Design and Operational Analysis of Automated Guided Vehicle-Based Goods-to-Person Order Picking and Sortation Systems

by

Francisco J. Aldarondo Valle

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Industrial and Operations Engineering) in the University of Michigan 2019

Doctoral Committee:

Professor Yavuz A. Bozer, Chair
Professor Ravi Anupindi
Associate Professor Cong Shi
Professor Mark Van Oyen
**Dedication**

Much has changed over the course of the years that I dedicated to pursuing a PhD and completing the research presented here. I married my wonderful wife, my daughter was born, my grandfather passed, I developed a strong affinity for teaching, and I made many friends and connections that will enrich my life and career for years to come. Now, my attire is vetted before I walk out the door in the morning, unicorns are very present in my life, I am delegated the responsibility of reciting poems and songs to the up and coming generation, I will seek to champion research-based teaching methods, and I have many more cities and countries to address Christmas postcards to.

This work is dedicated to my loving family, for their constant attention, unwavering love, weekly video calls, and unconditional support.
Acknowledgements

Throughout the writing of this dissertation I received great support and advice. First, I would like to thank my advisor, Prof. Yavuz A. Bozer, whose expertise was invaluable in completing the research presented here. Likewise, I want to thank my dissertation committee for their valuable time and feedback.

I would like to acknowledge my colleagues at the Industrial and Operations Engineering department for their help and support. A special thanks to:

Donald Richardson,
Neil Fernandez,

and the sponsors that supported this work:

Industrial and Operations Engineering Department
Rackham Merit Fellowship
MHEFI Foundation Scholarship
Rackham Travel Grant
INFORMS Minority Forum Travel Grant
# Table of Contents

Dedication .......................................................................................................................... ii

Acknowledgements ............................................................................................................ iii

List of Tables ...................................................................................................................... v

List of Figures ................................................................................................................... vii

List of Appendices ............................................................................................................ x

Abstract ............................................................................................................................. xi

Chapter 1 Introduction ....................................................................................................... 1

Chapter 2 A Simulation-Based Comparison of Two Goods-to-Person Systems in an Online Retail Setting ............................................................................................................. 5

Chapter 3 Expected Travel Distances in Automated Guided Vehicle-Based Order Picking Systems ......................................................................................................................... 47

Chapter 4 Heuristic Methods for the Location of Induction Points in Automated Guided Vehicle-Based Sortation Systems ......................................................................................... 82

Chapter 5 Conclusions and Future Research Opportunities .............................................. 120

Appendices .......................................................................................................................... 125
List of Tables

Table 2-1. Distribution of the number of line items/order ................................................................. 21
Table 2-2. Miniload system simulation results ...................................................................................... 27
Table 2-3. Miniload system simulation results for the buffers ............................................................. 28
Table 2-4. Kiva system simulation results ............................................................................................. 30
Table 2-5. Kiva system simulation results for the buffers ................................................................. 31
Table 2-6. Percentage breakdown of the average container delivery time ......................................... 34
Table 2-7. The miniload system with additional SKUs ........................................................................ 35
Table 2-8. The Kiva system with additional SKUs .............................................................................. 36
Table 2-9. Breakdown of AGV time per pod retrieval (in secs.) ............................................................ 37
Table 3-1. $E_{OUT}$ per pod for symmetric FA shapes .......................................................... 62
Table 3-2. $EDC$ per pod for symmetric FA shapes .............................................................................. 63
Table 3-3. $E[OUT]$ based on a rectangular FA and optimal shape factor ($b$) ..................................... 68
Table 3-4. Summary of $E[OUT]$ cases treated by method ............................................................. 76
Table 3-5: Abbreviations and notation ................................................................................................. 81
Table 4-1. Computational results for the EO-NEQ case ........................................................... 101
Table 4-2. Computational results for the EF-NEQ case .............................................................. 103
Table 4-3. Computational results for the EF-EQ case ..................................................................... 104
Table 4-4. Computational results with perimeter IPs and the EF-NEQ case .................................. 107
Table 4-5. Analytic expected travel distance expressions for two IPs – perimeter case ............ 112
Table H-1. $ETB$ results for asymmetric FA shapes. ................................................................. 145
Table H-2. $EOUT$ results for asymmetric FA shapes. .............................................................. 146
Table H-3. $EDC$ results for asymmetric FA shapes. ................................................................. 147
List of Figures

Figure 2-1. The miniload system (miniload AS/RS with a conveyor loop supplying the pick stations). ................................................................. 14

Figure 2-2. Two examples of a pick station. (Images are property of VanDerLande.) ............... 15

Figure 2-3. Flowchart of container flow in the Miniload system. .................................................. 17

Figure 2-4. The Kiva system (the forward area, the AGVs, and the pick stations). ..................... 18

Figure 2-5. Examples of the Kiva system (a) The forward area, (b) Kiva AGV traveling with a pod in the forward area, (c) A pick station. (Images are property of Kiva Systems.) .............. 18

Figure 2-6. Flowchart of container flow in the Kiva system. .......................................................... 20

Figure 2-7. Miniload rack with small, medium, and large SKU zones (Not to scale) ................. 23

Figure 2-8. System throughput as a function of the number of AGVs. ........................................ 37

Figure 2-9. Kiva system picker ergonomics. (a) Picker reaching up to pick from top shelf of the pod (Photo: Quiet Logistics). (b) Picker bending to pick from bottom shelf of the pod; note the step ladder (Photo: Amazon Fulfillment) .......................................................... 40

Figure 3-1. Typical AGV-OPS with PSs in the front. (Taken from Bozer and Aldarondo 2018). .................................................................................................................. 50

Figure 3-2. Asymmetric FA shapes. ......................................................................................... 56

Figure 3-3. $E_{OUT}$ per pod as a function of $N$. ........................................................................... 63

Figure 3-4. $E_{DC}$ per pod as a function of $N$. ............................................................................ 63
Figure 3-5. Percent error of curve fit approximation versus Monte Carlo sample for $E_{OUT}$ — 4/RECT/CLO/N .............................................................. 67

Figure 3-6. $E_{DC}$ per pod as a function of $N$. ................................................................. 72

Figure 3-7. FA layouts and PS configurations for usage experiment .................................... 74

Figure 3-8. Pick station usage measured by percentage of pods assigned to each PS .......... 74

Figure 3-9. CLO rule contour lines in a square with: (a) continuous PSs, (b) discrete uniformly located PSs, and (c) discrete even usage PSs ................................................................. 75

Figure 4-1. Aerial view of the AGV-SS. (Obtained from the China Global Television Network, CGTN, depicting a logistics center in China.) ................................................................. 91

Figure 4-2. (a) Close-up of the induction point. (Obtained from the China Global Television Network, CGTN, depicting a logistics center in China.) (b) Tilt tray discharges the item at the destination point. (Obtained from the China Global Television Network, CGTN, depicting a logistics center in China.) ................................................................. 92

Figure 4-3. Best heuristic solutions obtained on a $20 \times 20$ grid ..................................... 105

Figure 4-4. Best heuristic solutions obtained on a $50 \times 50$ grid ..................................... 106

Figure 4-5. Optimal allocation and the resulting service areas for given IP locations ............. 106

Figure 4-6. Best heuristic solutions obtained for perimeter IPs on a $20 \times 20$ grid............. 108

Figure B-1. Travel to the closest PS on the long side of the FA .......................................... 128

Figure B-2. Distance from all the points to the median point for $N = 4$ and $N = 5$ ............ 130

Figure D-1. Contour lines in a rectangular FA under CLO rule .......................................... 133

Figure F-1. Diamond shape divided into regions for $E_{OUT}$ calculation for PS on one side..... 139

Figure F-2. Diamond shape divided into regions for $E[OUT]$ calculation for PS on two sides. 140

Figure G-1. Line segments $a, b, d$ on the perimeter of the octagon .................................... 144
Figure I-1. Expected empty travel distance DP regions relative to the position of the IPs. ....... 149
Figure I-2. Expected full travel distance DP regions relative to the position of the IP. ............ 151
Figure J-1. Expected empty travel distance DP regions relative to the position of the IPs. ....... 154
Figure J-2. Expected empty travel distance DP regions relative to the position of the IPs. ....... 156
Figure J-3. Expected empty travel distance DP regions relative to the position of the IPs. ....... 158
List of Appendices

Appendix A Derivations for •/RECT/RAN/1 – $E_D C$. ................................................................. 126

Appendix B Derivations for 1•/RECT/CLO/N – $E_{OUT}$. ................................................................. 128

Appendix C Derivations for 2•/RECT/CLO/N – $E_{OUT}$. ................................................................. 131

Appendix D Derivations for 4/RECT/CLO/1 – $E_{OUT}$ and 4/RECT/CLO/2 – $E_{OUT}$ .......... 133

Appendix E Derivation for •/DIAM/RAN/1 – $E_{OUT}$ for a FA of unit area. ....................... 137

Appendix F Derivation for •/DIAM/CLO/1 – $E_{OUT}$ for a FA of unit area......................... 139

Appendix G Derivation for •/OCTA/RAN/1 – $E_{OUT}$ for a FA of area of $\pi$. ................. 142

Appendix H Tabulated results for asymmetric FA shapes. ...................................................... 145

Appendix I Derivation for expected travel distance for two IPs in a unit square - interior case. 148

Appendix J Derivation for expected travel distance for two IPs in a unit square - perimeter case.
............................................................................................................................................. 153
Abstract

Design and analysis of warehouse automation systems continues to be an active topic of interest both in academia and practice, especially in light of the significant increase in online retail sales. This work concerns the operational analysis and modeling of warehouse automation systems that rely on Automated Guided Vehicles (AGVs). Although, as a technology, AGV systems are not new, recently they have been used in “goods-to-person” order picking (OP) systems and in sortation systems.

AGV-based OP systems, also known as Robotic Mobile Fulfillment Systems (RMFS), were first introduced under the name of Kiva systems. A key component of an AGV-based OP system is the fleet of “robots” (or AGVs) that pick up the “pods” (or “racks”) and transport them to the appropriate pick station (PS), where a human picker picks the items ordered by customers.

We examine two types of well-known, goods-to-person order picking systems, namely, a miniload system and a Kiva system. Using a simulation model, we compare the performance of the two systems on the basis of expected throughput and expected container retrieval times to process the same set of customer orders. We also discuss some of the advantages and limitations of the above two systems.

The performance of AGV-based OP systems depends on the number of AGVs, which in turn depends on the time it takes the AGV to retrieve a pod from the storage area. We derive closed-form analytic expressions for the expected travel distance of the AGVs operating under two possible order assignment rules. Under the random assignment rule, an order is assigned to any PS with equal probability. Under the closest assignment rule, the order is assigned to the closest
PS. We also examine the impact of the shape of the storage area on the expected AGV travel distance and the impact of alternative PS configurations. The results offer valuable insight concerning the expected AGV travel distances, which are also needed for analytic design and performance evaluation models.

AGV-based sortation systems (AGV-SSs) have recently been adopted by some of the leading online retail companies. An AGV-SS performs the same basic function as a traditional, conveyor-based sortation system except that it relies on AGVs instead of conveyors to sort the items. In virtually all automated and semi-automated sortation systems, there is at least one “induction point” (IP) that serves as an identification and entry point for the items to be sorted by the system. In an AGV-SS, each (empty) AGV is loaded at the IP with an item to be sorted. The loading process may be manual or automated. Once the item is loaded onto the AGV, the AGV travels to the appropriate destination point (DP), where the item is automatically discharged, and the empty AGV travels to the same or another IP to pick-up the next item. Given the location of the DPs, we are concerned with determining the optimum location of the IPs in order to minimize the expected AGV travel distance. We propose heuristic methods for the problem and compare their performance through a computational study. We also present results for and compare interior versus perimeter placement of the IPs. Finally, we develop and solve an analytic model for the optimal location of two IPs in a continuous plane that represents the sortation region.
Chapter 1
Introduction

The U.S. Census Bureau (2016) reports that retail sales in the United States continue to show steady growth, approaching $5 trillion in 2015. Similarly, the percentage share of online retail sales has shown steady growth, reaching 7.26% (or $341 billion) in 2015. Equally striking is the fact that there is ample room for growth in online retail sales as a percentage of overall retail sales. Growth in the global retail market has also been equally impressive.

The above growth in online retailing has fueled not only the parcel delivery business but the order fulfillment business as well, where thousands of orders are now picked in fulfillment centers (FCs), largely by human pickers. In the current environment, the opportunity for automation in FCs is more present than ever.

At the low-end of automation are walk-and-pick OP systems, which are classified as “person-to-goods” OP since the pickers walk/travel to the pick face. At the higher-end of automation are “goods-to-person” (GTP) OP systems, where a material handling system brings the containers to the pickers. Generally speaking, GTP OP systems eliminate the need for the picker to walk/travel in the forward area (FA), which can significantly increase the pick rates and lead to labor cost savings. Furthermore, GTP OP systems can increase picking accuracy, yield better cube utilization inside the facility, and improve picker ergonomics (enVista 2017). In terms of implementation, GTP OP systems differ primarily by the type of material handling system used for bringing the containers to the pickers.
In Chapter 2, the performance of two types of GTP OP systems is compared via simulation. We focus on two specific types of such systems; namely, the miniload automated storage/retrieval-based OP system (or “miniload (OP) system” for short), and the Kiva system, an Automated guided vehicle (AGV)-based OP system (or “AGV-OPS” for short). In the miniload OP system, the SKUs are placed in containers (or trays) stored in racks. The containers are retrieved by a storage/retrieval machine and delivered to the appropriate pick station via a conveyor loop. In the AGV-OPS, the SKUs are stored in racks or “pods.” AGVs transport the racks to the pick stations. (Kiva Systems was acquired by Amazon in 2012 and renamed “Amazon Robotics.”)

In comparing the above two systems, we identify the design elements (i.e., the number of AGVs versus the number of miniload aisles) that yield the same throughput for picking the same set of orders. It is one of the few studies to directly compare two distinct OP systems. Whenever possible, the same or similar parameters are used. The two systems are compared on the basis of throughput, picker and material handling equipment utilizations, and order completion. While our primary purpose is to compare the system configurations that yield the same throughput, we also present approximate cost measures.

Since the introduction of Kiva systems, the emerging field of AGV-based OP has captured the attention of the online retail OP community, both in academia and industry. One key advantage of an AGV-OPS is its incremental scalability; that is, the user can add AGVs, pods, and/or PSs as needed, with minimal disruption. Other unique features are built-in redundancy and parallel processing. Multiple AGVs provide redundancy against breakdowns, since any pod can be transported by any AGV. Also, multiple pods can be moved concurrently to/from the PSs at any given time. Since Kiva was purchased by Amazon in 2012, there has been a proliferation of competing AGV-based OP systems developed by vendors worldwide. A few examples are
Swisslog’s CarryPick (Switzerland), Grey Orange’s Butler System (India), and Hitashi’s Racrew (Japan).

In Chapter 3, we take a closer look at AGV-OPS. The number of AGVs in the system plays a critical role not only because one must ensure that the pods are retrieved and stored at a rate needed to support the PSs but also because the AGVs represent a major component of the system cost. The AGVs retrieve and then store each pod, one at-a-time, using dual command (DC) cycles that consist of three legs. First, after depositing the current pod, the AGV travels empty from its current location to the location of the next pod to be retrieved (the “travel between” leg). It then picks up the pod and travels to the appropriate PS (the “out” leg). Once the item(s) are picked, the AGV returns the pod to the FA (the “back” leg), and it becomes available to retrieve another pod.

We study the expected travel distance of DC cycles in an AGV-OPS as a function of the order assignment rule, the shape of the FA, and the configuration of the PSs around the FA. We consider two order assignment rules, namely, random (RAN) and closest (CLO). Under RAN, all the pods associated with an order are assigned to one of the PSs with equal probability. Under CLO, the pods associated with an order are assigned to the PS that minimizes the total distance traveled to the PS. Once the item(s) are picked from a pod, it can be returned either to its original location or to any open location in the FA. We consider three “symmetric shapes” for the FA; i.e., a square, a diamond, and a circle. For each shape, we also consider its “asymmetric” counterpart; i.e., a rectangle, an elongated diamond, and an ellipse. Last, we consider five possible PS configurations. The PSs may be located on 1. The short side of the FA, 2. The long side of the FA, 3. The two short sides of the FA, 4. The two long sides of the FA, and 5. All four sides of the FA. For symmetric shapes, cases 1 and 2, and cases 3 and 4, are of course the same.
In Chapter 4, we study a relatively new type of unit load AGV, one which is equipped with a tilt tray, used for automatically transporting and sorting individual items, one at a time. Both unit load AGVs and tilt tray sortation conveyors have been used for decades. However, eliminating the conveyor, and using AGVs equipped with tilt trays is a new concept in sortation. We refer to this system as AGV-based sortation system (AGV-SS).

We focus on the design and analysis of AGV-SSs. More specifically, given a set of discrete DPs located in a square-shaped sortation region (SR), we develop heuristic procedures to investigate the problem of finding the optimal location of the IPs to minimize the expected full and empty AGV travel distances. We perform computational experiments to compare the performance of the heuristic procedures on several problem instances. We also investigate the optimal location of the IPs when they are restricted to the perimeter of the SR. Last, using a continuous approximation of the SR, we present exact analytic results for determining the optimum location of two IPs placed either inside or on the perimeter of the SR.

References


Chapter 2

A Simulation-Based Comparison of Two Goods-to-Person Systems in an Online Retail Setting

Abstract:

Design and analysis of order picking systems continues to be an active topic of interest both in academia and practice, especially in light of the significant increase in online retail sales. In this paper, we examine two types of well-known, goods-to-person order picking systems, namely, a miniload system and a Kiva system. Using a simulation model, we compare the performance of the two systems on the basis of expected throughput and expected container retrieval times to process the same set of customer orders. We also discuss some of the advantages and limitations of the two systems.

---

1 This Chapter was published as: Bozer, Y. A. and Aldarondo, F. J. 2018. “A Simulation-Based Comparison of Two Goods-to-Person Order Picking Systems in an Online.” *International Journal of Production Research* 56 (11), 3838-3858.
1. Introduction

According to Pricewaterhouse-Coopers (2014), retail is the largest private employer in the United States (US); it directly and indirectly supports 42 million jobs, and contributes $2.6 trillion annually to the US GDP. The US Census Bureau (2016) reports that retail sales in the US continue to show steady growth, approaching $5 trillion in 2015. Similarly, the percentage share of online retail sales has shown steady growth, reaching 7.26% (or $341 billion) in 2015. Equally striking is the fact that there is ample room for growth in online retail sales as a percentage of overall retail sales. Growth in the global retail market has been equally impressive. According to eMarketer (2014), retail sales worldwide topped $22 trillion in 2014, with e-commerce sales accounting for $1.3 trillion, led by China and the US. The group predicts that the “global retail market will see steady growth …, and in 2018, worldwide retail sales will … reach $28.3 trillion” with the percent share of e-commerce increasing to 8.8% (or $2.5 trillion) by 2018.

The above growth in online retailing has fueled not only the parcel delivery business but the order fulfillment business as well, where thousands of orders are now picked in warehouses (or fulfillment centers) largely by human pickers. As a result, order picking (OP) has become a focal point for many online retailers and academic researchers.

Nevertheless, OP is a labor-intensive activity and it accounts for 50% to 55% of the operating cost in a warehouse (Drury 1988, Frazelle 2002, p.148, and Tompkins et al. 2010, p.433). Searching for and extracting the items, including travel, may account for 80% of the time required to fill orders (Frazelle 2002, p.154). OP is particularly challenging for online retailers, which typically stock a large number of Stock Keeping Units (SKUs) and process many small orders as measured by the number of line items/order. Using the right OP method and investing in the right
equipment is a key decision for many online retailers since it impacts their ability to make timely and accurate deliveries in a cost-effective manner.

At the low-end of automation are walk-and-pick OP systems, which are classified as “person-to-goods” (PTG) OP since the pickers walk/travel to the pick face. At the higher-end of automation are “goods-to-person” (GTP) OP systems, where a material handling system brings the containers to the pickers. (There are also fully-automated OP systems with no human pickers.) According to enVista (2017):

“GTP systems have been around for decades, but recently, new technologies and software advancements have enabled significant improvements in functionality, efficiency, inventory management, and space utilization. … (A)n increasing number of distribution and warehousing professionals are starting to believe that PTG technologies and order fulfillment methods are somewhat outdated … Large warehouses, SKU proliferation and smaller order fulfillment windows are rendering PTG processes less effective due to increased picking travel, mispicks, and increased labor requirements. … However, many companies … are not been fully convinced that GTP integrated systems are reliable, and that they can effectively deliver acceptable ROI.”

In this study our objective is to compare the performance of two types of GTP OP systems; we are not concerned with PTG versus GTP. Nonetheless, increased picking rates and improved ergonomics are often stated as major advantages of GTP OP systems. For a more comprehensive comparison, the reader may refer to enVista (2017).

GTP OP systems differ primarily by the type of material handling system used for bringing the containers to the pickers. We focus on two specific types of such systems; namely, the Kiva system, and the miniload automated storage/retrieval (AS/R)-based OP system (or “miniload (OP) system” for short). In the Kiva system, the SKUs are stored in racks or “pods.” Automated guided vehicles (AGVs) transport the racks to the pick stations. (Kiva Systems was acquired by Amazon in 2012 and renamed “Amazon Robotics.”) One of the early industrial applications of Kiva involved a 600,000 sq.ft. Office Depot warehouse with 12,000 SKUs, 3,500 pods, and 430 AGVs, with separate pick and replenishment stations (Wulfraat, 2012). A more recent application of a
Kiva-type of system is the one installed for Lekmer.com, Scandinavia’s largest online toy retailer; it involves 1,500 pods and 65 AGVs (Swisslog, 2015).

In the miniload OP system, the SKUs are placed in containers (or trays) stored in racks. The containers are retrieved by a storage/retrieval (S/R) machine (named “S/R” for short) and delivered to the appropriate pick station via a conveyor loop. In some miniload systems, one picker is dedicated to each aisle; no conveyor loop is provided. Also, some miniload systems contain multiple shuttles; see, for example, Guller and Hegmanns (2014) for a description of such systems. Industrial applications of the miniload system can be found in VanDerLande (2017), where a 4-aisle system with 38,720 locations is reported to achieve 600 lines/hour per picker, and a 6-aisle system with 13,000 locations is installed in Netherlands for Tommy Hilfiger, a leading apparel and retail company.

In comparing the above two systems, we identify the design elements (i.e., the number of AGVs versus the number of miniload aisles) that yield the same throughput for picking the same set of orders. The primary tool used for the comparison is simulation. It is one of the few studies to directly compare two distinct OP systems. Whenever possible, the same or similar parameters are used. The two systems are compared on the basis of throughput, picker and material handling equipment utilizations, and order completion times as well as other relevant factors. While our primary purpose is to compare the system configurations that yield the same throughput, we also present approximate cost measures.

In the next section we present a brief review of the OP literature, including analytical results and simulation studies for miniload OP systems. A description of the two systems and the assumptions are presented in Section 3. Section 4 presents the numerical results, while
Section 5 compares the two systems based on other/qualitative factors. The summary results and possible future research directions are presented in Section 6.

2. Literature Review

OP has received considerable attention in the literature. For example, De Koster et al. (2007) cite 140 publications in OP, while Gu et al. (2007) cite 128 publications on warehousing, with 67 in OP. The selection of an OP method as a strategic decision is discussed in Gu et al. (2010) and Dallari et al. (2009). The latter study, derived from the statistical analysis of 67 warehouses, presents an OP system selection methodology that uses two key parameters: the number of SKUs in the picking area and the number of line items picked/day. The authors emphasize the need for design methodologies that require less detailed data and allow shorter “time-to-design” through analytic models. Another important aspect in the design stage is technology selection. As evidenced by the 155 publications cited in Roodbergen and Vis (2009), among the OP technologies available, AS/R systems have received significant attention.

Analytical results derived for GTP OP include Bozer and White (1990), who developed a two-server closed queueing model for a miniload system with two pick positions and one picker/aisle. Subsequent work by Bozer and White (1996) extended the above model to multiple pick positions/aisle and multiple aisles assigned to each picker. The above models estimate the expected throughput of the system and the expected utilization of the picker. Approximate throughput bounds were developed by Foley and Frazelle (1991) and by Foley et al. (2002, 2004) for a miniload system with two pick positions/aisle.

A miniload system with a U-shaped conveyor was studied by Park et al. (1999), who proposed a capacitated two-stage cyclic queueing model. The U-shaped conveyor provides additional buffer between the picker and the S/R. Cycle times are derived and closed-form
expressions for multiple performance measures are presented, assuming exponentially distributed pick times. An extension to the case where one picker serves multiple aisles is presented by Koh et al. (2005). Both of the above models are used for determining the steady-state behavior of the U-shaped conveyor and for buffer sizing.

A miniload system with multiple pick stations on a conveyor loop is analyzed by Claeys et al. (2016). The authors assume that new orders are always available to be released into the system so they can focus on the maximum throughput capability of the pick stations. The individual totes for an order are assumed to arrive randomly at each pick station, and they do not necessarily arrive in the sequence in which the orders were released (sometimes called “out-of-sequence arrivals”). Each pick station is modeled with multiple, short input buffers since it is assumed that each tote is within the reach of the picker. The authors assume that each pick station handles only one order at a time, meaning the picker must wait for a tote of the current order being processed even if totes for other orders are present in the buffer. Modeling each pick station as a single-server polling system (where the server represents the picker), the authors derive bounds for the mean order flow time, which yields bounds on the maximum throughput. Out-of-sequence arrivals are also considered by Andriansyah et al. (2014), who study an automated pick station that uses an integrated carousel mechanism to deal with such arrivals. Unlike Claeys et al. (2016), they assume that multiple orders can be processed simultaneously.

Hwang et al. (2002) examine an integrated system where the containers retrieved by a miniload AS/RS are delivered to an assembly line via an AGV system. The authors develop an analytic model to concurrently design the miniload AS/RS and the unit load size of the AGVs. Component parts are assumed to be delivered in kits prepared at a kitting station. We refer the
reader to Andriansyah (2011) and Roodbergen and Vis (2009) for further reading on analytical results for miniload OP systems.

Several researchers have used simulation models to study miniload OP systems. Medeiros et al. (1986) developed a simulation model for two pick positions/aisle, while Pulat and Pulat (1989) modeled the case with a U-shaped conveyor. Perry et al. (1984) modeled a miniload system with a conveyor loop. The authors present a heuristic design approach aimed at meeting the required throughput by manipulating several physical configuration variables and operating rules. They conclude that the conveyor loop is the bottleneck. Guller and Hegmanns (2014) also simulate a miniload system with a conveyor loop; however, they focus on a multishuttle system. Assuming Poisson arrival of orders, the authors study the expected order completion time and throughput as a function of the order profile (ranging, on average, from 1.9 to 2.5 line items/order). They observe that the completion time for a single-line order may be greater than that of a four-line order.

Raghunath et al. (1986) and Takakuwa (1996) developed modular simulation models that can represent multiple system configurations or operating policies. The latter paper covers the case where an AGV-loop serves as the interface between the AS/R system and the pick stations. Assuming rail-guided vehicles (RGVs) for the interface loop, Lee et al. (1996) use simulation to estimate the system throughput as they vary the number of RGVs. Orders are assumed to arrive randomly. The authors examine buffer sizing and RGV deadlocking as well.

Another modular simulation model for a miniload system with a conveyor loop and pick stations is presented by Andriansyah et al. (2011), who use a process algebra-based language. The authors demonstrate the modularity of the model by varying the control system logic and the number of miniload aisles. Their primary argument is that “model architecture and control structure are crucial for simulation studies of industrial scale AS/RS.” They assume that a tote
box is allowed to enter the conveyor loop only if it has “reserved” space on the loop. (This implies that simply waiting in the buffer without reserved space is not considered.) Also, each S/R machine is assumed to hold up to four tote boxes.

Onal et al. (2017) also use simulation; however, they focus on PTG OP. They study an online retailing warehouse with 400 SKUs and 3,240 bins. The authors show the impact of the “explosive storage” policy (where incoming materials are broken-up and distributed to multiple storage points in the warehouse) and other factors such as bins with comingled SKUs.

Despite numerous trade/commercial articles written about Kiva systems, there are relatively few refereed publications on the subject. In an earlier paper by Wurman et al. (2008), in addition to content aimed at the artificial intelligence community, the authors discuss the challenges associated with operating large OP systems and the drawbacks of carousels and AS/R-based systems. The authors, who were affiliated with Kiva Systems (the company), point out a number of advantages of the Kiva system, which we return to in Section 5. Although the paper is informative, it generally paints the Kiva system in a favorable light, and the limitations of the system are not addressed adequately. No analytic or simulation results are presented. A similar paper on Kiva systems, highlighting their advantages, is presented by Guizzo (2008).

An analytic model for Kiva systems, which is called “robotic mobile fulfillment” (RMF) by some researchers, is presented by Lamballais et al. (2017). Modeling the system as a queueing network, the authors present analytic results to estimate the expected system throughput and the average AGV utilization for single-line-item and multi-line-item orders. They also show that the throughput capacity of the system is impacted by the location of the pick stations but not by the shape of the forward area. Another study on RMF systems is presented by Zou et al. (2017). Using an analytic model, the authors show that the performance of the system can be improved by
assigning the pick stations to the AGVs so that the expected sojourn time of the retrieval transactions is minimized. The improvement is significant if the pick times at the stations have large differences.

A similar study for RMF systems is presented by Boysen et al. (2017), who focused on order processing and rack (or pod) retrieval sequencing for a single pick station. The authors consider the case where multiple orders are picked concurrently at the pick station, and the objective is to minimize the pod retrievals, given that multiple SKUs are located in each pod and the same SKU may appear in more than one of the orders being picked at the station. They present the problem as a formal decision problem and propose solution procedures, which are shown to reduce the number of AGVs required compared to simple decision rules that the authors claim are used in real-world warehouses. (Since the details of Kiva systems are not published, it is not clear how much improvement would be obtained over the retrieval sequencing logic used in Kiva systems.)

Another Kiva-based OP study is presented by Stowe (2016). Assuming that the downstream processes are not the bottleneck, the study focuses on maximizing the pick rate at the pick stations by segregating the SKUs by velocity (a well-known technique known as slotting) and through policies that increase picker retention. The latter scheme is based on the author’s observation that the number of person-hours worked by inexperienced pickers explains practically all of the variability in the pick rates, which argues for picker retention. Combined, the above two strategies are estimated to increase the throughput by 10%.

As the literature review indicates, a large majority of the refereed papers perform a detailed analysis of a specific type of OP system, while papers that directly and quantitatively compare two competing OP systems are very rare in the refereed literature. (By direct comparison we mean
that the same set of orders and the same pick times are used.) While detailed analyses of specific types OP systems will continue to enrich the OP literature, studies that compare two systems to better understand their relative strengths/weaknesses are needed as well. Such studies are also valuable from a practical standpoint since decision-makers must choose between competing systems when they install a new OP system.

3. System Description

We next present the details of the two systems.

3.1. The Miniload System

The main components of the miniload system are described below (Figure 2-1).

![Figure 2-1. The miniload system (miniload AS/RS with a conveyor loop supplying the pick stations).](image)
Miniload Aisles: Each aisle contains an aisle-captive S/R, and single-deep storage racks. Each aisle is attached to the loop by input/output (I/O) conveyors (see “miniload buffer” in Figure 2-1). The I/O point for the S/R is located at the lower front corner of the rack.

Pick Stations: The pick stations are attached to the loop by I/O conveyors (see “pick station buffer” in Figure 2-1). Each (active) pick station is assigned one picker. All the line items in an order are picked by the same picker. Containers arrive at a pick station through its input buffer. Once the item(s) are picked, the containers depart through its output buffer. No picker travel is necessary since the containers are automatically indexed, one at a time, towards the picker, and then taken away when picking is completed. Two examples are depicted in Figure 2-2. In some systems, at each pick station, additional conveyors attached to the loop deliver empty containers to the pick stations and/or transport the picked items to a sortation/packing area. Such conveyors are beyond the scope of our study.

Conveyor Loop: It transports the containers between the miniload aisles and the pick stations. The loop also provides additional buffering—if a container finds a full I/O buffer at a miniload aisle or pick station, the container recirculates on the loop.
Container Flow: A retrieval request queue is maintained for each miniload aisle. When an order is opened (Section 3.3.1), the appropriate retrieval request(s) are placed in the retrieval request queue of each S/R, which serves the requests on a FCFS basis. The request is completed when the S/R deposits the container in the miniload output buffer, where it waits to merge onto the loop. Once on the loop, the container is transported to the input buffer of the appropriate pick station. After the item(s) are picked, the container is moved to the pick station’s output buffer where it waits to merge onto the loop. Once on the loop, the container is returned to the same aisle via the miniload input buffer and stored back in the rack by the S/R. The container flow is depicted in Figure 2-3.

3.2. The Kiva System

The Kiva system (Figure 2-4) employs AGVs. Its main components are described below.

Forward Area and the Pods: The forward area consists of single-wide aisles and pods. Cross-aisles are added to minimize blocking and give the AGVs a clear path to/from the pick stations (Figures 2-4, 2-5(a), and 2-5(b)). (The shading in the forward area is explained later.) The pods are simple, multi-tier racks (Figure 2-5(b)) that are fairly short, which allows safe transportation by the AGVs while giving the pickers access (Figure 2-5(c)).

AGVs: The pods are moved, one at a time, between the forward area and the pick stations by the AGVs (Figures 2-5(b) and 2-5(c)), which navigate over a grid of two-dimensional barcodes (or QR Codes) embedded in the floor. The AGV spins in place to lift/lower a pod.
Input:
Parameters
Set of orders
Open initial orders.
Send each retrieval request to appropriate S/R machine retrieval queue.

For each retrieval request in each aisle:

<table>
<thead>
<tr>
<th>Question</th>
<th>Flowchart Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is S/R machine i busy?</td>
<td>Yes</td>
</tr>
<tr>
<td>Retrieval request for container k in aisle i waits for S/R machine.</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>S/R machine i starts retrieval trip for container k. Container k is now unavailable.</td>
<td></td>
</tr>
<tr>
<td>S/R machine deposits container k in the output buffer. Container k waits until it's first in buffer.</td>
<td></td>
</tr>
<tr>
<td>Picker pushes container k into output buffer. Container k waits until it's first in buffer.</td>
<td></td>
</tr>
</tbody>
</table>

Order available? |
Yes |
Open the order; send retrieval request to the appropriate S/R machine retrieval queue. |

No |
Wait for next tray storage. |

Wait until last tray is stored. |
END |

Close order when last item on that order is picked. |

Wait until last item on that order is picked. |

Last order? |
Yes |
Wait for next tray storage. |

No |
Scan order list for the next order with all containers available. |

Order available? |
Yes |
Open the order; send retrieval request to the appropriate S/R machine retrieval queue. |

No |
Wait for next tray storage. |

Figure 2-3. Flowchart of container flow in the Miniload system.
Figure 2-4. The Kiva system (the forward area, the AGVs, and the pick stations).

Figure 2-5. Examples of the Kiva system (a) The forward area, (b) Kiva AGV traveling with a pod in the forward area, (c) A pick station. (Images are property of Kiva Systems.)
**Pick Stations:** Each (active) pick station holds one picker. All the line items in an order are picked by the same picker. Once it is moved by an AGV to the appropriate pick station, the pod waits in the pick station buffer located ahead of the picker (Figures 2-4 and 2-5(c)). The AGV stays with the pod while it waits in the buffer.

**Container Flow:** A single retrieval queue is maintained for all the requests. When an order is opened (see Section 3.3.1), the appropriate retrieval requests are placed in the retrieval request queue. The requests are served, one at a time, on a FCFS basis by one of the AGVs. To serve a request, the AGV travels empty from its current location to the pod. It then picks up the pod and travels to the appropriate pick station. Once the item(s) are picked, the AGV returns the pod to the forward area and becomes available to serve another request. Two rules are considered to select an AGV when serving a request; the closest available (CA) rule and the first available (FA) rule. With CA, when there’s a retrieval request, if multiple AGVs are available, the AGV that is closest to the pod (to be retrieved) is used. If no AGVs are available, the first AGV to become available is used. With FA, the AGVs are numbered; when there’s a retrieval request, the AGVs are scanned starting with AGV number one, and the first available AGV is used. If no AGVs are available, the first AGV to become available is assigned to the request. The container flow is depicted in Figure 2-6.
Figure 2-6. Flowchart of container flow in the Kiva system.
3.3. Assumptions and Parameters

We aim for as much a balanced comparison as possible, while keeping the two simulation models simple. Our intent is a reasonably accurate but basic comparison of the two systems.

3.3.1. General Assumptions for Both Systems

1. **SKUs**: There are 6,000 SKUs. The SKUs are *equally* partitioned into three size categories (small, medium, and large). Each category experiences the same level of demand. We assume no stock-outs.

2. **Orders**: A batch of 2,400 customer orders is generated randomly for each replication. Both systems are simulated with the same set of orders. Each order may contain up to 10 line items, with an average of 2.325 line items/order (Table 2-1). A majority of the orders contain very few line items, which is typical in online retailing; see, for example, Lasgaa (2010), Napolitano (2013), and Onal et al. (2017). The number of pieces/line item is beyond the scope of our study. The pick time implicitly assumes that some line items may require multiple pieces.

   Table 2-1. Distribution of the number of line items/order.

<table>
<thead>
<tr>
<th>Number of line items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.45</td>
<td>0.25</td>
<td>0.10</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
</tr>
</tbody>
</table>

   To generate an order, we first sample the number of line items from Table 2-1. For each line item, we randomly choose an SKU size category (small, medium, or large). Given the size category, we then randomly sample an SKU. Each size category, and each SKU within a category, are *equally likely* to be selected.

3. **Trays**: The results are not limited to a specific type of container but to achieve sufficient detail, we assumed the SKUs are stored in “trays,” which is common for miniload systems.
Three tray configurations are assumed: tray type A holds four small-size SKUs, tray type B holds two medium-size SKUs, and tray type C holds one large-size SKU. Each tray measures 4 ft. (length) × 2 ft. (width) × 1 ft. (height).

4. **Picking:** There are four pick stations. Each pick station may have multiple orders open simultaneously. Picks are performed in the same sequence the containers are delivered at a pick station. The pick time/line item is uniformly distributed between 10 and 30 seconds, which includes the transition time between consecutive trays/pods.

5. **Order release:** To avoid flooding the system, we employ a pull-type mechanism like CONWIP (Spearman et al. 1990), which limits the number of orders that are opened and assigned to a pick station at one time. When an open order is completed (i.e., the last line item is picked), a new order is opened and assigned to the picker. To open a new order, the unpicked orders are scanned, and the first eligible order is selected. (An order is eligible if all the containers needed for that order are in the rack/forward area, and none of them are needed for orders that are already open.) If too many (or too few) orders are opened, it leads to congestion and it impacts the availability of the containers (or it results in picker idle time).

### 3.3.2. Miniload System Assumptions

1. **Storage rack and trays:** There are five aisles, with identical racks. Each rack is divided into three zones (Figure 2-7). A tray is equally likely to be located in any opening within its zone. Each opening measures 1.25 ft. (high) × 2.5 ft. (wide) × 4 ft. (deep). Since the number of bays and tiers are integers, the number of tray positions slightly exceeds the requirement. For example, the zone for small SKUs is seven tiers high and eight bays long, which yields 56 positions/rack, or $56 \times 2 \times 5 = 560$ positions across the five aisles. With
2,000 small SKUs, and 4 small SKUs/tray, there are actually 500 trays, which leaves space for future SKUs. (We also simulated the system with four aisles. When the fifth aisle is removed, the rack dimensions in the remaining four aisles are proportionally increased since the total number of trays is fixed.)

![Miniload rack with small, medium, and large SKU zones](image)

Figure 2-7. Miniload rack with small, medium, and large SKU zones (Not to scale).

2. **S/R machines**: There are five single-shuttle S/Rs (one in each aisle). Each S/R travels at a speed of 400 fpm (vertical) and 800 fpm (horizontal), which represent approximately 68% of the top speed for state-of-the-art miniload S/Rs (Swisslog 2016). No acceleration/deceleration effects are considered. The S/R container pickup/deposit (P/D) times are equal to 3 secs each. At the start of the simulation, each S/R is idle at its I/O point.

3. **S/R control logic**: The trays are retrieved on a FCFS basis. In retrieving and later storing the trays, the S/R follows an “opportunistic interleaving” policy (Bozer and Cho 2005); i.e., having completed a storage operation, if there is a retrieval request in the queue, the S/R performs a retrieval operation. If the retrieval request queue is empty, the storage request queue is checked; if it’s empty also, the S/R idles in the rack; if the storage request
queue is not empty, the S/R performs a storage operation. Likewise, having completed a retrieval operation, if there is a storage request in the queue, the S/R performs a storage operation. If the storage queue is empty, the retrieval request queue is checked; if it’s empty, the S/R idles at the I/O point; if the retrieval request queue is not empty, the S/R performs a retrieval operation.

4. **Conveyor loop:** The conveyor travels at 200 feet/minute and has a maximum capacity of 54 trays. The trays are moved on/off the loop by automated 90-degree transfer mechanisms, which are not simulated explicitly.

5. **Buffers:** Each I/O buffer holds up to five trays.

### 3.3.3. Kiva System Assumptions

1. **Pods:** Each pod consists of four shelves. Each shelf is equivalent to one miniload tray. A pod is assumed to hold SKUs of only the same size category; i.e., a pod holds either 16 small-size SKUs (four SKUs/shelf), or 8 medium-size SKUs (two SKUs/shelf), or 4 large-size SKUs (one SKU/shelf).

2. **Forward Area:** Pods in the forward area are zoned by size (Figure 2-4). A pod is equally likely to be stored in any position within its zone. To minimize congestion, cross-aisles are placed every five pods. There are 134 pods of small-size SKUs, 272 pods of medium-size SKUs, and 518 pods of large-size SKUs in the forward area. (Note that, as in the miniload system, we slightly exceed the minimum requirements. For example, we need $2,000/4 = 500$ trays for small SKUs but the pods provide $134 \times 4 = 536$ positions.)

3. **AGVs:** 50 AGVs (with no downtime or battery charging) forms the baseline. Also, congestion is not modeled, although it is likely. This assumption may favor the Kiva system, but it is partially rectified by the cross-aisles. Also, in traveling between two
points, an AGV follows the shortest rectilinear path. (This assumption may again favor the Kiva system since congestion-avoidance may force a vehicle to take a different route. However, congestion-avoidance is beyond the scope of our study.) Each AGV travels at a speed of 260 fpm or 3 mph (Wulfraat, 2012). No acceleration/deceleration effects are considered. The pod P/D times are equal to 5 secs each. (The AGV spins in place to pick-up/deposit a pod.) At the start of the simulation, all the AGVs are idle and located immediately outside the front of the forward area.

4. **AGV control logic:** Each AGV follows a dual command (DC) cycle; i.e., when picking is completed, the AGV travels from the pick station to the forward area to store the pod; it then travels to the next pod to be retrieved (which is termed “travel between” or TB), and subsequently returns to the pick station with the next pod. (Matching an AGV with the next pod, i.e., FA versus CA, was described in Section 3.2.) If there are no requests in the retrieval queue, the AGV idles at the last location where it stored a pod.

### 4. Simulation Model and Results

A simulation model was developed and validated for each system with ProModel, where model development starts with adding elements to the physical layout and then identifying the entity that flows through the layout. For the miniload system, the elements in the layout are: the retrieval queue, the miniload aisles, the miniload input and output buffers, the conveyor loop, the pick station input and output buffers, and the pick stations. For the Kiva system, the elements in the layout are: the retrieval queue, the forward area, the pick station input buffers, and the pick stations. The AGVs are modeled as a dynamic resource (since each AGV moves through the system but is not an entity/customer). The customer orders were created outside of Promodel as an excel file.
An extensive number of runs were made to validate the model and ensure that it runs according to the assumptions and descriptions stated in Section 3. These runs include but are not limited to: 1. Runs made with and without infinite-capacity buffers at the miniload and the pick stations, 2. Runs made with and without infinite speed for the conveyor loop and infinite speed for the S/R machines, 3. Comparison of the analytic value of the expected analytic S/R machine cycle time (Bozer and White, 1984) with the expected S/R machine cycle times obtained from the simulation model, 4. Comparison of the expected and simulated travel time values for the Kiva system for different sizes of the forward area, 5. Tracing the flow of individual trays (and pods) through each system to verify the container flow logic, and 6. Verifying conservation of flow by checking the number of trays (or pods) entering and exiting each element in the layout.

Five sets of orders were generated to simulate five replications for each system. Each system was evaluated with four pick stations, and 5 or 8 open orders/pick station. The miniload system was evaluated with 4 and 5 aisles, while the Kiva system was evaluated with 50 and 70 AGVs, using FA or CA dispatching.

For each system, we report the 1. Total elapsed time (in hours) to process all 2,400 orders (the simulation is stopped when the last tray/pod is returned to storage); 2. Throughput of the system (line items picked/hour); 3. Expected utilization of the pickers and the S/Rs or the AGVs; 4. Expected time (in mins) to deliver a tray/pod to the pick station (measured from the time an order is opened to the time individual trays/pods arrive at the pick station buffer); and 5. Expected time (in mins) to complete an order (measured from the time an order is opened to the time the last line item for that order is picked). We also measured the contents of the I/O buffers, the conveyor loop, and the retrieval queue(s).
Since throughput is a key measure, we focused on those configurations that yield a high expected picker utilization. However, equipment utilization is also relevant since the S/Rs and the AGVs are capital-intensive. The results for the miniload system are shown in Tables 2-2 and 2-3. Providing a fifth miniload aisle does not appear to be justified; it increases the throughput by at most 2%, while the expected S/R utilization decreases from about 92% to 73%. As anticipated, the number of open orders/pick station is an important parameter. With 5 open orders/pick station, we obtain satisfactory results; if it’s increased to 8, the throughput increases by only 1.7%, while the order completion time increases by about 55% and the loop content increases by 66%. The tray delivery time increases as well. (If the number of open orders/pick station is less than 5, the system throughput decreases noticeably.)

Table 2-2. Miniload system simulation results.

<table>
<thead>
<tr>
<th>No. of ML aisles</th>
<th>Open orders/picker</th>
<th>Total elapsed time (hrs)</th>
<th>Throughput (line items/hr)</th>
<th>S/R machine utilization</th>
<th>Picker utilization</th>
<th>Tray delivery time (mins)</th>
<th>Order completion time (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>Avg. 8.014</td>
<td>700.706</td>
<td>0.913</td>
<td>0.977</td>
<td>2.218</td>
<td>3.965</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>7.943, 8.086</td>
<td>698.307, 703.105</td>
<td>0.906, 0.920</td>
<td>0.975, 0.980</td>
<td>2.195, 2.241</td>
<td>3.930, 4.000</td>
</tr>
<tr>
<td>8</td>
<td>Avg. 7.843, 7.961</td>
<td>712.395</td>
<td>709.675, 715.115</td>
<td>0.925</td>
<td>0.995</td>
<td>3.602</td>
<td>6.213</td>
</tr>
<tr>
<td>5</td>
<td>95% CI</td>
<td>7.804, 7.961</td>
<td>712.395, 715.115</td>
<td>0.925</td>
<td>0.995</td>
<td>3.602</td>
<td>6.213</td>
</tr>
<tr>
<td>8</td>
<td>Avg. 7.866</td>
<td>713.921</td>
<td>0.733</td>
<td>0.995</td>
<td>2.015</td>
<td>3.891</td>
<td>6.125</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>7.790, 7.942</td>
<td>712.200, 715.642</td>
<td>0.731, 0.735</td>
<td>0.994, 0.996</td>
<td>3.851, 3.931</td>
<td>6.123, 6.257</td>
</tr>
<tr>
<td>5</td>
<td>Avg. 7.854</td>
<td>714.993</td>
<td>0.736</td>
<td>0.997</td>
<td>3.492</td>
<td>6.190</td>
<td>6.125</td>
</tr>
<tr>
<td>8</td>
<td>95% CI</td>
<td>7.779, 7.929</td>
<td>713.014, 716.971</td>
<td>0.732, 0.740</td>
<td>0.996, 0.997</td>
<td>3.446, 3.538</td>
<td>6.123, 6.257</td>
</tr>
</tbody>
</table>
All the buffers and the conveyor loop appear to have adequate capacity (Table 2-3). The pick station input buffers reach capacity (5 trays) but their average content is below 5 trays, provided we use 5 open orders/pick station. With 4 miniload aisles and 5 open orders/pick station, the loop is, on average, only 38% full, and its maximum content reaches 38 trays (70% full). That is, there is no indication that the conveyor loop is a bottleneck. Owing to technological advances, the conveyor in our study is faster than the conveyor used by Perry et al. (1984). With 8 open orders/pick station, however, the average loop occupancy increases to 65%, and its maximum content reaches its capacity (54 trays). The average content of the retrieval request queue (across the 4 miniload aisles) is 2.30 and 2.91 requests, respectively, for 5 and 8 open orders/pick station. For validation purposes, we also simulated the case with infinite buffer capacities, which means no recirculation occurs on the loop due to blocking. The expected travel time on the loop from the simulation model is equal to 36.52 secs. Assuming each miniload aisle and each pick station is equally likely to be used, the analytic expected travel time on the loop is equal to 37.50 secs.

The results suggest that balancing the workload between the pickers and the S/Rs is important. In our particular case, balance is achieved when the number of miniload aisles is equal
to the number of pickers. In general, the appropriate ratio depends on the cycle time of the S/Rs versus the pick times. (The picker may pick multiple line items from the same tray but given the total number of SKUs, the small order sizes, and the number of SKU’s/tray, that is a small probability.)

With four aisles, the average simulated S/R cycle time is 10.27 secs and 17.97 secs, respectively, for a single command (SC) and dual command (DC) trip, including the P/D times. Using Bozer and White (1984), and assuming randomized storage over the entire rack, the analytic expected values of SC and DC are equal to 10.33 secs and 19.08 secs, respectively. The simulation values are slightly smaller since there’s a higher probability of visiting the zones with small and medium SKUs. Recall that all the SKUs have equal velocity but the number of SKUs per tray depends on the SKU size. As explained in Bozer and Cho (2005), a SC trip is I/O − (storage or retrieval) − I/O, while a DC trip is I/O − (store old tray) − (retrieve new tray) − I/O. Given “opportunistic interleaving” (see Section 3.3.2), about 42% of the S/R trips are SC, and 58% are DC, yielding an average S/R cycle time of 9.523 seconds/tray. That is,

\[
E(S/R \text{ cycle time/tray}) = (\% \text{ SC trips}) \cdot E(\text{SC cycle time}) + (\% \text{ DC trips}) \cdot E(\text{DC cycle time}/2) \quad (1)
\]

\[
E(S/R \text{ cycle time/tray}) = (0.42)(10.27) + (0.58)(17.97/2) = 9.523 \text{ secs/tray} \quad (2)
\]

Note that the DC cycle time is divided by 2 because two trays are handled on each trip. Since each tray is moved twice per pick, the E(S/R workload/pick) = 2(9.523) = 19.05 secs/pick, and the total expected workload per S/R, say, \( W_{SR} \), is given by:

\[
W_{SR} = \frac{E(S/R \text{ workload/pick})(N)}{n_{SR}} = \frac{(19.05)(N)}{4} = 4.7625(N), \quad (3)
\]

where \( N = \) number of picks (or line items) in the batch, and \( n_{SR} = \) number of S/R machines.

Likewise, let \( W_p \) be the total expected workload per picker. That is,
\[ W_P = \frac{E(\text{pick time/line item})(N)}{n_P} = \frac{(20)(N)}{4} = 5.00(N), \]  

where \( n_P \) = number of pickers. A balanced workload implies that \( W_{RATIO} = W_{SR}/W_P = 1 \), which holds true for our case since \( 4.7625/5 \approx 1.0 \).

The results for the Kiva system are shown in Tables 2-4 and 2-5. Providing more than 50 AGVs does not appear to be justified. With 5 open orders/pick station, if the number of AGVs is increased from 50 to 70, the expected AGV utilization decreases from 0.953 to 0.739, while the expected picker utilization improves by only 0.3%, which is not statistically significant at 5%. Also, with 50 AGVs, increasing the number of open orders/pick station from 5 to 8 slightly decreases the expected picker utilization.

Table 2-4. Kiva system simulation results

<table>
<thead>
<tr>
<th>Dispatch</th>
<th>No. of AGVs</th>
<th>Open orders/picker</th>
<th>Total elapsed time (hrs)</th>
<th>Throughput (line items/hr)</th>
<th>AGV utilization</th>
<th>Picker utilization</th>
<th>Tray delivery time (mins)</th>
<th>Order completion time (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>50</td>
<td>5</td>
<td>Avg. 7.868</td>
<td>713.721</td>
<td>0.955</td>
<td>0.988</td>
<td>1.091</td>
<td>3.877</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.789, 7.947</td>
<td>711.015, 716.427</td>
<td>0.952, 0.958</td>
<td>0.987, 0.990</td>
<td>1.077, 1.106</td>
<td>3.840, 3.914</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.951</td>
<td>706.255</td>
<td>0.989</td>
<td>0.978</td>
<td>3.251</td>
<td>6.249</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.878, 8.025</td>
<td>703.312, 709.198</td>
<td>0.986, 0.992</td>
<td>0.975, 0.981</td>
<td>3.204, 3.297</td>
<td>6.188, 6.310</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>5</td>
<td>Avg. 7.853</td>
<td>715.096</td>
<td>0.744</td>
<td>0.990</td>
<td>0.736</td>
<td>3.870</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.779, 7.927</td>
<td>712.165, 718.026</td>
<td>0.740, 0.748</td>
<td>0.987, 0.993</td>
<td>0.728, 0.744</td>
<td>3.835, 3.905</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.868</td>
<td>713.697</td>
<td>0.976</td>
<td>0.988</td>
<td>1.623</td>
<td>6.164</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.808, 7.929</td>
<td>713.170, 714.224</td>
<td>0.973, 0.979</td>
<td>0.986, 0.991</td>
<td>1.590, 1.655</td>
<td>6.106, 6.223</td>
</tr>
<tr>
<td>FA</td>
<td>50</td>
<td>5</td>
<td>Avg. 7.878</td>
<td>712.826</td>
<td>0.953</td>
<td>0.987</td>
<td>1.279</td>
<td>3.893</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.803, 7.953</td>
<td>710.993, 714.659</td>
<td>0.949, 0.957</td>
<td>0.984, 0.990</td>
<td>1.257, 1.301</td>
<td>3.855, 3.930</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 8.038</td>
<td>698.624</td>
<td>0.988</td>
<td>0.967</td>
<td>3.457</td>
<td>6.301</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.977, 8.099</td>
<td>696.474, 700.775</td>
<td>0.985, 0.991</td>
<td>0.965, 0.970</td>
<td>3.397, 3.516</td>
<td>6.232, 6.371</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>5</td>
<td>Avg. 7.851</td>
<td>715.243</td>
<td>0.739</td>
<td>0.990</td>
<td>0.929</td>
<td>3.872</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.777, 7.926</td>
<td>712.944, 717.543</td>
<td>0.735, 0.743</td>
<td>0.988, 0.992</td>
<td>0.925, 0.932</td>
<td>3.837, 3.907</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.856</td>
<td>714.872</td>
<td>0.978</td>
<td>0.990</td>
<td>1.794</td>
<td>6.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.773, 7.938</td>
<td>711.698, 718.047</td>
<td>0.977, 0.980</td>
<td>0.987, 0.992</td>
<td>1.754, 1.835</td>
<td>6.105, 6.225</td>
</tr>
</tbody>
</table>

30
A closer examination reveals that this is due to the manner in which the AGVs are initially allocated to the pick stations. When an order is opened by the first pick station, all the AGVs needed for that order are dispatched. With 8 open orders, and 2.325 line items/order, the first pick station seizes the first 18 AGVs or so. (The number may be slightly less than 18 since, with a very small probability, some pods may contain two or more picks). Hence, the first three pick stations combined seize approximately all 50 AGVs, which forces the fourth station to wait. (With 5 open orders/pick station, we would dispatch $5 \times 2.325 \times 4 \approx 46$ AGVs.) When 70 AGVs are used, as
expected, 5 versus 8 open orders/pick station has no significant impact on the expected picker utilization. An alternative AGV allocation scheme could have avoided penalizing the last pick station, but regardless of the allocation scheme used, if too many orders are opened at once, it would create an AGV shortage.

The expected DC travel time is equal to approximately 80 secs for the FA rule. Assuming randomized storage over the forward area, it is straightforward to obtain the expected travel time analytically, which is equal to 84.07 secs. As before, the simulation value is slightly smaller since there’s a higher probability of visiting the zones with small and medium SKUs. (The expected DC travel time with CA is equal to 67 secs/trip. The CA rule reduces the expected TB time from 27.60 secs to 14.85 secs.)

Since each pod is handled twice, adding a P/D time of 10 secs, we obtain a total cycle time of 90 secs (FA) and 77 secs (CA) per trip. However, this reduction has no noticeable impact on the expected picker utilization because it is already quite high under FA, and the time gained by the AGV is essentially lost waiting at the pick station buffer. For example, as shown in Table 2-5, with 50 AGVs and 5 open orders/pick station, when we switch from FA to CA, there is no significant increase in the expected picker utilization, while the average content of the pick station buffer increases from 6.45 to 7.10 AGVs. With 70 AGVs, the expected picker utilization remains at 0.990, while the average buffer content increases from 7.45 to 8.13 AGVs.

We also consider balancing the workload in the Kiva system. With 50 AGVs, 5 open orders/pick station, and the FA rule, the average queue of 6.45 AGVs at each pick station corresponds to approximately 129 secs of expected waiting time in the pick station buffer. Hence, assuming one pick per pod, we have

\[ E(\text{AGV workload/pick}) = E(\text{DC cycle time/pod}) + E(\text{wait time}) \]
(Interestingly, the expected time an AGV waits in the pick station buffer exceeds the expected DC cycle time/pod.) Hence, the total expected workload per AGV, say, $W_{AGV}$, is given by:

$$W_{AGV} = \frac{E(AGV \text{ workload/pick})}{n_{AGV}} = \frac{(90 + 129)(N)}{50} = 4.38(N),$$

where $N = \text{number of picks (or line items) in the batch}$, and $n_{AGV} = \text{number of AGVs}$. As before, $W_p = 5.00(N)$. Thus, $W_{Ratio} = W_{AGV}/W_p = 4.38/5 = 0.876$, which is close to 1.0. However, the expected wait time at the pick station (which appears in the numerator) depends on the number of AGVs, with fewer AGVs generally resulting in a shorter queue. Therefore, the relationship between $n_{AGV}$ and $W_{AGV}$ is more complicated than it appears, which suggests that the notion of a balanced workload for the Kiva system needs further investigation.

We next compare the two systems. Although the expected picker utilization is slightly better for the Kiva system (0.977 versus 0.987, which is statistically significant at 5%), the expected throughput of the Kiva system with 50 AGVs, FA dispatching, and 5 open orders/pick station is essentially equal to the expected throughput of a 4-aisle miniload system with 5 open orders/pick station. That is, one miniload aisle yields about the same expected throughput as 12-13 AGVs.

One may also compare the order completion times and the container (tray or pod) delivery times. (The delivery time is measured from the instant an order is opened to the instant the tray/pod enters the pick station buffer. Waiting time in the buffer is not included.) Although the average order completion times are comparable (3.965 mins for the miniload system versus 3.877 mins for the Kiva system), the Kiva system provides considerably shorter average container delivery times (2.218 versus 1.279 mins), which would be a desirable feature for high-priority orders; however,
it does not yield a higher expected throughput since both systems maintain high expected picker utilizations by queueing the containers in front of each picker.

In the miniload system, each tray spends, on average, 1.186 mins (out of 2.218 mins) on the loop, including possible recirculation. This value drops to 0.726 mins if the pick station input buffer capacity is unlimited and no recirculation takes place. The above result suggests that the miniload container delivery time can be reduced if the input buffer capacity is increased but given that the loop is, on average, 38% full (with a maximum value of 70%), we do not see a compelling reason to increase the buffer sizes. Further insight is gained for both systems by examining the percentage breakdown of the container delivery times (see Table 2-6).

Table 2-6. Percentage breakdown of the average container delivery time.

<table>
<thead>
<tr>
<th>Miniload System</th>
<th>Kiva System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waiting in the retrieval queue</td>
<td>Waiting in the retrieval queue</td>
</tr>
<tr>
<td>Retrieved by the S/R machine</td>
<td>Retrieved by the AGV</td>
</tr>
<tr>
<td>Waiting to enter the conveyor</td>
<td>2.90%</td>
</tr>
<tr>
<td>Traveling on the conveyor loop</td>
<td>54.00%</td>
</tr>
</tbody>
</table>

The two systems were further evaluated by increasing the number of SKUs from 6,000 to first 12,000 and then 24,000. As before, the SKUs are equally partitioned into three size categories, and each category experiences the same level of demand. For the miniload system, the rack size is increased proportionally to accommodate the additional SKUs. For the Kiva system, however, we increased the number of trays per pod from 4 to 5 trays, and then we proportionally increased the size of the forward area. The results are shown in Table 2-7 for the miniload system. With 12,000 SKUs, 4 miniload aisles is sufficient to yield an expected picker utilization of 90% with 5 open orders per picker. However, 5 miniload aisles are required to obtain a near-100% picker utilization. Likewise, with 24,000 SKUs, a 96% picker utilization is achieved with 5 miniload aisles but a sixth miniload aisle is needed to achieve a near-100% picker utilization. Hence, as the
number of SKUs is increased, we need additional miniload aisles to maintain a near-100% picker utilization, although it’s interesting to note that increasing the number of open orders per station from 5 to 8 also yields a high picker utilization without requiring an additional aisle.

Table 2-7. The miniload system with additional SKUs.

<table>
<thead>
<tr>
<th>No. SKUs</th>
<th>No. of ML aisles</th>
<th>Open orders/picker</th>
<th>Total elapsed time (hrs)</th>
<th>Throughput (line items/hr)</th>
<th>S/R machine utilization</th>
<th>Picker utilization</th>
<th>Tray delivery time (mins)</th>
<th>Order completion time (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,000</td>
<td>4</td>
<td>5</td>
<td>Avg. 8.014</td>
<td>700.706</td>
<td>0.913</td>
<td>0.977</td>
<td>2.218</td>
<td>3.965</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.943, 8.086</td>
<td>698.307, 703.105</td>
<td>0.906, 0.920</td>
<td>0.975, 0.980</td>
<td>2.195, 2.241</td>
<td>3.930, 4.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.883</td>
<td>712.395</td>
<td>0.925</td>
<td>0.995</td>
<td>3.602</td>
<td>6.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.804, 7.961</td>
<td>709.675, 715.115</td>
<td>0.921, 0.930</td>
<td>0.994, 0.996</td>
<td>3.568, 3.637</td>
<td>6.152, 6.275</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>Avg. 7.866</td>
<td>713.921</td>
<td>0.733</td>
<td>0.995</td>
<td>2.015</td>
<td>3.891</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.790, 7.942</td>
<td>712.200, 715.642</td>
<td>0.731, 0.735</td>
<td>0.994, 0.996</td>
<td>1.987, 2.043</td>
<td>3.851, 3.931</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.854</td>
<td>714.993</td>
<td>0.736</td>
<td>0.997</td>
<td>3.492</td>
<td>6.190</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.779, 7.929</td>
<td>713.014, 716.971</td>
<td>0.732, 0.740</td>
<td>0.996, 0.997</td>
<td>3.446, 3.538</td>
<td>6.123, 6.257</td>
</tr>
<tr>
<td>12,000</td>
<td>4</td>
<td>5</td>
<td>Avg. 8.641</td>
<td>649.868</td>
<td>0.951</td>
<td>0.903</td>
<td>2.675</td>
<td>4.274</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 8.574, 8.708</td>
<td>646.283, 653.453</td>
<td>0.947, 0.955</td>
<td>0.897, 0.910</td>
<td>2.662, 2.688</td>
<td>4.248, 4.300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 8.353</td>
<td>672.324</td>
<td>0.977</td>
<td>0.937</td>
<td>4.320</td>
<td>6.579</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 8.261, 8.445</td>
<td>666.730, 677.918</td>
<td>0.970, 0.984</td>
<td>0.927, 0.947</td>
<td>4.287, 4.353</td>
<td>6.515, 6.643</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>Avg. 7.905</td>
<td>710.335</td>
<td>0.818</td>
<td>0.991</td>
<td>2.068</td>
<td>3.902</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.849, 7.961</td>
<td>708.800, 711.870</td>
<td>0.816, 0.820</td>
<td>0.990, 0.993</td>
<td>2.046, 2.091</td>
<td>3.875, 3.929</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.865</td>
<td>714.001</td>
<td>0.823</td>
<td>0.996</td>
<td>3.520</td>
<td>6.199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.805, 7.925</td>
<td>711.144, 716.858</td>
<td>0.821, 0.826</td>
<td>0.996, 0.997</td>
<td>3.500, 3.539</td>
<td>6.151, 6.247</td>
</tr>
<tr>
<td>24,000</td>
<td>5</td>
<td>5</td>
<td>Avg. 8.107</td>
<td>692.641</td>
<td>0.914</td>
<td>0.964</td>
<td>2.302</td>
<td>4.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 8.045, 8.170</td>
<td>688.427, 696.856</td>
<td>0.911, 0.917</td>
<td>0.960, 0.969</td>
<td>2.292, 2.313</td>
<td>3.977, 4.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.861</td>
<td>714.320</td>
<td>0.936</td>
<td>0.993</td>
<td>3.666</td>
<td>6.202</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.798, 7.925</td>
<td>711.479, 717.161</td>
<td>0.934, 0.939</td>
<td>0.992, 0.995</td>
<td>3.637, 3.695</td>
<td>6.150, 6.253</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>Avg. 7.874</td>
<td>713.204</td>
<td>0.773</td>
<td>0.993</td>
<td>2.030</td>
<td>3.888</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.812, 7.935</td>
<td>711.799, 714.610</td>
<td>0.770, 0.775</td>
<td>0.992, 0.994</td>
<td>2.023, 2.036</td>
<td>3.857, 3.919</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg. 7.836</td>
<td>716.615</td>
<td>0.776</td>
<td>0.997</td>
<td>3.491</td>
<td>6.191</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95% CI 7.768, 7.904</td>
<td>714.231, 718.999</td>
<td>0.774, 0.779</td>
<td>0.996, 0.997</td>
<td>3.455, 3.527</td>
<td>6.136, 6.246</td>
</tr>
</tbody>
</table>
For the Kiva system, increasing the number of trays/pod helps but the number of trays/pod is not doubled when the number of SKUs is doubled, leading to an increase in the size of the forward area and slight reduction in throughput. However, the results for the Kiva system (Table 2-8) show that additional AGVs are not required due to the larger forward area. Rather, the AGVs spend more time traveling, which results in less time waiting in the input buffer of the pick station as shown in Table 2-9. Nonetheless, the retrieval time per pod shows a slight increase as the number of SKUs is increased. The results depicted in Figure 2-8 show the change in the system throughput as a function of the number of AGVs, the dispatching rule (FA vs CA), and the number of SKUs.
Table 2-9. Breakdown of AGV time per pod retrieval (in secs.)

<table>
<thead>
<tr>
<th>Dispatch</th>
<th>No. of SKUs</th>
<th>No. of AGVs</th>
<th>Open orders/picker</th>
<th>Ave. storage travel time</th>
<th>Ave. empty travel time</th>
<th>Ave. retrieval travel time</th>
<th>Ave. time in PS input buffer</th>
<th>Ave. total time per pod</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>6,000</td>
<td>50</td>
<td>5</td>
<td>29.53</td>
<td>14.85</td>
<td>29.54</td>
<td>145.44</td>
<td>219.36</td>
</tr>
<tr>
<td></td>
<td>12,000</td>
<td>50</td>
<td>5</td>
<td>33.84</td>
<td>18.07</td>
<td>33.84</td>
<td>134.85</td>
<td>220.60</td>
</tr>
<tr>
<td></td>
<td>24,000</td>
<td>50</td>
<td>5</td>
<td>43.15</td>
<td>26.06</td>
<td>43.14</td>
<td>113.16</td>
<td>225.51</td>
</tr>
<tr>
<td>FA</td>
<td>6,000</td>
<td>50</td>
<td>5</td>
<td>29.28</td>
<td>27.70</td>
<td>29.29</td>
<td>128.29</td>
<td>214.56</td>
</tr>
<tr>
<td></td>
<td>12,000</td>
<td>50</td>
<td>5</td>
<td>33.89</td>
<td>34.73</td>
<td>33.89</td>
<td>117.70</td>
<td>220.21</td>
</tr>
<tr>
<td></td>
<td>24,000</td>
<td>50</td>
<td>5</td>
<td>43.02</td>
<td>48.77</td>
<td>43.01</td>
<td>94.49</td>
<td>229.29</td>
</tr>
</tbody>
</table>

Figure 2-8. System throughput as a function of the number of AGVs.

A related and interesting question is the methodology we used for sizing the two systems for comparison purposes. We basically fixed the number of pick stations, and then adjusted in each system the number of miniload aisles and the number of AGVs to reach an expected picker
utilization of nearly 100%. For the miniload system, this resulted in a simple search since the desired number of miniload aisles (4 in our case) is quickly identified. For the Kiva system, we experimented with the number of AGVs to determine the appropriate number (50 AGVs), which is the basis of the result shown in Figure 2-8. More work is needed to develop a more systematic and perhaps more comprehensive methodology to compare OP systems.

In terms of cost, there are no detailed cost models published to methodically compute the cost of either system as a function of the design elements. However, as a ballpark comparison, according to Wulfraat (2013), a miniload system costs $500,000 to $750,000 per aisle, plus the cost of the warehouse management/control system, and an additional $100,000 to $250,000 for “services.” A warehouse with a Kiva system of 50 to 100 AGVs, on the other hand, is estimated to cost between $2 to $4 million. The cost of interfacing the system to the warehouse management system, or the cost for “services,” is not indicated. Other factors that impact the usability of both systems and potentially their long-term performance/cost are presented in the next section.

5. Other Qualitative Factors

Wurman et al. (2008) and Wulfraat (2012) point out a number of advantages of the Kiva system over PTG OP systems such as improved picking rates and picker productivity, increased order accuracy, etc. Many of the advantages they state apply to other GTP OP systems as well. Also, some limitations of the Kiva system are not covered adequately. In this section, we compare the miniload system and the Kiva system with respect to qualitative (or hard to quantify) factors that we believe most decision-makers will take into account, beyond those results presented in the previous section. (Since both systems represent GTP OP, we will not state their advantages over PTG OP systems. For such a comparison, the interested reader may refer to enVista, 2017.)
**Flexibility and scalability:** Both systems offer flexibility in the number of pick stations that can be activated (or staffed). However, both systems are limited by the capacity of the material handling system that retrieves/stores the containers. Although the conveyor loop can be designed to accommodate a future miniload aisle (Figure 2-1), the Kiva system is more advantageous since it provides *incremental scalability* by letting the user add more AGVs and/or pods as needed, with minimal interruption. (Too many AGVs, however, may cause congestion.) Incremental scalability may be particularly attractive for an e-tailer on a growth path. Also, if growth forces relocation, the Kiva system is essentially portable while the miniload system would be costly to relocate.

**Parallel processing and built-in redundancy:** The Kiva system provides substantial parallel processing since a pod can be transported by any one of the AGVs, and at any given time, there could be a large number of pods being moved to/from the pick stations. (This advantage is reflected in the faster container delivery times in the Kiva system as shown in the previous section.) Multiple AGVs also provide redundancy against breakdowns and scheduled maintenance. If an S/R breaks down, however, all the trays in that aisle become inaccessible. If the conveyor loop breaks down, all the pick stations would be impacted, unless the trays are moved manually. In contrast, if an AGV breaks down, the impact on the Kiva system would be minimal. (Of course, frequent breakdowns or extended downtimes in either system would be unacceptable.)

**Ergonomic factors:** The Kiva system requires the picker to bend down or reach up to pick from the lower and upper levels of a pod (Figure 2-9). In contrast, each tray in the miniload system is presented to the picker at an appropriate height. Furthermore, the tray is tilted towards the picker, and the platform on which the picker stands can be adjusted to the picker’s height, making the picking process ergonomically more “operator-friendly”. Also, with the miniload system, the picker has an unobstructed view of the tray. To reduce picking errors, both systems can be
equipped with pick-to-light (or similar) technologies. An excellent paper comparing the ergonomics of a Kiva-based and AS/RS-based pick station is presented by Lee et al. (2017), who used digital human modeling to assess the risk factors for work-related musculoskeletal disorders for the pickers. The authors conclude that OP systems based on AS/RS pose lower risk factors for human pickers while the pick station in Kiva-based systems is in critical need of posture change. For further information on ergonomics in OP systems, the reader may refer to Grosse et al. (2015), who stress that, in the process of increasing the efficiency of OP systems, one must integrate human factors into the model in order to enhance performance and reduce the long-term cost.

Figure 2-9. Kiva system picker ergonomics. (a) Picker reaching up to pick from top shelf of the pod (Photo: Quiet Logistics). (b) Picker bending to pick from bottom shelf of the pod; note the step ladder (Photo: Amazon Fulfillment)

Footprint and cube utilization: The Kiva system results in poor cube utilization and a large footprint for the forward area since the pods cannot be stacked. Building a mezzanine would alleviate this limitation but it would also require elevators. This would increase the container delivery times and the cost of the system, especially if there is considerable weight on the mezzanine. Also, the elevators may become a bottleneck. The miniload system, on the other hand,
provides high cube utilization since the storage rack can be as tall as the building allows. A tall structure may be especially advantageous in areas with expensive/limited land. According to enVista (2017), “The real benefit of GTP systems … is unparalleled vertical space utilization. AS/RS systems can be constructed up to 100 feet high.”

Security of items: Both systems offer item security since access to the forward area is restricted. However, the miniload system offers additional security since the trays are stored in tall racks and each aisle is occupied by an S/R. Furthermore, with the appropriate tray design, items remain secure in the tray. With the Kiva pods, on the other hand, since the sides are open, items may fall out. Elastic bands can secure the items (see Figure 2-9b) but they may also interfere with picking, forcing the picker to pull the band with one hand while picking with the other.

Battery charging: According to Wulfraat (2012), the AGVs “run on rechargeable lead-acid batteries that are charged at frequent intervals … such that there is no battery change-out process required. The (AGVs) simply travel to designated charge stations every couple of hours where they receive a 5-minute battery recharge…. For the purposes of budgeting, assume that at any time, 5% of the (AGVs) will be (unavailable) … When developing the business case for this technology, it is important to note that the batteries … require periodic replacement. For batteries that undergo consistent daily usage, the typical battery life cycle is in the order of 1.5-2 years ….”

6. Conclusions and Future Research
Using simulation, we compare the performance of a miniload AS/R system and a Kiva system, in an online retail setting where most orders have very few line items. Given the same set of orders, and the same number of pick stations and picking parameters for both systems, we compare the configuration/size of each system that yields approximately the same throughput in line items picked/hour. Given the assumptions and parameters of the study, our primary conclusion is that a
miniload system with 4 aisles and a conveyor loop yields approximately the same throughput as a Kiva system with 50 AGVs. We also show numerically that, for both systems, a favorable balance between picker and equipment (miniload or AGV) utilization is achieved when the expected container retrieval-storage cycle time is approximately equal to the expected pick time/container. The results also indicate that the number of open orders/pick station is an important control parameter to ensure high picker utilization while avoiding congestion. Although we varied the number of SKUs and the number of miniload aisles/number of AGVs, our conclusions are limited to the configurations/parameter values we tested. Additional studies are needed to compare the above two systems and other OP systems under alternative parameter values. Last, we compare the two systems on other factors such as flexibility/scalability, parallel processing/redundancy, ergonomics, footprint/cube utilization, and the security of items.

We used three categories of SKUs but assumed randomized storage within each category. Future studies may consider slotting based on SKU velocities. Both systems would benefit from storing high-velocity SKUs closer to the pick stations (although storing the fast-moving SKUs in the same area may create AGV congestion in the Kiva system). Another direction to explore is the impact/timing of replenishment in both systems. To increase the cube utilization, one may also study the impact of installing a mezzanine and elevator(s) for the Kiva system. Another configuration to investigate would be the miniload system with the conveyor loop replaced by unit load AGVs to transport individual trays.

A majority of the studies in OP examine or optimize certain aspects of a specific type of system. While such studies are very valuable, more studies that directly and quantitatively compare two or more types OP systems would enhance the literature and at the same time provide additional insights for decision-makers and potential users. While the throughput capacity of the
system and capital cost/return-on-investment are key factors in selecting an OP system, other factors (in particular flexibility/scalability and ergonomics at the pick stations) should also be taken into account.

References


Chapter 3

Expected Travel Distances in Automated Guided Vehicle-Based Order Picking Systems

Abstract:

Automated Guided Vehicle (AGV)-based order picking (OP) systems, also known as Robotic Mobile Fulfillment Systems (RMFS), continues to receive attention both in industry and academia since their introduction under the name of Kiva systems. A key component of AGV-based OP systems is the fleet of “robots” (or AGVs) that pick up the “pods” (or “racks”) and transport them to the appropriate pick station (PS), where a human picker picks the items ordered by customers. Their performance depends on the number of AGVs, which in turn depends on the time it takes the AGV to retrieve a pod from the storage area. To aid system designers, we derive closed-form analytic expressions for the expected travel distance of the AGVs operating under two possible order assignment rules. Under the random assignment rule, an order is assigned to any PS with equal probability. Under the closest assignment rule, the order is assigned to the closest PS. We also examine the impact of the shape of the storage area on the expected travel distance of the AGVs and the impact of alternative PS configurations. The results offer valuable insight concerning the expected travel distances and are needed for analytic design and performance evaluation models.
1. Introduction

In goods-to-person (GTP) order picking (OP) systems, which have been in use since the 1960s (MHI 2018), a material handling system is utilized to retrieve the containers from the storage area, typically one at-a-time, and deliver them to the appropriate pick station (PS). Once picking is performed (often by a human working at the PS), the container is returned to the storage area by the same material handling system.

Generally speaking, GTP OP systems eliminate the need for the picker to walk/travel in the storage area, which can significantly increase the pick rates and lead to labor cost savings. Furthermore, GTP OP systems can increase picking accuracy, yield better cube utilization inside the facility, and improve picker ergonomics (enVista 2017).

Various types of material handling systems have been used to implement GTP OP. One of the common systems used is automated storage and retrieval (AS/R) systems. Although the term AS/R today refers to a variety of technologies, originally the term was used for storage systems with high-speed S/R machines operating in very narrow aisles (MHI 2018). Another type of material handling system used in GTP OP is horizontal or vertical carousels, which consists of a set of bins that revolve on a track. The reader may refer to Roodbergen and Vis (2009), among others, for a review of GTP OP systems. More recently, alternative types of GTP OP systems have been developed that are particularly suitable for online retail warehouses. One such example, and the focus of our paper, is the Robotic Mobile Fulfillment System (RMFS), which is an Automated Guided Vehicle (AGV)-based system that was introduced originally as a “Kiva system” and is now part of Amazon Robotics (Boysen, De Koster, and Weidinger 2018). The interested reader may refer to Boysen, De Koster, and Weidinger who present a good overview of the challenges faced by online retail warehouses and the warehousing systems suitable for such applications.
A typical configuration used for an AGV-based OP system is shown in Figure 3-1. In the storage area, which is known as the forward area (FA), multi-tier racks (also known as pods) are arranged in blocks that are accessible through narrow aisles and cross-aisles. PSs are placed along the perimeter of the FA. (The PSs in Figure 3-1 are located in the front of the system; however, depending on the implementation, the PSs can be placed on the opposite two sides or all four sides of the FA.) Each pod needed to fill an order is retrieved, one at-a-time, by one of the AGVs (which are also known as robots). The empty space between the FA and the PSs allows the AGVs to queue up in front of the PSs, waiting for the picker to pick the appropriate items from each pod. Once picking is completed, each pod is returned to the FA by the same AGV. A station may serve strictly as a PS, or it may also serve as a replenishment station as needed.

The number of AGVs in the system plays a critical role not only because one must ensure that the pods are retrieved and stored at a rate needed to support the PSs but also because the AGVs represent a major component of the system cost. In existing applications, the number of AGVs per installation ranges from 50 to several hundred (Bozer and Aldarondo 2018). Ackerman (2018) cites Amazon VP of Robotics saying over 100,000 AGVs are deployed over Amazon’s global fulfillment network.

One or more pods must be retrieved to fill each order, depending on the number of line items in the order. (Typically, one line item is picked from each pod but in some cases, two or more line items may be picked from the same pod.) The AGVs retrieve and then store each pod, one at-a-time, using dual command (DC) cycles that consist of three legs. First, after depositing the current pod, the AGV travels empty from its current location to the location of the next pod to be retrieved (the “travel between” leg). It then picks up the pod and travels to the appropriate PS (the “out” leg). Once the item(s) are picked, the AGV returns the pod to the FA (the “back” leg),
and it becomes available to retrieve another pod. A similar cycle occurs for replenishing a pod; i.e., an AGV retrieves the pod and transports it to the appropriate station where the pod is replenished and then returned to the FA by the same AGV.

Since the introduction of Kiva systems, the emerging field of AGV-based OP has captured the attention of the online retail order picking community, both in academia and industry. One key advantage of an AGV-based OP system is its incremental scalability; that is, the user can add AGVs, pods, and/or PSs as needed, with minimal disruption. Other unique features are built-in redundancy and parallel processing. Multiple AGVs provide redundancy against breakdowns, since any pod can be transported by any AGV. Also, multiple pods can be moved concurrently to/from the PSs at any given time (Bozer and Aldarondo 2018). Since Kiva became part of Amazon Robotics in 2012, there has been a proliferation of competing AGV-based OP systems developed by vendors worldwide. A few examples are Swisslog’s CarryPick (Switzerland), Grey Orange’s Butler System (India), and Hitashi’s Racrew (Japan). For brevity, in the remainder of the paper, we will refer to an AGV-based OP system simply as AGV-OPS.

Figure 3-1. Typical AGV-OPS with PSs in the front. (Taken from Bozer and Aldarondo 2018).
In this paper we study the expected travel distance of DC cycles in an AGV-OPS as a function of the order assignment rule, the shape of the FA, and the configuration of the PSs around the FA. We consider two order assignment rules, namely, random (RAN) and closest (CLO). Under RAN, all the pods associated with an order are assigned to one of the PSs with equal probability. Under CLO, the pods associated with an order are assigned to the PS that minimizes the total distance traveled to the PS. Once the item(s) are picked from a pod, it can be returned either to its original location or to any open location in the FA. We consider three “symmetric shapes” for the FA; i.e., a square, a diamond, and a circle. For each shape, we also consider its “asymmetric” counterpart; i.e., a rectangle, an elongated diamond, and an ellipse. Last, we consider five possible PS configurations. The PSs may be located on 1. The short side of the FA, 2. The long side of the FA, 3. The two short sides of the FA, 4. The two long sides of the FA, and 5. All four sides of the FA. For symmetric shapes, cases 1 and 2, and cases 3 and 4, are of course the same.

The results we derive offer valuable insight concerning the expected travel distances and ultimately the throughput capacity of the system as a function of the above parameters, and they would also be used as input for analytic design and performance evaluation models.

2. Literature Review

Although AGV systems, as a technology, date back to the 60s, their use in a GTP OP system, as developed by Kiva and their unique AGV/robot, is relatively new. Among the first and well-publicized applications of AGV-OPS are Staples (2007), Walgreens (2007), and Quiet Logistics (2009). The number of academic papers on AGV-OPS is limited but recently there has been a noticeable increase. In fact, Azadeh, De Koster, and Roy (2019) present a recent survey of academic studies on the subject. Among the first publications on AGV-OPS are general-audience
maga

zine articles (see, for example, Guizzo (2008), Scanlon (2009), Steiner (2009), Kopytoff (2012), and Mountz (2012)), professional magazine articles (see, for example, D’Andrea and Wurman (2008), and D’Andrea (2012)), expert reviews (Wulfraat (2012)), and refereed journal articles (Wurman (2009), and Enright and Wurman (2011)). These publications mostly describe the system and its operations, and they primarily discuss the benefits of AGV-OPS over person-to-goods OP systems. In the remainder of this section, we focus on studies involving the analytic modeling of AGV-OPS.

First, we consider academic studies that use queuing models to assess the throughput of AGV-OPS. Work by Roy et al. (2014) has been often cited as a working paper but recently it was published as Roy et al. (2019). Assuming randomized storage, the authors derive the first two moments of the AGV travel time based on dual command trips. They restrict the two pod locations (the pod being returned and the one being retrieved) to the same aisle, and the results are based on a system with a single PS located at the midpoint of one side of the FA. Given these assumptions, the authors develop queuing network models that estimate the system throughput for single-line orders. To account for congestion, each aisle is modeled as a single server queue with a service time proportional to the length of the aisle. This modeling approach implies that two or more AGVs cannot occupy the same aisle at one time. In addition, the authors assess the impact on throughput of the width-to-length ratio of the FA. They recommend that a system be designed with many short aisles in order to maximize throughput. This recommendation is perhaps, at least in part, a result of the above assumption where each aisle can accommodate only one AGV at a time. With no congestion, our results indicate that the above design policy may reduce the expected travel distance, but we also show instances when it may have the opposite effect.
Subsequent work by Nigam et al. (2014) extend Roy et al.’s travel time results to a 3 class-based storage system. To account for class-based storage, the authors develop multi-class closed queuing networks. They conclude that the closest-open location policy increases the system throughput, although it does not allow for the efficient use of the storage space, compared to a random storage policy.

Lamballais, Roy, and De Koster (2016) use a closed queuing network model to assess the maximum throughput of a multi-picker AGV-OPS by modeling each PS as a single-server system with dedicated AGVs. The authors claim that the system throughput is not sensitive to the length-to-width ratio of the FA but it is affected by the location of the PSs around the FA. In our study, we show that the impact of the width-to-length ratio and the location of the PSs on the expected travel time depends on the rule used to assign the pods to the PSs. We also note that the model in Lamballais, Roy, and De Koster (2016) relies on travel time distributions obtained numerically from detailed AGV travel rules (such as each aisle and cross-aisle allowing travel only in a single direction). Subsequently, Lamballais, Roy, and De Koster (2019) present a new type of semi-open queuing network (SOQN) to model AGV-OPS such that orders are matched only to those pods containing the appropriate Stock Keeping Unit (SKU) in the appropriate quantity. The authors study the optimal number of pods per SKU, the ratio of the number of PSs to the number of replenishment stations, and the replenishment level per pod. They show that the throughput increases when inventory is spread across multiple pods and when the ratio between the pick and replenishment stations is optimized. Finally, Yuan and Gong (2016, 2017) present very similar work; the authors develop open queuing network models that allow for multi-picker systems with dedicated and pooled AGVs. Their approach does not account for an AGV’s wait time at the PS queue, and the simultaneous possession of the AGV and the picker while picking is performed.
Next we consider academic studies that focus on operational decisions encountered in AGV-OPS. Boysen et al. (2017) determine how to batch the orders, their processing sequence, and the arrival sequence of the pods delivered by the AGVs. Their objective is to minimize the number of pod visits to a PS or conversely, the number of AGV trips required to fill a set of orders. The authors consider a single-PS system that can process a bounded number of customer orders in parallel, and they assume a fixed set of pods, with SKUs present in more than one pod, where each pod carries a sufficient quantity to satisfy any order. They provide a decomposition approach to solve the problem and conclude that an optimized order processing and pod sequencing scheme can reduce the number of AGVs needed. They also show that the diversity of SKUs stored in a pod can impact the performance of the system and the number of AGVs. Xiang et al. (2018) study the slotting problem (although they use the term “storage assignment problem”) in addition to the order batching problem. They present a mixed integer linear program to determine which SKUs to put into which pod(s) to maximize the SKU similarity, defined by the likelihood that the SKUs appear together in a customer order.

Yuan, Graves, and Cezik (2019) study the storage assignment problem by comparing turnover-based storage against random assignment. (The storage assignment problem is deciding which storage location within the FA to return a pod to, upon the completion of a pick or replenishment operation.) The authors consider two full turnover and class-based storage policies. They develop a fluid model to compute the expected travel distance of the AGVs to retrieve and store the pods. However, their model does not include empty AGV travel between two pod locations. (The expected travel time expressions we develop in our study are based on dual command cycles, which include travel between two pod locations.) Also, their model assumes that the distance from the $x$-th percentile closest pod to a PS is a linear function of $x$. Unfortunately,
the authors do not assess or quantify the error introduced by this assumption. Furthermore, the linear function measuring the distance between a PS and the pod fits a specific FA shape and, therefore, the model does not analytically extend to alternative FA shapes. Yuan, Graves, and Cezik assume that the pods are assigned to the closest PS, which is also explored in our study (in addition to the random assignment).

Zou et al. (2017) consider the assignment of PSs to the AGVs. They determine which PS to assign to an AGV in order to minimize the pod retrieval time. They propose an assignment rule based on pick times and design a neighborhood search algorithm to find a near-optimal assignment of pods to PSs. The authors also develop an SOQN model to assess the performance of the system under the random assignment rule and the pick times-based rule. The model allows an AGV to visit any PS according to a set of predetermined probabilities whose values are determined to reflect different assignment rules. The study also explores the impact of the pod block size. The authors claim that the optimal width of the pod block decreases with the width-to-length ratio of the FA. In a subsequent study, Zou et al. (2018) compare the performance of battery charging and battery swapping strategies via SOQN models.

Feng et al. (2018) develop an integer programming model to determine the optimal location of the PSs in an AGV-OPS with the objective of minimizing the total travel distance. Their discrete location model limits the set of PS candidate locations to points on the perimeter of the FA that align with the aisles and cross-aisles. The study considers both a traditional layout and a flying-V layout (Gue and Meller 2009) under the random storage assignment policy. The model considers the travel distance from any storage location to any PS in order to simultaneously optimize the location of the PSs and the allocation of storage locations to PSs.
3. Analysis of the Factors That Impact the Expected Travel Distance

Several factors may influence the distance traveled by an AGV in an AGV-OPS. In this section, assuming rectilinear AGV travel, we analyze the impact of the following factors: (1) The shape of the FA, (2) The order assignment rule, and (3) The configuration of PSs around the FA.

3.1. Shape of the FA

We consider three symmetric shapes, i.e., a square, a diamond, and a circle, as well as their asymmetric counterparts, i.e., a rectangle, an elongated diamond, and an ellipse. A rectangular FA is defined by its length and width. The area inside the rectangle constitutes the FA, and the PSs are located on the perimeter. An elongated diamond is obtained by rotating a rectangle 45 degrees. Since the AGV travels rectilinearly, its travel trajectory intersects the sides of the FA at 45-degree angles. As a result, although the elongated diamond consists of a traditional layout with aisles and cross-aisles, neither the aisles nor the cross-aisles are parallel to the sides FA. The ellipse, on the other hand, is defined by the length of its semi-major axis $a_1$ and the length of its semi-minor axis $a_2$ as shown in Figure 3-2(c).

![Figure 3-2. Asymmetric FA shapes.](image)

(a) Rectangle
(b) Elongated Diamond
(c) Ellipse
For each shape, the width-to-length ratio, i.e., W/L and a2/a1, for the rectangle/elongated diamond and the ellipse, respectively, is also considered. (When the above ratio is equal to one, we obtain of course the symmetric shapes, i.e., a square, a diamond, and a circle.) Although a rectangular FA is the most common shape in practice, we compare different FA shapes and width-to-length ratios to investigate alternative shapes and their impact on expected travel distances.

Varying the shape and/or the above ratio will change the area and/or perimeter of the FA. In comparing alternative shapes, one may keep the area constant, or keep the perimeter constant, or aim for a compromise between the first two options. We argue in favor of keeping the area constant with the following two observations:

1. There is only a 12% difference between the perimeter of a unit square and a unit circle. In contrast, there is a 27% difference between the area of a unit square and a circle of equivalent perimeter, i.e., a perimeter equal to four units.

2. There is a 25% difference between the perimeter of a unit square and a rectangle of unit area with a width-to-length ratio of 0.25. In contrast, there is a 18.5% difference between the perimeter of a unit square and a rectangle of area 0.90 with a width-to-length ratio of 0.25. That is, allowing for a 10% difference in the area reduces the difference in the perimeter by less than 10%.

Furthermore, by keeping the area constant, we ensure that different FA shapes can accommodate the same number of pods, which is important in terms of accommodating the same amounts of inventory. Of course, this also means that the perimeter may vary by shape and the width-to-length ratio. Since the PSs are located along the perimeter, shapes with larger perimeter can potentially accommodate more PSs. However, the perimeter is unlikely to become a constraint since the PSs are often positioned far apart.
3.2. Order Assignment Rules

The order assignment rule is concerned with selecting a PS for each pod required to fill an order. (Some orders require a single pod while others may require two or more pods. If multiple pods are required, they must all be assigned to the same PS in order to maintain order integrity.) We consider two order assignment rules, namely, RAN and CLO. Under RAN, a pod is equally likely to be taken to any PS. After picking is performed, the pod is returned to any open location in the FA. Alternatively, under CLO, the pod is assigned to the closest PS. If multiple pods are involved, they are assigned to the PS that minimizes the total distance traveled across all the pods. Once the items are picked, each pod is returned to their original location in the FA.

Recall from Section 1 that the AGVs perform DC cycles consisting of three legs. It is important to note that the order assignment rule affects only the second and third legs; i.e., travel from a pod location in the FA to a PS, and travel back to a pod location in the FA from the PS. We assume that travel between two pod locations in the FA is not impacted by the order assignment rule; please refer to Section 4.

3.3. Configuration of the PSs

In a typical AGV-OPS, the PSs are located on the perimeter of the FA. For a rectangle or elongated diamond, the perimeter is clearly composed of four sides. For the ellipse, on the other hand, we will define the “four sides” by dividing the ellipse into four regions as shown in Figure 3-2(c). The regions are delimited by diagonal lines that go through the center of the ellipse.

The configuration of the PSs along the perimeter of the FA will impact the expected travel distance, except for symmetric shapes under RAN rule. We consider the cases in which the PSs are located either on one side of the FA, or the two opposing sides, or around the entire perimeter (four sides). More specifically, we consider the five PS configurations below:
1S – the PSs are located along one of the two short sides of the perimeter.
1L – the PSs are located along one of the two long sides of the perimeter.
2S – the PSs are located along the two short sides of the perimeter.
2L – the PSs are located along the two long sides of the perimeter.
4 – the PSs are located along the entire perimeter.

3.4. Notation
To simplify our presentation, the A/B/C/D notation is used to specify the PS configuration (A), the shape of the FA (B), the order assignment rule (C), and the number of pods in an order (D). For example, 1S/RECT/RAN/1 refers to the case where the PSs are located on one of the short sides of a rectangular FA, in which the RAN rule is used, and each order contains a single pod. (The fourth parameter is needed since the expected travel distance may depend on the number of pods.) We use the symbol (N) for expressions that involve three or more pods.

4. The Expected Travel Distance Model
Recall that a DC cycle is composed of three legs, where \( TB \) = travel empty between two pods in the FA, \( OUT \) = travel out from the FA to a PS, and \( BACK \) = travel back into the FA. By definition, the expected DC distance, \( E[DC] \), is equal to:

\[
\]

Since \( E[OUT] = E[BACK] \) due to symmetry, we have:

\[
E[DC] = E[TB] + 2E[OUT].
\]

We first normalize the FA (Bozer and White 1984). Given \( W \) and \( L \), we define \( T \) (the “scaling factor”) and \( b \) (the “shape factor”) as follows:

\[
T = \max (W, L),
\]
The normalized FA, with dimensions 1 and $b$, simplifies the computation of the expected distances. In the case of an ellipse, the FA is normalized such that the semi-major axis has a length of 0.5, and the semi-minor axis has length of $b/2$. That is, given $a_1$ and $a_2$, we let:

$$T = 2 \max (a_1, a_2), \quad \text{and}$$

$$b = \frac{2 \min (a_1, a_2)}{T}. \quad (6)$$

Given the expected travel distances for a normalized FA (identified with a tilde), the results for the original FA can simply be obtained by multiplying with $T$. That is,

$$E[DC] = \tilde{E}[DC] \cdot T, \quad (7)$$

$$E[TB] = \tilde{E}[TB] \cdot T, \quad (8)$$

$$E[OUT] = \tilde{E}[OUT] \cdot T. \quad (9)$$

### 4.1. Assumptions

The following assumptions are used to derive the expected travel distances:

1. When traveling between any two points, an AGV follows the shortest rectilinear path.
2. Pods are equally likely to be located anywhere inside the FA. This assumption is justified for randomized storage, but it may not be applicable to turnover-based storage methods.
3. The PSs are modeled as continuous points that are equally likely to be located anywhere along the perimeter of the FA. Consider two systems with 16 PSs each, located equidistant on the perimeter. One system has a square shape with four PSs on each side. The other system is rectangular with $b = 2/3$, and 3 (5) PSs on each short (long) side. Modeling the PSs as continuous points overestimates $E[OUT]$ only by 4.1% and 4.4% for the square and
rectangular FA, respectively. As the number of PSs increases, this assumption becomes more tenable.

4. We do not consider the additional space between the FA and the PSs. The distance between the edge of the FA and the PSs is constant and it can be added to the expected distance results we derive.

5. Under RAN, a pod is equally likely to be returned to any location in the FA. Under CLO, a pod is returned to its original location in the FA. (Under RAN, if the pod is returned to its original location, our results would still hold.)

6. An available AGV is randomly assigned to any pod to be retrieved from the FA. That is, if there are multiple pods to be retrieved, there is no attempt to minimize $TB$ by assigning the AGV to the closest pod. Hence, $TB$ is always defined as the distance between two random points in the FA. Note that $TB$ is not impacted by the order assignment rule or the PS configuration.

7. Under CLO, depending on their locations, certain PSs may receive more pods than others. We will address this observation later in the paper, after deriving the expected distances.

8. The results are focused on expected travel distances as opposed to expected travel times since the latter would be impacted by the acceleration/deceleration of the AGVs and possible congestion in the system, which are factors beyond the scope of our study.

4.2. Symmetric FA Shapes

We consider first symmetric shapes, i.e. a square, a diamond, and a circle, of unit area, and present numeric results for $E(OUT)$ and $E(DC)$ to show the impact of the shape of the FA, RAN vs CLO, and the PS configuration as a function of the number of pods per order ($N$). The results, i.e., the
expected distance per pod for $E(OUT)$ and $E(DC)$, are shown in Table 3-1 (see Figure 3-2) and Table 3-2 (see Figure 3-3), respectively.

The results in Table 3-2 are obtained by doubling the results in Table 3-1 and adding the appropriate $E[TB]$ value. Recall that only the shape of the FA will impact $E[TB]$. For a FA of unit area, $E[TB]$ is equal to 0.667, 0.660, or 0.650, for a square, a diamond, and a circle, respectively (see Chapter 3.7.1 of Larson and Odoni 1981).

Since there is no distinction between the long and short sides of a symmetric shape, we consider only three PS configurations in this section, i.e., PSs on one side (1), two opposing sides (2), and all four sides (4). We also compare RAN and CLO; see Tables 3-1 and 3-2 for the percent difference in expected distances between the two rules. As described in Section 3.2, under CLO, all the pods for an order are assigned to a single PS. As a result, the expected distance per pod increases as $N$ increases. Under RAN, on the other hand, as anticipated, the expected distance per pod does not change with $N$.

Table 3-1. $E[OUT]$ per pod for symmetric FA shapes.

<table>
<thead>
<tr>
<th>PS config</th>
<th>FA shape</th>
<th>N = 1</th>
<th>N = 2</th>
<th>N = 3</th>
<th>N = 5</th>
<th>N = 10</th>
<th>N = 15</th>
<th>N = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SQRE</td>
<td>0.833</td>
<td>0.500</td>
<td>40%</td>
<td>1.667</td>
<td>1.333</td>
<td>20%</td>
<td>2.500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DIAM</td>
<td>0.825</td>
<td>0.707</td>
<td>14%</td>
<td>1.650</td>
<td>1.447</td>
<td>10%</td>
<td>2.475</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CIRC</td>
<td>0.814</td>
<td>0.508</td>
<td>38%</td>
<td>1.627</td>
<td>1.366</td>
<td>16%</td>
<td>2.441</td>
</tr>
</tbody>
</table>
Figure 3.3. $E[OUT]$ per pod as a function of $N$.

Table 3.2. $E[DC]$ per pod for symmetric FA shapes.

<table>
<thead>
<tr>
<th>PS config</th>
<th>FA shape</th>
<th>$N = 1$</th>
<th></th>
<th>$N = 2$</th>
<th></th>
<th>$N = 3$</th>
<th></th>
<th>$N = 5$</th>
<th></th>
<th>$N = 10$</th>
<th></th>
<th>$N = 20$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
<td>RAN CLO</td>
<td>% Diff</td>
</tr>
<tr>
<td>SQRE 2.333</td>
<td>1.667</td>
<td>29%</td>
<td>4.667</td>
<td>4.000</td>
<td>14%</td>
<td>7.000</td>
<td>6.000</td>
<td>14%</td>
<td>11.667</td>
<td>10.333</td>
<td>11%</td>
<td>23.333</td>
<td>21.212</td>
</tr>
<tr>
<td>DIAM 2.310</td>
<td>2.074</td>
<td>10%</td>
<td>4.620</td>
<td>3.214</td>
<td>9%</td>
<td>6.930</td>
<td>4.657</td>
<td>7%</td>
<td>11.549</td>
<td>10.689</td>
<td>7%</td>
<td>23.097</td>
<td>21.669</td>
</tr>
<tr>
<td>CIRC 2.361</td>
<td>1.799</td>
<td>26%</td>
<td>4.722</td>
<td>3.190</td>
<td>11%</td>
<td>7.083</td>
<td>6.222</td>
<td>12%</td>
<td>11.806</td>
<td>10.701</td>
<td>9%</td>
<td>23.611</td>
<td>22.087</td>
</tr>
<tr>
<td>SQRE 2.333</td>
<td>1.187</td>
<td>50%</td>
<td>4.667</td>
<td>3.333</td>
<td>29%</td>
<td>7.000</td>
<td>5.202</td>
<td>26%</td>
<td>11.667</td>
<td>9.303</td>
<td>20%</td>
<td>23.333</td>
<td>19.753</td>
</tr>
<tr>
<td>DIAM 2.310</td>
<td>1.367</td>
<td>41%</td>
<td>4.620</td>
<td>3.309</td>
<td>28%</td>
<td>6.930</td>
<td>5.337</td>
<td>23%</td>
<td>11.549</td>
<td>9.386</td>
<td>19%</td>
<td>23.097</td>
<td>19.883</td>
</tr>
<tr>
<td>CIRC 2.361</td>
<td>1.208</td>
<td>46%</td>
<td>4.722</td>
<td>3.555</td>
<td>25%</td>
<td>7.083</td>
<td>5.432</td>
<td>23%</td>
<td>11.806</td>
<td>9.781</td>
<td>17%</td>
<td>23.611</td>
<td>20.983</td>
</tr>
<tr>
<td>SQRE 2.333</td>
<td>1.000</td>
<td>57%</td>
<td>4.667</td>
<td>2.867</td>
<td>39%</td>
<td>7.000</td>
<td>5.193</td>
<td>26%</td>
<td>11.667</td>
<td>9.291</td>
<td>20%</td>
<td>23.333</td>
<td>17.736</td>
</tr>
<tr>
<td>DIAM 2.310</td>
<td>1.131</td>
<td>51%</td>
<td>4.620</td>
<td>2.867</td>
<td>35%</td>
<td>6.930</td>
<td>5.034</td>
<td>25%</td>
<td>11.549</td>
<td>9.021</td>
<td>22%</td>
<td>23.097</td>
<td>19.217</td>
</tr>
<tr>
<td>CIRC 2.361</td>
<td>1.133</td>
<td>52%</td>
<td>4.722</td>
<td>3.290</td>
<td>30%</td>
<td>7.083</td>
<td>5.253</td>
<td>26%</td>
<td>11.806</td>
<td>9.455</td>
<td>20%</td>
<td>23.611</td>
<td>20.519</td>
</tr>
</tbody>
</table>

Figure 3.4. $E[DC]$ per pod as a function of $N$. 

63
The expected distances shown in Tables 3-1 and 3-2 are obtained via either analytic results or Monte Carlo sampling (gray cells), which was used when closed-form expressions could not be obtained. The Monte Carlo results are based on 10,000 trips per replication and 10 replications. Tables 3-1 and 3-2 show $E[OUT]$ and $E[DC]$ averaged over the ten replications.

We obtain analytic $E[OUT]$ results for */SQRE/RAN/* and */DIAM/RAN/*. Under CLO, analytic results are obtained for */DIAM/CLO/1, */SQRE/CLO/1, and */SQRE/CLO/2. (The analytic results are derived in Appendix A through D.) For $N > 2$, only 1/SQRE/CLO/N is solved exactly. An analytic approximation is derived for 2/SQRE/CLO/N in Appendix E and discussed in Section 4.3. Likewise, an approximation for 4/SQRE/CLO/N is derived in Appendix F and discussed in Section 4.3.

Appendix G presents an approximate analytic result for $E[OUT]$ in */CIRC/RAN/*. The approximation is based on deriving the exact solution to $E[OUT]$ for an octagon-shaped FA with the same area as the circle. The octagon is constructed from a square (that envelops the circle) by removing the square’s four corners as shown in Figure G-1 in Appendix G. Each such corner consists of an isosceles triangle with sides $c$, $c$, and $\sqrt{2}c$. For a circle with a radius of 1 and area of $\pi$, $c$ is set equal to $\sqrt{2(1 - \pi/4)}$ such that the resulting octagon also has an area of $\pi$. The Monte Carlo results for $E[OUT]$ – based on 10,000 trips per replication and 10 replications – show that the approximation works well. In a FA of unit area, the mean (standard deviation) is equal to 0.813 (0.001) and 0.816 (0.001) for */CIRC/RAN/* and */OCTA/RAN/*, respectively. This represents an error of less than 1% when replacing the circle with an octagon of the same area.

From Figures 3-3 and 3-4, we note that, for RAN, the circle attains the smallest expected distances while the square yields the largest values. However, the difference is only about 1%. Under CLO, however, we observe larger differences in expected distances across the three shapes.
The square has the smallest values, while the diamond generally has the largest values for small \( N \) (less than 5) and the circle has the largest values for larger \( N \).

Under CLO, the expected distances are reduced significantly compared to RAN - as much as 80% (57%) for \( E[OUT] \) (\( E[DC] \)). However, the percent difference between RAN and CLO decreases as \( N \) increases. Since the pods are equally likely to be anywhere in the FA, as \( N \) increases, the expected distance per pod to any point on the perimeter (including the closest point) approaches the expected distance to a random point on the perimeter.

We observe in Table 3-2 that: 1. The diamond shows the smallest reduction in expected distances when CLO is used, while the square shows the largest reduction; we believe this is partially explained by the fact that in a square (diamond), the AGV travels in a straight (diagonal) path towards the sides of the FA; 2. For a given shape, the percent difference between RAN and CLO decreases significantly if the PSs are located on only one side of the FA; when there are PSs on all four sides, CLO performs better as expected; and 3. As noted above, for all FA shapes, the percent difference between RAN and CLO decreases for large values of \( N \). However, if the PSs are located on two or four sides of the FA, at least a 20% reduction can be obtained with CLO as long as \( N \leq 5 \). In most cases the difference between CLO and RAN is still significant (over 10%) at \( N = 20 \).

Under CLO, we note that the PS configuration and the number of pods play a significant role. As anticipated, the smallest expected distance values are obtained when the PSs are located on all four sides of the FA. For a square, the expected distance for \( N = 1 \) (\( N = 2 \)) is reduced by 40% (28%) when the PSs are located on all four sides instead of one.
4.3. Asymmetric FA Shapes

We extend the results in the previous section to asymmetric shapes, i.e., a rectangle, an elongated diamond, and an ellipse. With asymmetric shapes, one must consider the effect of the width-to-length ratio, which is expressed through the shape factor, $b$. Section 4.3.1 presents analytic results for a rectangular FA, while Section 4.3.2 presents a comparison of all three asymmetric FA shapes.

4.3.1. Analytic Expressions for a Rectangular FA

The expected distances are obtained for a rectangular FA as a function of $b$ and the area, $A$, of the FA. The equations are summarized in Table 3-3, which shows $\tilde{E}[OUT]$ under RAN and CLO for the six PS configurations listed in Section 3.2. Table 3-3 also shows $b^*$, i.e., the shape factor that minimizes $E[OUT]$ for each case.

For RAN, it suffices to obtain solutions for $N = 1$. For CLO, on the other hand, we derive expressions for $N = 1, N = 2$, and $N \geq 3$. Exact solutions are derived in Appendix B through D for 1•/RECT/CLO/1 and 1•/RECT/CLO/2. For $N \geq 3$, exact solutions are derived for 1•/RECT/CLO/N, while analytic approximations are derived for 2•/RECT/CLO/N. A curve fit approach is used for 4•/RECT/CLO/N.

The analytic approximation for 2S/RECT/CLO/N and 2L/RECT/CLO/N (see Appendix D) is presented in Equations (11) and (12). Both expressions can be decomposed into two parts as expressed in Equation (10). Travel that occurs orthogonal to the two sides of the FA with PSs is captured by $d_o$; i.e., orthogonal travel consists of the sum of the straight-line distances from all $N$ pods to one of two the opposing sides with PSs, which corresponds to the first two terms in Equations (11) and (12). Travel that occurs parallel to the two sides with PSs is captured by $d_p$, which corresponds to the third term in Equations (11) and (12). Parallel travel is an exact result, and it takes different forms for odd and even $N$ values. (See Appendix B for the derivation). We
impose the additional assumption that \( d_o \) follows a Normal distribution (see Appendix D) to obtain the approximation in Equations (11) and (12).

\[
\tilde{E}[OUT] = d_o + d_p
\]  

(10)

\[
\tilde{E}[OUT] \cong \begin{cases} 
\frac{N}{2} - \left( \frac{N}{6\pi} \right)^{1/2} + b \left[ \frac{N - 1}{4} \right], & N \text{ odd} \\
\frac{N}{2} - \left( \frac{N}{6\pi} \right)^{1/2} + b \left[ \frac{N^2}{4(N + 1)} \right], & N \text{ even}
\end{cases}
\]  

(11)

\[
\tilde{E}[OUT] \cong \begin{cases} 
b \left[ \frac{N}{2} - \left( \frac{N}{6\pi} \right)^{1/2} \right] + \frac{N - 1}{4}, & N \text{ odd} \\
b \left[ \frac{N}{2} - \left( \frac{N}{6\pi} \right)^{1/2} \right] + \frac{N^2}{4(N + 1)}, & N \text{ even}
\end{cases}
\]  

(12)

We measure the approximation error in Equation (11) via Monte Carlo sampling based on 10,000 samples. The percent error is computed for \( 3 \leq N \leq 30 \) and \( b \) values of \((1.00, 0.75, 0.50, 0.25)\). All the percent error values are smaller than 1\%.

![Figure 3-5. Percent error of curve fit approximation versus Monte Carlo sample for \( E[OUT] - 4/\text{RECT/CLO/N} \).](image_url)
The curve fit approximation for 4/RECT/CLO/N derived in Appendix E is presented in Equation (13). Monte Carlo sampling results indicate that $\bar{E}[OUT]$ for a given $b$ value can be fitted with a linear function of $N$, with an $R^2$ value of 0.999. Likewise, the intercept, $C_1(b)$, and the slope, $C_2(b)$, of this linear function can be fitted with a linear function of $1/b$, with an $R^2$ of 0.999.

$$
\bar{E}[OUT] \cong C_1(b) + C_2(b) N = \left(-0.4633 + \frac{0.2285}{b}\right) + \left(0.4661 + \frac{0.2492}{b}\right) N \quad (13)
$$

We measure the approximation error in Equation (13) via Monte Carlo sampling based on 10,000 samples. The percent error is computed for $3 \leq N \leq 30$ and $b$ values ranging from 0.25 to 0.95. (An exact result was derived for $b = 1$ in the previous section.) The percent error values (see Figure 3-5) are mostly within 4%.

Table 3-3. $\bar{E}[OUT]$ based on a rectangular FA and optimal shape factor ($b^*$).

<table>
<thead>
<tr>
<th>Rule</th>
<th>PS config.</th>
<th>$N$</th>
<th>$\bar{E}[OUT]$; $b^*$ is based on $E[OUT] = \sqrt{A/b} \bar{E}[OUT]$</th>
<th>$b^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAN</td>
<td>1S</td>
<td>1</td>
<td>$\frac{1}{2} \cdot b + \frac{b}{3}$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1L</td>
<td>1</td>
<td>$\frac{b}{2} \cdot \frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td></td>
<td>2S</td>
<td>1</td>
<td>$\frac{1}{2} \cdot b + \frac{b}{3}$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2L</td>
<td>1</td>
<td>$\frac{b}{2} \cdot \frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>$\frac{(b^2 + 3b + 1)}{3(1 + b)}$</td>
<td>1</td>
</tr>
<tr>
<td>CLO</td>
<td>1S</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\frac{1}{3}$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$N$</td>
<td>$\frac{N}{2} + b \left(\frac{N-1}{4}\right)$ $N$ odd</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( \frac{N}{2} + b \left( \frac{N^2}{4(N+1)} \right) ) ( N ) even</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>1L</strong></td>
<td>1</td>
<td>( \frac{b}{2} )</td>
<td>( \rightarrow 0 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>( b + \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( N \left( \frac{b}{2} \right) + \frac{N-1}{4} ) ( N ) odd</td>
<td>( \frac{N-1}{2N} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( N \left( \frac{b}{2} \right) + \frac{N^2}{4(N+1)} ) ( N ) even</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>2S</strong></td>
<td>1</td>
<td>( \frac{1}{4} )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>( \frac{2}{3} + \frac{b}{3} )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( \frac{N}{2} - \sqrt{\frac{N}{6\pi}} + b \left( \frac{N-1}{4} \right) ) ( N ) odd</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( \frac{N}{2} - \sqrt{\frac{N}{6\pi}} + b \left( \frac{N^2}{4(N+1)} \right) ) ( N ) even</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>2L</strong></td>
<td>1</td>
<td>( \frac{b}{4} )</td>
<td>( \rightarrow 0 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>( \frac{2b}{3} + \frac{1}{3} )</td>
<td>( \frac{1}{2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( b \left( \frac{N}{2} - \sqrt{\frac{N}{6\pi}} \right) + \frac{N-1}{4} ) ( N ) odd</td>
<td>( \rightarrow \frac{1}{2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( b \left( \frac{N}{2} - \sqrt{\frac{N}{6\pi}} \right) + \frac{N^2}{4(N+1)} ) ( N ) even</td>
<td>( \rightarrow \frac{1}{2} )</td>
<td></td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>1</td>
<td>( \frac{b(3-b)}{12} )</td>
<td>( \rightarrow 0 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>( \frac{(b^3 - 5b^2 + 30b + 20)}{60} )</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( N )</td>
<td>( \left( -0.4633 + \frac{0.2285}{b} \right) + \left( 0.4661 + \frac{0.2492}{b} \right) N )</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
4.3.2. Comparison of asymmetric FA shapes

In this section we present numeric results to compare the expected distances across the three asymmetric FA shapes and we discuss the impact of the FA shape, the number of pods \(N\), and order assignment rules, RAN and CLO. In contrast to Section 4.2, here we consider five PS configurations (see Section 3.3) as well as the shape factor, \(b\).

We compute the expected distances via analytic expressions or Monte Carlo sampling. Analytic expressions were derived only for the rectangle and are presented in Table 3-3. All the Monte Carlo results are based on 10,000 trips per replication and 10 replications. Figure 3-6 shows \(E[DC]\) per pod as a function of \(N\). (Figure 3-6 shows the average values across 10.) Appendix H presents the Tables for \(E[OUT]\), \(E[TB]\), and \(E[DC]\). All the results are based on FA shapes with a unit area.

Based on the results in Figure 3-4 and Tables H1 through H3 in Appendix H, we comment on the effects of each parameter on the expected distances:

**FA shape:** The effect of the shape is more evident with small \(b\) values. In general, ELDI is the worst performing shape, while ELIP values consistently match or outperform the RECT. The effect of shape is also compounded by the PS configuration, where PS configurations 1S and 2S have a particularly negative impact on \(E[DC]\) in ELDI.

**PS configuration:** The effect of the PS configuration is more evident with small \(b\) values. The largest \(E[DC]\) values are observed under PS configurations 1S and 2S, while the smallest \(E[DC]\) values are observed under configuration 4 and the CLO rule, or configurations 1L and 2L under the RAN rule.

**Width-to-length ratio:** \(E[DC]\) consistently increases as \(b\) becomes smaller. The increase in \(E[DC]\) is a result of the increase in \(E[TB]\) as \(b\) becomes smaller (see Table H2 in Appendix H).
In general, $E[OUT]$ also increases with smaller $b$ values, except for •L/*/CLO/1 and 4/*/CLO/1. A specific instance of this exception is reported for the RECT shape, where the closed form $E[OUT]$ results in Table 3-3 show that $E[OUT]$ is minimized when $b = 0$. This exception is created by our analytic model and may not have practical relevance. Note the exception happens in cases that allow the AGV to travel to the long sides of the FA, namely •L and 4 PS configurations. As $b$ goes to 0, the two short sides become infinitely short, the two long side become infinitely long and the FA resembles a line. Any point in this FA will travel an infinitely small amount to visit the closest point in a long side.

Order assignment rule: As it is the case with symmetric shapes, $E[DC]$ is lower under CLO compared to RAN. Under CLO, $E[DC]$ can be reduced by as much as 65% compared to RAN, as is the case in 4/RECT/*/1 with $b = 0.25$. Furthermore, we note that $E[DC]$ is less sensitive to the effect of $b$ under CLO. Table H2 in Appendix H shows that the percent difference between CLO and RAN increases as $b$ decreases. While $E[DC]$ increases under both CLO and RAN, it increases at a higher rate under RAN than under CLO, resulting in increasing percent difference values as $b$ decreases. We again note that as $N$ increases, the gap between CLO and RAN decreases.
Figure 3-6. \( E[DC] \) per pod as a function of \( N \).
5. Pick Station Usage

As we remarked in Section 4.1 (see assumption 7), under CLO, depending on their locations, certain PSs may receive more pods than others. In practice, this could lead to an uneven workload across the PSs, which will negatively impact the throughput of the system.

To address the above concern, we perform a numerical experiment. We simplify this analysis by limiting to the case with N=1. We argue that orders that require a single pod are commonplace in AGV-OPS due to the characteristic order profile in online retail (Boysen, De Koster, and Weidinger 2018) and efforts to reduce pod retrieval through improved slotting (Xiang, Liu, and Miao 2018).

We consider six FA layouts, each with a total of 16 PSs, as shown in Figure 3-7. (Note that, on each side of the FA, the PSs are located equidistant and uniformly.) Random points are sampled within the FA and the closest PS is selected for a total of 10,000 sampled points per replication and 10 replications. Figure 3-8 shows the percentage of pods that are assigned to each PS (numbered from 1 to 16) for each shape shown in Figure 3-7. (With 16 PSs, even usage would result in each PS being assigned 6.25% of the pods.)

In Figure 3-8 we observe that the range of usage values across the PSs is smaller for symmetric shapes relative to their asymmetric counterparts. When comparing (a)symmetric shapes, the (elongated) diamond achieves more even usage. Considering all shapes, the ellipse shows the largest range of usage values across the PSs. It is also apparent from Figure 3-8 that the PSs with the largest usage values are those located towards the center of each side; see, for example, PS 6 and PS 8 at the center of the two long sides, and PS 2 and PS 10 at the center of the two short sides of the rectangle. PSs located close to the corners have lower usage values. For example, in the rectangle, PSs 1, 3, 4, 8, 9, 11, 12, and 16 have the lowest usage values.
Figure 3-7. FA layouts and PS configurations for usage experiment.

Figure 3-8. Pick station usage measured by percentage of pods assigned to each PS.
In some cases, it is possible to achieve even usage across the PSs under CLO. Consider, for example, a square. For simplicity, we will restrict the PS configuration to be symmetric across all four sides of the FA. Under the continuous approximation of the PS locations, the contour lines (denoting which side of the FA a pod will be assigned to) would be diagonal lines that cross through the center of the square and end at its corners (as shown in Figure 3-9(a)). The same contour lines would also apply to discrete PSs, so long as the PS configuration is symmetric across the four sides and symmetric within each side. (A PS configuration is symmetric within a side if the PSs on the left half of that side form a mirror image of the PSs on the right half of the side as shown in Figure 3-9(c)). Additional contour lines appear between discrete PSs within the triangular areas in Figure 3-9(a) that denote which PS a pod is assigned to (as shown in Figure 3-9(b)). We argue that the PSs can be located such that the portion of the FA assigned to each PS is equalized as shown in Figure 3-9(c). Repeating the Monte Carlo experiment for the 16 PSs located as in Figure 3-9(c), we confirm that each PS is assigned approximately 6.25% of the pods.

![Figure 3-9](image_url)

Figure 3-9. CLO rule contour lines in a square with: (a) continuous PSs, (b) discrete uniformly located PSs, and (c) discrete even usage PSs.

We further explore the impact on $E[OUT]$ by comparing the results from Figures 3-9(b) and 3-9(c) with a Monte Carlo sample of 10,000 trips per replication and 10 replications. Based on a square of unit area, the 95% confidence interval for $E[OUT]$ is computed as (0.2197, 0.2201)
and (0.2214, 0.2218) for Figure 3-9(b) and Figure 3-9(c), respectively. Hence, there is a statistically significant difference in $E[OUT]$ between the PS configurations in Figures 3-9(b) and 3-9(c). However, the PS configuration in Figure 3-9(c) increases $E[OUT]$ by less than 1% while it achieves even usage across the PSs. While this subject needs further investigation, we believe that in most cases, when picking discrete locations for the PSs, the user can achieve near-even usage across the PSs without creating a significant increase in $E[OUT]$.

6. Conclusions and Future Research

We derive closed-form analytic expressions for the expected travel distance in an AGV-based fulfillment system operating under two order assignment rules. Exact and approximate results are presented for different PS configurations and FA shapes as a function of the width-to-length ratio ($b$). Table 3-4 presents a summary of the cases treated in this paper and classifies the results as exact or approximate analytic or Monte Carlo. Table 3-4 pertains to $E[OUT]$ results only, while an exact $E[TB]$ expression was derived for the rectangle FA and exact results were derived by Larson and Odoni (1981) for all three symmetric shapes.

Table 3-4. Summary of $E[OUT]$ cases treated by method.

<table>
<thead>
<tr>
<th>Symmetric shapes ($b = 1$)</th>
<th>Exact</th>
<th>Approximate</th>
<th>Monte Carlo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*/SQRE/RAN/• 1/SQRE/CLO/•</td>
<td>2/SQRE/CLO/N&gt;2 4/SQRE/CLO/N&gt;2 2/SQRE/CLO/1</td>
<td>*/DIAM/CLO/N&gt;1 3/SQRE/CLO/N&gt;2 */CIRC/CLO/1</td>
</tr>
<tr>
<td></td>
<td>2/SQRE/CLO/{1,2,3} •</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/SQRE/CLO/{1,2} •</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*/DIAM/RAN/•</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*/DIAM/CLO/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymmetric shapes ($b &lt; 1$)</td>
<td><em>/RECT/RAN/• 1</em>/RECT/CLO/•</td>
<td>2*/RECT/CLO/N&gt;3 4*/RECT/CLO/N&gt;2</td>
<td>*/ELDI/• 3/SQRE/CLO/N&gt;2 2/SQRE/CLO/N&gt;2</td>
</tr>
<tr>
<td></td>
<td>2*/RECT/CLO/{1,2,3} •</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4*/RECT/CLO/{1,2} •</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*/ELIP/•</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We perform numerical experiments to assess the effect of the FA shape, order assignment rule, PS configuration, width-to-length ratio ($b$), and batch size ($N$) on the expected travel distance of DC cycles. We conclude that order assignment rule has the biggest impact on expected DC travel distance. In almost all cases considered CLO results in a lower $E[DC]$ than RAN, up to 65% lower. PS configuration can also have a significant impact on $E[DC]$, specially in narrow FA shapes (small $b$ values) and cases with small $N$ values. Take for example •/ELDI/CLO/1 with $b = 0.25$, the 1S configuration performs worst with $E[DC] = 3.93$ and 2L configuration performs best with $E[DC] = 1.28$, that is around 67% reduction in travel. The width-to-length ratio appear to compound the effect of PS configuration. Notice that the symmetric version of the previous case, •/DIAM/CLO/1, only experiences a 34% reduction in $E[DC]$ comparing 1S configuration against 2L configuration. Nonetheless, reducing the value of $b$ consistently increases $E[DC]$.

Based on the results of the numerical experiment we would recommend a square-shaped FA with PS on all four sides and the use of CLO rule. The CLO rule could potentially be implemented in an existing system, where FA shape and PS configuration modification may be restricted. The results of this study suggest that the PS configuration should be carefully considered in early stage design. Locating PS on all sides of the FA or the long sides, rather than the short sides is beneficial. We can attribute that to better coverage of the FA provided by PS that are located farther apart. Our results also suggest that the geometric shape of the FA has less impact on the $E[DC]$ than the width-to-length ratio. Designers should favor symmetric shapes, i.e., small width-to-length ratio, independent of the FA shape. In this study we consider three convex FA shapes, while the study of non-convex FA shapes could be addressed in future research.

We further explore the feasibility of the implementation of the CLO rule, in light of the possibility of uneven use of PS. We conclude that in a square-shaped FA, PSs can be located to
achieve an even usage (balanced allocation of pods to PSs), while suffering little impact on the $E[DC]$. Future research can investigate methods to optimize the location of PSs that considers even PS usage. Another consideration related to the location of the PSs is finding alternatives to locating the PSs in the perimeter of the FA. For example, what can be gained by locating the PSs in the center of the FA.

It is important to emphasize that our conclusions are based on the assumption that every pod is equally to be retrieved (justified under randomized storage) and pods are modeled continuously over the FA. Future research could extend the results presented here in a system with turnover-based storage, like the one considered in Yuan, Graves, and Cezik (2019).

References


### Notation

**Table 3-5: Abbreviations and notation**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>Forward area</td>
</tr>
<tr>
<td>PS</td>
<td>Pick station</td>
</tr>
<tr>
<td>OUT</td>
<td>AGV travel from FA to PS</td>
</tr>
<tr>
<td>TB</td>
<td>AGV travel inside FA</td>
</tr>
<tr>
<td>BACK</td>
<td>AGV travel from PS to FA</td>
</tr>
<tr>
<td>DC</td>
<td>Dual command</td>
</tr>
<tr>
<td>AGV</td>
<td>Automated guided vehicle</td>
</tr>
<tr>
<td>RAN</td>
<td>Random order assignment</td>
</tr>
<tr>
<td>CLO</td>
<td>Closest order assignment</td>
</tr>
<tr>
<td>SQRE</td>
<td>Square</td>
</tr>
<tr>
<td>RECT</td>
<td>Rectangle</td>
</tr>
<tr>
<td>DIAM</td>
<td>Diamond</td>
</tr>
<tr>
<td>ELDI</td>
<td>Elongated diamond</td>
</tr>
<tr>
<td>CIRC</td>
<td>Circle</td>
</tr>
<tr>
<td>ELIP</td>
<td>Ellipse</td>
</tr>
<tr>
<td>OCTA</td>
<td>Octagon</td>
</tr>
</tbody>
</table>

### Abbreviations and notation:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>Forward area</td>
</tr>
<tr>
<td>PS</td>
<td>Pick station</td>
</tr>
<tr>
<td>OUT</td>
<td>AGV travel from FA to PS</td>
</tr>
<tr>
<td>TB</td>
<td>AGV travel inside FA</td>
</tr>
<tr>
<td>BACK</td>
<td>AGV travel from PS to FA</td>
</tr>
<tr>
<td>DC</td>
<td>Dual command</td>
</tr>
<tr>
<td>AGV</td>
<td>Automated guided vehicle</td>
</tr>
<tr>
<td>RAN</td>
<td>Random order assignment</td>
</tr>
<tr>
<td>CLO</td>
<td>Closest order assignment</td>
</tr>
<tr>
<td>SQRE</td>
<td>Square</td>
</tr>
<tr>
<td>RECT</td>
<td>Rectangle</td>
</tr>
<tr>
<td>DIAM</td>
<td>Diamond</td>
</tr>
<tr>
<td>ELDI</td>
<td>Elongated diamond</td>
</tr>
<tr>
<td>CIRC</td>
<td>Circle</td>
</tr>
<tr>
<td>ELIP</td>
<td>Ellipse</td>
</tr>
<tr>
<td>OCTA</td>
<td>Octagon</td>
</tr>
</tbody>
</table>

- **FA**: Forward area
- **PS**: Pick station
- **OUT**: AGV travel from FA to PS
- **TB**: AGV travel inside FA
- **BACK**: AGV travel from PS to FA
- **DC**: Dual command
- **AGV**: Automated guided vehicle
- **RAN**: Random order assignment
- **CLO**: Closest order assignment
- **SQRE**: Square
- **RECT**: Rectangle
- **DIAM**: Diamond
- **ELDI**: Elongated diamond
- **CIRC**: Circle
- **ELIP**: Ellipse
- **OCTA**: Octagon

**Abbreviations**:

- **FA**: Forward area
- **PS**: Pick station
- **OUT**: AGV travel from FA to PS
- **TB**: AGV travel inside FA
- **BACK**: AGV travel from PS to FA
- **DC**: Dual command
- **AGV**: Automated guided vehicle
- **RAN**: Random order assignment
- **CLO**: Closest order assignment
- **SQRE**: Square
- **RECT**: Rectangle
- **DIAM**: Diamond
- **ELDI**: Elongated diamond
- **CIRC**: Circle
- **ELIP**: Ellipse
- **OCTA**: Octagon

**Notations**:

- **$A$**: Area
- **$N$**: Number of pods
- **$b$**: Shape factor
- **$T$**: Scaling factor
- **$d_o$**: Travel that occurs orthogonal to the side(s) of the FA that contains PSs
- **$d_p$**: Travel that occurs parallel to the side(s) that contain PSs
- **$C_1(b)$**: Linear function parameter (intercept)
- **$C_2(b)$**: Linear function parameter (slope)
- **$E[OUT]$**: Normalized expected travel distance of the OUT leg
Chapter 4
Heuristic Methods for the Location of Induction Points in Automated Guided Vehicle-Based Sortation Systems

Abstract: Automated Guided Vehicle-based sortation systems (AGV-SSs) have recently been adopted by some of the leading online retail companies. An AGV-SS performs the same basic function as a traditional, conveyor-based sortation system except that it relies on AGVs instead of conveyors to sort the items. In virtually all automated and semi-automated sortation systems, there is at least one “induction point” (IP) that serves as an identification and entry point for the items to be sorted by the system. In an AGV-SS, each (empty) AGV is loaded at the IP with an item to be sorted. The loading process may be manual or automated. Once the item is loaded onto the AGV, the AGV travels to the appropriate destination point (DP), where the item is automatically discharged, and the empty AGV travels to the same or another IP to pick-up the next item. Given the location of the DPs, this study is concerned with determining the optimum location of the IPs in order to minimize the expected AGV travel distance. One of the distinctive characteristics of the AGV-SSs deployed so far is the uniform arrangement of the DPs such that they form a grid. We propose heuristic methods for the problem and through a computational study we compare their performance. We also present results for and compare interior versus perimeter placement of the IPs. Finally, we derive and solve an analytic model for the optimal location of two IPs in a continuous plane that represents the sortation region.
1. Introduction

Customer expectations have changed in the online retail sector. A company’s fulfillment capabilities are closely related to the customer’s experience and overall satisfaction. With rapid growth in online retailing, the opportunity for automation in fulfillment centers (FCs) and distribution centers (DCs) is more present than ever. The primary justification for automation in warehousing, at least initially, was the elimination or reduction of travel time required of the warehouse workers, improvements in inventory and picking accuracy, and the reduction of fulfillment cycle times. In the current environment, in addition to the above benefits expected of automation, FCs and DCs need more flexibility and scalability, while reducing their reliance on temporary and an unreliable worker pool (Futch 2017). Today, various types of systems are available to automate FC and DC operations. There are systems that can load, unload, sort, pick, transport, store, deliver, and audit.

In this paper we study a relatively new type of unit load automated guided vehicle (AGV), one which is equipped with a tilt tray, used for automatically transporting and sorting individual items, one at a time. Both unit load AGVs and tilt tray sortation conveyors have been used for decades. However, eliminating the conveyor, and using AGVs equipped with tilt trays is a new concept in sortation.

In a tilt tray conveyor system (see Tompkins et al. 2010, p. 230), each item to be sorted is loaded (automatically or manually) onto an individual tray at an “induction point” (IP). Multiple trays are “linked” together to form a conveyor loop. Once the item is placed on a tray, it is transported by the conveyor, and when it reaches the appropriate destination point (DP), the tray is tilted to automatically discharge the item via gravity (often into a gravity chute). In an AGV-
based sortation system (AGV-SS), the conveyor loop is eliminated and AGVs equipped with tilt trays are used for sorting the items.

Automated sortation has historically relied on conveyors, which have been available for a long time and have been adapted to many variations, well-suited for different applications. (The interested reader may refer to Tompkins et al. (2010) to view a variety of conveyor-based sortation systems and their description.) Romaine (2018) points to the fact that conveyors are fixed whereas AGVs are mobile. In light of that, one of the key differences between conveyor-based and AGV-based sortation systems is the floor space they require. Conveyors need more dedicated floor space than a fleet of AGVs. Also, the fixed-path nature of conveyors makes them more difficult to expand or scale up. Although one may install a larger conveyor loop with future expansion in mind, a larger loop would significantly increase the initial cost of the system and require even more floor space. AGV-SSs, on the other hand, can be expanded or scaled up by adding more AGVs as needed (or if needed). These systems “can be obtained at a lower capital investment and be expanded as needed, which reduces the initial investment and improves overall return on investment” (Futch, 2017).

Being a relatively new concept, currently there are not many applications of AGV-SSs in the U.S. A recent example of an AGV-SS is the t-Sort system introduced by Tompkins Robotics and described by Futch (2017):

“It performs much like a traditional automated sortation system, such as a tilt tray or crossbelt sorter. However, it uses completely independent robots, which are the equivalent of having a tilt tray with no rail. They can go to any divert and induction station autonomously along the shortest path. Robots, chutes or receiving receptacles, and induction stations can be added modularly at any time with no interruption to operations. t-Sort is a means of automated order fulfillment for units to complete an order and parcel shipping operations, allowing for better planning, implementation, controls efficiency, flow of goods and customer service. The system is flexible and can be deployed in any size DC/FC.”
In the summer of 2019, Amazon announced its implementation of an AGV-SS called Pegasus. The system uses AGVs equipped with belt conveyors to sort packaged orders. Leonard (2019) reports that 800 Pegasus AGVs were deployed in the Denver, Colorado sortation center in October 2018. As described by Leonard, “Human workers scan packages and place them on top of the Pegasus system, which then drives to the opening of the correct chute where the conveyor belt will shoot the package on the slide.” Amazon’s VP of Robotics is quoted as saying that the technology can reduce the number of missorts by 50% over other sortation solutions. A spokesperson is quoted by Heater (2019) as saying: “Pegasus drive units help reduce sort errors, minimize damage, and speed up delivery times”. They also claim the use of robotics at “sort centers” will increase the building capacity.

Other applications that have been reported recently include Flipkart’s deployment of 100 AGVs at its Soukya sortation center, a facility that employs about 1,000 workers. Ahaskar (2019) reports that the system can process up to 4,500 shipments in an hour and can be scaled up to 5 times with slight adjustments. It is claimed that the AGV system can improve process efficiency by 60% and enhance the throughput of the existing facility without an expansion or increase in the headcount.

In this paper we focus on the design and performance of AGV-SSs. More specifically, given a set of discrete DPs located in a square-shaped sortation region (SR), we develop heuristic procedures to investigate the problem of finding the optimal location of the IPs to minimize the expected full and empty AGV travel distances. We perform computational experiments to compare the performance of the heuristic procedures on several problem instances. We also investigate the optimal location of the IPs when they are restricted to the perimeter of the SR. Last,
using a continuous approximation of the SR, we present exact analytic results for determining the optimum location of two IPs placed either inside or on the perimeter of the SR.

2. Literature Review

Being a relatively new system, there are very few academic/refereed publications on AGV-SSs. To the best of our knowledge, our study is the first academic study to address the optimal location of the IPs in an AGV-SS. Mei and Zhou (2018) propose a multi-AGV path planning algorithm and they test its effectiveness via simulation. Mauro (2017) considers the design of a system that allows multiple AGVs to carry a single large item. An auction-like algorithm is proposed and tested in an agent-based simulation. In addition, the study investigates the effect using dedicated travel paths for loaded and unloaded AGVs. Heuvel (2018) performs a series of simulation experiments to test the feasibility of using exact and greedy algorithms to solve two operation decisions: the allocation of items to AGVs and path planning. Three SR layouts are used in the experiments: DPs along three sides of a rectangular SR, DPs arranged as a grid, and DPs arranged to form aisles. In all three layout alternative, IPs are located along one of the sides of the SR.

The optimum location of the IPs generally falls within the subject of location theory, which has been studied extensively in the literature. Since a comprehensive review of the location literature is well beyond the scope of our paper, we focus only on those location studies that are closely related to our problem. More specifically, we are interested in location problems on a grid of uniformly arranged demand points also known as existing facilities or customer locations. We also investigate papers related to location-allocation problems (LAPs) with continuous demand regions, which can be a suitable approximation to uniformly arranged discrete DPs. One of the challenging aspects of the IP location problem is allocating an equal number of DPs to each IP, with fractional assignments prohibited or allowed. We search for studies on the capacitated LAP
and those that equalize the number of customers allocated to each facility. In addition, we briefly discuss those studies that influenced our choice of solution methods.

The grid-based location problem (GBLP) is introduced by Noor-E-Alam, Mah, and Doucette (2012) to solve a light post location problem as an Integer Linear Program (ILP). Noor-E-Alam and Doucette (2014) solved a GBLP considering fixed costs associated with opening facilities and showed that the ILP model is efficient in solving small to moderate-sized problems. Heuristic methods are proposed for large-scale instances. Faiz and Noor-E-Alam (2018) explore the optimal location of data centers. The grid-based approach is used to approximate a continuous demand region. They present two Mixed Integer Linear Programs for the capacitated single-source data center location problem. One model provides optimal locations, capacities, and demand allocation in the absence of existing data center. The second model considers satisfying new demand in the presence of existing data centers.

Brimberg and Drezner (2013) propose a two-phase approach for the classical p-median problem in a plane. The first phase finds a good initial solution by heuristically solving a discretized problem that is obtained by overlaying a grid of candidate sites over the area containing the demand points. The second phase implements an improved version of the well-known alternate location and allocation algorithm presented by Cooper (1963). The authors test Steepest Descent (SD), Tabu Search (TS), and Simulated Annealing (SA) based heuristic methods on the grid and obtain the best results with SA. In this paper, we also evaluate SA and we present an alternative scheme to perturb the current solution. As part of our computational study, we compare our perturbation scheme with the one proposed by Brimberg and Drezner.

Treating the IPs as capacitated facilities is one way to enforce an equal allocation among the IPs. Berman et al. (2009) investigate the facility location problem with the goal of achieving
equalized loads. Their formulation minimizes the maximum weight assigned to each facility, without any distance minimization considerations. Their objective is to “find the smallest possible capacity that will have a feasible solution to the capacitated p-median or p-center problem.” Several heuristic algorithms were used to solve the resulting models, including SD and TS. Marin (2011) considers a discrete location problem that balances the difference between the minimum and maximum number of customers assigned to each facility. Aardal et al. (2015) provide an approximation algorithm for the capacitated k-facility location problem with fixed costs. The algorithm relaxes the constraint on the number facilities to be opened ($k$) and may result in up to twice the number of facilities being opened, i.e., $2k$ facilities.

The IP location problem we present is similar to the rectilinear version of the capacitated multi-facility weber problem (CMFWP). Aras, Orbay, and Altinel (2008) propose two MILP formulations that exploit a property under the rectilinear metric shown by Hansen (1980). The property states that an optimal solution exists where all facilities are within the convex hull of the demand points. Furthermore, facility locations occur at the intersection points of vertical and horizontal lines drawn through demand points. Several other algorithms have been proposed for the general CMFWP. Boyaci, Kuban, and Aras (2013) present an approximation subgradient algorithm that relaxes the capacity constraint. They solve instances with 10 facilities and 250 demand points. Akyuz, Altinel, and Oncan (2014) propose two branch-and-bound (BB) algorithms. One algorithm performs a BB search on the allocation variable space based on the formulation presented by Sherali, Ramachandran, and Kim (1992). The second BB algorithm searches on the location variable space. The authors solve small problems of 6 facilities and 30 demand points or 10 facilities and 20 demand points. All of the above studies either solve small problem instances optimally or they propose heuristics for larger problems.
As Brimberg et al. (2000) point out, the computational time is a limiting factor for the solution of large problems even for the uncapacitated multisource weber problem. The authors present computational experiments to compare the performance of various heuristics including TS, Variable Neighborhood Search, and Genetic Algorithms. While they conclude that no method is best in all instances, they state that relocation-based methods (i.e., those that seek to improve the solution by varying the location of the new facilities) are generally more efficient than reallocation-based methods (i.e., those that seek to improve the solution by changing the allocation of the customers to the new facilities). Numerous subsequent studies have implemented heuristic and metaheuristic approaches to solve the CMSWP. Zainuddin and Salhi (2007) use a modified Alternating Transportation-Location (ATL) heuristic (Cooper 1972). Luis, Salhi, and Nagy (2011) and Hosseininezhad, Salhi, and Jabalameli (2015) consider the CMSWP with fixed costs. Manzour-al-Ajdad, Torabi, and Eshghi (2012) implement a SA-based method that perturbs the allocation variable. In this paper, we investigate the performance of a SD- and SA-based location search method in addition to alternating methods motivated by Cooper (1963) for the uncapacitated case and Cooper (1972) for the capacitated case.

A LAP with discrete demand points uniformly arranged – such as a grid of DPs – can be approximated by a continuous demand region of uniform demand. For example, Campbell (1990) derives an analytic solution to the optimal terminal location problem in the context of freight consolidation. The study solves the one-dimensional case where demand is uniformly distributed over a line. Even spacing between the terminals that slightly sink towards the center of the line, is found to be optimal. As a result, a slightly larger service region is allocated to the two outermost terminals.
Suzuki and Drezner (1997) formulate a continuous model for the airline hub location problem. They aim to minimize the travel distance when customers must travel to the closest hub and then to their destination. They present exact analytic solutions for special cases with two or three hubs in a square with uniform demand, and for the case with two hubs in a rectangle. The square with two hubs is restricted to two cases where both hubs are equidistant from the center of the square. In one case, the hubs must lay in a horizontal line that goes through the center of the square. In the second case, a diagonal line goes through the center of the square. In this paper we generalize the result for two IPs in a square, by removing the restrictions placed on the locations of the IPs. We extend the result to consider uneven service areas.

Ghaffarinasab, Woensel, and Minner (2018) implement an iterative Weizfeld-type algorithm and a particle swarm algorithm to solve a planar hub location-routing problem. They plot solutions for problem instances with two through ten hubs. The overall structure of their solutions resembles the solutions we obtain in this paper, although there are noticeable differences in the shape of the regions based on the Euclidean metric (their case) versus the rectilinear metric (our case).

3. System Description and the Model

An AGV-SS consists of three primary hardware components: a fleet of AGVs, a (small) set of IPs, and a (large) set of DPs. The items to be sorted enter the system through one of the IPs, where they are scanned/identified and placed on an AGV by an operator or robot. Subsequently, each item is transported, one at a time, by one of the AGVs from the IP to the appropriate DP. The AGVs are equipped with a mechanism (such as a tilt tray or belt conveyor) that automatically discharges the item when the AGV reaches the DP. The DP may be an opening in the floor...
connected to a slide/gravity chute or it may be a container (such as a bin or tote box) large enough to accumulate the items that have been assigned to it.

The aerial view of an AGV-SS installed in a logistics center in China is shown in Figure 4-1. The system has three IPs and many DPs as shown in the Figure, where arrows 1, 2, and 3 show an IP, an AGV, and a DP, respectively. Note that the items being sorted are generally small-to-medium sized parcels, no larger than what the tilt tray can accommodate. Also note that the DPs are arranged in a grid pattern within the SR. A close-up of the IP is shown in Figure 4-2a. The items to be sorted are typically brought to the IPs via overhead belt conveyors that lead to gravity chutes where the items wait to be processed by the operator. (Use of overhead conveyors allow more flexibility in locating the IPs.) Note the large number of items that may accumulate in the gravity chute. A close-up of an item being discharged at the DP is shown in Figure 4-2b. The system depicted in Figure 4-1 can reportedly sort up to 18,000 items/hr, and it is claimed that it reduces the manual labor cost by 70%.

![Figure 4-1. Aerial view of the AGV-SS. (Obtained from the China Global Television Network, CGTN, depicting a logistics center in China.)](image)

91
A given AGV performs two trips for each item sorted. After the AGV has discharged the previous item at a DP, it travels empty to one of the IPs to pick up the next item. We denote this first component as “empty travel.” At the IP, a new item is placed on the AGV, and the AGV transports the item to the appropriate DP. We denote this second component as “full travel.” During both trips, empty or full, the AGV follows the shortest rectilinear path between two points.

Once an AGV discharges an item, it must be dispatched to an IP to pick up the next item. Although several dispatching options are available, among the simple ones, one may consider sending the AGV to: (a) A randomly selected IP; (b) The closest IP; or (c) The IP with the most items waiting. In order to reduce empty AGV travel, we will select option (b). Option (c) may also be used but, in general, since there is a large number of items waiting at each IP, dispatching the AGV to reduce empty travel seems more desirable.

The assumptions we make to study the IP location problem are listed as follows:

1. The number of IPs is given. The IPs may be located either in the interior or on the perimeter of the SR.
2. There is sufficient work at each IP. That is, at each IP there is at least one item waiting to enter the system at all times.

3. The number of DPs and their locations are given. Furthermore, the DPs are arranged in a two-dimensional, uniform grid with one DP located at each intersection point of the grid. The number of DPs is significantly larger than the number of IPs.

4. All the DPs have equal “demand.” That is, an item picked up at any IP is equally likely to be taken to any DP.

5. An empty AGV is assigned to the closest IP.

6. In traveling between any two points in the SR, the AGV follows the shortest rectilinear path. Acceleration/deceleration of the AGV and possible congestion in the SR are not taken into account.

If we draw a vertical and horizontal line through each DP, we would obtain the Hanan grid for the SR. It is well-known in location theory that, with rectilinear distances, the optimum location of a new facility overlaps with one of the intersection points in the Hanan grid (Hansen, 1980). Hence, we know that an optimal solution exists where each IP is located at one of the intersection points obtained by drawing vertical and horizontal lines through each DP. Therefore, by assumption 3, each DP is a candidate site for an IP. For implementation purposes, an IP can be located adjacent to a DP. In fact, an AGV might not be necessary to transport items destined to an adjacent DP.

Note, from assumption 4, that it is straightforward to compute the expected full AGV travel distance out of any IP. Also, in assumption 5, if there is a tie, it is broken randomly based on a carefully selected allocation. For example, if IPs 1 and 2 are equidistant from DP 10, and the allocation scheme is 30-70, then each time the AGV becomes empty at DP 10, it is dispatched to
IP 1 (IP 2) with probability 0.30 (0.70). In such instances, we say that DP 10 is “shared” between IPs 1 and 2.

We next introduce our notation as follows:

Sets:

\[ i \in I: \text{index for candidate IP site. } I = \{1, \ldots, N\} \]
\[ j \in J: \text{DP index. } J = \{1, \ldots, N\} \]

Parameters:

\[ d_{ij}: \text{distance between candidate IP site } i \text{ and DP } j. \]
\[ \bar{f}_i: \text{average (full) AGV travel distance out of candidate IP site } i. \text{ We have } \bar{f}_i = \frac{1}{N} \sum_j d_{ij}. \]
\[ m: \text{the number of IPs to locate.} \]

Decision variables:

\[ x_{ij}: \text{fraction of AGV trips from DP } j \text{ allocated to candidate IP site } i. \]
\[ y_i: \text{equals 1 if an IP is located at candidate IP site } i, 0 \text{ otherwise.} \]

The proposed formulation for the IP location problem is presented as follows:

\[
\min z = \frac{1}{N} \sum_j \sum_l x_{ij} (d_{ij} + \bar{f}_i) \tag{1}
\]

s.t.

\[
\sum_l x_{ij} = 1 \quad \forall j \tag{2}
\]
\[
\sum_j x_{ij} = \frac{N}{m} \quad \forall i \tag{3}
\]
\[
x_{ij} \leq y_i \quad \forall i, j \tag{4}
\]
\[
\sum_l y_i = m \tag{5}
\]
\[
y_i \in \{0,1\} \quad \forall i \tag{6}
\]
\[
x_{ij} \geq 0 \quad \forall i, j \tag{7}
\]
The objective function (Equation 1) computes the expected empty plus full AGV travel distance per item sorted. For each DP, Equation 1 computes the weighted sum of the distance to every candidate IP site $i$ and the expected full travel distance out of IP site $i$. The weights are determined by the fractional allocation used for each IP site at DP $j$. Finally, the total weighted distance is averaged over all the DPs. Equation 2 ensures that the fraction of AGV trips out of each DP sums up to 1. Equation 3 ensures that an equal number of AGV trips is allocated to each IP. Otherwise, some IPs will receive more AGVs than the other IPs, which would upset the balance of the system. If there is sufficient work at each IP, then each IP must receive its “fair share” of AGVs, which is enforced through Equation 3. We will refer to this constraint as the “equal service area” (ESA) constraint, since the DPs assigned to each IP can be considered as its service area. Equation 4 ensures that DPs are allocated to an IP site only if an IP is placed at that site. Equation 5 ensures that exactly $m$ IPs are placed at the IP sites. Finally, Equations 6 and 7 are the binary and the non-negativity constraints.

4. Alternative Solution Methods

Many studies in the location literature have been performed on the type of location problem presented in Section 3. Some of these studies were reviewed in Section 2. Sherali and Nordai (1980) show that the LAP with discrete candidate sites is NP-hard, even if all the demand points are located on a straight line. Cooper (1963) states that capacitated LAPs generally contain many optimal or near-optimal solutions.

The lack of a sharp, unique optimum creates an opportunity to develop a heuristic procedure that is likely to identify an optimal or near-optimal solution. Thus, we focus our efforts on finding good heuristic solutions to the IP location problem. We pursue the well-known alternating heuristic proposed by Cooper (1963, 1972). Starting with a set of initial locations, the
heuristic alternates between solving the allocation and location sub-problems until no improvement is possible. We also pursue “relocation heuristics” that seek to improve the solution by methodically altering the location of the IPs. One SD-based and two SA-based relocation heuristics are presented and evaluated.

4.1. Generating Initial Solutions
The solution methods described in the following subsections are implemented with two initial solution generation schemes: random start (RS) and circular start (CS). With RS, as the name implies, we randomly select $m$ candidate IP sites as the initial solution. Different RS initial solutions are used in each replication. A set of RS initial solutions is generated a priori, and the same set is used across all the methods. CS, on the other hand, consists of a unique initial solution for each value of $m$. It is motivated by the observation that, in general, good solutions to the IP location problem are obtained when the IPs are spread evenly across the SR. With CS, the selection of initial candidate sites is derived from the optimal packing of $m$ circles into a square; see Eckard (2010), among others. Given the optimal packing of $m$ circles, we select the candidate IP sites that most closely approximate the centroid of each circle. (We use the circle packing results from the literature.)

4.2. The Location and Allocation Subproblems
Given a fixed allocation of the DPs to the IPs, the problem reduces to $m$ rectilinear single-facility location problems that can readily be solved using the median location method (Love et al. 1988). On the other hand, given a fixed set of IP locations, the problem reduces to a pure allocation problem (and Equations 5 and 6 become redundant). We address the allocation problem for the three cases described below.
1. Minimize the empty AGV travel distance only, without the ESA constraint (EO-NEQ) – For this case, which serves as a benchmark, the objective function is given by Equation 8. Relaxing the ESA constraint (Equation 3), the optimum allocation is obtained simply by assigning each DP to the closest IP.

\[ z_{EO} = \frac{1}{N} \sum_{j} \sum_{i} x_{ij} \cdot d_{ij} \]  

(8)

2. Minimize the empty plus full AGV travel distance, without the ESA constraint (EF-NEQ) – For this case, the objective function is given by Equation 1. Relaxing the ESA constraint (Equation 3), the optimum allocation is obtained simply by assigning, one at a time, each DP \( j \) to the IP with the minimum \( d_{ij} + \bar{f}_i \) value. In the case of a tie, the DP is assigned to the lowest indexed IP.

3. Minimize the empty plus full travel distance, with the ESA constraint (EF-EQ) – This case represents the full and original problem described in Section 3. With the ESA constraint, the optimal allocation is obtained through the well-known transportation problem, where the unit costs are given by \( d_{ij} + \bar{f}_i \) for each arc.

4.3. The Alternating Heuristic (AH)

This method is based on Cooper’s well-known approach. It is initialized with a set of \( m \) candidate IP sites chosen as the initial solution. Given the IP locations, the allocation procedure optimally allocates each DP to the IPs. Given the solution to the allocation problem, the IP locations are re-optimized using the median location method. The heuristic alternates between the above two subproblems until either there is no change to the IP locations or until \( z \) remains unchanged for five consecutive iterations. The current solution is reported as the best solution.
4.4. The Steepest Descent (SD) Heuristic

The SD-based heuristic is motivated by Brimberg and Drezner (2013). In each iteration, up to 8 possible new locations are evaluated for each IP, leading to a maximum of $8m$ candidate solutions evaluated in each iteration. The 8 new locations are based on the 8 neighboring points on the grid. (An IP located on the perimeter has 5 neighbors, except for corner IPs, which have only three neighbors.) The optimal allocation must be obtained in order to compute $z$ for each candidate solution. Following the SD logic, the candidate solution that produces the largest reduction in $z$ is selected and the procedure is repeated. The heuristic stops when either no further improvement is possible by changing the IP locations or when $z$ remains unchanged for five consecutive iterations. The current solution is reported as the best solution.

4.5. Simulated Annealing (SA)-based Heuristic

Two SA-based heuristic procedures are implemented based on the annealing scheme presented in Tompkins et al. (2010). (For brevity, we do not present the annealing scheme here; the reader may refer to Tompkins et al., p. 348 for the details.) As previously stated, the SA-based heuristic seeks to reduce the value of $z$ by altering the IP locations (i.e., it is a relocation heuristic).

In each iteration, SA generates a trial solution by perturbing the current location of the IPs. We propose a new perturbation scheme, denoted by BA. It considers each IP in the current solution, one at a time, and with probability 0.50, the IP remains at its current location, and with probability 0.50, the IP is moved to one of the randomly-selected neighboring points on the grid. Note that more than one IP location may be altered under the BA scheme but an IP never moves by more than one grid.

We also implement the perturbation scheme proposed by Brimberg and Drezner (2013), denoted by BD. Under the BD perturbation scheme, the trial solution is generated by randomly
selecting one of up to $8m$ candidate solutions described in Section 4.4. Under the BD scheme, only one IP location is altered in each iteration, and the new IP location is always a neighboring point on the grid. In the remainder of the paper, SA-BA (SA-BD) refers to the SA-based heuristic using the BA (BD) perturbation scheme.

Other studies, see, for example, Aras, Ainel, and Orbay (2007), and Manzour-al-Ajdad (2012), use SA to perturb, in each iteration, the allocation of the DPs to the IPs. Considering the finding of Brimberg et al. (2000) that relocation heuristics are more efficient than reallocation heuristics, we opted not to use SA to perturb the allocation of the DPs to the IPs.

5. Computational Results and Comparison of the Heuristics

In this section we present the computational results and compare the performance of the heuristics described in the previous section. In Section 5.1, each IP may be located at any DP, including the interior points. In Section 5.2, the IPs are restricted to the perimeter of the SR. The appropriate case depends on the application and potential constraints placed on the location of the IPs.

5.1. Interior IP Locations

Given four heuristic procedures, and two possible starting solutions, we consider a total of eight solution approaches. For each approach we run 40 replications. (Only one replication is used for the SD heuristic with CS, and the AH heuristic with CS, since these approaches are deterministic.) For each approach, we report the median, minimum, and maximum $z$ value across the replications. We also report the number of times the best $z$ value is attained as well as the median runtime per replication (in seconds).

For each solution approach we consider three cases and present the results in Table 4-1, Table 4-2, and Table 4-3, respectively, for EO-NEQ, EF-NEQ, and EF-EQ, which were described in Section 4.2. The results are based on nine instances, i.e., $2 \leq m \leq 10$, in a $20 \times 20$ SR with
400 DPs. The “% Best known” in each Table represents the percent of time the heuristic approach yielded the best solution we were able to obtain for that problem instance.

From Table 4-1 we observe that SA-BD with RS most often attained the highest value for the “% Best known,” with values above 28% in all instances. In addition, it consistently obtains the best median $z$ value. However, SA-BD with RS resulted in the longest run times in all instances except for $m = 10$. The shortest run times are reported for AH. This is particularly relevant since the percent difference between the median $z$ values across all the solution approaches is less than 3.5%. In addition, AH with CS generally yields maximum $z$ values smaller than those obtained with SA-BD with RS.

From Table 4-1 we also observe that CS generally results in shorter run times. The impact on the run times is most noticeable for larger values of $m$. We note that CS reduced the median run times by more than 50% for SD, and more than 70% for SA, for $m \geq 8$. This suggests that, as a starting point, CS works quite well in terms of run times. However, in terms of solution quality, CS is unfortunately not consistent across all instances. This may lead to longer run times and poorer solutions when compared to RS. For SD, CS consistently reduces the median run times but yields mixed results in solution quality. (The “% Best known” is 100% for $m = 4, 5, 6$, and 0% otherwise.) For SA-BD, CS increased the median run times compared to RS for $m = 6$ and $m = 7$.

The results in Table 4-2 show a sharp reduction in the run times compared to Table 4-1. The performance of SD with RS is comparable to the performance of SA-BD with RS. Generally, SD shows a higher percent of replications attaining the best-known $z$ value. As was noted in Table 4-1, in Table 4-2 we again observe a small percent difference (less than 2%) for the median $z$ values across all the approaches.
Table 4-1. Computational results for the EO-NEQ case.

<table>
<thead>
<tr>
<th>Method</th>
<th>SD</th>
<th>AH</th>
<th>SA-BD</th>
<th>SA-BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>RS</td>
<td>CS</td>
<td>RS</td>
<td>CS</td>
</tr>
<tr>
<td>% Best known</td>
<td>50%</td>
<td>0%</td>
<td>35%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>5.912</td>
<td>5.918</td>
<td>5.950</td>
<td>5.933</td>
</tr>
<tr>
<td>Min</td>
<td>5.912</td>
<td>5.918</td>
<td>5.912</td>
<td>5.933</td>
</tr>
<tr>
<td>Max</td>
<td>5.918</td>
<td>5.918</td>
<td>5.968</td>
<td>5.933</td>
</tr>
<tr>
<td>% Best known</td>
<td>57%</td>
<td>0%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>4.965</td>
<td>4.950</td>
<td>5.043</td>
<td>5.000</td>
</tr>
<tr>
<td>Min</td>
<td>4.950</td>
<td>4.950</td>
<td>4.975</td>
<td>5.000</td>
</tr>
<tr>
<td>Max</td>
<td>5.020</td>
<td>4.950</td>
<td>5.540</td>
<td>5.000</td>
</tr>
<tr>
<td>% Best known</td>
<td>30%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>28%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>15%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>3.310</td>
<td>3.252</td>
<td>3.553</td>
<td>3.300</td>
</tr>
<tr>
<td>% Best known</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>22</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>26</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Due to the additional computational effort required to solve the transportation problem for each trial solution, the run times observed in Table 4-3 are larger. To control the run times, we imposed a maximum run time of 999 seconds per replication for each approach. The run time restriction clearly affected SA-BA the most, since the large majority of the replications for \( m \geq 5 \) terminated due to the run time limit. The relative performance of the solution approaches is similar to what we observed in Table 4-1, where SA-BD with RS generally yields slightly superior results in solution quality. SD yields comparable results with shorter run times, except for \( m = 10 \). As in Table 4-1, the run times in Table 4-2 for AH are very competitive and the median \( z \) values are at most only 2.3% larger than those obtained with SA.

With the addition of the ESA constraint, we expect the \( z \) values in Table 4-3 to be greater than or equal to those reported in Table 4-2. The percent difference between the best-known \( z \) values in Table 4-2 and Table 4-3 are 0.00%, 0.36%, 0.00%, 0.41%, 0.21%, 0.18%, 0.08%, 0.23%, 0.30% for \( m = 2 \) through 10, respectively, which is very small. This is not an entirely surprising result since without the ESA constraint, there is still a strong incentive to “spread out” the IPs fairly evenly in order to minimize the expected empty and the expected (empty + full) AGV travel times, which results in service areas that are not significantly different across the IPs. From a practical perspective, the above result also suggests that the IP location problem can be first solved without the ESA constraint (to avoid long run times), and the solution can subsequently be corrected by reallocating a few DPs.

Figure 4-3 depicts the best solutions from Table 4-3 for each \( m \) value. The DPs, shown as dots in the Figure, are shaded based on their allocation to the IPs, which are shown as diamonds. (Some of the DPs are not visible since they overlap with the IPs. In practice, such IPs would be placed immediately adjacent to the DPs, and the operator may drop the item into the DP.)
Table 4-2. Computational results for the EF-NEQ case.

<table>
<thead>
<tr>
<th>Method</th>
<th>SD</th>
<th>AH</th>
<th>SA-BD</th>
<th>SA-BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>RS</td>
<td>CS</td>
<td>RS</td>
<td>CS</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>18.300</td>
<td>18.300</td>
<td>18.348</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>18.300</td>
<td>18.300</td>
<td>18.300</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>18.300</td>
<td>18.300</td>
<td>18.703</td>
</tr>
<tr>
<td>% Best known</td>
<td>100%</td>
<td>100%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>17.308</td>
<td>17.308</td>
<td>17.346</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>17.308</td>
<td>17.308</td>
<td>17.328</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>17.712</td>
<td>17.308</td>
<td>18.136</td>
</tr>
<tr>
<td>% Best known</td>
<td>97%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>45%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Best known</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>15.808</td>
<td>15.820</td>
<td>15.912</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>15.796</td>
<td>15.820</td>
<td>15.834</td>
</tr>
<tr>
<td>% Best known</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>15.595</td>
<td>15.593</td>
<td>15.741</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>15.587</td>
<td>15.593</td>
<td>15.611</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>15.701</td>
<td>15.593</td>
<td>15.988</td>
</tr>
<tr>
<td>% Best known</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>15.467</td>
<td>15.451</td>
<td>15.605</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>15.451</td>
<td>15.451</td>
<td>15.486</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>15.496</td>
<td>15.451</td>
<td>15.747</td>
</tr>
<tr>
<td>% Best known</td>
<td>7%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>15.354</td>
<td>15.348</td>
<td>15.488</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>15.335</td>
<td>15.348</td>
<td>15.361</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>15.419</td>
<td>15.348</td>
<td>15.707</td>
</tr>
<tr>
<td>% Best known</td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Run time</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4-3. Computational results for the EF-EQ case.

<table>
<thead>
<tr>
<th>Method</th>
<th>SD</th>
<th>AH</th>
<th>SA-BD</th>
<th>SA-BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Best known</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Best known</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Best known</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Best known</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Best known</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

104
Figure 4-4 depicts the best solutions we obtained for a larger problem consisting of a $50 \times 50$ grid with 2,500 DPs. Although such a large number of DPs is not likely to exist in practice, we present Figure 4-4 to show the results for a “near-continuous” case (where the SR may be treated as a continuous area) and at the same time compare the results to Figure 4-3. The structure of the solution, i.e., the IP locations and the DP clusters, remain mostly unchanged compared to Figure 4-3, which implies that a larger problem does not alter the basic solution structure, and there may be an opportunity to develop solutions for a continuous SR.

![Figure 4-3. Best heuristic solutions obtained on a $20 \times 20$ grid](image)

It is instructive to note that, for a given set of IP locations, some solutions may result in a disconnected service area (such as the one shown in Figure 4-5(b)), even if the DP-to-IP allocation is solved to optimality through the transportation problem. In some cases, multiple DPs are equidistant to at least two IPs (which is not uncommon in problems with rectilinear distances). In
such cases, it is possible to reallocate the DPs by performing even swaps across the IPs, without increasing the $z$ value. Figure 4-5(a), which is also optimum in terms of the DP-to-IP allocation, represents an alternative solution with no disconnected service areas.

Figure 4-4. Best heuristic solutions obtained on a $50 \times 50$ grid.

Figure 4-5. Optimal allocation and the resulting service areas for given IP locations.
5.2. IP locations restricted to the perimeter

Depending on the application, in some cases the IPs may be restricted to the perimeter of the SR. We present the computational results for this case with the ESA constraint imposed. As before, we solve nine instances of the problem, i.e., \(2 \leq m \leq 10\) in a fixed grid of \(20 \times 20\) with 400 DPs. Narrowing our attention to the two best-performing heuristics from Table 4-3, i.e., SA-BD and SD, we present the results in Table 4-4, which are based on 40 replications with initial solutions based on RS for each approach.

The results in Table 4-4 suggest that there is little to no difference in the quality of the solutions between the two methods. However, the SD run time may be three to ten times faster.

The best solutions are plotted in Figure 4-6. We note that, when restricted to the perimeter, the IPs tend to concentrate towards the center of the sides, avoiding the corners. In addition, in some cases (see \(m = 3\)), the IP locations are not symmetric.

Table 4-4. Computational results with perimeter IPs and the EF-NEQ case.

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median</td>
<td>Min</td>
<td>Max</td>
<td>% Best known</td>
<td>Run time</td>
<td>Median</td>
<td>Min</td>
<td>Max</td>
<td>% Best known</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>25.040</td>
<td>24.000</td>
<td>27.300</td>
<td>15%</td>
<td>5</td>
<td>25.040</td>
<td>24.000</td>
<td>27.300</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.943</td>
<td>20.120</td>
<td>23.528</td>
<td>13%</td>
<td>100</td>
<td>20.943</td>
<td>20.120</td>
<td>23.528</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4-6. Best heuristic solutions obtained for perimeter IPs on a 20 × 20 grid.

In Table 4-4 we also compare the best $z$ values obtained for the interior case (Table 4-3) with the best values obtained for the perimeter case. Furthermore, we break AGV travel down to empty and full travel for each case as shown in Table 4-4. As anticipated, the expected empty AGV travel decreases with the addition of more IPs. However, there is a noticeable decrease in the rate of reduction starting with $m = 5$. Unless more IPs are needed for throughput reasons, the above result suggests that no significant benefits in empty AGV travel is achieved with 6 or more IPs. Considering the cost associated with each IP, the above result can be used as a general design guideline.

We also observe in Table 4-4 that full AGV travel increases with the number of IPs, which at first may appear to be a counterintuitive result. If only full travel is considered, the optimum
location of each IP is the center of the square, regardless of the number of IPs. That is, full travel is minimized when each IP is located at the center. However, when empty travel is considered, then as the number of IPs is increased, the IPs move away from the center and spread across the square in order to reduce empty AGV travel, resulting in an increase in full travel. (Recall that expected full travel is based on the expected distance to all the DPs, not just those in the service area of an IP.)

The ratio of the best $z$ values between perimeter and interior IPs reveals a significant advantage for interior IP locations. In all instances shown in Table 4-4, perimeter IP solutions are at least 24% larger in expected empty + full travel distance per item sorted. Considering larger $m$ values beyond those explored in Table 4-4, we found strong empirical evidence that the above ratio converges to approximately 1.4 as $m$ goes to infinity.

We provide an analytic approximation as $m$ goes to infinity for a continuous SR. Obviously, as $m$ goes to infinity, the service areas become infinitely small, and the optimal interior IP locations are obtained by placing an IP at each DP. (In the continuous model, a DP is a point inside the square.) As a result, expected empty AGV travel goes to zero. Expected full AGV travel, on the other hand, can be computed as the expected distance between two random points inside a square. For a unit square, Larson and Odoni (1981) show that the expected distance equals $2/3$. For the perimeter IP case, we approximate the empty travel distance as the distance form a random point in the square to the closest point on the perimeter. Expected full AGV travel is estimated as the shortest rectilinear distance from a random point in the square to a random point on the perimeter. For the unit square, in Chapter 3 we show that the above two distances are equal to $1/6$ and $5/6$, respectively. Hence, based on our approximation, we conclude that the above ratio is equal to 1.5 as $m$ goes to infinity.
6. Analytic Results with Two IPs

6.1. Interior Case ($m = 2$)

We derive closed-form analytic expressions for the expected empty and full AGV travel based on a continuous representation of the SR. More specifically, we consider a unit square with uniformly distributed DPs. (Results based on a unit square are defined as normalized expected distances and are denoted by a tilde “~.” The normalized expected distances can be scaled back to an $L \times L$ square simply by multiplying the normalized results by $L$.) With an origin located at the lower left-hand corner, let $(x_1, y_1)$ and $(x_2, y_2)$ denote the $(x, y)$ coordinates of IP 1 and 2, respectively. Without loss of generality, suppose $x_2 \geq x_1$. We consider the case when $y_1 \geq y_2$. (The case when $y_1 < y_2$ is covered in Appendix I.) The normalized expected empty travel distance, $\tilde{E}[e]$, and the normalized expected empty plus full travel distance, $\tilde{E}[e + f]$, are shown in Equations 9 and 10, respectively, as a function of the two IP coordinates. Please see Appendix I for the derivation.

\[
\tilde{E}[e] = \left(\frac{1}{12}\right) \left(6 - 3(-2 + x_2^2 - 3y_2^2 + 4x_1 - 2y_2x_1 + x_1^2 + 4y_1 + 2y_2y_1 - 2x_1y_1 - 3y_1^2 + 2x_2(y_2 - x_1 + y_1))
+ 2(x_2^3 + 3x_2x_1^2 + 3x_2^2(y_2 - x_1 + y_1) - x_1^2(-6 + 3y_2 + x_1 + 3y_1))\right)
\]  

Equation 9 attains a minimum value of $3/8$ at $(x_1, y_1) = (1/2, 3/4)$ and $(x_2, y_2) = (1/2, 1/4)$. Such an IP configuration splits the square horizontally into two equal halves, and it places an IP at the center of each half. We note that the optimal solution satisfies the ESA constraint.
\[\bar{E}[e + f] = \left(\frac{1}{12}\right)\left(24 + 14x_2^3 - 6x_2^4 + 6y_2^3 - 24x_1 + 15x_1^2 + 10x_1^3 - 6x_1^4 - 24y_1 \right.
\]
\[+ 6x_1y_1 - 6x_1^2y_1 + 27y_1^2 + 6x_1y_1^2 - 6x_1^2y_1^2 - 6y_1^3
\]
\[+ y_2^2(3 - 6x_1 + 6x_1^2 + 6y_1)
\]
\[- 3x_2^2(3 - 6y_2 + 2y_2^2 + 6x_1 - 4x_1^2 - 2y_1 - 2y_1^2)
\]
\[- 6y_2(-3x_1 + 3x_1^2 + y_1 + y_1^2) + 6y_2(-3x_1 + 3x_1^2 + y_1 + y_1^2)
\]
\[+ 6x_1(-3y_2 + y_2^2 + 3x_1 - x_1^2 - y_1(1 + y_1))\]  

(10)

The minimum value of Equation 10 is found via the Nelder-Mead method implemented in Mathematica (Wolfram 2019). A minimum value of 0.9167 is attained at \((x_1, y_1) = (0.5, 0.667)\) and \((x_2, y_2) = (0.5, 0.333)\). Such an IP configuration once again splits the square horizontally into two equal halves, and it places the IPs one-third of the way from the top and bottom sides. The optimal solution satisfies the ESA constraint.

6.2. Perimeter Case \((m = 2)\)

We must consider three cases; two IPs on one side, two IPs on adjacent sides, and two IPs on the opposing sides of the square. Table 4-5 summarizes the normalized expected travel distance expressions obtained for \(\bar{E}[e]\) and \(\bar{E}[e + f]\). The third column, labeled “Min,” shows the minimum value of each expression, and the last column shows the optimum solution. (Grey cells indicate results obtained via the Nelder-Mead method.) The derivations for Equations 11 through 15 are presented in Appendix J.

In deriving Equations 11 and 12 we assume, without loss of generality, that the two IPs are located on the top or bottom side of the square, and that \(x_1 \leq x_2\) (with an origin located at the lower left-hand corner of the unit square). Likewise, in deriving Equations 13 and 14 we assume, without loss of generality, that IP 1 is on the bottom side, and IP 2 is on the right side, of the square.
We let \((x, 0)\) and \((1, y)\) be the \(x\) and \(y\)-coordinates of IP 1 and 2, respectively. We assume \(1 - x \leq y\).

Last, in deriving Equations 15 and 16 we assume, again without loss of generality, that the two IPs are on the top and bottom sides. We let \((x_1, 0)\) and \((x_2, 1)\) be the \(x\) and \(y\)-coordinates of IP 1 and 2, respectively, and further assume that \(x_1 \leq x_2\).

Table 4-5. Analytic expected travel distance expressions for two IPs – perimeter case.

<table>
<thead>
<tr>
<th>Case</th>
<th>Normalized expected travel distance</th>
<th>Min</th>
<th>Optimal((x_1, x_2)) or ((x, y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 side</td>
<td>(\bar{E}[e] = \frac{1}{2} \left[ 1 + x_1^2 + \frac{(x_2 - x_1)^2}{2} + (1 - x_2)^2 \right] )</td>
<td>5/8</td>
<td>(\left(\frac{1}{4}, \frac{3}{4}\right))</td>
</tr>
<tr>
<td></td>
<td>(\bar{E}[e + f] = \frac{1}{4} \left[ 8 + 2x_1^3 - 8x_2 + 9x_2^2 - 2x_2^3 - 2x_1x_2(1 + x_2) + x_1^2(1 + 2x_2) \right] )</td>
<td>17/12</td>
<td>(\left(\frac{1}{3}, \frac{2}{3}\right))</td>
</tr>
<tr>
<td>Adj. sides</td>
<td>(\bar{E}[e] = \frac{1}{12} \left[ 11 - 2x_1^3 - 12y - 6xy + 9y^2 + x^2(3 + 6y) \right] )</td>
<td>0.527</td>
<td>((0.347, 0.742))</td>
</tr>
<tr>
<td></td>
<td>(\bar{E}[e + f] = \frac{1}{12} \left[ 23 + 10x_1^3 - 6x_2^4 - 24y + 27y^2 - 6y^3 - 6xy(1 + y) + x^2(-3 + 6y + 6y^2) \right] )</td>
<td>1.301</td>
<td>((0.421, 0.645))</td>
</tr>
<tr>
<td>Opp. sides</td>
<td>(\bar{E}[e] = \frac{1}{12} \left[ 9 - 2x_1^3 - 6x_2 + 3x_2^2 + 2x_2^3 + x_1^2(3 + 6x_2) - 6x_1(1 - x_2 + x_2^2) \right] )</td>
<td>1/2</td>
<td>(\left(\frac{1}{2}, \frac{1}{2}\right))</td>
</tr>
<tr>
<td></td>
<td>(\bar{E}[e + f] = \frac{1}{12} \left[ 21 + 10x_1^3 - 6x_1^4 - 12x_2 + 3x_2^3 + 14x_1^2 - 6x_1^2(2 - 3x_2 + 3x_2^2) + 3x_1^2(1 - 2x_2 + 4x_2^2) \right] )</td>
<td>5/4</td>
<td>(\left(\frac{1}{2}, \frac{1}{2}\right))</td>
</tr>
</tbody>
</table>
From the results shown in Table 5, the two IPs are located optimally when they are placed at the center of the opposing sides of the square. We extend this observation and present two properties as follows:

**Property 1:** Given any solution with two IPs on the same side, one can always construct a better solution with two IPs on opposite sides. Therefore, a single-sided solution can never be optimal.

**Proof:** Let \((x_1, 0)\) and \((x_2, 0)\) denote the \(x\) and \(y\)-coordinates of the two IPs, where \(x_1 \leq x_2\). From Equation 12 we obtain the normalized expected empty plus full travel distance, \(\tilde{E}_S[e + f]\). We construct a new solution with two IPs on opposite sides, i.e., \((x_1, 0)\) and \((x_2, 1)\), and compute the normalized expected empty plus full travel, \(\tilde{E}_O[e + f]\), from Equation 15. The difference between the two is given by:

\[
\tilde{E}_S[e + f] - \tilde{E}_O[e + f] = \frac{1}{4} - \frac{x_1^3}{3} + \frac{x_1^4}{2} + x_1(x_2 - 1)^2 - x_2 - x_1^2(x_2 - 1)x_2 + x_2^2 - \frac{5x_3^2}{3} + \frac{x_2^4}{2},
\]

which attains a minimum value of \(7/96\) \((> 0)\), when \((x_1, x_2) = (0, 1/2)\).

(A similar proof can be constructed for the case with two IPs on adjacent sides.) Although Property 1 establishes that a single-sided solution can never be optimal, with Property 2 we show that a single-sided solution may perform better than a 2-sided solution with poorly placed IPs.

**Property 2:** Property 1 does not imply that any solution with two IPs on opposite sides of the square is better than a single-sided solution.

**Proof:** At \((0,0)\) and \((1,1)\), \(\tilde{E}_O[e + f] = 5/3\), which represents a maximum value of \(\tilde{E}_O[e + f]\). On the other hand, the minimum on \(\tilde{E}_S[e + f]\) provided in Table 5 is equal to \(17/12 < 20/12 = 5/3\).
Furthermore, from the results presented in Figure 4-6, we conjecture that no optimal solution with \( m \) or fewer IPs will have more than \( \lceil m/4 \rceil \) IPs on one side of the square. (\( \lceil x \rceil \) denotes the ceiling function.) As shown in Figure 4-6, for \( m \leq 4 \), no side has more than one IP, for \( m \leq 8 \), no side has more than two IPs, and for \( m \leq 10 \), no side has more than three IPs.

6.3. Expected Distances and the Number of AGVs

A key decision in the design of an AGV-SS is the number of AGVs required since it represents a major cost component of the system and at the same time, it impacts the performance of the system. The number of AGVs must be determined in order to sustain a certain sortation rate (expressed as the number of items sorted per unit time), which is a function of the AGV travel time per item or the “AGV cycle time.” If the AGV queueing time at the IPs and the DPs is negligible, then the AGV cycle time is simply equal to the expected empty plus full travel time. (The expected time required to place an item on the AGV at an IP, and the expected time required to discharge the item at a DP, can be added to the travel time.) Assuming negligible impact due to AGV acceleration/deceleration and possible congestion, the expected travel times can be obtained from the expected travel distances derived in our study. Thus, based on our expected distance results, it would be straightforward to construct a rudimentary model to estimate the number of AGVs required. Such a model would be very useful in practice during the early stages of system design.

7. Conclusions and Future Research

We present a computational study to compare the performance of various heuristics to solve the IP location problem in an AGV-SS. Other studies have used exact and approximate algorithmic approaches to solve only small to moderate-sized instances of problems similar to the one we study. Hence, we concentrate our efforts on finding good heuristic solutions, and we select methods that have been successfully tested in other studies.
Based on our computational results, SA-BD with CS, and SD with CS, are robust in terms of solution quality but they require longer run times. On the other hand, AH consistently results in significantly shorter runtimes, with objective function values that are only about 2% higher than the best solutions. We observe only a small percent difference in the objective function value when the ESA constraint is introduced; it is less than 0.4% in all instances. Hence, a practical solution strategy would be to first solve the IP location problem without the ESA constraint, and then modifying the solution to satisfy the ESA constraint by reallocating a few DPs. In addition, we show that the structure of the solution – for a given $m$ value – does not change significantly as more DPs are added.

We also treat a special case of the problem where the candidate IP sites are restricted to the perimeter of the SR. We solve the problem for $2 \leq m \leq 10$ using SA-BD with RS, and SD with RS. The SD run times are 3 to 10 times shorter, and it yields comparable results in solution quality to the SA-BD solutions. We find that the best objective function values for perimeter IPs are more than 24% larger than those obtained with interior IPs. Empirical evidence strongly suggests that the above percentage converges to 40% as $m$ goes to infinity. We present a continuous analytic model to approximate the above percentage and show that it converges to 50% as $m$ goes to infinity.

We analyze the individual contribution of empty and full travel to the DC cycle travel distance. As anticipated, the expected empty AGV travel decreases with the addition of more IPs. However, there is a noticeable decrease in the rate of reduction starting with $m = 5$. Unless more IPs are needed for throughput reasons, the above result suggests that no significant benefits in empty AGV travel is achieved with 6 or more IPs. Considering the cost associated with each IP, the above result can be used as a general design guideline.
We present continuous analytic expressions for $E[e + f]$ when two IPs are to be located in a squared sortation region. The optimal IP locations based on the analytic models match the best heuristic results from our computational study. This is further evidence that the solution structure may not change significantly in problems with a larger number of DPs. Future research can investigate this claim, considering larger problem instances.

References


Chapter 5
Conclusions and Future Research Opportunities

In this chapter we recapitulate the main results of this work. We summarize conclusions from Chapters 2, 3, and 4 and highlight opportunities for future research.

In Chapter 2, we compare the performance of a miniload AS/R system and an AGV-OPS using simulation. We compare the configuration/size of each system that yields approximately the same throughput in line items picked/hour. Given the assumptions and parameters of the study, our primary conclusion is that a miniload system with 4 aisles and a conveyor loop yields approximately the same throughput as an AGV-OPS with 50 AGVs. We also show numerically a strongly intuitive result that, for both systems, a favorable balance between picker and equipment (miniload or AGV) utilization is achieved when the expected container retrieval-storage cycle time is approximately equal to the expected pick time per container. The results also indicate that the number of open orders per pick station is an important control parameter to ensure high picker utilization while avoiding congestion. Although we varied the number of SKUs and the number of miniload aisles/number of AGVs, our conclusions are limited to the configurations and parameter values we tested. Additional studies are needed to compare the above two systems and other OP systems under alternative parameter values. Last, we compare the two systems on other factors such as flexibility/scalability, parallel processing/redundancy, ergonomics, footprint/cube utilization, and the security of items. In summary, an AGV-OPS possesses advantages such as modular scalability and flexibility that would be particularly valuable to an online retail business.
but at the expense of inferior cube utilization and relatively poor ergonomics, which has strong implications for performance and long-term cost.

We used three categories of SKUs but assumed randomized storage within each category. Future studies may consider slotting based on SKU velocities. Both systems would benefit from storing high-velocity SKUs closer to the pick stations (although storing the fast-moving SKUs in the same area may create AGV congestion in the Kiva system). Another direction to explore is the impact/timing of replenishment in both systems. Another configuration to investigate would be the miniload system with the conveyor loop replaced by unit load AGVs to transport individual trays.

In Chapter 3, we derive closed-form analytic expressions for the expected travel distance of dual command cycles \(E[DC]\) in an AGV-OPS operating under two order assignment rules. Under the random assignment rule (RAN), an order is assigned to any PS with equal probability. Under the closest assignment rule (CLO), the order is assigned to the closest PS. Exact and approximate results are presented for different PS configurations and FA shapes as a function of the width-to-length ratio. The results offer not only valuable insight concerning the expected travel distances, but they would also be needed in analytic design and performance evaluation models where the expected travel distances are used as input to the model.

We perform numerical experiments to assess the effect of the FA shape, order assignment rule, PS configuration, width-to-length ratio, and batch size on the expected travel distance. We conclude that order assignment rule has the biggest impact on \(E[DC]\). In almost all cases considered, CLO results in a lower \(E[DC]\) than RAN, up to 65% lower. The effect of PS configuration on \(E[DC]\), is most noticeable in narrow FA shapes (small width-to-length ratio) and
cases with small $N$ values. Nonetheless, the effect of reducing the width-to-length ratio is consistently an increase in $E[DC]$.

Based on the results of the numerical experiment we would recommend a square-shaped FA with PS on all four sides and the use of CLO rule. The CLO rule could potentially be implemented in an existing system, where FA shape and PS configuration modification may be restricted. In early stage design, the results of this study suggest the PS configuration should be carefully considered. Locating PS on all sides of the FA or the long sides, rather than the short sides is beneficial. We can attribute that to better coverage of the FA provided by PS, since they are farther apart. Our results also suggest that the geometric shape of the FA has less impact on the expected travel distance than the width-to-length ratio. Designers should favor symmetric shapes, that is when width-to-length ratio is small. In considering these insights, we note that this study considers three convex FA shapes, while the study of non-convex FA shapes remains an open research question.

We further explore the feasibility of implementing the CLO rule, in light of the possibility of uneven use of PS. We conclude that, in a square-shaped FA, PSs can be located to achieve an even usage (balanced allocation of pods to PSs), while suffering little impact on the $E[DC]$. Future research can investigate methods to optimize the location of PSs that considers even PS usage. Another consideration related to the location of the PSs is finding alternatives to locating the PSs in the perimeter of the FA. For example, what can be gained by locating the PSs in the center of the FA.

It is important to emphasize that our conclusions are based on the assumption that every pod is equally likely to be retrieved (justified under randomized storage) and pods are modeled
continuously over the FA. Future research could extend the results presented here in a system with turnover-based storage.

In Chapter 4, we perform a computational study to compare the performance of heuristics for the IP location problem for AGV-SS. Other studies have used exact and approximate algorithmic approaches to solve only small to moderate instances of problems similar to the one we present. Hence, we concentrate our efforts on finding good heuristic solutions and select methods that have been successfully tested in other studies.

A big challenge to scaling the problem instances solved, i.e. increasing the number of DP or the number of IPs, is enforcing equal allocation among the IPs. We propose a model with the objective of minimizing the expected empty + full AGV travel distance \( E[e + f] \), that treats the IPs as capacitated facilities. We refer to the facility capacity constraint as equal service area (ESA) constraint.

In the computational study we solve problem instances with and without the inclusion of the ESA constraint. We observe a small percent increase in \( E[e + f] \) after the introduction of the ESA constraint; less than 0.4% in all instances. Hence, a practical solution strategy can consist of solving the IP location problem without the ESA constraint. The solution can be modified to satisfy the constraint with a subsequent reallocation step. We did not investigate this approach. In addition, we observe that the structure of the solution – for a set number of IPs – does not change significantly when considering a larger problem instance (larger number of DPs). We propose an approach to solving large problem instances approximately by scaling the results we present.

We treat a special case of the IP location problem where the set of candidate IP sites is restricted to the candidate sites in the perimeter of the sortation region (we denote this as “perimeter case”). Based on the best heuristic solutions, \( E[e + f] \) for perimeter cases are over 24% larger.
than “interior cases” (when IPs can be located anywhere in the sortation region). Strong empirical evidence suggests this ratio converges to 40% as the number of IPs goes to infinity. We present a continuous analytic model that approximates the ratio and converges to 50% as the number of IPs goes to infinity.

An analysis of the individual contribution of empty and full travel to total $E[e + f]$ is performed. As anticipated, the expected empty AGV travel decreases with the addition of more IPs. However, there is a noticeable decrease in the rate of reduction starting with 5 IPs. Unless more IPs are needed for throughput reasons, the above result suggests that no significant benefits in empty AGV travel is achieved with 6 or more IPs. Considering the cost associated with each IP, the above result can be used as a general design guideline.

We present continuous analytic expressions for $E[e + f]$ when two IPs are to be located in a squared sortation region. The optimal IP locations based on the analytic models match the best heuristic results from our computational study. This is further evidence that the solution structure may not change significantly in problems with a larger number of DPs. Future research can investigate this claim, considering larger problem instances.
Appendices
Appendix A
Derivations for */RECT/RAN/1 – \( E[DC] \).

For each PS configuration described in Section 3.3, we derive the expected DC distance in a normalized \((b \times 1)\) rectangular FA. First, we consider \( \bar{E}[OUT] \). Given the assumptions in Section 4.1, and the RAN rule, \( \bar{E}[OUT] \) is defined as the expected distance between a random point inside the rectangle and a random point on the perimeter. The expected distance is easily obtained by conditioning on the location of the point on the perimeter.

A.1. The 1L (1S) PS configuration

By assumption 1, AGV travel is either parallel or orthogonal to the pick side. The parallel travel distance \((d_p)\) and the orthogonal travel distance \((d_o)\) are independent random variables. Hence, \( d_o \) is \( \sim U(0, b) \) or \( \sim U(0, 1) \) if the pick side is of length 1 or \( b \), respectively. Furthermore, \( d_p \) is the distance between two random points that are uniformly and independently distributed between \((0, 1)\) or \((0, b)\) if the pick side is of length 1 or \( b \), respectively. Hence, we obtain:

\[
\bar{E}[OUT] = \frac{1}{2} + \frac{b}{3} \quad \text{for 1S} \tag{A1}
\]

\[
\bar{E}[OUT] = \frac{b}{2} + \frac{1}{3} \quad \text{for 1L} \tag{A2}
\]

We now consider \( \bar{E}[TB] \). By assumption 5, a pod is returned by the AGV to any point inside the rectangle. Since the next pod to be visited by the AGV is equally likely to be anywhere in the rectangle, \( \bar{E}[TB] \) is obtained by the rectilinear distance between two points in the rectangle. Consider the travel distance parallel to 1-side \((d_1)\) and the travel distance parallel
to $b$-side ($d_b$). The random variable $d_1$ ($d_b$) is the distance between two uniformly distributed random points in $[0,1)$ $([0,b])$. The resulting expression is,

$$
\bar{E} [TB] = d_1 + d_b = \frac{1}{3} + \frac{b}{3}
$$

(A3)

A.2. The 2L (2S) PS configuration

The results are the same as those obtained in A.1.

A.3. The 4 PS configuration

The probability that a point on the perimeter is sampled from the side of length $b$ or $1$ is $b/(1+b)$ or $1/(1+b)$, respectively. Using the results in A.1, we obtain:

$$
\bar{E} [OUT] = \frac{b}{1+b} \left( \frac{1}{2} + \frac{b}{3} \right) + \frac{1}{1+b} \left( \frac{b}{2} + \frac{1}{3} \right) = \frac{(b^2 + 3b + 1)}{3(1+b)}.
$$

(A4)

$\bar{E} [TB]$ follows from A.1. Hence, the expected DC distance as defined in Section 4 is given by:

$$
\bar{E} [DC] = 2\bar{E} [OUT] + \bar{E} [TB] = \frac{3b^2 + 8b + 3}{3(b+1)}.
$$

(A5)
First, we address the PS configuration 1L. Consider a $b \times 1$ rectangle as depicted in Figure B1.

![Figure B-1. Travel to the closest PS on the long side of the FA](image)

**B.1. $N = 1$:**

Consider, for example, pod 2 in Figure B1. The distance to the closest PS is depicted by the vertical line between points $(X_2, Y_2)$ and $(X_2, 0)$. Designating the vertical distance by the random variable $Y$, due to assumptions 2 and 3 (section 4.1), we have $Y \sim U(0, b)$. Hence, the expected travel distance is given by:

$$\bar{E}[OUT] = E[Y] = b/2.$$  \hspace{0.5cm} (B1)

**B.2. $N = 2$:**

From Figure B1, it is clear that:
\[ \tilde{E}[\text{OUT}] = E[Y_1 + Y_2 + |X_1 - X_2|] = 2 \left( \frac{b}{2} \right) + \frac{1}{3}. \] (B2)

**B.3.** \( N \geq 3 \):

We define \( d_o \) and \( d_p \) as in Appendix A, so that \( \tilde{E}[\text{OUT}] = E[d_o] + E[d_p] \). In Figure B1, \( d_o \) is the total travel distance orthogonal to the bottom of the rectangle, and \( d_p \) is the total travel distance parallel to the bottom of the rectangle. More specifically, \( d_p \) is the total distance in the \( x \)-axis from all the pods to the median \( x \)-coordinate.

a. Orthogonal travel distance \((d_o)\) follows from the result in item 1.

\[
E[d_o] = N \cdot E[Y] = N \left( \frac{b}{2} \right) \tag{B3}
\]

b. Parallel travel distance \((d_p)\) can be computed from the order set \( \{X(1), X(2), \ldots, X(N)\} \). The expected distance between consecutive uniformly distributed points \( X(i) \) and \( X(i+1) \) in \([0,1]\) is \( 1/(N + 1) \). Figure B-2 depict the distance from each point \( X(i) \) to the median point.

It is clear that \( E[d_p] \) has the form,

\[
E[d_p] = \frac{1}{N + 1} f(N). \tag{B4}
\]

where \( f(N) \) can be obtained for \( N \) odd and \( N \) even as shown below:

\[
f(N) = \begin{cases} 
\sum_{i=1}^{N-1} 2i, & N \text{ odd} \\
\left( \sum_{i=1}^{N-2} 2i \right) + \frac{N}{2}, & N \text{ even} 
\end{cases} \tag{B5}
\]

Hence,
\[ E[d_p] = \begin{cases} \frac{N - 1}{4}, & N \text{ odd} \\ \frac{N^2}{4(N + 1)}, & N \text{ even} \end{cases} \]  \hspace{1cm} (B6)

Figure B-2. Distance from all the points to the median point for \( N = 4 \) and \( N = 5 \).

The results for the 1S PS configuration are given as follows:

**B.4.** \( N = 1 \):

\[ \bar{E}[OUT] = 1/2. \]  \hspace{1cm} (B7)

**B.5.** \( N = 2 \):

\[ \bar{E}[OUT] = 1 + \frac{b}{3}. \]  \hspace{1cm} (B8)

**B.6.** \( N \geq 3 \)

\[ \bar{E}[OUT] = \begin{cases} \frac{N}{2} + b \frac{N - 1}{4}, & N \text{ odd} \\ \frac{N^2}{2} + b \frac{N^2}{4(N + 1)}, & N \text{ even} \end{cases} \]  \hspace{1cm} (B9)
Appendix C
Derivations for 2•/RECT/CLO/N – \( \bar{E}[OUT] \).

First, we derive the expression for the 2L configuration. Consider the \((b \times 1)\) rectangle depicted in Figure B1.

C.1. \( N = 1 \):

Recall that \( Y \sim U(0, b) \). Hence, the expected distance is given by:

\[
\bar{E}[OUT] = E[\min(Y, b - Y)] = E[E[\min(Y, b - Y) | Y]]
\]

\[
= \frac{1}{2} E\left[ \min(Y, b - Y) | Y \leq \frac{b}{2} \right] + \frac{1}{2} E\left[ \min(Y, b - Y) | Y > \frac{b}{2} \right]
\]

\[
= \frac{1}{2} \left[ E[Y | Y \leq \frac{b}{2}] + E[b - Y | Y > \frac{b}{2}] \right] = \frac{b}{4}
\]  

(C1)

C.2. \( N = 2 \):

a. Orthogonal travel distance \((d_o)\):

\[
E[d_o] = E[\min(Y_1 + Y_2, 2b - Y_1 - Y_2)] = E[Y_1 + Y_2 | Y_1 + Y_2 < b].
\]  

(C2)

Letting \( Z = (Y_1' + Y_2') \), if \( Y_i' \sim U(0,1) \), then \( Z \) follows an Irwin-Hall distribution. For the special case with two uniform random variables, \( Z \) takes the form of a triangular distribution with a pdf, \( f(z) \), given by:

\[
f(z) = \begin{cases} 
  z, & 0 \leq z \leq 1 \\
  2 - z, & 1 \leq z \leq 2
\end{cases}
\]  

(C3)

Using \( Z \) to compute \( E[d_o] \),

\[
E[d_o] = b * E \left[ Z | Z < \frac{N}{2} \right], \text{ and}
\]  

(C4)
\begin{equation}
E \left[ Z \mid Z < \frac{N}{2} \right] = \int_{0}^{\frac{N}{2}} z \frac{f(z)}{F \left( \frac{N}{2} \right)} dz = \frac{2}{3}.
\end{equation}

(C5)

b. Parallel travel distance \((d_p)\):

\[ E[d_p] = 1/3 \]

(C6)

C.3. \(N \geq 3\):

a. Orthogonal travel:

We let \(Z = \sum_{i=1}^{n} Y_i'\) and proceed as in C.2. However, we approximate \(Z\) with \(Z'\), which follows a Normal distribution with mean \(\mu = \frac{n}{2}\), and standard deviation \(\sigma = \left(\frac{N}{12}\right)^{\frac{1}{2}}\).

\[ E[d_o] = b \ast E \left[ Z \mid Z < \frac{N}{2} \right] \approx b \ast E \left[ Z' \mid Z' < \frac{N}{2} \right] \]

(C7)

\[ E \left[ Z' \mid Z' < \frac{N}{2} \right] = \frac{N}{2} - \left(\frac{N}{6\pi}\right)^{\frac{1}{2}} \]

(C8)

b. Parallel travel: Same as \(N = 2\), 1L PS configuration in Appendix B.

The results for 2S are as follows:

C.4. \(N = 1\):

\[ \tilde{E}[OUT] = \frac{1}{4} \]

(C9)

C.5. \(N = 2\):

\[ \tilde{E}[OUT] = \frac{2}{3} + \frac{b}{3} \]

(C10)

C.6. \(N \geq 3\):

\[ \tilde{E}[OUT] = \begin{cases} 
\frac{N}{2} - \left(\frac{N}{6\pi}\right)^{\frac{1}{2}} + b \left[ \frac{N - 1}{4} \right], & N \text{ odd} \\
\frac{N}{2} - \left(\frac{N}{6\pi}\right)^{\frac{1}{2}} + b \left[ \frac{N^2}{4(N + 1)} \right], & N \text{ even}
\end{cases} \]

(C11)
Appendix D
Derivations for 4/RECT/CLO/1 – $\bar{E}[OUT]$ and 4/RECT/CLO/2 – $\bar{E}[OUT]$.

Consider a $(b \times 1)$ rectangle as shown in Figure D1, which depicts the four regions defined by the contour lines obtained under the CLO rule. The closest PS to any pod in region 1 is located on the left side, while the closest PS to any pod in region 2 is located on the top side, and so on.

![Figure D-1. Contour lines in a rectangular FA under CLO rule.](image)

**D.1. $N = 1$:**

By conditioning on whether the pod is in a triangular region or a trapezoidal region, it follows that:

$$\bar{E}[OUT] = \bar{E}[OUT|\text{triangle}]P(\text{triangle}) + \bar{E}[OUT|\text{trapezoid}]P(\text{trapezoid}),$$

(D1)

where the two probabilities are given by:

$$P(\text{triangle}) = \frac{b}{2}, \text{ and}$$

(D2)
\[ P(\text{trapezoid}) = 1 - \frac{b}{2}. \quad \text{(D3)} \]

It is straightforward to compute the expected travel distance for each region:

\[ E[\text{OUT}|\text{triangle}] = \frac{b}{6}, \quad \text{and} \]

\[ E[\text{OUT}|\text{trapezoid}] = \frac{b(3 - 2b)}{6(2 - b)}. \quad \text{(D5)} \]

Substituting into Equation (D1) and simplifying, we obtain:

\[ E[\text{OUT}] = \frac{1}{12} b(3 - b) \quad \text{(D6)} \]

**D.2.** \( N = 2 \):

Let \( d \) be the distance traveled by the two pods under CLO; i.e., \( \bar{E}[\text{OUT}] = E[d] \). Then,

\[ d = \min \left( \begin{array}{c} Y_1 + Y_2 + |X_1 - X_2|, \\ X_1 + X_2 + |Y_1 - Y_2|, \\ 2b - Y_1 - Y_2 + |X_1 - X_2|, \\ 2 - X_1 - X_2 + |Y_1 - Y_2| \end{array} \right) \quad \text{(D7)} \]

We rewrite \( d \) as:

\[ d = 2 \cdot \min \left( \begin{array}{c} \min(Y_1, Y_2), \\ \min(X_1, X_2), \\ \min(1 - Y_1, 1 - Y_2), \\ \min(1 - X_1, 1 - X_2) \end{array} \right) + |Y_1 - Y_2| + |X_1 - X_2|, \quad \text{(D8)} \]

or alternatively \( d = 2 \min(d_1, d_2) + C \), were \( d_i \) is the distance from pod \( i \) to the closest point on the perimeter of the rectangle, and \( C \) is the rectilinear distance between the two pods. Recall that the expected distance between two points is given by:

\[ E[C] = \frac{b}{3} + \frac{1}{3}. \quad \text{(D9)} \]

When \( b = 1 \), \( d_i \) follows the triangular distribution with a pdf, \( f(z) \), given by:
\[ f(z) = \begin{cases} 
4 - 8z, & 0 \leq z \leq \frac{1}{2}, \\
0, & \text{otherwise}
\end{cases} \text{ yielding (D10)} \]

\[ E[\min(d_1, d_2) \mid b = 1] = \frac{1}{10}. \] (D11)

For \( b \leq 1 \), we can obtain the distribution of \( d_1 \) and \( d_2 \) by conditioning on the location of the two pods. Let \( A_1 \) be the event that either pod is in either triangular region (i.e., both pods are in region 1, or both pods are in region 3, or one pod is in region 1 and the other pod is in region 3). Likewise, let \( A_2 \) be the event that either pod is in either trapezoidal region, and \( A_3 \) be the event that one pod is in a triangular region and the other pod is in a trapezoidal region. The probability for each event is given by:

\[ P(A_1) = \left( \frac{2}{2} \right) \left( \frac{b}{2} \right)^2, \] (D12)

\[ P(A_2) = \left( \frac{2}{2} \right) \left( 1 - \frac{b}{2} \right)^2, \text{ and} \] (D13)

\[ P(A_3) = \left( \frac{2}{1} \right) \left( \frac{b}{2} \right) \left( 1 - \frac{b}{2} \right). \] (D14)

For \( A_1 \), we have

\[ f(z) = \begin{cases} 
\frac{8}{b^2} \left( \frac{b}{2} - z \right), & 0 \leq z \leq \frac{b}{2} \\
0, & \text{otherwise}
\end{cases} \] which yields \( (D15) \)

\[ E[\min(d_1, d_2) \mid A_1] = \frac{b}{10}. \] (D16)

For \( A_2 \), we have,

\[ g(z) = \begin{cases} 
\frac{4 - 8z}{(2 - b)b'}, & 0 \leq z \leq \frac{1}{2} \\
0, & \text{otherwise}
\end{cases} \] which yields \( (D17) \)
\[ E[\min(d_1, d_2) \mid A_2] = \frac{b(b^2 - 25b + 20)}{30(b - 2)^2}. \]  

(D18)

For \( A_3 \), without loss of generality, let \( d_1 \) follow \( f(\cdot) \) and \( d_2 \) follow \( g(\cdot) \), we have:

\[ E[\min(d_1, d_2) \mid A_3] = \frac{b(3b - 5)}{20(b - 2)}. \]  

(D19)

By conditioning on events \( \{A_i, i \in (1, 2, 3)\} \) and the probabilities in (D12) – (D14) we obtain:

\[ E[\min(d_1, d_2)] = E[E[\min(d_1, d_2) \mid A_i]] = \frac{b}{60}(b^2 - 5b + 10). \]  

(D20)

Finally, the expression from \( E[OUT] \) follows:

\[ E[d] = E[2 \min(d_1, d_2) + C] = 2 \ E[\min(d_1, d_2)] + E[C] \]  

(D21)

\[ E[OUT] = \frac{(b^3 - 5b^2 + 20b + 10)}{30} \]  

(D22)
Appendix E
Derivation for •/DIAM/RAN/1 – $E[OUT]$ for a FA of unit area.

We derive $F(z)$, the cdf of the travel distance in the horizontal direction from any point inside the diamond to a point on the perimeter. Without loss of generality, we assume the point on the perimeter falls in the upper left-hand side of the diamond, with an $x$-coordinate measured from an origin placed at the leftmost point of the diamond. By definition, we have:

$$F(z | a \leq \frac{\sqrt{2}}{4}) = \begin{cases} 
4az, & 0 \leq z < a \\
4a^2 + (3a + z)(z - a), & a \leq z < \frac{\sqrt{2}}{2} - a \\
\frac{1}{2} + \left(\frac{3\sqrt{2}}{2} - z - a\right)(z - \sqrt{2} + a), & \frac{\sqrt{2}}{2} - a \leq z \leq \sqrt{2} - a
\end{cases}$$

(E1)

and

$$f(z | a \leq \frac{\sqrt{2}}{4}) = \begin{cases} 
4a, & 0 \leq x < a \\
2(a + x), & a \leq x < \frac{\sqrt{2}}{2} - a \\
2(\sqrt{2} - x - a), & \frac{\sqrt{2}}{2} - a \leq x \leq \sqrt{2} - a
\end{cases}$$

(E2)

$$F(z | a > \frac{\sqrt{2}}{4}) = \begin{cases} 
4az, & 0 \leq z < \frac{\sqrt{2}}{2} - a \\
2\sqrt{2}a - 4a^2 + \left(z - \frac{\sqrt{2}}{2} + a\right)(\sqrt{2} - 2z + 2a), & \frac{\sqrt{2}}{2} - a \leq z < a \\
4\sqrt{2}a - 4a^2 - 1 + (2\sqrt{2} - 3a - x)(x - a), & a \leq z \leq \sqrt{2} - a
\end{cases}$$

(E3)

and

$$f(z | a > \frac{\sqrt{2}}{4}) = \begin{cases} 
4a, & 0 \leq x < \frac{\sqrt{2}}{2} - a \\
2(\sqrt{2} - 2x), & \frac{\sqrt{2}}{2} - a \leq x \leq a \\
2(\sqrt{2} - x - a), & a \leq x \leq \sqrt{2} - a
\end{cases}$$

(E4)

Hence,
\[ E \left[ z \mid a \leq \frac{\sqrt{2}}{4} \right] = F(a) \int_{0}^{a} zf(z)dz + \left( F\left( \frac{\sqrt{2}}{2} - a \right) - F(a) \right) \int_{a}^{\frac{\sqrt{2}}{2} - a} zf(z)dz \]  
\[ + \left( F\left( \frac{\sqrt{2}}{2} - a \right) - F\left( \frac{\sqrt{2}}{2} - \frac{a}{2} \right) \right) \int_{\frac{\sqrt{2}}{2} - a}^{\frac{\sqrt{2}}{2} - \frac{a}{2}} zf(z)dz \]  
(E5)

and \[ E \left[ E \left[ z \mid a \leq \frac{\sqrt{2}}{4} \right] \right] = \int_{0}^{\frac{\sqrt{2}}{4}} \left( E \left[ z \mid a \leq \frac{\sqrt{2}}{4} \right] \ast \frac{4}{\sqrt{2}} \right) da = \frac{73}{96\sqrt{2}} \approx 0.5377 \]  
(E6)

while \[ E \left[ E \left[ z \mid a > \frac{\sqrt{2}}{4} \right] \right] = \frac{13}{32\sqrt{2}} \approx 0.2873 \]  
(E7)

\[ \therefore E[OUT] = 2E[z] = E \left[ z \mid a \leq \frac{\sqrt{2}}{4} \right] + E \left[ z \mid a > \frac{\sqrt{2}}{4} \right] \approx 0.825 \]  
(E8)
Appendix F

Derivation for */DIAM/CLO/1 – \( E[OUT] \) for a FA of unit area.

We consider the three PS configuration in a symmetric FA, i.e., PSs on one side of the FA, PSs on two opposing sides of the FA, and PSs on all sides of the FA.

**F.1.** PSs on one side of the FA:

Without loss of generality, we compute the expected distance to the closest point on the upper left-hand side of the diamond of unit area. (As before, the origin is placed at the leftmost point of the diamond as shown in Figure F1.) We obtain the expected distance by conditioning on the location of the pod inside the diamond. Three cases are identified based on triangular regions, labelled A1, A2, and A3, as shown in Figure F1. Given the triangular regions, we have:

\[
E[OUT] = \frac{1}{4} E[OUT](x, y) \in A_1] + \frac{1}{2} E[OUT](x, y) \in A_2] + \frac{1}{4} E[OUT](x, y) \in A_3]
\] (F1)

![Figure F-1. Diamond shape divided into regions for \( \bar{E}[OUT] \) calculation for PS on one side.](image)
\[ E[OUT|(x, y) \in A_1] = \int_0^{\frac{\sqrt{2}}{2}} \int_0^{x} \left( (x - y) \cdot \frac{1}{x} \right) dy \cdot 4x \, dx = \frac{1}{3\sqrt{2}} \] (F2)

\[ E[OUT|(x, y) \in A_2] = \int_{\frac{\sqrt{2}}{2}}^{\sqrt{2} - x} \int_0^{\frac{\sqrt{2}}{2}} \left( (x - y) \cdot \frac{1}{\sqrt{2} - x} \right) dy \cdot \left( \frac{8}{\sqrt{2}} - 4x \right) \, dx = \frac{1}{\sqrt{2}} \] (F3)

\[ E[OUT|(x, y) \in A_3] = \int_{\frac{\sqrt{2}}{2}}^{\sqrt{2} - x} \int_0^{\frac{\sqrt{2}}{2}} \left( (x + y) \cdot \frac{1}{\sqrt{2} - x} \right) dy \cdot \left( \frac{8}{\sqrt{2}} - 4x \right) \, dx = \frac{5}{3\sqrt{2}} \] (F4)

**F.2. PSs on two opposing sides of the FA:**

We obtain the expected distance by again conditioning on the location of the pod inside the diamond. We identify two cases based on the triangular regions, labelled A1 and A4, as shown in Figure F2. Given the two regions, we have:

\[ E[OUT] = \frac{1}{2} E[OUT|(x, y) \in A_1] + \frac{1}{2} E[OUT|(x, y) \in A_4] \] (F5)
\[ E[OUT|(x, y) \in A_4] = \int_{\frac{\sqrt{2}}{2}}^{\frac{3\sqrt{2}}{4}} \int_{\frac{\sqrt{2}}{2} - x}^{\sqrt{2} - x} \left( (x - y) \cdot \frac{1}{3\sqrt{2}/2 - 2x} \right) dy \cdot \left( 12\sqrt{2} - 16x \right) dx \]

\[ = \frac{\sqrt{2}}{3} \]

**F.3. PSs on all four sides of the FA:**

We obtain the expected distance by again conditioning on the location of the pod inside the diamond. We only need to consider one case given by the triangular region labelled A1. The result below follows from Equation (F2):

\[ E[OUT] = E[OUT|(x, y) \in A_1] = \frac{1}{3\sqrt{2}} \]  

(F7)
Appendix G
Derivation for */OCTA/RAN/1 – $E[\text{OUT}]$ for a FA of area of $\pi$.

We derive $F(z)$, the CDF of the travel distance from any point inside the octagon (with area of $\pi$ and parameter $c$) to a point in the perimeter. The parameter $c$ in $F(z)$ –shown in Figure G-1– is chosen to construct an octagon with area of $\pi$ by removing the four corners of a square with area of 4. The removed corners form isosceles triangles with two sides of length $c$. $F(z)$ is determined by conditioning on the location of the point along the perimeter. We choose, without loss of generality, the perimeter point from one of three line segments labelled ($a$, $b$, and $c$) in Figure G-1. The perimeter point can be determined by its $y$-coordinate (the origin is in the center point of the octagon). We have:

$$f(z|y \leq 1-c) = \begin{cases} \frac{2z}{\pi}, & 0 \leq z < 1 - c - y \\ \frac{3z - y - c + 1}{2\pi}, & 1 - c - y \leq z < 1 - c + y \\ \frac{z - c + 1}{\pi}, & 1 - c + y \leq z < 1 + c - y \\ \frac{z - y - c + 3}{2\pi}, & 1 + c - y \leq z < 1 + c + y \\ \frac{2}{\pi}, & 1 + c + y \leq z < 2 \\ \frac{-2z + 6}{\pi}, & 2 \leq z < 3 - c - y \\ \frac{-z + y + 3}{\pi}, & 3 - c - y \leq z < 3 - c + y \end{cases} \quad (G1)$$

$\therefore E[z|y \leq 1-c] = \frac{2(c - 5)(c^2 - 2)}{3\pi} \quad (G2)$
\[ f \left( z \left| 1 - c \leq y < \frac{1}{2} \right. \right) = \begin{cases} \frac{2z}{\pi}, & 0 \leq z < 2(y + c - 1) \\ \frac{3z + 2y + 2c - 2}{2\pi}, & 2(y + c - 1) \leq z < 2y \\ \frac{z + y}{\pi}, & 2y \leq z < 2(1 - y) \\ \frac{z + 2}{2\pi}, & 2(1 - y) \leq z < 2(2 - y - c) \\ \frac{-c + 2}{\pi}, & 2(2 - y - c) \leq z < 2 \\ \frac{z + c - 4}{\pi}, & 2 \leq z < 2 + 2(1 - c) \end{cases} \quad (G3) \]

\[ \therefore E \left[ z \left| 1 - c \leq y < \frac{1}{2} \right. \right] = \frac{50 - 27c - 15c^2 + 6c^3}{6\pi} \quad (G4) \]

\[ f \left( z \left| \frac{1}{2} \leq y < 1 - \frac{c}{2} \right. \right) = \begin{cases} \frac{2z}{\pi}, & 0 \leq z < 2(y + c - 1) \\ \frac{3z + 2y + 2c - 2}{2\pi}, & 2(y + c - 1) \leq z < 2(1 - y) \\ \frac{-c + 2}{\pi}, & 2(1 - y) \leq z < 2y \\ \frac{z + 2}{2\pi}, & 2y \leq z < 2(2 - y - c) \\ \frac{-c + 2}{\pi}, & 2(2 - y - c) \leq z < 2 \\ \frac{z + c - 4}{\pi}, & 2 \leq z < 2 + 2(1 - c) \end{cases} \quad (G5) \]

\[ \therefore E \left[ z \left| \frac{1}{2} \leq y < 1 - \frac{c}{2} \right. \right] = \frac{50 - 31c - 9c^2 + 4c^3}{6\pi} \quad (G6) \]

\[ P(y \leq 1 - c) = \frac{2 - 2c}{2 + (\sqrt{2} - 2)c} \quad (G7) \]

\[ P \left( 1 - c \leq y < \frac{1}{2} \right) = \frac{\sqrt{2} (2c - 1)}{2 + (\sqrt{2} - 2)c} \quad (G8) \]

\[ P \left( \frac{1}{2} \leq y < 1 - \frac{c}{2} \right) = -\frac{\sqrt{2} (c - 1)}{2 + (\sqrt{2} - 2)c} \quad (G9) \]
$$E[OUT] = E[z|y \leq 1 - c] * P(y \leq 1 - c) + E[z|1 - c \leq y < \frac{1}{2}] * P\left(1 - c \leq y < \frac{1}{2}\right) + E[z|\frac{1}{2} \leq y < 1 - \frac{c}{2}] * P\left(\frac{1}{2} \leq y < 1 - \frac{c}{2}\right)$$

$$\therefore E[OUT] = \frac{80 + (46\sqrt{2} - 96)c - (24 + 17\sqrt{2})c^2 + (48 - 23\sqrt{2})c^3 + 8(\sqrt{2} - 1)c^4}{6(2 + (\sqrt{2} - 2)c)\pi}$$

For a FA of area $\pi$, $c = \sqrt{2(1 - \pi/4)}$ and $E[OUT] \approx 1.446$. 

Figure G-1. Line segments $a, b, d$ on the perimeter of the octagon.
Appendix H
Tabulated results for asymmetric FA shapes.

Table H-1. $E[TB]$ results for asymmetric FA shapes.

<table>
<thead>
<tr>
<th>FA shape</th>
<th>RECT</th>
<th></th>
<th>ELDI</th>
<th></th>
<th>ELIP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>E[TB]</td>
<td>0.674</td>
<td>0.707</td>
<td>0.833</td>
<td>0.706</td>
<td>0.742</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td>0.670</td>
<td>0.689</td>
<td>0.810</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table H-2. $E[OUT]$ results for asymmetric FA shapes.

| PS config. | FA shape | b   | N = 1 | | N = 2 | | N = 3 |
|------------|----------|------|-------|-----------------|-----------------|-----------------|
|            |          |      | RN    | CP   | % Diff. | RN    | CP   | % Diff. | RN    | CP   | % Diff. |
| 1S         | RECT     | 0.75 | 0.866 | 0.577 | 33%     | 1.732 | 1.443 | 17%     | 2.598 | 2.165 | 17%     |
|            | ELDI     | 0.75 | 0.906 | 0.842 | 7%      | 1.811 | 1.706 | 6%      | 2.717 | 2.462 | 9%      |
|            | ELIP     | 0.75 | 0.860 | 0.524 | 39%     | 1.720 | 1.482 | 14%     | 2.579 | 2.086 | 19%     |
| 1L         | RECT     | 0.75 | 0.818 | 0.433 | 47%     | 1.636 | 1.251 | 24%     | 2.454 | 1.876 | 24%     |
|            | ELDI     | 0.75 | 0.818 | 0.614 | 25%     | 1.637 | 1.304 | 20%     | 2.455 | 1.956 | 20%     |
|            | ELIP     | 0.75 | 0.760 | 0.490 | 36%     | 1.519 | 1.245 | 18%     | 2.279 | 1.892 | 17%     |
| 2S         | RECT     | 0.75 | 0.866 | 0.289 | 67%     | 1.732 | 1.058 | 39%     | 2.598 | 1.704 | 34%     |
|            | ELDI     | 0.75 | 0.932 | 0.384 | 59%     | 1.864 | 1.131 | 39%     | 2.795 | 1.842 | 34%     |
|            | ELIP     | 0.75 | 0.851 | 0.301 | 65%     | 1.701 | 1.033 | 39%     | 2.552 | 1.734 | 32%     |
| 2L         | RECT     | 0.75 | 0.818 | 0.217 | 74%     | 1.636 | 0.962 | 41%     | 2.454 | 1.531 | 38%     |
|            | ELDI     | 0.75 | 0.827 | 0.239 | 71%     | 1.655 | 0.906 | 45%     | 2.482 | 1.484 | 40%     |
|            | ELIP     | 0.75 | 0.768 | 0.246 | 68%     | 1.536 | 0.940 | 39%     | 2.303 | 1.517 | 34%     |
| 4          | RECT     | 0.75 | 0.845 | 0.162 | 81%     | 1.691 | 0.805 | 52%     | 2.536 | 1.502 | 41%     |
|            | ELDI     | 0.75 | 0.823 | 0.233 | 72%     | 1.646 | 0.841 | 49%     | 2.469 | 1.533 | 38%     |
|            | ELIP     | 0.75 | 0.794 | 0.208 | 74%     | 1.587 | 0.927 | 42%     | 2.381 | 1.463 | 39%     |

Table H-3. $E[DC]$ results for asymmetric FA shapes.

<table>
<thead>
<tr>
<th>BN config.</th>
<th>FA shape</th>
<th>$b$</th>
<th>N = 1</th>
<th>N = 2</th>
<th>N = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RN</td>
<td>CP</td>
<td>% Diff.</td>
</tr>
<tr>
<td>0.75</td>
<td>RECT</td>
<td>0.75</td>
<td>2.410</td>
<td>1.830</td>
<td>24%</td>
</tr>
<tr>
<td>0.50</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.490</td>
<td>2.240</td>
<td>10%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.380</td>
<td>1.780</td>
<td>25%</td>
</tr>
<tr>
<td>0.50</td>
<td>RECT</td>
<td>0.75</td>
<td>2.310</td>
<td>1.540</td>
<td>33%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.280</td>
<td>1.920</td>
<td>16%</td>
</tr>
<tr>
<td>0.50</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.200</td>
<td>1.480</td>
<td>32%</td>
</tr>
<tr>
<td>0.25</td>
<td>RECT</td>
<td>0.75</td>
<td>2.410</td>
<td>1.250</td>
<td>48%</td>
</tr>
<tr>
<td>0.50</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.390</td>
<td>1.490</td>
<td>38%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.300</td>
<td>1.240</td>
<td>46%</td>
</tr>
<tr>
<td>0.50</td>
<td>RECT</td>
<td>0.75</td>
<td>2.100</td>
<td>1.100</td>
<td>52%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.230</td>
<td>1.270</td>
<td>43%</td>
</tr>
<tr>
<td>0.50</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.240</td>
<td>1.110</td>
<td>51%</td>
</tr>
<tr>
<td>0.25</td>
<td>RECT</td>
<td>0.75</td>
<td>2.360</td>
<td>0.998</td>
<td>58%</td>
</tr>
<tr>
<td>0.50</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.390</td>
<td>1.150</td>
<td>52%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.320</td>
<td>1.060</td>
<td>55%</td>
</tr>
<tr>
<td>0.50</td>
<td>RECT</td>
<td>0.75</td>
<td>2.510</td>
<td>1.000</td>
<td>60%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELDI</td>
<td>0.75</td>
<td>2.530</td>
<td>1.140</td>
<td>55%</td>
</tr>
<tr>
<td>0.25</td>
<td>ELIP</td>
<td>0.75</td>
<td>2.320</td>
<td>1.030</td>
<td>56%</td>
</tr>
</tbody>
</table>
Appendix I
Derivation for expected travel distance for two IPs in a unit square - interior case.

We derive the expected empty and full travel distance in a unit square. First, we consider expected empty travel distance, \( \bar{E}[e] \). We assume an AGV travels to the closest IP after each DP visit. (This assumption may cause a solution to violate the ESA constraint.) We determine the DP to IP allocation by drawing Voronoi partitions as in Figure I-1. The travel distance expression as a function of the IP location is easily obtained by conditioning on the position of the DP relative to the IP. Let \((x_j, y_j), j \in \{1,2\}\) be the coordinates of the IPs, when the origin is located in the bottom left corner of the unit square. Without loss of generality, suppose that \(x_1 \leq x_2\). We must consider 4 cases:

1. \(y_1 \geq y_2\) and \(x_2 - x_1 \leq y_1 - y_2\). That is, IP 2 is to the right and under IP 1 and their horizontal distance is less than their vertical distance.

2. \(y_1 \geq y_2\) and \(x_2 - x_1 > y_1 - y_2\).

3. \(y_1 < y_2\) and \(x_2 - x_1 \leq y_1 - y_2\).

4. \(y_1 < y_2\) and \(x_2 - x_1 > y_1 - y_2\).

We consider case 1. Cases 2 through 4, yield the same results by symmetry. Figure I-1 shows five DP regions relative to \((x_j, y_j)\). The regions are labeled \(r_i, i \in \{1,2,3,4,5\}\) with area \(A_i\). It follows that for IP 1:

\[
A_1 = x_1(1 - y_1)
\]  
\[
A_2 = (1 - x_1)(1 - y_1)
\]
\[ A_3 = x_1 \cdot \frac{y_1 + x_2 - x_1 - y_2}{2} \]  
\[ A_4 = (1 - x_1) \frac{x_1 + y_1 - x_2 - y_2}{2} \]  
\[ A_5 = \frac{(x_2 - x_1)^2}{2} \]  
\[ \tilde{E}[e \mid r_1] = \frac{x_1}{2} + \frac{(1 - y_1)}{2} \]  
\[ \tilde{E}[e \mid r_2] = \frac{1 - x_1}{2} + \frac{1 - y_1}{2} \]  
\[ \tilde{E}[e \mid r_3] = \frac{x_1}{2} + \frac{y_1 + x_2 - x_1 - y_2}{4} \]  
\[ \tilde{E}[e \mid r_4] = \frac{(1 - x_1)}{2} + \frac{x_1 + y_1 - x_2 - y_2}{4} \]  
\[ \tilde{E}[e \mid r_5] = \frac{y_1 + x_2 - x_1 - y_2}{2} + \frac{2(x_2 - x_1)}{3} \]  

Figure I-1. Expected empty travel distance DP regions relative to the position of the IPs.
We repeat the process for IP 2 and obtain total empty travel as follows:

\[
\tilde{E}[e] = \sum_{i=1}^{5} \tilde{E}[e \mid r_i] \cdot A_i
\]

\[
= (1/12) \left( 6 - 3(-2 + x_2^2 - 3y_2^2 + 4x_1 - 2y_2x_1 + x_1^2 + 4y_1 + 2y_2y_1 - 2x_1y_1 \\
- 3y_1^2 + 2x_2(y_2 - x_1 + y_1)) \\
+ 2(x_2^3 + 3x_2x_1^2 + 3x_2^2(y_2 - x_1 + y_1) - x_1^2(-6 + 3y_2 + x_1 + 3y_1)) \right)
\]

(I11)

Now, we consider expected full travel distance from IP \( j \), \( \tilde{E}[f \mid j] \). (For convenience, we drop the \( j \) condition until Equation I115). By continuous DP assumption in Chapter 4 Section 6, \( \tilde{E}[f] \) is the expected distance between the IP, which can be any point inside the square and all points inside the square. The travel distance expression as a function of the IP location is easily obtained by conditioning on position of the DP relative to the IP. Let \((x, y)\) be the coordinates of the IP, when the origin is located in the bottom left corner of the unit square. Figure I-1 shows four DP regions relative to \((x, y)\). The regions are labeled \( r_i, i \in \{1, 2, 3, 4\} \) with area \( A_i \), where \( \sum A_i = 1 \). It follows that:

\[
\tilde{E}[f \mid r_1] = \frac{x}{2} + \frac{1-y}{2}
\]

(I12)

\[
\tilde{E}[f \mid r_2] = \frac{1-x}{2} + \frac{1-y}{2}
\]

(I13)

\[
\tilde{E}[f \mid r_3] = \frac{1-x}{2} + \frac{y}{2}
\]

(I14)

\[
\tilde{E}[f \mid r_4] = \frac{x}{2} + \frac{y}{2}
\]

(I15)
\[ \mathcal{E}[f | j] = \sum_{i=1}^{4} \mathcal{E}[f | r_i] \cdot A_i = x_j^2 - x_j + y_j^2 - y_j + 1 \]  

\[ (I16) \]

Now we compute total full travel from the conditional expression in Equation I12-I16. Notice that the probability AGVs travel to IP 1 is proportional to the area of its service region \((S_1)\), i.e. 

\[ S_1 = \sum_i A_i, \text{ from Equations I1 through I5.} \]

It follows:

\[ \mathcal{E}[f] = \mathcal{E}[f | 1] \cdot S_1 + \mathcal{E}[f | 2] \cdot S_2 \]

\[ = \left(\frac{1}{2}\right) \left( -(x_2^2 - x_1^2)(-x_2 + x_2^2 - y_2 + y_2^2 + x_1 - x_1^2 + y_2 - y_2^2) \right) \]

\[ + (2 + x_2^3 + y_2^3 - 2 x_1 + x_1^2 + x_1^3 - 2 y_2 - x_2^2 y_1 + 3 y_1^2 + x_1 y_1^2) \]  

\[ - y_1^2 + y_2^2(-1 - x_1 + y_1) + x_2^2(-1 + y_2 - x_1 + y_1) \]

\[ + x_2(-2y_2 + y_2^2 + 2x_1 - x_1^2 - y_1^2) - y_2(-2x_1 + x_1^2 + y_1^2) \) \].

Finally, the expected empty and full travel distance, is given by:
\[ E[e + f] = \tilde{E}[e] + \tilde{E}[f] \]

\[
= \frac{1}{12} \left( 24 + 14x_2^3 - 6x_2^4 + 6y_2^3 - 24x_1 + 15x_1^2 + 10x_1^3 - 6x_1^4 \right)

- 24y_1 + 6x_1 y_1 - 6x_1^2 y_1 + 27y_1^2 + 6x_1 y_1^2 - 6x_1^2 y_1^2 - 6y_1^3

+ y_2^2 (3 - 6x_1 + 6x_1^2 + 6y_1) \quad \text{(I18)}

- 3x_2^2 (3 - 6y_2 + 2y_2^2 + 6x_1 - 4x_1^2 - 2y_1 - 2y_1^2)

- 6y_2 (-3x_1 + 3x_1^2 + y_1 + y_1^2) + 6y_2 (-3x_1 + 3x_1^2 + y_1 + y_1^2)

+ 6x_1 (-3y_2 + y_2^2 + 3x_1 - x_1^2 - y_1 (1 + y_1)) \right).
Appendix J
Derivation for expected travel distance for two IPs in a unit square - perimeter case.

We consider three cases: two IPs on one side, on adjacent sides, and on opposite sides of the unit square. First, we treat two IPs on one side. Without loss of generality, we assume that the IP are located in the bottom side of the unit square. Let $x_1$ and $x_2$ be the $x$-coordinates of IP 1 and IP 2 respectively when the origin is at the bottom left corner of the square. We proceed to compute $\tilde{E}[e]$ by conditioning on the position of DP relative to the IPs. Figure J-1, shows four DP regions labelled $r_i$, $i \in \{1, 2, 3, 4\}$ with area $A_i$, where $\sum A_i = 1$. It follows that:

\[
\tilde{E}[e | r_1] = \frac{x_1}{2} + \frac{1}{2} \quad (J1)
\]
\[
\tilde{E}[e | r_2] = \frac{x_2 - x_1}{4} + \frac{1}{2} \quad (J2)
\]
\[
\tilde{E}[e | r_3] = \frac{x_2 - x_1}{4} + \frac{1}{2} \quad (J3)
\]
\[
\tilde{E}[e | r_4] = \frac{1 - x_2}{2} + \frac{1}{2} \quad (J4)
\]

\[
\tilde{E}[e] = \sum_{i=1}^{4} \tilde{E}[e | r_i] \cdot A_i = \frac{1}{2} \left( 1 + x_1^2 + \frac{(x_2 - x_1)^2}{2} + (1 - x_2)^2 \right) \quad (J5)
\]
Figure J-1. Expected empty travel distance DP regions relative to the position of the IPs.

We compute the expected full travel distance for each IP $j$, $\bar{E}[e | j]$, by conditioning on the position of DPs relative to the IP. $\bar{E}[e | 1]$ is $\frac{x_1}{2} + \frac{1}{2}$ with probability $x_1$ for DPs to the left and $\frac{1-x_1}{2} + \frac{1}{2}$ with probability $(1 - x_1)$ for DPs to the right. Following the same reasoning for IP 2, we obtain:

$$\bar{E}[f | j] = x_j^2 - x_j + 1. \quad (J6)$$

Notice that the Voronoi line between the service region of IP 1 and 2 is the vertical line at $\frac{x_1+x_2}{2}$.

Hence the area of service region 1 and 2, are given by:

$$S_1 = \frac{x_1 + x_2}{2} \quad \text{and} \quad (J7)$$

$$S_2 = 1 - \frac{x_1 + x_2}{2}. \quad (J8)$$

Using Equations J6 through J8, we get:

$$\bar{E}[f] = \bar{E}[f | 1] \cdot S_1 + \bar{E}[f | 2] \cdot S_2 \quad (J9)$$

$$= (1/2)(2 + x_1^3 + x_1^2(x_2 - 1) - 2x_2 + 3x_2^2 - x_1x_2^2 - x_2^3).$$

And finally expected empty and full travel is given by:
\[ \tilde{E}[e + f] = \tilde{E}[e] + \tilde{E}[f] \]
\[ = (1/4)(8 + 2x_1^3 - 8x_2 + 9x_2^2 - 2x_2^3 - 2x_1x_2(1 + x_2)) \]
\[ + x_1^2(1 + 2x_2)). \]

Next we consider 2 IPs on adjacent sides. Without loss of generality, we assume that IP 1 is located in the bottom side and IP 2 on the right side of the unit square. Let \( x \) and \( y \) be the \( x \)-coordinates of IP 1 and \( y \)-coordinate of IP 2, respectively. (The origin is at the bottom left corner of the unit square). Furthermore, suppose that \( 1 - x \leq y \). (The same results are obtained when \( 1 - x > y \), by symmetry.) We proceed to compute \( \tilde{E}[e] \) by conditioning on the position of the DPs relative to the IPs. Figure J-2, shows 6 DP regions labelled \( r_i, i \in \{1,2,3,4,5,6\} \) with area \( A_i \), where \( \sum A_i = 1 \).

It follows that:

\[ A_1 = x \left( \frac{y}{2} + \frac{1-x}{2} \right) \] (J11)
\[ A_2 = \frac{(1-x)^2}{2} \] (J12)
\[ A_3 = (1-x) \left( \frac{y}{2} - \frac{1-x}{2} \right) \] (J13)
\[ A_4 = A_2 \] (J14)
\[ A_5 = \frac{y + x - 1}{2} \] (J15)
\[ A_6 = 1 - y \] (J16)

\[ \tilde{E}[e \mid r_1] = \frac{x}{2} + \frac{y - x + 1}{4} \] (J17)
\[ \tilde{E}[e \mid r_2] = \frac{y + x - 1}{2} + \frac{2(1-x)}{3} \] (J18)
\[ \tilde{E}[e \mid r_3] = \frac{1-x}{2} + \frac{y + x - 1}{4} \] (J19)
\[ \bar{E}[e \mid r_4] = \bar{E}[e \mid r_2] \]  
\[ \bar{E}[e \mid r_5] = \frac{y + x - 1}{4} + \frac{1}{2} \]  
\[ \bar{E}[e \mid r_6] = \frac{1 - y}{2} + \frac{1}{2} \]  
\[ \bar{E}[e] = \sum_{i=1}^{6} \bar{E}[e \mid r_i] \cdot A_i = \left( \frac{1}{12} \right) \left( 11 - 2x^3 - 12y - 6xy + 9y^2 + x^2(3 + 6y) \right) \]

Figure J-2. Expected empty travel distance DP regions relative to the position of the IPs.

We compute the area of service region 1 and 2 as follows:

\[ S_1 = \sum_{i=1}^{3} A_i \]  and \[ S_2 = \sum_{i=4}^{6} A_i. \]  

From Equations J6 and J24 we compute expected total full travel as follows:

\[ \bar{E}[f] = \bar{E}[f \mid 1] \cdot S_1 + \bar{E}[f \mid 2] \cdot S_2 \]
\[ = (1/2) \left( 2 + x^3 - x^4 - 2y + 3y^2 - xy^2 - y^3 + x^2(y^2 - 1) \right). \]

And finally expected full and empty travel from Equations J23 and J25 is given by:
\[ E[e + f] = E[e] + E[f] \]
\[ = (1/12) \left( 23 + 10x^3 - 6x^4 - 24y + 27y^2 - 6y^3 - 6xy(1 + y) \right) \tag{J26} \]
\[ + x^2(-3 + 6y + 6y^2) \].

Next we consider two IPs on opposite sides. Without loss of generality, we assume that IP 1 is located in the bottom side and IP 2 on the top side of the unit square. Let \( x_1 \) and \( x_2 \) be the \( x \)-coordinates of IP 1 and 2, respectively when the origin is at the bottom left corner of the unit square. We proceed to compute \( E[e] \) by conditioning on the position of DP relative to the IPs. Figure J-3, shows 6 DP regions labelled \( r_i, i \in \{1,2,3,4,5,6\} \) with area \( A_i \), where \( \sum A_i = 1 \). It follows that,

\[ A_1 = x_1 \frac{1 + x_2 - x_1}{2} \tag{J27} \]
\[ A_2 = \frac{(x_2 - x_1)^2}{2} \tag{J28} \]
\[ A_3 = (1 - x_1) \frac{1 - x_2 + x_1}{2} \tag{J29} \]
\[ A_4 = A_2 \tag{J30} \]
\[ A_5 = x_2 \frac{1 - x_2 + x_1}{2} \tag{J31} \]
\[ A_6 = (1 - x_2) \frac{1 + x_2 - x_1}{2} \tag{J32} \]
\[ E[e \mid r_1] = \frac{x_1}{2} + \frac{1 + x_2 - x_1}{4} \tag{J33} \]
\[ E[e \mid r_2] = \frac{1 - x_2 + x_1}{2} + \frac{2(x_2 - x_1)}{3} \tag{J34} \]
\[ E[e \mid r_3] = \frac{1 - x_1}{2} + \frac{1 - x_2 + x_1}{4} \tag{J35} \]
\[ E[e \mid r_4] = E[e \mid r_2] \tag{J36} \]
\[ \tilde{E}[e | r_5] = \frac{x_2}{2} + \frac{1 - x_2 + x_1}{4} \]  
\[ \tilde{E}[e | r_6] = \frac{1 - x_2}{2} + \frac{1 + x_2 - x_1}{4} \]

\[ \tilde{E}[e] = \sum_{i=1}^{6} \tilde{E}[e | r_i] \cdot A_i \]

\[ = \frac{1}{12} \left( 9 - 2x_1^3 - 6x_2 + 3x_2^2 + 2x_2^3 + x_1^2(3 + 6x_2) \right. \\
\left. - 6x_1(1 - x_2 + x_2^2) \right) \]

Figure J-3. Expected empty travel distance DP regions relative to the position of the IPs.

We compute the area of service region 1 and 2 as follows:

\[ S_1 = \sum_{i=1}^{3} A_i \text{ and } S_2 = \sum_{i=4}^{6} A_i. \]  

(J40)

From Equations J6 and J40 we compute expected total full travel as follows:

\[ \tilde{E}[f] = \tilde{E}[f | 1] \cdot S_1 + \tilde{E}[f | 2] \cdot S_2 \]

\[ = 1 + x_1^3 - \frac{x_1^4}{2} - x_2^2 + x_1^2(x_2 - 1)x_2 + x_2^3 - \frac{x_2^4}{2} + x_1 \left( x_2 - x_2^2 - \frac{1}{2} \right). \]  

(J41)

And finally expected full and empty travel from Equations J39 and J41 is given by:
\[ \tilde{E}[e + f] = \tilde{E}[e] + \tilde{E}[f] \]

\[ = (1/12)(21 + 10x_1^3 - 6x_1^4 - 12x_2 + 3x_2^3 + 14x_2^3 - 6x_2^4) \]

\[ - 6x_1(2 - 3x_2 + 3x_2^2) + 3x_1^2(1 - 2x_2 + 4x_2^2)) \]