Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse Through Effective Utilization of Vertical Lift Modules

by

Stephen Douglas Racca
Sc.M. Brown University, 2014
B.S.M.E. University of New Hampshire, 2013
ASMET New Hampshire Institute of Technology, 2010
Submitted to the Department of Mechanical Engineering in partial fulfillment of the requirements for the degree of Master of Engineering in Manufacturing at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2015

© Stephen Douglas Racca, MMXV. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature redacted

Stephen Douglas Racca
Department of Mechanical Engineering
August 10, 2015

Signature redacted

Stephen C. Graves
Abraham J. Siegel Professor of Management Science
Thesis Supervisor

Signature redacted

David E Hardt
Professor of Mechanical Engineering
Chairman, Committee for Graduate Students
Abstract

This thesis deals with improving the operational efficiency of automated part storage devices, in this case Vertical Lift Modules (VLM). This was accomplished by using dynamic slotting to maximize and maintain high material throughput, eliminating the need for periodic reslotting. Multiple VLMs can be used in parallel picking operations to improve material throughput. Common industry practice is to periodically reslot items once an unbalanced workload is obvious. This thesis investigates a method to avoid periodic reslotting by using incoming parts as a means to maintain a balanced workload amongst VLMs. Three different part allocation strategies are compared, namely Randomization, Snake and Order Grouping to determine their effectiveness and respective feasibility. The three strategies are then crafted into logical systems that could be used to strategically place received parts and eliminate the need for periodic reslotting. The Snake method was found to be the most well suited for this particular situation due to the small order sizes. This method provides a 35% savings in pick time, which is equivalent to approximately 733 hours annually.

Thesis Supervisor: Stephen C. Graves
Title: Abraham J. Siegel Professor of Management Science
Acknowledgments

I would like to Thank our advisor Stephen Graves, you helped us stay on track for this project and always asked the right questions to keep us moving in the right direction. We couldn’t ask for a friendlier and more knowledgable advisor.

Thank you to the MengM Faculty, Jose, David Hardt, Brian Anthony, for allowing me the privilege to study at MIT. Also, Thank you for managing this course and maintaining relationships with such a great local companies. This whole experience has been a dream come true.

Special Thank you to our supervisor Dan Martin, for allowing us the freedom to develop our own project and for your consistent engagement and support along the way. You have been a tremendous help and great mentor throughout the project.

Thank you to the management team, Tom Faulkner, Ron Dognazzi, Gaetano Peritore, Mark Goodhue, Robbie Roberts, Anthony Lafata, Maria Doyle, James Firth for being so supportive with the project. Additionally, I would like to thank all the warehouse staff, Kyle, Allen, Vinny and Geena and others for all your help and encouragement for this project. We couldn’t have done it without you.

Thanks to Sean Willems for your advice and sharing your experiences. Speaking with you helped us learn what really goes into solving these industry problems. Thanks for taking your personal time to come and tour the warehouse.

This project was an unforgettable learning experience and I want to thank my teammates, Grace and Paramveer for working through this project with me. It was truly a pleasure working with you both. Grace you are an amazing person and unbelievable hard worker it was great getting to know you this year. I wish you both all the best.

Most importantly, I wanted to thank my Mom and Dad for always supporting me and pushing me follow my passion, wherever it led me. You provided me with the best upbringing and I am eternally grateful to everything you’ve done for me. You have always been the most supportive and best parents on the planet. I wouldn’t be here without you. I love you.

Thanks Nana and Papa for always being on my side and supporting me through my seemingly endless college career. Also thanks for pointing me towards engineering in the first place I wouldn’t be where I am today without you. I love you.

Joe, your the most loyal and protective brother I could have, thanks for everything, I love you.
This page intentionally left blank
# Contents

1 Introduction 13

1.1 Semiconductor Industry Overview 14

1.2 Applied Materials and Varian Semiconductor Equipment Associates Background 16

1.3 Current Operations of the Warehouse 17

1.3.1 Storage Locations 18

1.3.2 Picking Process 21

1.4 Problem Statement 21

1.4.1 Motivation 21

1.4.2 Problem Context 23

1.4.3 Summary of Contributions 24

1.5 Thesis Organization 26

2 Preliminary Analysis 27

2.1 Problem Identification and Current VLM Performance 27

2.2 Causes of Poor VLM performance 31

2.3 Possible PACI Assignment Strategies 32

2.3.1 Randomization 32

2.3.2 Snake Frequency Balancing 32

2.3.3 Order Grouping 33

3 PACI Assignment Analysis 35

3.1 Experimental Method 36

3.1.1 Order Variance 38

3.1.2 Daily Variance 39

3.1.3 Common Industry Practice 41

3.2 Time Savings Estimate 41

3.3 Discussion of Results 43

4 VLM Sustainability 45

4.1 Dynamic Slotting Strategies 46

4.1.1 Randomization 47

4.1.2 Snake Frequency Balancing 48

4.1.3 Order Grouping 49

4.2 Cost-Benefit Analysis of Dynamic Slotting Strategies 50
## Conclusion and Recommendations

5.1 Benefits Discussion ............................................. 56
5.2 Final Recommendations ........................................... 57
5.3 Future Work ......................................................... 57

A Raw Variance Data ................................................. 61
List of Figures

1-1 Semiconductor value chain, Applied Materials fits into the chain in the bottom left under Equipment Vendors .............................. 15
1-2 Warehouse Floor plan shows various picking areas, Kitting rooms, and consolidation areas .............................. 18
1-3 1-3a gives a description of a single VLM Bay; 1-3b VLM terminology; VLMs, PODs and Bays .............................................. 19
1-4 Shows the ergonomic differences between GL and VLM picking and the optimal picking zone ......................................... 19
1-5 Pick-consolidate flow chart .............................................. 22
2-1 Picking rates and total picks from the VLM and Fixed picking areas. Pick rates correspond to columns and left axis and number of picks corresponds to the overlaid lines and the right axis .............................. 28
2-2 Total picks per POD from June 2014 - May 2015 units in thousands. Note: POD 1 has 2 times the picks of POD 3 ......................... 28
2-3 The original PACI distribution per POD. The quantity of uniquely assigned PACIs is shown in left axis and the percentage of the total cumulative pick frequency per POD is shown as the red line corresponding to the right axis .............................................. 29
2-4 Fishbone analysis used for problem identification .............................. 31
3-1 The lines per order of 9 months worth of order data. 35% (14,888) of orders have more than 2 lines ......................... 37
3-2 Relative $\bar{\sigma}^2_{POD}$ per order averaged per month. Includes only orders with two or more lines ........................................ 39
3-3 Relative $\bar{\sigma}^2_{Bay}$ per order averaged per month. Includes orders with two or more lines are in the same POD. ....................... 40
3-4 Relative $\bar{\sigma}^2_{POD}$ per day, averaged per month ................. 40
3-5 Relative $\bar{\sigma}^2_{Bay}$ per day, averaged per month ....................... 41
3-6 Cumulative Distribution of pick frequencies for different strategies including original .......................... 42
3-7 Total picks per POD for 9 month period of orders. Shows pick distributions if the arrangement strategies had been followed ............... 43
4-1 The logic required to implement the Randomization strategy into the software ......................................................... 47
4-2 The logic required to implement the Snake Frequency Balancing strategy into the software ........................................ 48
4-3 The logic required to implement the Order Grouping strategy in the software ......................................................... 49

A-1 Raw POD Variance per order averaged per month, where a specific order has 2 or more lines ........................................ 61
A-2 Raw Bay Variance per order averaged per month, where 2 or more lines are in the same POD ........................................ 62
A-3 Raw POD Variance per day averaged per month .................. 62
A-4 Raw Bay Variance per order averaged per month .................. 63
List of Tables

1.1 Bin sizes used in the VLMs for part storage. .............................................. 20

2.1 Time study results from busy period. Cutoffs exclude orders outside the stated time and percent excluded shows the percentage of orders that was excluded by the cutoff. The amount of time total pick time is assumed to be $\text{Max(PickTime)} + \text{Consolidation time}$ .................................. 30

2.2 Example showing how the Order Grouping method arranges parts in an order with 27 lines. The ideal POD and Bay assignments are shown for each part. ................................................................. 33

3.1 Example variance calculation used for a unique order with 27 lines organized by Order Grouping method. The number of parts picked from each POD and each Bay within that POD are shown. $\sigma^2_{POD}$ and $\sigma^2_{Bay}$ are calculated and averaged across either orders or days; shown in bold. ................................................................. 38

3.2 Shows pick time savings with relative percentage time savings from the Original pick distribution over a 9 month period. Time savings is estimated using an averaged pick rate of $30 \frac{\text{picks}}{\text{hr}}$. ........................................... 43

4.1 Implementation cost estimate from IT department using logic diagrams. The partial Implementation is the cost of implementing the system without modifying the PACI and using only a POD assignment. 52

4.2 Summary of benefits from each strategy. The % Pick Time Savings and % improvements are averaged over 9 months. ............................ 53
Chapter 1

Introduction

In the semiconductor manufacturing equipment industry, late deliveries to the customers are costly. In order for customer orders to be delivered on time, semiconductor equipment manufacturers need an efficient flow of material from the warehouse to the production floor. However, maintaining an efficient material flow is a challenge for companies in this industry because semiconductor capital equipment is highly complex and made up of more than 100,000 parts.

In this thesis we examine an operational efficiency problem at a semiconductor capital equipment manufacturing facility, Varian Semiconductor Equipment and Associates (VSEA). VSEA manufactures ion implantation equipment and also provides spare parts to its existing customers. For this company, as for many other warehouses that manage a high amount of stock keeping units (SKUs), maintaining an efficient material flow and delivering parts on time are key issues.

This section provides an overview of the semiconductor industry, Applied Materials, Inc., and where they fit in the semiconductor value chain. It goes on to provide a brief overview of the current warehouse operations and material flow through the warehouse, followed by describing the operational inefficiencies discovered at the warehouse. This work is part of a team project focused on improving the operational efficiency of the warehouse and improving the consistency of part delivery times. The
specific focus of this thesis is to determine the most effective part allocation strategy that can be used to improve the throughput of automated storage and retrieval systems known as Vertical Lift Modules (VLM).

VLMs typically double the picking rates compared to manual picking. To obtain higher throughputs a series of VLMs can be used in parallel picking operations. Upon initial loading, part velocity is used to distribute the parts evenly across the VLMs to balance the expected workload. This improvement gained from the initial balancing typically diminishes over time due to changing part demand, requiring periodic rebalancing. Dynamic slotting eliminates periodic reslotting by slotting every received part to maintain a balance as time progresses. Different part allocation strategies are compared to determine the most well suited strategy for this warehouse. Snake was discovered as the best strategy for the situation at VSEA and will provide an approximate pick time savings of 35%.

1.1 Semiconductor Industry Overview

In 2014, the worldwide semiconductor market revenue was $340.3 billion [1] and the global semiconductor capital equipment spending was $65.3 billion [2]. The semiconductor industry is a key enabler and driver for technological progress. Our smartphones, computers and electronic gadgets are running at greater processing speeds with larger storage capacity. Advancements in electronics with greater storage capacity, enhanced data transmission and processing capabilities have enabled the creation and development of the Internet of Things (IoT). The IoT market is driven by cloud computing and the growing interconnectedness of machines and electronic devices [3].

The semiconductor industry can be categorized into several sub industries: semiconductor materials and equipment suppliers, semiconductor foundries, integrated device manufacturers (IDMs), semiconductor electronic design and manufacturing services and original equipment manufacturers (OEMs) [4]. The semiconductor value
chain is summarized in Figure 1-1. OEMs produce electronic devices and hardware that are sold to end users. Examples of OEMs include Apple, Dell, Cisco and Seagate. Semiconductor electronic design and manufacturing services include electronics manufacturing service (EMS) and original design manufacturers (ODM), which, test, manufacture and distribute electronic components for OEMs. IDMs are semiconductor companies that design, manufacture and sell integrated circuit devices. The key IDM players are Intel, Samsung Electronics, Qualcomm, Micron Technology and SK Hynix [5]. Semiconductor foundries focus solely on mass producing chips and the top two players in this sector are Taiwan Semiconductor Manufacturing Company (TSMC) and United Microelectronics (UMC) [6]. Semiconductor manufacturing is a highly complex process and requires specialized equipment. The major semiconductor capital equipment manufacturers include Applied Materials, ASML, Tokyo Electron, KLA-Tencor and Lam Research [7]. Semiconductor capital equipment manufacturers sell equipment to the foundries and IDMs.
1.2 Applied Materials and Varian Semiconductor Equipment Associates Background

Applied Materials, Inc., founded in 1967 in Santa Clara, California, is the leading producer of semiconductor capital equipment, services and software for the global semiconductor, flat panel display, solar photovoltaic industries [6]. It offers a wide portfolio of products such as chemical vapor deposition, physical vapor deposition, etch and ion implantation.

Varian Semiconductor Equipment and Associates (VSEA) is a wholly owned subsidiary of Applied Materials based in Gloucester, Massachusetts that specializes in designing and manufacturing ion implantation\(^1\) equipment. VSEA was incorporated in 1999 and acquired by Applied Materials in 2011.

The four main product lines are medium current, high current, high energy and ultra-high dose. A detailed description of the different production lines and products can be found in previous research work [8, 9]. These products are sold in low volume (not more than 300 tools a year) but highly customizable. There are a high number of unique parts referred to as stock-keeping units (SKUs), over 20,000 SKUs. There are three warehouses known as Buildings 5, 70 and 80 that manage these parts. The majority of the parts are stored in Building 80 and will be the focus of this thesis. Building 5 is the main building with the production floor and it has two primary storage locations: the Supermarket (SMKT), which stores parts required for sub-assemblies and Module (MOD), which houses larger components and material kits required to build a module. The bulkiest parts are stored in Building 70. These parts include enclosures, large paneling and other protective equipment typically only installed once the machine is delivered. Improving the operational

---

\(^1\)Ion implantation is the fundamental process used to fabricate semiconductor devices. The process introduces dopants in the wafer surface altering the electrical conductivity properties. Ions from an ion source are electrostatically accelerated to a high energy (10 – 500keV), and then selectively filtered through electro magnets and finally implanted onto the wafer substrate. The ion implantation process is characterized by the dose and penetration of the dopant and these characteristics are dependent on the ion beam current.
efficiency in Building 80 will be the primary focus of this thesis and the complimentary theses [10, 11].

1.3 Current Operations of the Warehouse

This section will provide an overview of the warehouse operations including storage locations and current picking methods. It will also discuss many terminologies used throughout the remainder of the thesis. Material flows through the warehouse in a clearly defined sequence with each step being monitored by scanning barcodes. The information is managed using a software known as SAP Extended Warehouse Management (EWM). Parts are first received at the receiving area, sized accordingly and if new, assigned a default Put Away Control Indicator (PACI)\(^2\), and then put away into their respective storage locations. When an order is released to the warehouse, parts are picked from their respective locations. After all the picks have been completed they are consolidated\(^3\) in the consolidation area. The consolidated pallets or boxes of parts are moved to the loading dock for shipping.

Orders come in three different types which are sales, production and transfers. Production orders can either be classified as shop orders or z-picks. A shop order is a list of standard assemblies required for a machine whereas a z-pick refers to the customizable options a customer can choose. Parts required for the assembly of a machine are ordered 24 hours in advance of a laydown\(^4\). Transfers are parts issued from the warehouse to replenish the supermarket inventory\(^5\) in the main building. Sales orders are replacement parts for existing machines, and are shipped directly to the customer [12, 9].

\(^2\)A PACI tells EWM where to start looking for a open bin during the part put-away search, in the VLMs a sequential numerical order is followed (e.g. PACI of VL01 results in a search sequence of VL01-01 → VL01-02 → VL01-03 → VL02-01 → ...etc.)

\(^3\)Consolidation involves both physically grouping parts as well as electronic grouping of the Handling Units (HUs) of each part. In EWM the parts are grouped in a hierarchal structure using one as a master HU.

\(^4\)Laydown signifies the start of assembling a new machine

\(^5\)The supermarket inventory is an inventory of parts for sub-assemblies that are assembled into sub-modules of the ion implantation equipment.
1.3.1 Storage Locations

Material is stored in the warehouse in three distinct storage locations, namely High Rack, GL and VLMs as illustrated in Figure 1-2. The High Rack section contains the bulkiest parts which are picked with the aid of cherry pickers and forklifts. Parts in GL are less bulky and manually picked from shelves. The VLMs are automated storage and part retrieval systems that mechanically deliver trays of parts to be manually picked at an access area shown in Figure 1-3a. While the term VLM can be used to describe either a POD or a Bay for the purposes of the remainder of this thesis, and to stay in accordance with the terminology used at the company VLM refers to a POD. A single Bay is shown in Figure 1-3b. A VLM or POD is made up of 2 or 3 Bays of vertically arranged trays for storage, a tray delivery mechanism and a control system. Upon request from the software a tray is automatically brought to the access area (picking zone) and a location indicator guides the picker to the part to be picked. The operator picks or replenishes the stock and then the tray is returned to its original position after confirmation. Figure 1-4 highlights the ergonomic differences in picking from the VLM and GL areas; the golden zone signifies the optimal picking location where the most efficient picks are made.

The VLM picking area at VSEA is comprised of five PODs with one operator assigned.
Trays are stored on both sides of the VLM unit. Trays contain bins; trays are 96” x 24” inches in size. Parts are stored in seven different bin types. Access area (picking zone)

Figure 1-3: 1-3a gives a description of a single VLM Bay; 1-3b VLM terminology; VLMs, PODs and Bays

1 VLM = 1 POD
3 Bays per POD

Figure 1-4: Shows the ergonomic differences between GL and VLM picking and the optimal picking zone

to each. Four of the PODs (PODs 1, 2, 3 and 5) have three Bays and one (POD 4) has two Bays. PODs 4 and 5 are also shorter and thus have slightly less storage capacity. There are 40 – 50 trays per Bay and each tray (96” x 33”) can fit 144 of the smallest size bin (4” x 5.5”). The part locations on the trays are assigned a grid location relative to the smallest bin size; 24 locations along the width (A – X) and 6 bins locations along the depth (01 – 06). There are seven distinct bin sizes shown in Table 1.1 with a total of approximately 25,000 bins across all five PODs. The number of bins, types of bins and number of trays for each Bay can be uniquely configured. An example of a VLM storage identification number is shown using VL01-02-03-A01 in 1.1.
Table 1.1: Bin sizes used in the VLMs for part storage.

<table>
<thead>
<tr>
<th>Bin type</th>
<th>Width [in]</th>
<th>Length [in]</th>
<th>Height [in]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA04</td>
<td>4</td>
<td>5.5</td>
<td>4</td>
</tr>
<tr>
<td>BB04</td>
<td>4</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>BC04</td>
<td>8</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>BD04</td>
<td>8</td>
<td>16.5</td>
<td>4</td>
</tr>
<tr>
<td>BF04</td>
<td>8</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td>BG04</td>
<td>12</td>
<td>16.5</td>
<td>4</td>
</tr>
<tr>
<td>BG08</td>
<td>12</td>
<td>16.5</td>
<td>8</td>
</tr>
</tbody>
</table>

VLMs increase pick rates by bringing parts to pickers, by reducing travel time spent between picks and by improving work safety. Typically each VLM setup has multiple Bays (ideally 3) so as to reduce the waiting time for the parts to be delivered to the access area [13]. While the picker is picking from one of the Bays, the remaining Bays search and retrieve the next part in preparation to be picked. This time savings from always having parts ready to be picked translates into higher pick rates at the VLM. The pick rate quoted from the initial estimate was \( \frac{60 \text{ picks}}{\text{hr}} \), which is three times faster than the GL pick rate \( \frac{20 \text{ picks}}{\text{hr}} \). This is critical for delivering parts on time since the faster parts can be picked, the sooner orders can be consolidated and ready for shipping. In addition, unlike picking from the shelves of GL which may require pickers to climb ladders, picking from VLMs is much safer since the parts are brought to the picker at the access area. Last but not least, VLMs have a much smaller footprint for the same storage capacity as conventional shelving since parts are stored vertically in VLMs. This reduction in floor space can be up to 85% [14].
1.3.2 Picking Process

A picking process flow chart illustrating the picking operations at the warehouse is shown in Figure 1-5. The warehouse first receives orders either from the production floor or sales and a designated employee releases the parts in groups periodically known as waves\(^6\) to the warehouse pickers. The parts are then picked in parallel from the three storage areas. VLM consolidation involves grouping the parts that belong to the same consolidation group together. A consolidation group is a group of parts that has to be delivered as one package and typically corresponds to z-pick kit codes or shop orders. Similarly GL consolidation and High Rack consolidation involves putting together parts of the same consolidation group at their respective consolidation areas. Once all the parts of a consolidation group from the three different picking areas have been picked, the three are finally staged for trucking. There are two different methods that parts can get picked from in the VLMs: these are Pick and Pass and Pick and Consolidate (also known as parallel picking). The difference between the two methods being one is a serial process from POD to POD, and one is a parallel process of simultaneously picking from all PODs associated with an order. The comparison between these processes are outside of the scope of this thesis and are discussed in detail in a complimentary thesis [10].

1.4 Problem Statement

1.4.1 Motivation

Part delays can cause major backups in the production line and may eventually lead to missed or late machine shipments. Due to the high cost of entry the semiconductor Industry is primarily customer driven with few major customers. With such a small customer base, missed or late shipments are particularly concerning for VSEA as they stand apart by focusing on customer service. The customer service is so good

\(^6\)Waves can consist of up to eight orders
in-fact, they allow customers to change any configuration and even cancel an order at any point until the shipment date. This high level of service is very helpful for a customer but can be risky for a company. Since customers routinely make configuration changes, VSEA must reduce risk and WIP inventory by pushing the actual production to the last possible moment. In order to minimize production time and push it to the last possible moment, VSEA and other MIT teams have focused on implementing lean practices to reduce order lead times and improve assembly times [9, 15, 8, 16, 12]. While this past work has helped drastically reduce lead times, it helped make visible other potential improvements to further refine the production system. Consistent part delivery is critical to fully realizing all of the lean production methods that have been developed. Improvement to the part delivery times would allow production planners to develop more accurate and reliable laydown schedules and more carefully allocate labor resources.

Due to operational inefficiencies the warehouse typically struggles to keep up with the high part demand during busy periods. Presently VSEA uses a part delivery goal of 24 hr to monitor the warehouse performance. The time starts when an order is
sent and stops $\frac{1}{2} \ hr$ after the parts are on the truck. During the last busy period in December 2014, more than 40% of the orders during the last delivery time exceeded the 24 hr delivery goal. Late part deliveries lead to production schedulers ordering parts earlier in an attempt to compensate for the anticipated late delivery times. This increases WIP and helps further increase the delays due to there being more front loaded picks instead of the usual steady demand.

1.4.2 Problem Context

A key issue faced in the warehouse is that during busy periods up to 40% of orders are not delivered on time. The declining warehouse performance has been linked to inefficiencies in the picking and part storage methodologies in the VLM and GL picking areas. These problems increase picking time and delay arrival to the consolidation area. Effective part allocation strategies were used in the warehouse to strategically place parts to help improve the picking throughput and get the parts into the consolidation areas faster. There is presently a heavy reliance on the GL manual picking area which is identified as the bottleneck in the system. Along with the GL bottleneck, the VLMs exhibit poor and inconsistent picking performance with an uneven work distribution across both the PODs and Bays. The uneven distribution across the PODs is a common and hardly addressed problem with automated storage devices which, lose effectiveness over time with changing part demand.

This project involves multiple components to improve the delivery times including relocating parts between the VLM and GL picking areas, reallocating parts in the VLM and investigating 2 different widely known picking strategies, pick and pass, and parallel picking, to determine the most effective method for this warehouse. A new part removal process was generated to allow slow moving parts to be efficiently removed from the VLMs to make room for fast moving GL parts. Three different part allocation strategies are analyzed to determine the most beneficial strategy to implement. The three strategies are used to develop the logic required for implement-

\textsuperscript{7}Gathered from raw data on SAP and distilled picking times per order.
tation into the warehouse software; the costs and benefits are compared to ensure the most beneficial yet cost effective strategy is implemented.

This document goes in detail comparing the different part allocation strategies in the VLMs and comparing them against actual picking data. Cost benefit analysis was performed on the different strategies to validate the most effective option for the warehouse.

1.4.3 Summary of Contributions

The 2015 MengM MIT-VSEA Project is captured in a set of theses namely

**Thesis 1:** Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Strategic Allocation of Parts [11]

**Thesis 2:** Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Effective Utilization of Vertical Lift Modules

**Thesis 3:** Improving and Maintaining the Efficiency of a Semiconductor Equipment Manufacturing Warehouse [10]

Toor's [11] thesis goes in greater depth explaining the problem identification process. It focuses on moving and identifying 2,052 FISH parts (parts that are not picked in past 1 year and are not projected to be picked in next 6 months) to GL area in the most efficient manner possible. It also goes further to identify fast moving parts in the GL area that could be fit into VLM and allocating these fast movers to specific VLMs to balance the workload distribution of VLMs. It also identifies the fast moving GL parts that can not fit into the VLMs and uses appropriate slotting techniques in the GL area to improve picking efficiency. Finally, the thesis discusses how the implementation process was carried out.

This thesis concentrates on balancing the workload distribution of all 5 VLMs to maximize and maintain the throughput of the VLMs. The thesis goes into greater depth to evaluate three different part allocation strategies: Random, Snake, and Order.
Grouping strategies. The random strategy as the name implies randomly distributes parts amongst VLMs and Bays, the snake strategy distributes parts in VLMs based on its pick frequency, while the kit code strategy focuses on arranging parts that get ordered together evenly among all 5 PODs. The thesis also discusses, which strategy would be the most effective strategy to implement into a dynamically self sustaining system to maintain the system performance.

Fong’s [10] thesis discusses the picking strategies and the sustainability and implementation of the workload balancing strategies. Pick and pass, and the pick and consolidate strategy. Pick and pass strategy in the instance where there is a larger consolidation area is ineffective due to the insignificant consolidation time savings compared to the actual consolidation times. This strategy involves a serial picking process that progresses through the PODs in an ascending order (e.g. POD 1, POD 2, POD 3, etc.). Pick and consolidate strategy picks all parts from VLMs in a parallel process and finally all the individual totes are consolidated. The thesis talks about pros and cons of both these strategies and identifies which strategy would be best at what point of time. The remaining portion of this thesis talks about the long term VLM sustainability including workload balancing and sustaining what was implemented through this process.
1.5 Thesis Organization

Chapter 1: Introduction

Provides a general overview of the semiconductor industry and where Applied Materials fit into the value chain. Furthermore, it goes into detail describing the problem statement and describes how this project was divided amongst the participants and their respective theses.

Chapter 2: Preliminary Analysis

Describes the preliminary analysis conducted to further validate the project and gain an understanding of the current performance of the system. Also discusses possible solutions correct the VLM workload unbalance using PACI assignment strategies.

Chapter 3: Static PACI Assignment Analysis

Compares the different PACI assignment strategies using variance as a metric and a time saving estimate. The strategies are compared without including the cost or real complexities of implementation and looks from a performance standpoint.

Chapter 4: Maintaining Balanced Workload

The different assignment strategies are used to generate the logic required for a software upgrade capable of dynamically updating the PACI to automatically balance and sustain the workload across the VLMs. The different implementation feasibilities and costs are discussed between the strategies.

Chapter 5: Conclusion and Recommendations

The final conclusions and recommendations are summarized for the different allocation methods. The randomization method is recommended based on the performance improvement and ease of implementation into the current system.
Chapter 2

Preliminary Analysis

2.1 Problem Identification and Current VLM Performance

This project began by first mapping out the manufacturing process in its entirety and then focusing on the most prevalent issues. Throughout the process a common problem kept arising, that the part delivery time is inconsistent and unreliable during busy periods. Inconsistent delivery times can cause a variety of downstream issues. In an effort to tackle this problem at the source the warehouse became the focus.

Through more worker interviews and recovering historical data from the warehouse, some issues became clear. A steadily declining picking rate shown in Figure 2-1 was observed. This picking rate figure also shows that more picks where coming from the fixed picking areas than VLM. The highest pick rate displayed was 35 picks per hour, which is a mere 58% of the $60 \text{ picks/hr}$ rate that was in the original estimate.

The declining pick rate was cause for a major concern so the pick distribution across the PODs was calculated and is shown in Figure 2-2. During visits and interviews POD 1 seemed to work a noticeably higher amount then the other PODs and furthermore in a single POD a noticeable amount of parts are located in a single bay.
Figure 2-1: Picking rates and total picks from the VLM and Fixed picking areas. Pick rates correspond to columns and left axis and number of picks corresponds to the overlaid lines and the right axis causing the worker to wait for the trays.

Figure 2-2: Total picks per POD from June 2014 - May 2015 units in thousands. Note: POD 1 has 2 times the picks of POD 3

This interesting distribution led to more research and interviews with warehouse management and the IT department to map out the logical flow of the software for both putting parts away and picking. The cause for the skewed distribution has to do with the put away logic. In EWM parts are assigned a Put Away Control Indicator
(PACI) which tells the software where to start looking for an open bin. The search goes in a sequential order starting from the PACI assigned POD, and moving in a sequential numerical order looking for the next open bin that can house the part in the PODs and Bays (e.g. PACI=VL02 Search sequence is VL02-01 → VL02-02 → VL02-03 → VL03-01 → ...etc.). The issue with the current arrangement is that the PACI assignment is a defaulted process, meaning for each part the PACI is set to POD 1 by default. The initial balancing of the system was done by a local consulting group upon installation of the VLMs however, with part demand changes many of the parts in the VLMs have very little or no demand, there are referred to as FISH parts (First In Still Here). Furthermore, the PACI is only set the first time a part is introduced into the system. The original PACI distribution is shown along with the percentages of cumulative pick frequencies in Figure 2-3. The pick frequency percentage calculation involved summing the pick frequency for the parts with each respective PACIs and calculating the relative percentage. The noticeably skewed distribution shows more then 40% of the cumulative pick frequency are assigned to POD 1 because all of the high frequency and new parts are assigned a PACI of VL01 causing the skewed distribution in picks shown in figure 2-2.

![Figure 2-3](image.png)

**Figure 2-3:** The original PACI distribution per POD. The quantity of uniquely assigned PACIs is shown in left axis and the percentage of the total cumulative pick frequency per POD is shown as the red line corresponding to the right axis.
Further investigation was performed to determine if improving the VLM efficiency would be beneficial to the company. Running at the current efficiency, the VLMs still were thought to have the fastest pick rate. To investigate this point, a time study was performed using individual order data obtained during the busy periods to see the average time taken to complete orders through the various picking areas. The results of this study are summarized in Table 2.1. The part delivery times in this analysis showed that there was some part shortages that caused some major delays for over 2 weeks. To avoid the outliers skewing the results, orders taking over 72 hours were neglected, the corresponding percentage of neglected parts is shown. This study used consolidation groups for the orders and found the difference in times of the parts in the order segregated by different picking and consolidation areas. Upon inspection it is clear the GL area is the bottleneck and the VLM is the second runner up.

**Table 2.1:** Time study results from busy period. Cutoffs exclude orders outside the stated time and percent excluded shows the percentage of orders that was excluded by the cutoff. The amount of time total pick time is assumed to be $Max(PickTime) + Consolidation\ time$

<table>
<thead>
<tr>
<th></th>
<th>Total Average [hr]</th>
<th>3 Day Cutoff [hr]</th>
<th>1 Day Cutoff [hr]</th>
<th>3 Day % Excluded</th>
<th>1 Day % Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLM</td>
<td>41.9</td>
<td>11.8</td>
<td>7</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>GL</td>
<td>49.5</td>
<td>19.1</td>
<td>9.4</td>
<td>16</td>
<td>40</td>
</tr>
<tr>
<td>RK</td>
<td>27.9</td>
<td>12.7</td>
<td>8.2</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>CO80</td>
<td>24.3</td>
<td>3.1</td>
<td>1.1</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Total Pick Time</td>
<td>73.8</td>
<td>22.2</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The focus of this thesis will be the VLM inefficiencies and balancing the workload. Therefore the simultaneous corrections and improvements made in the GL area will be discussed from a higher level. The cumulative part distributions were plotted for both the GL and VLM areas and it was found that 90% of the total picks come from 25% of the unique SKUs and of those, approximately half are located in the GL area. This large percentage indicated a high reliance on the GL picking area for parts that could potentially be located in the VLMs. To aid in the part transfer process the team manually inspected over 1,400 unique SKUs to gauge the necessary bins sizes...
required in the VLMs. The reallocation of parts and slotting strategies used for this portion of the project can be found in more detail in [11]. A summary of the problem identification process was compiled in a Fishbone diagram shown in Figure 2-4.

### 2.2 Causes of Poor VLM performance

The causes of the VLM inefficiency was discovered by using a root cause analysis shown in Figure 2-4. The PACI is an overlooked parameter that can be used to improve and maintain balanced workload and maximize the throughput among PODs. Currently, there is no set strategy used to assign PACIs aside from a default VL01. A common and widely known problem with automated storage picking systems is a declining throughput at part demand changes occur and more parts are received into
Other causes of poor picking performance are the heavy reliance on the GL picking area than the VLMs; and a confusion between the most effective picking method, improvements are discussed in [11, 10].

2.3 Possible PACI Assignment Strategies

Possible solutions were identified to help rebalance the VLMs. These allocation strategies will be used to distribute the frequently picked parts among the PODs. The hypotheses are described below. When developing these possible solutions an effort was made to investigate the different levels of complexities, ranging from the most complex to the simplest feasible solution.

2.3.1 Randomization

Randomization as the name implied uses a uniform random PACI assignment strategy. This method would help to spread out the parts uniformly amongst the PODs. This method has a large benefit as it is very simple to implement without needing any additional information about the parts or orders.

2.3.2 Snake Frequency Balancing

The Snake method focuses on distribution based on the known part frequency data. This method involves arranging a list of the parts by descending pick frequency and numbering them with a numerical sequence (e.g. 1, 2, 3, 4, 5, 4, 3, 2, 1, 1, 2, ... etc.). This method has the advantage of balancing the pick frequency across the VLMs exceptionally well; however, there is no guarantee that on a per order basis the order will be balanced.
2.3.3 Order Grouping

This strategy required developing an algorithm capable of balancing the parts across the PODs for each order. This program assigns parts based on the historical order data and takes the groups of parts that have been picked together and allocates them evenly across PODs. The algorithm takes into account parts which have previously been assigned a PACI and locates the other parts around that to maintain a balanced per order workload across PODs and Bays. Small orders, having two or less lines, are assigned random placement among PODs to account for the majority of small orders and help to uniformly balance the workload. For larger orders this method has the advantage of micromanaging on a per order level to ensure that every order will be evenly distributed amongst the PODs and Bays. One additional feature of this strategy is that it reads and assigns parts in order so orders that are more frequent can be given a higher priority. This strategy is the most inherently accurate at balancing the picks; however, a downside is that it is computationally demanding and may be cumbersome to integrate into the existing system.

Table 2.2: Example showing how the Order Grouping method arranges parts in an order with 27 lines. The ideal POD and Bay assignments are shown for each part.

<table>
<thead>
<tr>
<th></th>
<th>POD 1</th>
<th>POD 2</th>
<th>POD 3</th>
<th>POD 4</th>
<th>POD 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bay 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Bay 2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bay 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>2</td>
</tr>
</tbody>
</table>
Chapter 3

PACI Assignment Analysis

This chapter performs an analysis using historical picking data to determine the potential improvement that could have been achieved if parts had been relocated using the three strategies. Variance of the number of picks was used to quantify the amount of imbalance each PACI assignment strategy would provide on a daily and per order basis. The variance for the different strategies are normalized relative to the original to focus on trends. The raw variance data used to calculate the relative graphs are shown in appendix A. Two methods are used to compare the different part allocation strategies including, variance per order (orders > 2 lines), variance across days (all orders). The variance metrics are averaged per month. Furthermore, comparison between the strategies using typical industry practice is done by using cumulative pick frequency. All of the methods are used to compare the strategies and determine the most well suited for this particular situation. The analysis goes into estimating the potential time savings VSEA could expect once the strategy is fully implemented. The PACI, as described in section 2.2, is used to identify where the put-away search sequence begins for new incoming parts. The preliminary analysis discovered the original PACI assignment has no fixed strategy behind the placement of incoming parts. All the parts are assigned to VL01 by default, which leads to parts not being evenly distributed across the PODs; this results in an unbalanced workload across the PODs. The possible PACI assignment strategies investigated in
this section are introduced in section 2.3 and are Randomization, Snake, and Order Grouping.

3.1 Experimental Method

This analysis was performed by using some key assumptions to ensure a fair comparison was made between strategies. A key assumption is that every part can be relocated in the VLMs, and put into its ideal location based on the respective PACI assignment strategy. Exact space constraints for parts in PODs were neglected for the sake of this analysis as a strategic comparison was the main focus. In reality when, fewer parts will be able to fit into POD 4 and 5 due to the smaller physical size of these PODs. The exact tray and grid location for the parts was also neglected in this analysis because it is assumed balancing the workload between the PODs and Bays will provide the most benefit. Arranging parts per tray is more beneficial in single bay VLM systems where reducing the machine travel time is the objective. In this case the focus is to balance the picks across the PODs and Bays so a tray is ready and the worker never waits for the machine.

The experiment began by first compiling a master list of 12588 parts satisfying our criteria of what should be located in the VLM. We then used each strategy to assign all of the parts in the master list. This gave us the ideal POD and Bay locations for each part for the respective strategies. The assignments were used to represent the current VLM inventory. We then obtained actual picking data for the VLM and GL areas from the SAP records for the past 9 months. This data base consisted of 42,881 orders, with a total of 136,294 lines$^1$. The lines per order distribution is shown in Figure 3-1. The order distribution data base consists of 65% single line orders with 14,888 orders having 2 or more lines.

By using the order data base and the part location assignments for each strategy, we

---

$^1$Lines is common terminology used in the industry referring to the physical number of unique parts or SKUs
were then able to calculate the number and location of the picks for each strategy. The relative variance was used to compare the strategies against the original distribution, where a lower value represents a more balanced workload. Due to the specific order distribution at VSEA, and the desire to include as much data as possible, all orders having two or more parts are included. The variance for small orders is taken as a function of number of parts in an order. This is because a 2 line order should have a variance of zero as long as the parts are in separate Bays. For an example of the per order variance calculation, a balanced order with 2 lines has a variance \( \text{Var}(1,1) = 0 \), similarly a 3 line order may have a variance \( \text{Var}(1,1,1) = 0 \), an unbalanced 3 line order could have a variance \( \text{Var}(2,1,0) = 0.66 \), etc. The per Day variance was calculated by passing all the picks in a day and calculating the variance of the number of picks across the PODs and Bays within each POD. The highest throughput is obtained by having a balanced workload across the PODs, and across Bays within each POD. This variance per order, or per Bay was averaged per month for the different allocation strategies. The results were compared by showing a monthly average variance per order. The variance, shown in Eq 3.1, was used to compare the different strategies for both POD, Bay, and daily Variances.

**Figure 3-1:** The lines per order of 9 months worth of order data. 35% (14,888) of orders have more than 2 lines...
\[
\sigma^2 = \frac{1}{N} \sum_{i=1}^{n} (X_i - \mu)^2, \tag{3.1}
\]

where, \(N\) is the number of lines in an order, \(n\) is the number of orders, \(X_i\) is the number of picks from a POD, Bay, or Day, and \(\mu\) is the average of the quantities assigned to the PODs or Bays.

### 3.1.1 Order Variance

An example of the POD and Bay variance calculation per order is shown in Table 3.1. The POD Variance \((\sigma^2_{POD})\), shown in bold, calculates the variance across the number of unique lines picked from each POD (e.g. \(\sigma^2_{POD}(6,5,5,5,6) = 0.24\)). The Bay Variance calculates the variance across the number of lines picked for each POD (e.g. \(\sigma^2_{Bay}(2,2,2,3) = 0\), or \(\sigma^2_{Bay}(2,1,2) = 0.22\)). Bay Variance is important, a high Bay Variance indicates a skewed pick distribution across a POD. In this case the picker has idle time and must wait for a tray between picks, reducing the pick rate. Orders with less than 2 lines are neglected in this analysis. Bay Variances are averaged across all PODs and orders. Likewise, POD Variances are averaged across all orders.

**Table 3.1:** Example variance calculation used for a unique order with 27 lines organized by Order Grouping method. The number of parts picked from each POD and each Bay within that POD are shown. \(\sigma^2_{POD}\) and \(\sigma^2_{Bay}\) are calculated and averaged across either orders or days; shown in bold.

<table>
<thead>
<tr>
<th>Bay</th>
<th>POD 1</th>
<th>POD 2</th>
<th>POD 3</th>
<th>POD 4</th>
<th>POD 5</th>
<th>(\sigma^2_{POD})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bay 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td><strong>0.24</strong></td>
</tr>
<tr>
<td>Bay 2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bay 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_{Bay})</td>
<td>0</td>
<td><strong>0.22</strong></td>
<td><strong>0.22</strong></td>
<td><strong>0.25</strong></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The distribution of POD and Bay Variances per month are shown in Figure 3-2 and 3-3 respectively. The data shows a clear reduction in variance by using any of the distribution methods. The Order Grouping method shows a clear advantage boasting a 55% reduction in POD Variance however, the Random and Snake methods still
achieve 33% reduction. When looking at the variance across the Bays, the reduction for all the strategies will provide an expected 70-75% reduction in Bay Variance over the original distribution. These reductions in variance will greatly help reduce the uneven workload and help to improve the utilization and picking throughput of the VLMs.

![Graph showing variance across different strategies](image)

**Figure 3-2**: Relative $\sigma_{POD}^2$ per order averaged per month. Includes only orders with two or more lines.

### 3.1.2 Daily Variance

The per order variance comparison is valuable but it neglects 65% of the total orders. To take all the orders into account, the Daily Variance was calculated and averaged per month. The daily picks from each POD and Bay are counted and the variances are calculated using the same method shown in Table 3.1. The average variance helps illustrate the workload balance on a per month basis, across PODs and Bays as shown in Figures 3-4 and 3-5, respectively. This comparison accounts for all orders and shows a more balanced workload using the random and snake methods on a daily basis. Daily Relative Bay Variance shows an 85% improvement in workload balance by using the randomization or snake strategy. The improvement in workload balancing
Figure 3-3: Relative $\sigma^2_{Bay}$ per order averaged per month. Includes orders with two or more lines are in the same POD.

from the Order Grouping strategy is, 35%. The Daily Relative POD Variance results show a similar improvement for all strategies of 77-87% with randomization at the top.

Figure 3-4: Relative $\sigma^2_{POD}$ per day, averaged per month.
3.1.3 Common Industry Practice

Common practice in the industry is to distribute the pick frequency across the PODs to balance the workload. This balancing method is similar to the Snake strategy, by focusing on pick frequency to locate the parts. The PACI assignment strategies are compared by a percentage of cumulative pick frequency in Figure 3-6. As noted in section 2.1, the original distribution shows a spike in pick frequencies for POD 1. This analysis method shows the advantage behind the Snake and Random strategy due to the uniform distribution of the pick frequencies. The lower percentage in POD 4 comes from only having 2 bays. The Order Grouping strategy shows a disadvantage with handling small orders, it is much better suited for distributing larger orders. For this scenario with the majority of the orders being single lines an apparent trend toward POD 5 exists for the Order Grouping strategy. The Random or Snake strategies would be the distribution method of choice if this were the only metric used.

3.2 Time Savings Estimate

The potential time savings was estimated to provide more insight into the benefits of the different strategies. A crucial assumption, is the total number of picks in a
specific POD is directly proportional to the pick time. When parallel picking, the total pick time takes as long as the POD with the most picks (e.g., the slowest POD). Furthermore, for the purposes of this analysis all picks are assumed to require the same picking effort. This analysis compares the slowest POD resulting from each strategy the number of picks and total estimated pick times are calculated. The pick distribution per POD for each strategy is shown in Figure 3-7.

The estimated time savings is shown in Table 3.2. The reduction in picks between strategies was converted directly into a time savings by using an average pick rate of $30 \frac{picks}{hr}$. The relative percentages are shown and illustrate the estimated monthly time savings. By using the different methods a time savings of approximately 35% is possible by the randomization method. The time savings estimate compares the actual picking to the potential picking if the respective strategies had been in place. The time savings estimate shows that the randomization strategy will provide the most uniform pick distribution of picks amongst the PODs. The Order Grouping method shows a skewed distribution, because the method focuses on balancing on a per order level, rather than batches of orders. To account for the large majority of single line orders Order Grouping method randomly places order with two or fewer
Figure 3-7: Total picks per POD for 9 month period of orders. Shows pick distributions if the arrangement strategies had been followed.

Table 3.2: Shows pick time savings with relative percentage time savings from the Original pick distribution over a 9 month period. Time savings is estimated using an averaged pick rate of $30\text{picks/hr}$.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Pick Time [hr]</th>
<th>% Reduction Pick Time</th>
<th>Monthly Time Savings [hr]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>1552</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Randomization</td>
<td>1000</td>
<td>35.5</td>
<td>61.3</td>
</tr>
<tr>
<td>Snake</td>
<td>1004</td>
<td>35.3</td>
<td>60.8</td>
</tr>
<tr>
<td>Order Grouping</td>
<td>1280</td>
<td>17.5</td>
<td>30.2</td>
</tr>
</tbody>
</table>

3.3 Discussion of Results

The strategies have been compared and all strategies show their advantages in respective scenarios. Order Grouping allows for the most balanced workload and minimum variance per order across the Bays and PODs. Randomization and Snake provide a more uniform pick distribution to reduce daily workload unbalance and perform comparably when using cumulative pick distribution. When the pick frequency dis-
tribution is compared between the strategies the Random and Snake strategies show they are more balanced than the Order Grouping and Original distributions.

It became clear during this analysis that a strategy like Order Grouping provides worse overall performance than a much simpler strategy like randomization can bring. The orders at VSEA are typically small with 65% having one line. Order Grouping would be better suited in a facilities with larger order sizes. The Randomization strategy provides a 35.5% time savings improvement over the original method while Order Grouping obtains 17.5%.

To conclude, either the randomization or snake strategy will be the most effective at balancing the overall workload amongst the PODs in this particular warehouse setting. This will help to minimize the variance between PODs and improve the throughput of the VLMs. When comparing the different strategies it seems Order Grouping has the advantage interns of ensuring a balanced workload however with very small orders randomization has the overall advantage. Snake performs just as well as pure randomization in all the trials. The commonly practiced method in the industry uses pick frequency and in this case, Randomization and Snake seem to be the better performing methods.
Chapter 4

VLM Sustainability

Maintaining a balanced VLM workload requires developing a system that can self balance as new parts are received. This chapter discusses how the different PACIs assignment strategies could be implemented in a dynamic way to enable a self-maintaining system. Periodic reslotting by manually moving parts to balance the workload between the VLMs is very labor intensive and costs valuable time. Since the initial workload is so skewed, the first stage of this project was to transfer parts between the GL and VLM areas to help obtain a better balanced starting point [11]. To further improve upon this starting point, a dynamic PACI assignment strategy is required. The intelligence used in this strategy is derived from the PACI strategies that were investigated in chapter 3. Each of the three strategies were used to design the logical requirements necessary to implement the strategies into the existing system. The costs and complexities of each possible strategy are compared to determine the most suitable solution for this situation.
4.1 Dynamic Slotting Strategies

Slotting is the activity of determining the most appropriate storage location for each part in the warehouse, by extension, in the VLMs. In our work, the objective of slotting is to maximize the VLMs’ picking efficiency by balancing the workload across the five PODs. Re-slotting is performed when the picking efficiency of the VLMs is declining or when the workloads across the five PODs become skewed. In the industry, the common practice for re-slotting involves periodically reviewing the workload distribution across the VLMs – in our work, we call this static slotting. If the distribution progressively becomes skewed, parts have to be manually transferred from one VLM to another in order to balance the workload. Static slotting is not only labor intensive, but the approach is more of a reactive one rather than a preventive one i.e. the solution is implemented only after the problem has occurred. In this section, we propose to implement a preventive solution – dynamic slotting. Unlike static slotting, dynamic slotting focuses on determining the most appropriate storage location for each part so as to minimize both the imbalance in workload across the five PODs and the imbalance in workload across the Bays within a POD. By employing a dynamic slotting solution, we postulate that re-slotting the VLMs (manually moving parts to re-balance the workload) would be kept to the minimum.

At VSEA, the warehouse can receive hundreds of parts on any given day, most of which need a PACI assignment. This section looks at implementing dynamic slotting into the EWM software at VSEA. The implementation strategy involves changing the software logic in the mainframe. To determine which strategy works best in this situation, logical flow charts were created to determine the programming requirements and cost estimates. At present, the PACI of a part is based purely on POD assignment. However, having only a POD assignment is inadequate; a Bay assignment is also important for ensuring that the workload across the Bays within a POD is balanced since this relates to the pick rate of the VLMs. In this section, referring to the flow diagrams in figures 4-1, 4-2 and 4-3, a method is considered to be fully implemented if the PACI includes both POD and Bay assignments. The method is considered
partially implemented if only POD assignments are allowed. We also propose an alternative dynamic slotting method that is not based on PACI assignment, but rather the cumulative expected pick frequencies of the PODs.

4.1.1 Randomization

The randomization method shown in Figure 4-1 can be implemented into a dynamic system easily. The system needs a random number generating function which can generate a number between 1 and 5. Once this number is generated the PACI is assigned and put-away search sequence can commence. The PACIs in this situation will not be permanent and will randomly select a new PACI every time a part is received.

![Flowchart](image)

**Figure 4-1:** The logic required to implement the Randomization strategy into the software
4.1.2 Snake Frequency Balancing

The exact snake frequency balancing method which, uses pick frequency and assigns parts to PODs in a sequential order as discussed in section 2.3.2 would be excessively complicated to implement directly. In order to capture the essence of the strategy in an implementable manner the logic diagram shown in Figure 4-2 was created. This method is in the middle ground in terms of complexity and implementability. The focus of this is workload balancing and it involves calculating the current workload distribution of the PODs by calculating the total picks for the last 6 months. The sequence of PODs will then be arranged in ascending order from most to least utilized and this order will be used as the put-away search sequence with the lowest utilized POD being the PACI assignment. If the Bay assignment is possible, then each Bay within the chosen POD will be ranked in-terms of total picks in the last 6 months and the Bay with the least amount of picks in the last 6 months will be the assigned Bay.

Figure 4-2: The logic required to implement the Snake Frequency Balancing strategy into the software
4.1.3 Order Grouping

The Order Grouping method, discussed in Section, is by far the most computationally intensive strategy to implement. The logic diagram is shown in Figure 4-3. This would require knowledge of the most frequently orders, parts associated with those orders and the current location of every other part in that order. The current part locations per POD will be counted and the POD and Bay housing the fewest number of parts for that particular order will be assigned the new part. For example if an order has 10 parts with only 9 currently in the VLMs distributed amongst the PODs e.g. [2, 2, 1, 2, 2] then when the 10th part arrives it should be assigned to POD 3 to make the distribution e.g. [2, 2, 2, 2, 2]. This will ensure that every order will be spread evenly across the PODs. The analysis will be repeated for every part received.

**Figure 4-3:** The logic required to implement the Order Grouping strategy in the software
4.2 Cost-Benefit Analysis of Dynamic Slotting Strategies

A cost-benefit analysis of the dynamic slotting methods is performed in this section based on the following criteria:

**Programming costs** The logic flow charts were submitted to the IT department to get a cost estimation for implementing the PACI-based dynamic slotting methods. There are two quotations provided: i) for full implementation and ii) for partial implementation. Currently, PACIs are for a specific POD however, improving the resolution to a specific bay would be helpful to balancing the workload with each POD. The full implementation costs include the costs of implementing both the POD and Bay assignments. The partial implementation costs include the cost of using the current system with only the POD assignments.

**Complexity** The complexity of a new dynamic slotting method is measured as Easy, Medium, or Hard, based on the number of steps required for implementation.

**% Pick time savings** The percentage pick time savings is measured over a 9-month period and is calculated based on the difference in number of picks between the slowest POD in the original slotting method and the slowest POD in the dynamic slotting method. It is worth noting that the slowest POD has the highest number of picks. For example, the highest number of picks is 46,500 in POD 1 for the Original method and the highest number of picks for randomization is 30,000 in POD 2. The % time savings between Randomization and original is the relative difference, 35.5%.

**% Improvement $\sigma^2_{POD}$ per order** The percentage improvement in the POD Variance per order quantifies how balanced the workload is across the five pods for each dynamic slotting method is per order. They are taken relative to the original workload distribution and averaged over a 9 month period.

**% Improvement $\sigma^2_{Bay}$ per order** The percentage improvement in the Bay Variance per order quantifies how balanced the workload is across the 3 bays within a POD for
each dynamic slotting method is per order. They are taken relative to the original workload distribution and averaged over a 9 month period.

% Improvement $\sigma^2_{POD}$ per day The percentage improvement in the POD Variance per day quantifies how balanced the workload is on a daily basis across the five pods for a particular dynamic slotting method, relative to the original workload distribution and averaged over a 9 month period.

% Improvement $\sigma^2_{Bay}$ per day The percentage improvement in the Bay variance per day quantifies how balanced the workload is on a daily basis across the 3 Bays within a POD for each dynamic slotting method, relative to the original workload distribution and averaged over a 9 month period.

The last five metrics were used to compare the three slotting methods (Randomization, Snake Frequency Balancing and Order Grouping) in 3. The analysis specifies PACI assignments to each part and the picking performance is quantified over a 9 month period. In this chapter, we extend the results obtained in the static case to be implemented in a dynamic way and holistically compare the different strategies.

The programming costs scale with the number of days required to develop, configure and test the new software. The daily programming cost can range between $800 – $1000, in this estimate the upper limit was used. The number of days required to develop the new dynamic slotting software depends on two factors – i) whether a full implementation or partial implementation is desired and ii) the complexity of the dynamic slotting method. From Table 4.1, the cost difference between a full and partial implementation is $3,000. It is estimated that three more days are needed to develop the software application for a full implementation rather than partial implementation.

We highly recommend a full implementation because Bay assignments are critical in maintaining a balanced workload across the Bays within a POD, which is important for balancing the workload for each picker. If a partial implementation is put in place, the consequence is that picks will not be evenly spread across the Bays in a POD and
Table 4.1: Implementation cost estimate from IT department using logic diagrams. The partial Implementation is the cost of implementing the system without modifying the PACI and using only a POD assignment.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Full Implementation [$]</th>
<th>Partial Implementation [$]</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization</td>
<td>28,000</td>
<td>25,000</td>
<td>Easy</td>
</tr>
<tr>
<td>Snake</td>
<td>33,000</td>
<td>30,000</td>
<td>Medium</td>
</tr>
<tr>
<td>Order Grouping</td>
<td>38,000</td>
<td>35,000</td>
<td>Hard</td>
</tr>
</tbody>
</table>

the machines could become the bottleneck. If all the parts are located in a single Bay the worker will need to wait for the tray to arrive between picks. Therefore, the costs of the dynamic slotting methods are compared based on the full implementation costs as shown in table 4.2.

Depending on the complexity of the dynamic slotting methods, the time necessary to develop a new dynamic slotting software varies between 10 and 20 days. Typically, configuration of the new software takes 5 days long. An additional 10 days is added due to the testing required prior to actual program implementation. From table 4.2, the order grouping method is the most complex and the randomization method is the least complex. In general, each of the three strategies require a new basic software capable of assigning a strategic PACI to incoming parts. However, both the snake frequency balancing and order grouping methods are slightly more complex because they require additional programs to support the fundamental POD and Bay assignments. The Snake Frequency balancing method requires an additional program to be written that calculates the utilizations of the pods and bays. The Order Grouping method requires a separate program that classify which parts are frequently ordered together by the production floor and customers.

The complexity differences among the strategies equate to an approximate price difference of $5,000 between each of the different methods. Consequently, the randomization strategy is the cheapest to implement and the order grouping method is the most expensive.

Comparing the improvements in the workload distributions among the three methods,
Table 4.2: Summary of benefits from each strategy. The % Pick Time Savings and % improvements are averaged over 9 months.

<table>
<thead>
<tr>
<th>Dynamic slotting method</th>
<th>Randomization</th>
<th>Snake</th>
<th>Order Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost [$]</td>
<td>28,000</td>
<td>33,000</td>
<td>38,000</td>
</tr>
<tr>
<td>% Pick Time Savings</td>
<td>35.5</td>
<td>35.3</td>
<td>17.5</td>
</tr>
<tr>
<td>% Improvement $\sigma_{POD}^2$ per Order</td>
<td>70.1</td>
<td>69.3</td>
<td>75.3</td>
</tr>
<tr>
<td>% Improvement $\sigma_{Bay}^2$ per Order</td>
<td>33.4</td>
<td>33.1</td>
<td>54.8</td>
</tr>
<tr>
<td>% Improvement $\sigma_{POD}^2$ per Day</td>
<td>87.0</td>
<td>85.0</td>
<td>76.7</td>
</tr>
<tr>
<td>% Improvement $\sigma_{Bay}^2$ Per Day</td>
<td>87.1</td>
<td>87.4</td>
<td>35.6</td>
</tr>
</tbody>
</table>

we find that the randomization method gives the highest % Pick Time Savings of 35.5%. In most instances, the randomization method performs better than the snake frequency balancing method, except for the % improvement in $\sigma_{Bay}^2$ per day, the latter method performs marginally better (by 0.3%) than the former method. Although the Order Grouping method shows the greatest improvements in the variances among PODs and Bays when measured on a per order basis, it has the weakest performance on a daily basis. This is caused by the majority of the orders at VSEA have a single line, the effectiveness of the Order Grouping method in maintaining a balance workload distribution on a daily basis is limited. Thus, the Order Grouping method has the lowest % pick time savings over the 9-month period of 17.5%. The Order Grouping method may be more effective if majority of the orders are large-sized i.e. having several lines (preferably greater than five lines).

From the cost-benefit analysis summarized in table 4.2, the randomization strategy provides the most benefit at the lowest cost. The least well suited to this company is the Order Grouping method because most of the orders are small. In addition, developing the software for the order grouping method is highly complex. The final software would also require an unnecessary amount of computing power to intelligently assign each received part.

The snake frequency balancing strategy is also a possible option. Although the performance of the snake strategy is not as good as the randomization method, the
differences in performances between both methods are marginal. We consider the overall performance of the snake strategy to be relatively good. However, the snake strategy requires an additional $5,000 of upfront investment to create the program to measure the utilization of the PODs and Bays. The key question here is – is the additional investment worth it?

While one may immediately see this as an additional benefit for monitoring how balanced is the workload across the PODs and Bays, it is worth noting that this can be done easily by retrieving the total number of picks per POD and per Bay from EWM. We expect that it takes no longer than a couple of minutes to obtain the data. Hence, the additional $5,000 is not justified if it is seen as a method to monitor the workload distribution of the PODs and Bays.

However, the additional $5,000 can be justified if it will result in a faster improvement in the workload balance across the PODs and Bays. Since the workload distribution is currently skewed towards POD 1, even if a dynamic slotting strategy is implemented, some time is needed for the workload to be re-balanced. We expect that the snake frequency balancing method will attain this balanced workload faster than randomization because it puts away a part based on the part’s expected pick frequency. A fast moving part will be put away into the Bay that has the lowest expected utilization and this will raise the workload of the Bay. Therefore, we expect that the workload will become more balanced within a shorter period of time if the snake frequency balancing strategy is used.

We recommend the company adopt the snake frequency balancing dynamic slotting method primarily because the workload imbalance issue can be resolved within a shorter period of time. Moreover, the predicted improvement for this strategy is comparable to the randomization approach and the additional benefits would help ensure a faster re-balancing.

---

1The expected utilization of a POD or Bay is measured based on the cumulative expected number of picks i.e. the sum of the expected number of picks of all the parts in the POD.
Chapter 5

Conclusion and Recommendations

Our work is focused on improving the operational efficiency of the warehouse so as to consistently deliver parts on time to the production floor. If parts are not delivered on time, machine lay-downs on the production floor are delayed and the likelihood of a missed shipment increases. Because late shipments are costly to the company, it is critical for the warehouse to meet the target part delivery time of 24 hours. However, during the busy periods when the warehouse is faced with heavy workloads, the warehouse struggles to keep up with the target part delivery times.

The delivery times are improved by the simultaneous reduction in reliance on GL and improvement in picking efficiency at the VLMs. From our preliminary analysis, we find that picking an order from GL takes the longest time (i.e. GL is the bottleneck). Therefore, our solution is to reduce the workload on GL by increasing the number of VLM picks. This can be achieved by moving fast moving parts from the GL into the VLMs. The primary motivation for increasing picking from VLMs is that it is more efficient than picking from GL shelves so the throughput is higher. The predicted pick rate of the VLMs is 60 picks per hour as compared to that of GL is only 20 picks per hour. Based on the work done by [11], 735 fast moving GL parts that had
expected pick frequencies greater than 11 and were able to fit into the VLM bins were moved into the VLMs. An additional 265 small sized GL parts with expected pick frequencies greater than five would also be moved into the VLMs. The workload is expected to increase by 13% as fast moving parts are moved from the GL to the VLMs [11].

Since VSEA bought the VLMs, the picking efficiency of the VLMs has not been meeting expectations. There are two ways to improve the picking efficiency: i) effective utilization of VLMs by balancing the workload across the VLMs and ii) using an efficient picking method.

The main issue addressed is the workload imbalance problem. The distribution of picks across the five pods are skewed towards the first pod. In addition, despite the increase in workload on the VLMs, we also estimate a 23% time savings to pick an order once the workload across the pods is balanced. The picking methods were compared and the most effective is the pick and consolidate method. Pick and consolidate results in a 20 minute (8%) per order time savings.

5.1 Benefits Discussion

A 35% time savings can be obtained by using a Randomization or Snake strategy to balance the VLM workload. Due to shifting the reliance from the GL to the VLMs this 35% will decrease to 23%. The additional time savings will provide additional time for workers to perform cycle counts and put-aways. The industry expects a steady ramp up in demand over the forecasted future so the improvements in the warehouse will help to ensure on time part deliveries. The industry standard practice for slotting automated part storage systems will be improved by implementing the dynamic slotting framework in the system. This dynamic slotting framework coupled with the FISH removal process will eliminate the need for periodic reslotting the VLMs and allow for easy removal of parts when pick demand changes. The FISH removal process is can also be applied to another process known as Excess and Obsolete
E&O, where parts greater than 3 years old are removed from inventory.

5.2 Final Recommendations

The final recommendations in this thesis should be taken in conjunction with Fong and Toor [11, 10].

- The fast moving parts in the GL area need to be moved into the VLMs and the larger parts should be slotted to put them into the golden picking zone.
- The Snake Dynamic slotting strategy should be implemented into the system to ensure the most rapid improvement in performance.
- The pick and consolidate strategy should be always used during the busy periods.
- Keep the FISH Removal method and periodically identify and pull out out FISH parts.
- Periodically monitor the FISH growth rate to ensure its kept in control.

5.3 Future Work

During the analysis a noticeable amount of part shortages aided in causing late deliveries. The ordering methodologies should be investigated to see the possibilities for improvements. Follow up on tracking VLM performance improvements should be monitored over time to help determine the effectiveness of the dynamic slotting strategy. Many of the parts that could have fit into the VLMs were in far too large of quantities to be moved in so lot sizing methods for fast moving GL parts that couldn’t fit should be reviewed. Consider shifting the delivery time goal to be on the production floor the next morning.
Bibliography


Appendix A

Raw Variance Data

Figure A-1: Raw POD Variance per order averaged per month, where a specific order has 2 or more lines.
**Figure A-2:** Raw Bay Variance per order averaged per month, where 2 or more lines are in the same POD.

**Figure A-3:** Raw POD Variance per day averaged per month.
Figure A-4: Raw Bay Variance per order averaged per month.