $Q$ that minimize the sum of the costs of the two subproblems. As before, we find the solution by searching from numerous initial points.

For Model 2, all the processes are the same except that location-dependent constraints are now (3.3c, 3.3f).

To confirm the quality of the feasible solutions for both Model 1 and 2, we compared the cost of each feasible solution with its corresponding lower bound for each of 81 scenarios.

\(^8\)Since constraints (3.3c-3.3f) are now redundant, the remaining relevant constraints and the objective function, can be seen to decompose into two independent subproblems: one for $Q$ and the second for $(r, f_B, f_G)$.
involving a wide variety of conditions; see Chapter 4\(^9\). In every case, the difference between the two costs for the ring-radial network averaged less than 1.5%, and it never exceeded 3.4%. For the grid network, the average difference was less than 1%. This confirms that the feasible solutions, found in the above way are indeed near-optimal.

### 3.5 Summary

We built two CA models to design rotationally symmetric transit networks for circular and grid cities. These models minimize the generalized cost which includes agency cost and passenger cost. Since the models are too complex to solve directly, we proposed a two-step method which can find the near-optimal solutions. We will apply the models to investigate the general structure patterns of transit networks in the next chapter.

---

\(^9\)The search process can be completed in less than 1 sec for each case.
Chapter 4

Parametric Analysis for Transit Networks

The procedures of Chapter 3.4 are now used to design systems for 81 different scenarios and to look for general insights, see Section 4.1.

The scenarios in question will also be used in Secs. 4.2 and 4.3. The scenarios are constructed by changing the values of the three parameters that are most likely to differ greatly across cities: the city’s surface area $S$ ($50,225,400 \text{ km}^2$); the demand density $\rho$ ($50,200,600 \text{ trips/h/km}^2$); and the value of time $\mu$ ($5,20,50 \text{ $/h}$). In addition, three different transit modes are considered: Bus, BRT and Metro. For simplification, we only consider the uniform passenger demand spatially as the worst case for public transit. Thus, there are 81 scenarios in total. The remaining parameters, which characterize the transit modes, are taken from [4]\(^{1}\). They are summarized in Table 4.1.

Table 4.1: Technology parameters for the considered scenarios.

<table>
<thead>
<tr>
<th>Mode</th>
<th>$L$ ($$/\text{km-h})$</th>
<th>$V$ ($$/\text{veh-km})$</th>
<th>$M$ ($$/\text{veh-h})$</th>
<th>$v$ ($\text{km/h})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>9</td>
<td>2</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>BRT</td>
<td>90</td>
<td>2</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Metro</td>
<td>900</td>
<td>6</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>$\tau$ (sec/stop)</td>
<td>$\tau'$ (sec/p)</td>
<td>$w$ (km/h)</td>
<td>$C_{pax}$ (pax/veh)</td>
<td></td>
</tr>
<tr>
<td>Bus</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td>BRT</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>Metro</td>
<td>45</td>
<td>0(^{\dagger})</td>
<td>2</td>
<td>1200</td>
</tr>
</tbody>
</table>

\(^{\dagger}\)Zero is used because the vehicle’s fixed dwell time is already sufficiently long to accommodate boardings.

\(^{1}\)The values of $\tau'$ assume that fares are collected offline, see [4].
4.1 System Structure Insights

The BRT scenario with $S = 225 \text{ km}^2$, $\rho = 200 \text{ trips/h/km}^2$, and $\mu = 20 \text{ \$/h}$ will be used as the base case for illustration purposes. We will find that the near-optimal configurations of Model 1 and Model 2 are quite different. This happens because the demand for travel, i.e. the spatial distribution of the passenger-miles traveled per unit time in different parts of a city, is very different for our two city types. Figure 3 shows the distribution of travel along rings as a function of $x/R$. Strikingly different among the two cities is the demand for travel along a ring for $x \approx R$, which reaches a minimum at $x = R$ for the circular city and a maximum for the square city. This difference in demand has a profound impact on the structures of the transit systems designed to accommodate it as we shall now see.

Optimum System Structure: Model 1

Figure 4.2a shows the main design variables obtained with the procedure of Chapter 3.4 for the base case scenario. Note how the ideal headways and spacings vary with distance from the city center even though travel demand is uniformly distributed. (Optimal designs would, of course, still feature spatial variations in headways and spacings in cases where demand varies with distance from the city center.) Figure 4.2b presents part of a network designed to have features similar to the ideal of Fig. 4.2a. In this figure, the thickness of each line represents its vehicle flows, with thicker lines denoting higher flows (lower headways).

![Figure 4.1: Fraction of total vehicle kms flowing along a circular (square) crown of unit width and radius $x/R$](image)
CHAPTER 4. PARAMETRIC ANALYSIS FOR TRANSIT NETWORKS

Application of Model 1 to other scenarios involving ring-radial street networks produces outcomes that are similar to the ones in Fig. 4.2. The following discusses those features of Fig. 4.2 that apply to all the scenarios studied and can therefore be considered general.

Radial routes in the central district: The spacing of radial routes increases with \( x \) inside the central district; e.g., as shown by the solid curve in Fig. 4.2a for \( 0 < x < r \approx 3.07 \) km in the base case. The rate of increase is only modest, however, so the radial routes need to bifurcate to match the ideal spacings approximately; e.g., as shown in Fig. 4.2b. The headways on these lines should grow with \( x \) as well, e.g., as shown by the bold dashed line of Fig. 4.2a for the base case. This growth is to be expected from the vehicle conservation rule (1), since at every bifurcation outbound transit vehicles diverge, thereby increasing the headway along the branching routes.

Circular routes in the central district: The spacing of the circular ring lines diminishes with \( x \) inside the central district. This is shown by the solid, thin curve in Fig. 4.2a for the base case. This occurs because with a ring-radial network the passenger demand for using the rings increases with \( x \) near the city center, see Fig 4.1. On the other hand, and like their radial counterparts, headways on the circular rings (dashed, thin curve) grow with \( x \).

![Figure 4.2: Ring-radial network example for \( S = 225 \) km²](image)

(a) Ideal features of an optimum ring-radial network  
(b) Actual network with features resembling the idea

Size of the central district: Just as very few origin-destination pairs benefit from rings at \( x \approx 0 \), very few origin-destination pairs would use rings around \( x \approx R \), see Fig 4.1. For this reason, circular routes with larger radii serve little purpose and should not be provided. As a result, the central districts in all cases studied turn out to be fairly small, with \( r/R \) rarely exceeding 50% – this ratio is 36% in Fig. 4.2a, so the central district covers only 13.1% of the area.

Radial routes at the boundary and in the periphery: In all cases studied the spacing between radial lines declines discontinuously at the boundary of the central district where \( x = r \). See the jump in the bold solid line of Fig. 4.2a for the base case. This discontinuity
is accommodated in actual plans by bifurcating the radial lines at the edge of the central district, e.g. as shown in Fig. 4.2b. Concurrent with this jump there must be in all cases a sudden increase in radial-line headways at $x = r$, as required by the vehicle conservation rule (1). This jump can also be seen in Fig. 4.2a. Finally, in the periphery itself, the radial routes turn out in all cases to be more closely spaced on average than in the central district, even though the spacing always increases slightly with $x$. Figure 4.2a illustrates the effect for the base case. The just discussed differences between radial service in the periphery and the central district are partly due to the lack of supporting ring routes in the periphery.

Circulator route on the boundary: In all cases studied, the service frequency on this route is considerably higher than on neighboring ring routes. For the base case this is shown by the black square in Fig. 4.2a, which is significantly below the thin dashed line. This happens because as Fig. 4.1 suggests the boundary route attracts a great deal of demand.

**Optimum System Structure: Model 2**

Outcomes from the application of Model 2 to our base case scenario are displayed in Figs. 4.3a and b. Application of Model 2 to the other 80 scenarios produces similar outcomes. All the cases exhibit the following common features.

**Size of the central district:** The travel concentration phenomenon arising in the middle rings of ring-radial networks does not occur with street grids because these networks spread passenger traffic more evenly in space, see Fig 4.1. As a result, Model 2 always exhibits a much larger central district than Model 1 for the same parameter set. For example, for the base case scenario shown in Fig. 4.3, the central district is about four times larger than for Model 1, extending now to $x/R = r/R \approx 93\%$, and covering 87% of the area.

**Routes in the central district and circulator on its boundary:** By construction, all the routes in the interior of the central district have fixed headway and spacing. This is shown by the horizontal lines of Fig. 4.3a and the even grid of Fig. 4.3b. On the boundary, fewer passengers transfer than on the boundary of Model 1 because the periphery is now much smaller. As a result, it is never necessary to provide extra circulator buses to accommodate the extra transfers on these routes. It is found that in all cases studied it is optimal to set $H_B = H_g$. This feature is highlighted for the base case by the black square of Fig4.3a, which lies on top of the grid line headway.

**Radial routes at the boundary and in the periphery:** As regards the radial routes in the periphery, and as illustrated for the base case by the slanted lines of Fig. 4.3a, the effects are similar to those of Model 1: (i) radial spacings that abruptly diminish at $r$ to compensate for the abrupt discontinuation of double-coverage service; (ii) radial headways that abruptly expand at that location as outbound vehicles distribute themselves across branching routes; and (iii) headways and spacings that expand gradually as the routes approach the city boundary.
4.2 Comparisons between Ring-radial and Grid Cities

Costs are now compared. For ease of interpretation we will use average unit costs (per trip) rather than total cost. Furthermore, to be currency-independent, we shall use as the monetary unit the amount of money equivalent to one hour of travel.

Figure 4.4 compares the optimum costs of Models 1 and 2 for the 81 cases. Note how the ring-radial networks always result in lower generalized costs. Differences in the costs of the two structures are typically between 9% and 13%, though differences as high as 21.5% are also observed. The finding means that the ring-radial street layouts are always more transit-friendly than grids. This happens for two reasons. First, ring-radial networks allow for more direct travel than do grids; e.g. see [7]. And second, because ring-radial networks tend to concentrate travel demand on certain rings (see Fig. 4.1), which is a good thing for collective transportation because it, unlike automobile travel, becomes more efficient when traffic is spatiotemporally concentrated (e.g. see [4]).

4.3 Model robustness

As in [11] we now examine the suitability of the two CA models for designing transit routes to fit over city-street networks that are different from what the models assume. This is again done using the same 81 cases of Section 4.2. For each case, each hybrid ring-radial transit design was adapted to conform to a grid street network, and each hybrid grid transit design was adapted to a ring-radial street network.\footnote{In each case, the design variables from Model 1(2) were chosen to match the optima generated from Model 2(1). These design variables were redefined so that they would apply to both city-street structures.} The costs of the adapted and original designs...
Figure 4.4: Parametric comparisons

Robustness is revealed by seeing how the costs for a given transit morphology change when the design is adapted. Curiously, it is found that the costs of hybrid grid transit systems always decline slightly (by 1-2%) when they are adapted to sit on ring radial street networks, but the reverse is not true. This happens partly because the origins and destinations of circular cities with ring-radial street networks are closer to one another than those of square cities with street grids, so that any adaptation to ring-radial streets helps reduce the average user distance traveled.

In order to exclude these structural effects from our robustness tests we compared the average costs of the two adapted designs with the average of the two originals. Figure 4.5 displays the results. As expected, controlling for city structure, the adaptations add to the costs, but only slightly – 2.7% on average. So the CA method appears to be robust. The differences between a real street network and the best-matching idealized street network (grid or radial) are likely to be smaller than the differences underlying Fig 4.5. This means that if one fits a hybrid grid transit system to a street network roughly resembling a grid, or a ring-radial system to a street network with an approximate ring-radial structure the model prediction errors should be smaller than those of Fig. 4.5.

These variables were the: transit network’s total length; total vehicle distance traveled each hour; and area of the central district. These three variables co-determine the transit networks’ spatial layout and determine the headways.
4.4 Comparisons between CA and Non-CA Models

The costs arising with the two CA models are now compared against the costs produced by comparable non-CA models with three variables \((r, s_g, H_g)\).\(^3\) The results illustrate that the average cost of the CA models is 3.11% lower than that of the non-CA models for Model 1 and 3.35% for Model 2. This comparison reveals that the benefits of spatial variations in headways and spacings is significant but limited. We expect similar benefits to hold when the models are applied in reality because our tests show that the non-CA models are not more robust than the CA models.\(^4\) It thus appears that the first order of business when designing a transit system is to get the line spacings, headways and central district sizes right.

4.5 Summary

The two models presented in this dissertation use CA methods to fit transit systems atop a city’s streets so as to minimize the generalized cost of a trip. The methods lend themselves nicely to the design of systems with vehicle headways and line spacings that can vary with location in the city, so as to better serve its travel demand. Model 1 was formulated for cities with ring-radial street networks. It allows spacings and headways to vary optimally over an entire city, and extends the model in [38] by allowing for a hybrid structure of transit

---

\(^3\)The non-CA models are identical to the CA models, except that the line spacing and headway (i.e., \(s_c, H_c\), and \(s_r\) for model 1 and \(s_r\) for model 2) are homogeneous over the entire city.

\(^4\)Cost increases by about 6% on average when a non-CA model is fit to the "wrong" city.
lines. Moreover, Model 1 repairs the flaws of Vaughan’s model by constraining the tours of radial-line vehicles. Model 2 was developed for cities with grid-street networks. It generalizes [4] by allowing headways and spacings to vary within a city’s periphery.

Application of the models to a wide range of scenarios reveals that Model 1 always has a lower cost than Model 2 on comparable cities. This suggests that ring-radial networks are considerately more friendly to transit than are grids. It was also found that the CA models improve cost, albeit only by about 3%, with non-CA counterparts. Robustness tests suggest that the CA models can be used confidently on cities that have networks resembling the ideal, and that the benefits of the CA models continue to hold when the models are applied to real networks.
Chapter 5

Conclusions

Encouraging people to use public transit can reduce both urban congestion and pollution. However, designing a highly efficient transit system is not easy. There are problems at both the macro and micro scales. This dissertation proposes solutions for two problems in order to improve transit systems, one at the macro scale and another at the micro scale. At the micro scale, it furnished an operating strategy to reduce the delays of buses at signalized intersections. At the macro scale, it provided analytical models to design transit networks for circular and grid cities.

The contributions of the dissertation are summarized in Section 5.1. Directions for future work are discussed in Section 5.2.

5.1 Contributions

The primary contributions of this dissertation can be summarized as follows:

1. A strategy which combines the TSP with midblock pre-signal. This strategy converts the right-most lane to bus lane. The pre-signal is applied to compensate the loss of one lane for private vehicles. The pre-signal and main intersections signal are coordinated to make the whole strategy feasible.

2. Analytical models to optimize the signal plans for pre-signal and main-intersection signals. Additional analytical models to estimate the Pareto frontiers between expected bus delays and car capacities. Finding that deploying bus priority schemes and pre-signals concurrently can not only reduce bus delays, but also increase car capacities.

3. Continuous approximation models to design hybrid transit networks for circular and grid cities. Strategies that can achieve sub-optimal close to global-optimal for these models.

4. Typical optimal network patterns, which are illustrated with the CA models. A hybrid network structure is preferred for both circular and grid cities. Comparisons of the cost
of circular and grid cities reveal that circular cities are always more transit friendly than grid cities (i.e. ring and radial street networks are better for bus systems than are grids).

5. The cost of transit networks can be reduced by allowing transit line spacing and headway to vary in space.

To be sure, all of the models we present in this thesis are idealized. Yet these models represent a step toward a better understanding of transit systems at the micro and macro levels.

5.2 Future Work

A long list of research opportunities can build upon the present work. Some of these opportunities are summarized below:

1. Micro level transportation simulation models can be built to reveal the benefits of the bus signal priority schemes proposed in this study.

2. An adaptive transit signal priority strategy can be researched as a way to further decrease bus delays.

3. Research can explore the possibility of different transit modes being combined into a single transit system. For instance, it might be possible to use the BRT/metro as the transit mode for the radial line system, and use normal bus/ flexible bus as the feeder mode for the ring line system.

4. The addition of time variations would add one more dimension to the transit system. Thus, the value of the headway function could vary with time for a short period (i.e. on a daily basis). The value of both the headway function and spacing function could vary with time over a long period (i.e. on a yearly basis).
Appendix A

A List of Notations for Chapter 2

Table A.1: Notations for Chapter 2 (part (a))

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>Cycle length of the main signal.</td>
</tr>
<tr>
<td>$G$</td>
<td>Green duration of the main signal for the subject approach, including a through-moving phase and a protected left-turning phase.</td>
</tr>
<tr>
<td>$G_L, G_T$</td>
<td>Durations of main signal green phases for left-turning and through-moving traffic, respectively.</td>
</tr>
<tr>
<td>$R_L, R_T$</td>
<td>Durations of main signal red phases.</td>
</tr>
<tr>
<td>$g_L, g_T$</td>
<td>Durations of pre-signal green phases for left-turning and through-moving traffic, respectively.</td>
</tr>
<tr>
<td>$N$</td>
<td>Total number of car lanes at the main signal for left-turning and through-moving traffic.</td>
</tr>
<tr>
<td>$N_L, N_T$</td>
<td>Numbers of left-turning and through-moving car lanes at the main signal, respectively.</td>
</tr>
<tr>
<td>$n$</td>
<td>Total number of car lanes at the pre-signal for left-turning and through-moving traffic.</td>
</tr>
<tr>
<td>$n_L, n_T$</td>
<td>Numbers of left-turning and through-moving car lanes at the pre-signal, respectively.</td>
</tr>
<tr>
<td>$d$</td>
<td>Distance between the pre-signal and the main one.</td>
</tr>
<tr>
<td>$d_{min}$</td>
<td>Minimum required distance between the pre-signal and the main one.</td>
</tr>
</tbody>
</table>
Table A.2: Notations for Chapter 2 (part (b))

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l$</td>
<td>Left-turning ratio.</td>
</tr>
<tr>
<td>$t_y$</td>
<td>Yellow phase duration.</td>
</tr>
<tr>
<td>$t_m$</td>
<td>Maximum extended green time.</td>
</tr>
<tr>
<td>$t_a$</td>
<td>Predicted bus arrival time to the intersection.</td>
</tr>
<tr>
<td>$t_L$</td>
<td>The green time loss for through-moving traffic in a cycle due to bus signal priority.</td>
</tr>
<tr>
<td>$t_e$</td>
<td>The extra green time added to a $G_T$ phase to accommodate the arriving bus(es).</td>
</tr>
<tr>
<td>$t_R$</td>
<td>The redundant green time of a $G_T$ phase.</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of buses arriving between time 0 and $t_m$, which will enjoy signal priority.</td>
</tr>
<tr>
<td>$q_s$</td>
<td>Capacity of a car lane in the subject approach.</td>
</tr>
<tr>
<td>$v_f$</td>
<td>Free-flow speed of vehicles.</td>
</tr>
<tr>
<td>$w$</td>
<td>Backward wave speed of the subject approach.</td>
</tr>
<tr>
<td>$q_L, q_T$</td>
<td>Saturation flows of the left-turning and through-moving traffic at the pre-signal, respectively.</td>
</tr>
<tr>
<td>$w_L, w_T$</td>
<td>The shock wave speeds of the tail of a jammed queue for left-turning and through-moving traffic, respectively, in the sorting area.</td>
</tr>
<tr>
<td>$Q_M(N)$</td>
<td>Maximum vehicle capacity that can be achieved at an intersection with $N$ lanes and no pre-signal.</td>
</tr>
<tr>
<td>$Q_P(N, n)$</td>
<td>Maximum vehicle capacity of an intersection with $N$ lanes at the main signal and $n$ lanes at the pre-signal.</td>
</tr>
<tr>
<td>$E[W_{B,i}]$</td>
<td>Expected bus delay in Scenario $i (i = 1, 2, 3, 4)$.</td>
</tr>
<tr>
<td>$E[Q_{T,i}]$</td>
<td>Expected through-moving car-carrying capacity in Scenario $i (i = 1, 2, 3, 4)$.</td>
</tr>
<tr>
<td>$q_A$</td>
<td>Vehicle inflow per through-moving lane in Baseline Scenario 1.</td>
</tr>
<tr>
<td>$\lambda_b$</td>
<td>Bus arrival rate.</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Value of PCE of a bus.</td>
</tr>
</tbody>
</table>
Appendix B

Derivation of the Minimum Length of Sorting Area

We develop an upper bound of the worst-case queue length of cars in the sorting area. Without loss of generality, we consider the case shown in Figure 2.3b, i.e., when a through-moving car queue is followed by a small left-turning queue. The maximum possible length of a through-moving queue can be calculated by geometry: \( \frac{g_{T,n}}{N} \left( \frac{1}{w} + \frac{1}{v_f} \right) \). To calculate the length of the left-turning queue, we first find in turn that: 

\[
t_1 = g_T + g_L - (G_T + R_L + t_y + G_L);
\]

and 

\[
t_2 = t_1 - \frac{g_T(N-n_T)}{N} = g_T + g_L - (G_T + R_L + t_y + G_L) \text{ (see Figure 2.3b)}.
\]

Thus, the length of the left-turning queue is 

\[
n_{L,N} - \max \left\{ \frac{g_T n_T}{N-n_T} + g_L - (G_T + R_L + G_L + 2t_y), 0 \right\} \left( \frac{1}{w} + \frac{1}{v_f} \right).
\]

The total queue length of this worst case is 

\[
\left( \frac{g_{T,n}}{N} + \frac{n_L}{N-n_T} \max \left\{ \frac{g_T n_T}{N-n_T} + g_L - (G_T + R_L + G_L + 2t_y), 0 \right\} \right) \left( \frac{1}{w} + \frac{1}{v_f} \right).
\]

By symmetry, the queue length when a left-turning queue is followed by a small through-moving queue is 

\[
\left( \frac{g_{L,n}}{N} + \frac{n_T}{N-n_T} \max \left\{ \frac{g_L n_L}{N-n_L} + g_T - (G_L + R_T + G_T + 2t_y), 0 \right\} \right) \left( \frac{1}{w} + \frac{1}{v_f} \right).
\]

Combining these two cases, we have equation (2.4).

As a final note, it can be shown (but is skipped here for brevity) that the signal alterations of the main and pre-signal under the priority scheme will not further increase the maximum vehicle queue length.

\[\text{Note that this upper bound is not always tight, but the tight upper bound would have a much more complicated analytical form.}\]
Appendix C

Derivation of Equation (2.12)

We plot the queueing diagrams of through-moving vehicle counts at the main signal in Figures C1(a)-(c) for three cases. In each figure, the dashed line with the slope $q_{AT}$ represents the virtual vehicle counts should neither signal be present. The dotted curve represents the virtual vehicle counts should only the pre-signal be present, where the slope is $q_{SN}$ at the beginning of the $g_T$ phase, $q_{AN}$ after queued vehicles have all discharged, and 0 otherwise. The solid curve represents the actual vehicle counts, where the slope is $q_{SN}$ at the beginning of the $G_T$ phase, and reduces to $q_{AN}$ when queued vehicles in the sorting area have all discharged. The offset between the start times of $g_T$ and $G_T$ is $g_T - G_T + \max\{0, g_L - (R_L + G_L + t_y)\}$; see Figures C1(a) and (b). The shaded area represents the total vehicle delay per cycle.

The case of Figure C1(a) occurs when $q_{SN}(G_T - \max\{0, g_L - (R_L + G_L + t_y)\}) < q_{AN}T$. By geometry, we find that the trapezoidal area is $\left(\frac{q_{AN}T}{2q_{SN}} + \frac{T}{2} - G_T + \max\{0, g_L - (R_L + G_L + t_y)\}\right)q_{AN}T$. Thus the average vehicle delay is $\frac{q_{AN}N}{2q_{SN}}T - G_T + \max\{0, g_L - (R_L + G_L + t_y)\}$.

The second case (Figure C1(b)) occurs when $q_{SN}(G_T - \max\{0, g_L - (R_L + G_L + t_y)\}) \geq q_{AN}T$. Again by geometry, we find the average vehicle delay is $\frac{q_{SN}}{2T(q_{SN} - q_{AN}T)}(T - G_T + \max\{0, g_L - (R_L + G_L + t_y)\})^2$.

The delay formula for the third case (Figure C1(c)) is more complicated, but it can be approximated by the above formula for the second case. Thus, we combine this case with the second case.

Finally, using $q_{AN} = Q_{T,2} + \lambda_b\delta$, we have Equation (2.12).
Figure C.1: Queueing diagram of through-moving vehicles in Baseline Scenario 2
Appendix D

Calculation of $E[t_L]$

Let $t_e$ be the extra green time added to a $G_T$ phase to accommodate the arriving bus(es) ($0 \leq t_e \leq t_m$). This is also the time truncated from the next $G_T$ phase. Since each $G_T$ includes a redundancy of $t_R = G_T \frac{q_t^{gr}}{N_T}$, the actual green time loss for through-moving traffic is: $t_L = \max\{0, t_e - t_R\}$.

Since $t_e \leq t_m$, if $t_R \geq t_m$, there will be no capacity loss and $t_L = 0$. Next we consider the case where $t_R < t_m$. We first find the distribution of $t_e$.

Suppose there are $M$ buses arriving to the intersection between 0 and $t_m$. The $M$ follows a Poisson distribution with mean $\lambda_b t_m$. When $M = 0, t_e = 0$. When $M \geq 1$, we have:

$t_e|M = \max_{1 \leq i \leq M} \{t_{ai}\}$.

It has been shown that the rearrangement of $t_{a1}, t_{a2}, ..., t_{aM}$ in the ascending order is statistically equivalent to the order statistics of $M$ random variables that are uniformly distributed in $[0, t_m]$ ([41]). Thus, $t_e$ for a given $M$ is equivalent to the $M^{th}$ order statistic of those uniformly-distributed random variables. We have the cumulative distribution function of $t_e$ given $M$: $F_{t_e|M}(s) = \Pr(t_e \leq s|M) = \Pr(t_{a1}, ..., t_{aM} \leq s|M) = \left(\frac{s}{t_m}\right)^M$. Thus, the probability density function of $t_e|M$ is: $f_{t_e|M}(s) = M \frac{s^{M-1}}{t_m^M}, m \geq 1$. Thus, we have:
APPENDIX D. CALCULATION OF $E[t_L]$

$$E[t_L] = E[E[t_L|M]] = \sum_{i=0}^{\infty} \Pr(M = i)E[t_L|M = i]$$

$$= \sum_{i=1}^{\infty} \frac{(\lambda_b t_m)^i}{i!} e^{-\lambda_b t_m} \int_{s=t_R}^{t_m} f_{t_L|M=i}(s)(s - t_R)ds$$

$$= \sum_{i=1}^{\infty} \frac{(\lambda_b t_m)^i}{i!} e^{-\lambda_b t_m} \int_{s=t_R}^{t_m} \frac{s^{i-1}}{t_m^{i}}(s - t_R)ds$$

$$= \sum_{i=1}^{\infty} \frac{(\lambda_b t_m)^i}{i!} e^{-\lambda_b t_m} \left( t_m - t_R + \frac{1}{i+1} \left( \frac{t_R^{i+1}}{t_m^i} - t_m \right) \right)$$

$$= (t_m - t_R)(1 - e^{-\lambda_b t_m}) + \sum_{i=1}^{\infty} \frac{(\lambda_b t_m)^i}{(i+1)!} e^{-\lambda_b t_m} \left( \frac{t_R^{i+1}}{t_m^i} - t_m \right)$$

$$= (t_m - t_R)(1 - e^{-\lambda_b t_m}) + \frac{1}{\lambda_b} \left( \sum_{i=1}^{\infty} \frac{(\lambda_b t_R)^{i+1}}{(i+1)!} e^{-\lambda_b t_m} - \sum_{i=1}^{\infty} \frac{(\lambda_b t_m)^{i+1}}{(i+1)!} e^{-\lambda_b t_m} \right)$$

$$= (t_m - t_R)(1 - e^{-\lambda_b t_m}) + \frac{1}{\lambda_b} \left( e^{-\lambda_b (t_m-t_R)}(1 - e^{-\lambda_b t_R} - \lambda_b t_R e^{-\lambda_b t_R}) - (1 - e^{-\lambda_b t_m} - \lambda_b t_m e^{-\lambda_b t_m}) \right)$$

$$= t_m - t_R + \frac{1}{\lambda_R} \left( e^{-\lambda_b (t_m-t_R)} - 1 \right)$$

Therefore, we get: $E[t_L] = \max\{0, t_m - G_T + \frac{g_T n_T}{N-1} + \frac{1}{\lambda_b} \left( e^{-\lambda_b \max\{0, t_m-G_T+\frac{g_T n_T}{N-1}\}} - 1 \right)\}$
Appendix E

Model 1 for Ring-radial Networks

We derive each of the components for Model 1. The notation is as follows.

Table E.1: Notations for ring-radial networks

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>Radius of the city’s central district</td>
</tr>
<tr>
<td>$R$</td>
<td>Radius of the city boundary</td>
</tr>
<tr>
<td>$x$</td>
<td>Distance to the city center</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Angle between the origin and destination in polar coordinates centered at the city center</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>Travel demand for the entire city (p/h) in the rush hour</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Vehicle lost time at each station due to acceleration and deceleration</td>
</tr>
<tr>
<td>$\tau'$</td>
<td>Boarding time per passenger</td>
</tr>
<tr>
<td>$C_{pax}$</td>
<td>Vehicle passenger-carrying capacity</td>
</tr>
<tr>
<td>$s$</td>
<td>Station spacing throughout the city</td>
</tr>
<tr>
<td>$v_B$</td>
<td>Cruising speed on the outermost (boundary) ring line</td>
</tr>
<tr>
<td>$s_c(x)$</td>
<td>Ring line spacing at distance $x$</td>
</tr>
<tr>
<td>$s_r(x)$</td>
<td>Radial line spacing at $x$</td>
</tr>
<tr>
<td>$v_{cc}(x)$</td>
<td>Commercial speed on ring line at $x$</td>
</tr>
<tr>
<td>$v_{cr}(x)$</td>
<td>Commercial speed on radial line at $x$</td>
</tr>
<tr>
<td>$v_c(x)$</td>
<td>Cruising speed on ring line at $x$</td>
</tr>
<tr>
<td>$v_r(x)$</td>
<td>Cruising speed on radial line at $x$</td>
</tr>
<tr>
<td>$H_c(x)$</td>
<td>Headway on ring line at $x$</td>
</tr>
<tr>
<td>$H_r(x)$</td>
<td>Headway on radial line at $x$</td>
</tr>
<tr>
<td>$P_o(x)$</td>
<td>Probability that the origin lies within rings of radii $(x, x + dx)$</td>
</tr>
<tr>
<td>$P_d(x)$</td>
<td>Probability that the destination lies within rings of radii $(x, x + dx)$</td>
</tr>
<tr>
<td>$O(x)$</td>
<td>Maximum expected number of passengers in a vehicle serving the ring at $x$</td>
</tr>
<tr>
<td>$Q$</td>
<td>Total flow of radial-line buses across every ring</td>
</tr>
</tbody>
</table>
APPENDIX E. MODEL 1 FOR RING-RADIAL NETWORKS

Local Cost for the Length of the Transit Lines: \( y_L(x) \)

Consider a pair of rings of radii \((x, x+dx)\). The area between those rings is \(2\pi x dx\), and contains \(\frac{2\pi x}{s_r(x)}\) radial lines, each of length \(dx\). Therefore, the length of the network’s radial lines within the ring pair is \(\frac{2\pi x}{s_r(x)}\) \(dx\). The area similarly contains \(\frac{2\pi x}{s_c(x)}\) ring lines, each of length \(2\pi x\). The length of the ring lines is therefore \(\frac{2\pi x}{s_c(x)}\) \(dx\). The local cost of that length, \(y_L(x)\), is the length of the transit lines in the area divided by the area’s width \(dx\); that is:

\[
y_L(x) = \frac{\frac{2\pi x}{s_r(x)} + \frac{2\pi x}{s_c(x)}}{2\pi x} \text{ in the central district, and } \frac{2\pi x}{s_r(x)} \text{ in the periphery.}
\]

Local Cost for the Vehicle-distance Traveled per Hour: \( y_V(x) \)

Consider again the ring lines of area \(2\pi x dx\). The \(y_V(x)\) is obtained by multiplying the local cost of ring-line and radial-line distances, \(\frac{2\pi x}{s_c(x)}\) and \(\frac{2\pi x}{s_r(x)}\), with their corresponding transit flows, \(\frac{1}{H_c(x)}\) and \(\frac{1}{H_r(x)}\), and multiplying by the factor 2 to account for the bi-directional travel on each line. Thus, the local cost for the vehicle-distance traveled per hour \(y_V(x)\) is:

\[
y_V(x) = \frac{4\pi x}{s_r(x)H_r(x)} + \frac{4\pi x}{s_c(x)H_c(x)} \text{ in the central district; and } \frac{4\pi x}{s_r(x)H_r(x)} \text{ in the periphery.}
\]

Global Cost for the Vehicle-distance Traveled per Hour: \( F_V \)

For the global cost, we consider only the cost on the outermost (boundary) ring line. Thus, the vehicle-distance traveled per hour on that ring is: \(\frac{4\pi r}{H_B}\).

Local Cost for the Fleet Size during the Rush: \( y_M(x) \)

The \(y_M(x)\) is obtained from the result of E.2 and the vehicles’ commercial speed at distance \(x\). Thus, \(y_M(x)\) is:

\[
y_M(x) = \frac{4\pi x}{s_r(x)H_r(x)v_{cr}} + \frac{4\pi x}{s_c(x)H_c(x)v_{cc}} \text{ in the central district; and } \frac{4\pi x}{s_r(x)H_r(x)v_{cr}} \text{ in the periphery.}
\]

Global Cost for the Fleet Size during the Rush: \( F_M \)

Much as in E.4, \(F_M\) is obtained from the result of E.3 and the vehicle’s commercial speed at the outermost (boundary) ring. Thus, \(F_M\) is:

\[
F_M = \frac{4\pi r}{H_Bv_{cb}}.
\]

Local Cost for the Patron’s Average Waiting Time: \( y_W(x) \)

We divide the problem into six cases, as shown in Fig. E1. For the first case, the trip’s origin and destination both lie in the periphery and \(\theta < 2\), see case 1 in the figure. In this case, a patron takes a radial line, transfers at the outermost ring and transfers again to another radial line to reach her destination. The patron would thus encounter an expected waiting time \(\left(\frac{H_r(x)}{2} + \frac{H_r(y)}{2}\right)\) at radial-line stations. For calculating local cost, we do not consider
APPENDIX E. MODEL 1 FOR RING-RADIAL NETWORKS

Figure E.1: cases for average waiting time per trip

the waiting time at the outermost ring (This is part of the global cost). The local cost for this case is therefore: 

$$(P_o(x) \int_r^R P_d(y)dy + P_d(x) \int_r^R P_o(y)dy) \frac{2}{\pi} H_r(x)$$

for $r < x < R$.

For the second case, the origin and destination again lie in the periphery, but $\theta > 2$, see case 2 in Fig. E1. The patron in this case travels by two radial lines and transfers at the city center as shown in the figure, such that the local cost for the waiting time is:

$$(P_o(x) \int_r^R P_d(y)dy + P_d(x) \int_r^R P_o(y)dy)(1 - \frac{2}{\pi} H_r(x))$$

for $r < x < R$.

For the third case, the origin and destination both lie in the central district (i.e. $0 < x < r$) and $\theta < 2$; see case 3 in Fig. E1. The patron travels by ring and radial lines. Her local cost for waiting time is:

$$\frac{H_r(x)}{2} P_o(x) \frac{2}{\pi} \int_x^r P_d(y)dy + \frac{H_r(x)}{2} P_d(x) \frac{2}{\pi} \int_x^r P_o(y)dy$$

if $x < y$ and
Global cost for the patron’s average waiting time: $F_W$

For global cost, we only consider the cost happened at the outermost ring line. Thus, the trip’s origin and destination both lie in the periphery area and $\theta < 2$. and the waiting time at the outermost ring line (i.e. $F_W$) is: 

$$F_W = \int_r^R P_o(x)dx \int_r^R P_d(y)dy \frac{2 H_s(x)}{\pi}$$

**Local Cost for Access Time: $y_A(x)$**

Transfer stations reside at all intersections of ring and radial lines, and intermediate, non-transfer stations are spaced at $s$ over the network; see Fig. E2. We use $\frac{s_c}{s+s_c}$ as the average station spacing along the rings, and $\frac{s_r}{s+s_r}$ along the radials. These continuum approximations are reasonable when $s_c$ and $s_r$ are both larger than $s$. Thus, a patron who originates within a

---

The fourth case is like the third, except that $\theta > 2$; see case 4 in Fig. E1. The local cost for this case is: 

$$(P_o(x)dx \int_0^r P_d(y)dy + P_d(x)\int_0^r P_o(y)dy)(1 - \frac{2}{\pi}) \frac{H_s(x)}{2}$$

for $0 < x < r$.

For the fifth case, the origin is in the central district, the destination is in the periphery (or vise versa), and $\theta < 2$, see case 5 in Fig. E1. The local costs for waiting time are: 

$$(P_d(x)\int_0^r P_o(y)dy + P_o(x)\int_0^r P_d(y)dy) \frac{H_s(x)}{2}$$

for $r < x < R$; and 

$$(P_o(x)\int_r^R P_d(y)dy + P_d(x)\int_r^R P_o(y)dy) \frac{H_s(x)}{2}$$

for $0 < x < r$.

The sixth case is like the fifth, except that $\theta > 2$; see case 6 in Fig. E1. The local costs for waiting time are: 

$$(P_d(x)\int_0^r P_o(y)dy + P_o(x)\int_0^r P_d(y)dy)(1 - \frac{2}{\pi}) \frac{H_s(x)}{2}$$

for $r < x < R$; and 

$$(P_o(x)\int_r^R P_d(y)dy + P_d(x)\int_r^R P_o(y)dy) \frac{H_s(x)}{2}$$

for $0 < x < r$.

The sum of the six cases above is the local cost for the waiting time per trip $y_W(x)$:

$$\frac{H_s(x)}{\pi} P_o(x) \left(\int_x^R P_d(y)dy\right) + \frac{H_s(x)}{\pi} P_d(x) \left(\int_x^R P_o(y)dy\right) + \frac{H_s(x)}{\pi} P_o(x) \left(\int_0^x P_d(y)dy\right) + \frac{H_s(x)}{\pi} P_d(x) \left(\int_0^x P_o(y)dy\right) + \left(1 - \frac{2}{\pi}\right) \frac{H_s(x)}{2} P_o(x) + \left(1 - \frac{2}{\pi}\right) \frac{H_s(x)}{2} P_d(x)$$

for $0 < x < r$; and 

$$\left(P_o(x) + P_d(x)\right)$$

for $r < x < R$. 

---

**Figure E.2: transit stations in the central district**

\[
\frac{H_s(x)}{\pi} P_o(x) \left(\int_x^R P_d(y)dy\right) + \frac{H_s(x)}{\pi} P_d(x) \left(\int_x^R P_o(y)dy\right) + \frac{H_s(x)}{\pi} P_o(x) \left(\int_0^x P_d(y)dy\right) + \frac{H_s(x)}{\pi} P_d(x) \left(\int_0^x P_o(y)dy\right) + \left(1 - \frac{2}{\pi}\right) \frac{H_s(x)}{2} P_o(x) + \left(1 - \frac{2}{\pi}\right) \frac{H_s(x)}{2} P_d(x) \text{ for } 0 < x < r; \text{ and } \\
\left(P_o(x) + P_d(x)\right) \text{ for } r < x < R.
\]
central district is expected to travel distance $\frac{1}{4} \left( s_r + \frac{s_{ss}}{x + s_r} \right)$ to access a radial line, and distance $\frac{1}{4} \left( s_r + \frac{s_{ss}}{s + s_r} \right)$ to access a radial line. Average walking distance is $\frac{1}{4} (s_r + s)$ in the periphery.

We divide the problem into six cases, as shown in Fig. E2. In the first case: the trip’s origin and destination both lie in the central district; the origin lies closer to the city center than does the destination; and $\theta < 2$; see case 1 in the figure. A patron takes a ring line, and transfers to a radial line to reach her destination. The access time from her origin to the nearest station on the ring line is: $\frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right)$, and the access time from the nearest stop on the radial line to the destination is: $\frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right)$. The local cost for this case is therefore: $P_o(x) \int_x^r P_d(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right) dy + P_d(x) \int_0^x P_o(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy$ for $0 < x < r$.

The second case is like the first, except that the destination is closer to the city center than is the origin; see case 2 in Fig. E2. The local cost for this case is: $P_o(x) \int_0^x P_d(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy + P_d(x) \int_x^r P_o(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right) dy$ for $0 < x < r$.

For the third case, the origin and destination again lie in the central district, but $\theta > 2$; see case 3 in Fig. E2. The patron in this case, travels by two radial lines as shown in the figure, such that the local cost for access time is: $P_o(x) \int_0^x P_d(y) \left( 1 - \frac{2}{\pi} \right) \frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy$ for $0 < x < r$, where the first term is the expected access time from the origin (located at $x$) to the radial line, and the second term is the expected access time from the radial line to the destination.

For the fourth case, the origin is in central district, the destination is in the periphery (or vice-versa), and $\theta < 2$; see case 4 in Fig. E2. The patron in this case travels along the ring line, and transfers to the radial line that leads to her the destination. The local cost for access time at the origin (or destination) is therefore either: $P_o(x) \int_x^r P_d(y) dy \left[ \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right) \right]$, or $P_d(x) \int_x^r P_o(y) dy \left[ \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) \right]$ for $0 < x < r$.

The fifth case is like the fourth, except that $\theta > 2$; see case 5 in Fig. E2. This patron travels only by radial lines, and the local cost for access time in the central district is:

$\left( P_d(x) \int_x^r P_o(y) dy + P_o(x) \int_x^r P_d(y) dy \right) \left[ \left( 1 - \frac{2}{\pi} \right) \frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) \right]$ for $0 < x < r$.

For convenience, the access times that occur in the peripheries were thus far ignored for cases 4 and 5. Those peripheral access times are instead accounted for in the sixth case; see case 6 in Fig. E2. The local cost is $(P_o(x) + P_d(x)) \left( \frac{1}{4} s_r(x) + \frac{1}{4} s \right) \frac{1}{w}$ for $r < x < R$.

The sum of the six cases above is the entire local cost for access time, $y_A(x)$: $P_o(x) \int_x^r P_d(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right) dy + P_o(x) \int_0^x P_d(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy + P_d(x) \int_x^r P_o(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_c(x) \right) dy + P_d(x) \int_0^x P_o(y) \frac{2}{\pi w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy + P_d(x) \int_x^r P_o(y) \left( 1 - \frac{2}{\pi} \right) \frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy + P_d(x) \int_0^x P_o(y) \left( 1 - \frac{2}{\pi} \right) \frac{1}{w} \left( \frac{s_{ss}(x)}{4 \pi s + s_r(x)} + \frac{1}{4} s_r(x) \right) dy$. It seems there is a missing term that needs to be added for the complete expression.
Figure E.3: Cases for average access time per trip

\[
\begin{align*}
(P_d(x) \int_x^R P_o(y)dy + P_o(x) \int_x^R P_d(y)dy) & \left[ \frac{2}{\pi w} \left( \frac{ss_c(x)}{4s+sr(x)} + \frac{s_c(x)}{4} \right) \right] \\
& \left( 1 - \frac{2}{\pi} \right) \frac{1}{w} \left( \frac{ss_c(x)}{4s+sr(x)} + \frac{s_c(x)}{4} \right)
\end{align*}
\]
for \(0 < x < r\); and \((P_o(x) + P_d(x)) \left( \frac{1}{4} sr(x) + \frac{1}{4} s \right) \frac{1}{w}\), for \(r < x < R\).

**Ring-line Commercial Speed:** \(v_{cc}(x) = \frac{1}{\left( \frac{1}{v_c(x)} + \tau \left( \frac{1}{s} + \frac{1}{sr(x)} \right) \right)} + \tau' \Lambda \left[ P_o(x) \int_x^R P_d(y)dy + P_d(x) \int_x^R P_o(y)dy \right] \frac{s_c(x)H_c(x)}{2\pi^2 x} \)

The commercial speed in the central district, \(v_{cc}(x)\), is determined by: 1) the cruising time per unit distance, \(\frac{1}{v_c(x)}\); 2) the time lost at stations per unit distance due to acceleration and deceleration, \(\tau \left( \frac{1}{s} + \frac{1}{sr(x)} \right)\); and 3) the time spent boarding patrons per unit distance, \(\tau' \Lambda \left[ P_o(x) \int_x^R P_d(y)dy + P_d(x) \int_x^R P_o(y)dy \right] \frac{s_c(x)H_c(x)}{2\pi^2 x}\).

When calculating commercial speed (as well as the maximum expected number of passengers in a vehicle), the travel demand during the rush, \(\Lambda\), is used to account for the worst
case condition. We fix $\Lambda$ to be 2.5 times the average day-long travel demand.

### Commercial Speed on Outermost Boundary Ring:

$$v_{cB} = 1/(\frac{1}{v_B} + \tau(\frac{1}{s} + \frac{1}{s_r(r)}) + \tau'\Lambda[\int_r^R P_o(y)dy \int_y^R P_d(y)dy] H_B^{r^2/2\pi^2r)}$$

Much as in E.9 and E.10, the commercial speed on the boundary route has three components: 1) cruising time, $\frac{1}{v_B}$; 2) lost time at transit stations, $\tau(\frac{1}{s} + \frac{1}{s_r(r)})$; and 3) passenger boarding time, $\tau'\Lambda[\int_r^R P_o(y)dy \int_y^R P_d(y)dy] H_B^{r^2/2\pi^2r}$.

### Local Cost for Expected In-vehicle Travel Time: $y_T(x)$

Consider a pair of rings of radii $(x, x + dx)$, as shown by the grey-shaded swath in Fig. E4a. We obtain the in-vehicle travel time on that swath weighted by the proportion of patrons who travel along it and who cross it. We do this for the swaths at each $x < R$, and divide the in-vehicle travel time on that swath by the width of the swath, $dx$, to obtain $z_T(x)$. We do not consider the in-vehicle travel time on the outermost boundary ring line here. It is instead included in the global cost for in-vehicle travel time, $F_r$.

The problem is divided into three cases. The first of these pertains to patrons who use only radial lines (i.e. $\theta > 2$), as exemplified in Fig. E4a. Note from the two example trips in that figure how patrons might cross a swath once or twice. The time spent crossing a swath at distance $x$ is $\frac{dx}{v_{cr}(x)}$. Multiplying by the proportion of patrons who make the trip, we obtain:

$$(1 - \frac{2}{\pi})\left[\int_x^R P_o(y)dy \int_y^R P_d(y)dy + \int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right] \frac{1}{v_{cr}(x)}$$

for $0 < x < R$.

The second case pertains to patrons who use radial lines and the outermost boundary ring, as shown in Fig. E4b. Much as in the previous case, patrons on radial lines will cross ring-shaped swaths once or twice. Hence, the weighted in-vehicle travel time spent at $x$ except for the travel time on the boundary ring is:

$$\frac{2}{\pi} \int_x^R P_o(y)dy \int_y^R P_d(y)dy \frac{2}{v_{cr}(x)} + \frac{2}{\pi} \left(\int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right) \frac{1}{v_{cr}(x)}$$

for $r < x < R$.

The third case pertains to patrons whose origins and/or whose destinations lie in the central district, as exemplified for one case in Fig. E4c. The in-vehicle time spent on a ring at $x$ is: $(P_d(x) \int_x^R P_o(y)dy) \frac{2x}{\pi} \frac{1}{v_{cr}(x)}$. These patrons cross a ring-shaped swath no more than once. Thus, the local cost for in-vehicle time for this case is:

$$\frac{2}{\pi} \left(\int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right) \frac{1}{v_{cr}(x)}$$

for $0 < x < r$.

The sum of the three cases is the local cost for expected in-vehicle travel time, $y_T(x)$:

$$(1 - \frac{2}{\pi})\left[\int_x^R P_o(y)dy \int_y^R P_d(y)dy + \int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right] \frac{1}{v_{cr}(x)} +$$

$$\frac{2}{\pi} \left(\int_x^R P_o(y)dy \int_0^x P_d(y)dy + \frac{2}{\pi} \left(\int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right) \frac{1}{v_{cr}(x)}\right)$$

for $r < x < R$; and

$$(1 - \frac{2}{\pi})\left[\int_x^R P_o(y)dy \int_y^R P_d(y)dy + \int_x^R P_o(y)dy \int_0^x P_d(y)dy + \int_x^R P_d(y)dy \int_0^x P_o(y)dy\right] \frac{1}{v_{cr}(x)}$$

for $x$.
APPENDIX E. MODEL 1 FOR RING-RADIAL NETWORKS

(a) the first case for expected in-vehicle travel time

(b) the second case for expected in-vehicle travel time

(c) the third case for expected in-vehicle travel time

Figure E.4: In-vehicle travel time

\[ \int_0^x P_o(y)dy \frac{1}{v_{cr}(x)} + \frac{2}{\pi} \left( \int_0^R P_o(y)dy \int_0^x P_d(y)dy + \int_0^R P_d(y)dy \int_0^x P_o(y)dy \right) \frac{1}{v_{cr}(x)} + \]
\[ (P_o(x) \int_0^R P_d(y)dy + P_d(x) \int_0^R P_o(y)dy) \frac{2x}{\pi} \frac{1}{v_{cz}(x)} \] for \( 0 < x < r \).

Global Cost for Expected In-vehicle Travel Time: \( F_T \)

For global cost, we consider only the cost at the outermost boundary ring. Thus, the trip’s origin and destination both lie in the periphery and \( \theta < 2 \). The in-vehicle travel time at the boundary ring, \( F_T \), is:

\[ \int_r^R P_o(x)dx \int_r^R P_d(x)dx \frac{2r}{\pi} \frac{1}{v_{cz}}. \]

Global Cost for Expected Number of Transfers:

\( F_{eT} = 1 + \frac{2}{\pi} \int_r^R P_o(x)dy \int_r^R P_d(x)dy \)

All patrons must transfer at least once each trip. Those whose origins and destinations both lie in the periphery must transfer twice, and the added expectation is given in the second term of the above expression.
Vehicles Passenger-carrying Capacity Constraint except on Boundary Ring

The expected maximum number of patrons on board a vehicle should be constrained to be less than the vehicle’s passenger-carrying capacity. The total number onboard all vehicles on a ring line at distance \( x \) is: \( \Lambda(\int_P(x) \int_P P_d(y)dy + P_d(x) \int_P P_o(y)dy) \frac{2}{\pi} \). The flow of transit vehicles on that ring is: \( \frac{2}{(s_c(x)H_c(x))} \). And the ratio of the average trip length to the length of the ring line is: \( \frac{\pi}{2\pi x} \). Hence, the expected maximum on that ring line is:

\[
\Lambda(\int_P(x) \int_P P_d(y)dy + P_d(x) \int_P P_o(y)dy) \frac{2}{\pi} \frac{1}{2/(s_c(x)H_c(x))} \frac{x}{2\pi x}.
\]

In similar fashion, the expected maximum number onboard a radial line at \( x \) is:

\[
\Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \]

in the inbound direction, and

\[
\Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \]

outbound.

Thus, the maximum of expected number of passengers in a vehicle at \( x \), \( O(x) \), is:

\[
\max \left[ \Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \right] \Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \].

And the vehicle’s passenger-carrying capacity constraint (except at the boundary ring) is: \( O(x) - C_{pax} < 0 \).

Vehicle’s Passenger-carrying Capacity Constraint on Boundary Ring

Much like in the E.15, the total number onboard all vehicles on the outermost boundary ring is:

\[
\Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \]

And the ratio of the average trip length to the length of the ring line is: \( \frac{1}{2\pi} \). Hence, the expected maximum number of passengers in a vehicle on the boundary ring, \( O_B \), is:

\[
\Lambda(\int_P(x) \int_P P_d(y)dy \frac{1}{\pi} + \int_P P_0(y)dy(1 - \frac{2}{\pi})) \frac{1}{2\pi x/(s_r(x)H_r(x))} \frac{1}{2\pi} \].

And the vehicle’s passenger-carrying capacity constraint on that ring is: \( O_B - C_{pax} < 0 \).

Vehicle Conservation for Radial-line Buses

The vehicle conservation constraints are derived from Equation (3.1) and the total flow of radial-line buses across every ring is obtained from the radial-line bus flow at the outmost ring, i.e. \( Q = \frac{2\pi R}{s_r(R)H_r(R)} \). Thus, the vehicle conservation constraint for radial-line buses is:

\[
Q = \frac{2\pi x}{s_r(x)H_r(x)}, \forall x \in [0, R].
\]
Appendix F

Grid Transit Network Model

We begin by introducing some new terms not used in Appendix F.

Table F.1: Notations for grid networks

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Ratio of central district size to the city size, i.e. $\alpha = \frac{r}{R}$</td>
</tr>
<tr>
<td>$s_g$</td>
<td>Line spacing in the central district</td>
</tr>
<tr>
<td>$v_{ci}$</td>
<td>Commercial speed in the central district</td>
</tr>
<tr>
<td>$v_g$</td>
<td>Vehicle cruising speed in the central district</td>
</tr>
<tr>
<td>$H_g$</td>
<td>Headway in the central district</td>
</tr>
</tbody>
</table>

Local Cost for Length of the Transit Lines: $y_L(x)$
Since line spacing varies within the periphery, the length of transit lines at $x$ is: $4 \cdot \frac{2x}{s_r(x)^2} s_r(x) = \frac{8x}{s_r(x)}$. This holds where $x > r$ for $y_L$ and all $y$ below.

Global Cost for Length of the Transit Lines: $F_L$
Since line spacing is $s_g$ in the central district, the global cost for the total length of the transit lines in that district, $F_L$, is: $\frac{4r^2}{s_g^2} \frac{4s_g}{2} = \frac{8r^2}{s_g}$.

Local Cost for Vehicle-distance Traveled per Hour: $y_V(x)$
The average ratio of the total distance traveled to the perpendicular travel distance is $\frac{3}{2}$, as shown in [4]. Thus, the vehicle-distance traveled per hour at $x$ in the periphery (i.e. $y_V(x)$) is: $2 \cdot 4 \cdot \frac{2x}{s_r(x)H_r(x)} \frac{3}{2} = \frac{24x}{s_r(x)H_r(x)}$. 
APPENDIX F. GRID TRANSIT NETWORK MODEL

Global Cost for Vehicle-distance Traveled per Hour: $F_V$

Since headway in the central district is $H_g$, the vehicle-distance traveled per hour in that district is: $\frac{16r^2}{s_Hg}$. The vehicle-distance traveled per hour on the outermost boundary ring is: $2\frac{8r}{H_B}$. The sum of the two components is the global cost for vehicle-distance traveled per hour, $F_V$: $\frac{16r^2}{s_Hg} + \frac{16r}{H_B}$.

Local Cost for Fleet Size in the Rush: $y_M(x)$

As in Appendix E, fleet size is the ratio of the total distance traveled per hour to the commercial speed. Thus, $y_M(x)$ is: $\frac{24r}{sr(x)Hr(x)v_c(x)}$.

Global Cost for Fleet Size in the Rush: $F_M$

Much as in F.5, the global cost for fleet size in the rush, $F_M$, is: $\frac{16r^2}{s_Hg_vca} + \frac{16r}{H_Bv_cB}$.

Global Cost for Expected Number of Transfers per Trip: $F_{cT}$

$F_{cT} = 1 + \frac{1}{2}\left(\int _r P_o(y)dy \int _r P_d(y)dy\right)$

This expected value is estimated as in [4], except that we modify it to account for the possibility of non-uniformly distributed demand for travel.

Local Cost for Waiting Time: $y_W(x) = (P_o(x) + P_d(x))\frac{H_r(x)}{2}$

The local cost for waiting time includes the time waiting at the origin stop at $x$ (i.e. $P_o(x)\frac{H_r(x)}{2}$), and the time at the destination stop (i.e. $P_d(x)\frac{H_r(x)}{2}$).

Global Cost for Waiting Time:

$F_W = \int _0^{d/2} \left(P_o(x) + P_d(x)\right)\frac{H_g}{2}dx + \frac{1}{4}\left(\int _r P_o(y)dy \int _r P_d(y)dy\right)\left(\frac{H_g}{2} + \frac{H_B}{2}\right)$

The waiting time in the central area is: $\int _0^{d/2} \left(P_o(x) + P_d(x)\right)\frac{H_g}{2}dx + \frac{1}{4}\left(\int _r P_o(y)dy \int _r P_d(y)dy\right)\frac{H_g}{2}$.

The waiting time on the boundary route is $\frac{1}{4}\left(\int _r P_o(y)dy \int _r P_d(y)dy\right)\frac{H_B}{2}$. The sum is the global cost for waiting time.

Local Cost for Access Time: $y_A(x) = (P_o(x) + P_d(x))\left(\frac{sr(x)}{4w} + \frac{s}{4w}\right)$

The expected access distance in the periphery is $\frac{sr(x)}{4w} + \frac{s}{4w}$. This value, weighted by the travel demand, is the local cost for access time, $y_A(x)$.
APPENDIX F. GRID TRANSIT NETWORK MODEL

Global Cost for Access Time: \( F_A = (P_o(x) + P_d(x))(\frac{ss_g}{s+sg} + \frac{s_g}{4w}) \)

As in Appendix E, \( \frac{ss_g}{s+sg} \) is used as the average stop spacing in the central district. Thus, the expected access distance in that district is: \( \frac{ss_g}{s+sg} \cdot \frac{1}{4} + \frac{s_g}{4} \). The global cost for access time is obtained by weighting the latter value by the travel demand.

Local Cost for Expected In-vehicle Travel Time: \( y_T(x) \)

The local cost for in-vehicle travel time per trip, \( y_T(x) \), is: \( (P_o(x) + P_d(x))(x - r)^{\frac{3}{2}v_{cr}(x)} \), where the factor \( \frac{3}{2} \) is the same ratio used in F.3.

Global Cost for Expected In-vehicle Travel Time:

\[
F_T = 2R(\frac{\alpha}{12})\left[(11 - \alpha^2)(1 - \alpha^4) + 8\alpha^4 \right) \frac{\frac{1}{2}v_{ci}}{2\alpha^2} + \left( \int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{1}{3}3v_{cB} \]

The expected in-vehicle travel time per trip inside the central district is obtained by multiplying the result in \([4]\) for uniformly-distributed travel demand, \( 2R(\frac{\alpha}{12})\left[(11 - \alpha^2)(1 - \alpha^4) + 8\alpha^4 \right) \frac{\frac{1}{2}v_{ci}}{2\alpha^2} \). Hence, we find that

\[
v_{ci} = 1/(\frac{1}{v_q} + \tau(\frac{1}{s_g} + \frac{1}{s}) + \tau' \Lambda(\alpha^2 + 1 + \frac{1}{4}(\int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{H_s}{16\alpha^2})
\]

Commercial speed \( v_{ci}, v_{cB}, \) and \( v_{cr}(x) \)

The commercial speed in the central district, \( v_{ci} \), is determined by (i) cruising time, \( \frac{1}{v_q} \); (ii) the time lost at stations due to acceleration and deceleration, \( \tau(\frac{1}{s_g} + \frac{1}{s}) \); and (iii) the time spent boarding patrons, \( \tau' \Lambda(\alpha^2 + 1 + \frac{1}{4}(\int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{H_s}{16\alpha^2}) \). Hence, we find that

\[
v_{ci} = 1/(\frac{1}{v_q} + \tau(\frac{1}{s_g} + \frac{1}{s}) + \tau' \Lambda(\alpha^2 + 1 + \frac{1}{4}(\int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{H_s}{16\alpha^2})
\]

Commercial speed on the boundary route, \( v_{cB} \), is also dependent on the above three factors. Hence, \( v_{cB} = 1/(\frac{1}{v_B} + \tau(\frac{1}{s} + \frac{1}{s_{cr}}(x)) + \tau' \Lambda(\int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{H_B}{54\alpha^2}) \)

Commercial speed in the periphery, \( v_{cr}(x) \), is also dependent upon the above three factors and varies with \( x \). Hence, \( v_{cr}(x) = 1/(\frac{1}{v_{cr}(x)} + \tau(\frac{1}{s} + \tau' \Lambda(\frac{8x}{4R^2} H_g s_{cr}) \frac{H_{cr}}{24\alpha^2}) \).

Vehicle’s Passenger-carrying Capacity Constraint on Boundary Route

As in the Appendix E, the total number onboard all vehicles on the boundary line is: \( \Lambda \left( \int_{r}^{R} P_o(y)dy \int_{r}^{R} P_d(y)dy \right) \frac{1}{4} \). The flow of transit vehicles on that boundary is: \( \frac{2}{v_{cr}(x)} \). And the ratio of the average trip length to the length of the boundary line is: \( \frac{1}{12} \). Hence,
the expected maximum number of passengers in a vehicle on the boundary line, $O_B$, is:

$$\Lambda \left( \int_r^R P_o(y)dy \int_r^R P_d(y)dy \right)^{\frac{1}{4}} H^{\frac{1}{12}}. $$

And the vehicle’s passenger-carrying capacity constraint on the boundary is: $O_B - C_{pax} < 0$.

The passenger-carrying capacity constraints for the central district and the periphery are as in [4].

**Vehicle Conservation Constraint**

In the periphery, the vehicle conservation constraint is the same as in the ring-radial network, see appendix E.

In the central district, the total vehicle flow is $\frac{8r}{s_g H_y}$. Because all the buses travel to/from the city’s boundary, the vehicle conservation constraint in the central district is $Q = \frac{8r}{s_g H_y}$. 
Bibliography


