Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Strategic Allocation of Parts

by

Paramveer Singh Toor

Bachelor of Science in Aerospace Engineering
University of California, Los Angeles, 2014

Bachelor of Arts in Business Economics
University of California, Los Angeles, 2014

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

Signature redacted

Author...

Paramveer Singh Toor
Department of Mechanical Engineering
August 10, 2015

Signature redacted

Certified by...

Stephen C. Graves
Abraham Siegel Professor of Management Sciences
Thesis Supervisor

Signature redacted

Accepted by.................................

David E. Hardt
Ralph E. and Eloise F. Cross Professor of Mechanical Engineering
Chairman, Committee for Graduate Student

This page left blank intentionally.

Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Strategic Allocation of Parts

by

Paramveer Singh Toor

Submitted to the Department of Mechanical Engineering on August 10, 2015

in partial fulfillment of the requirements for the degree of

Master of Engineering in Manufacturing

Abstract

The work addresses the operational inefficiency problem in a semiconductor equipment manufacturing warehouse of Applied Material's Varian Semiconductor Business Unit. At Varian, the target part delivery time from the warehouse to the production floor is 24 hours. However, during busy periods, parts are not delivered on time. Late part delivery from the warehouse to the production floor could delay the machine laydown date, which in turn could result in late or missed shipment of tools to the customers, which can be very costly. To improve the efficiency and the reliability of the warehouse, picking efficiency is to be improved. Parts from the warehouse are picked from three picking locations- Vertical Lift Modules (VLMs), GL, and RK. VLMs are automated machines, while GL and RK are manual picking zones. Picking an order from GL takes the most amount of time. The overall picking efficiency at the warehouse can be improved by partially shifting the workload from GL to the VLMs, and by further improving the picking efficiency at the VLMs. The workload from GL to the VLMs is shifted by transferring fast moving parts from GL to the VLMs. The picking efficiency of the VLMs is improved by balancing the workload of all five VLM pods, and by employing a more efficient 'pick-and-consolidate' picking strategy. The workload at GL is decreased by 25% and the workload at VLMs is increased by 13%. Despite the increase in workload at VLMs, 23% time savings could be achieved by balancing the utilization of all five VLM pods. Additional time savings of 20 minutes per order (8%) could be achieved by using 'pick-and-consolidate' picking strategy over 'pick-and-pass' picking strategy.

Thesis Supervisor: Stephen C. Graves
Abraham Siegel Professor of Management Sciences

This page left blank intentionally.

Acknowledgements

I would like to express my deep gratitude to Professor Stephen C. Graves, my thesis supervisor, and Mr. Daniel Martin, my supervisor at Applied Materials, for their patient guidance, enthusiastic encouragement, and useful critiques of this work. Thank you Daniel Martin and Thomas Faulkner for giving me this opportunity to work under your guidance and for having faith in me.

I would like to thank the faculty and the leadership of the Master of Engineering program at MIT for offering me an opportunity to further improve my skills and knowledge.

Likewise, I must thank Dean Vijay Dhir, Professor Ivan Catton, Professor Mohamed Abdou, and Damian Toohey for believing in my abilities and for furnishing recommendations for my admission to MIT.

I would also like to thank my teammates, Grace and Steve, for their hard work throughout the project. Working with both of you was a great learning experience.

I must thank Ronald Dognazzi, Maria Doyle, Anthony Lafata, James Firth, Mark Goodhue, Gaetano Peritore, Robbie Roberts, and everyone else at Varian for supporting this project.

Above all, I thank Waheguru for giving me the strength, the ability, and the vision to carry out this project, and also for giving me parents, relatives, and friends, who have all been very supportive and encouraging throughout my life.

Once again "THANKS" to all of you.

This page left blank intentionally.

Table of Contents

This page left blank intentionally.

List of Figures

This page left blank intentionally.

List of Tables

This page left blank intentionally.

1. Introduction

Varian is a semiconductor business unit of Applied Materials, Inc. that designs, develops, manufactures, and services ion implantation equipment, which are used in the fabrication of semiconductor chips. Ion implantation equipment introduces dopants in the semiconductor wafers to modify the electrical properties of these wafers. Due to fierce competition, semiconductor capital equipment companies provide high service levels to customer; the tools have to be delivered on time to customers to prevent delays in the wafer fabrication schedule of semiconductor fabs. Semiconductor capital equipment is highly complex and made of many parts, for instance an ion implantation tool is made up of more than 100,000 parts. An efficient material flow from the warehouse to the semiconductor equipment manufacturing plant thus becomes critical in ensuring that parts are delivered to the production floor on time so that the production schedules can be met.

This section provides an overview of the semiconductor industry, background of Applied Materials, Inc. and Varian Semiconductor Equipment Associates (VSEA), and discusses where Applied Materials, Inc. and VSEA fit in the semiconductor value chain. It goes on to provide a very brief overview of warehouse operations and material flow through the warehouse, followed by describing the operational inefficiency problem in the warehouse. This work is part of a team project on improving the warehouse picking efficiency to reduce part delivery times from the warehouse to the production floor. The specific focus of this thesis is to discuss in detail the problem identification process, root cause analysis and corresponding solutions, and reducing the imbalance in the picking process by decreasing the workload at the GL shelves.

1.1. Semiconductor Industry Overview

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 memories and additional functionalities. In particular, advancements in sensing modules in terms of greater data storage capacity, enhanced data transmission, and processing capabilities have enabled the creation and development of Internet of Things (IoT) devices, which consumers and businesses use to monitor, track, and process data. In fact the IoT market is driven by cloud computing and the growing interconnectedness of machines and electronic devices [1].

The semiconductor value chain is summarized in Figure 1. The industry can be categorized into several semiconductor sectors: semiconductor materials and equipment suppliers, semiconductor foundries, integrated device manufacturers (IDMs), semiconductor electronic design and manufacturing services, and original equipment manufacturers (OEMs) [2]. 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 manufacturer (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 [3]. 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) [4]. 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 [5]. Essentially, semiconductor capital equipment manufacturers' customers are semiconductor foundries and IDMs.

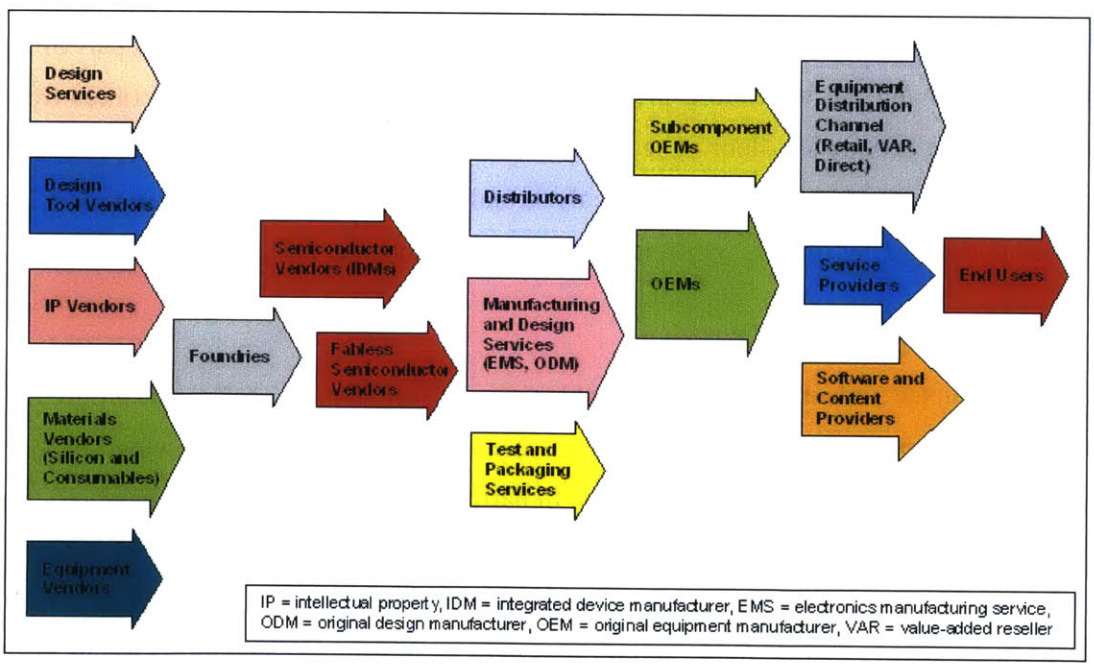


Figure 1: Semiconductor value chain; Applied Materials fits into the chain in the bottom left under equipment vendors

In 2014, the worldwide semiconductor market revenue was $340.3 billion [6] and the global semiconductor capital equipment spending was $65.3 billion [7].

1.2. Applied Materials & VSEA Background

Applied Materials, Inc., founded in 1967 in Santa Clara, California, is the leading producer of semiconductor equipment, services, and software for the global semiconductor, flat panel display, solar photovoltaic, and related industries [4]. It offers a wide portfolio of products for chemical vapor deposition, physical vapor deposition, etch, and ion implantation.

Varian Semiconductor Equipment and Associates (VSEA) is a wholly owned subsidiary of Applied Materials, Inc. based in Gloucester, Massachusetts and specializes in designing and manufacturing ion implantation equipment. Ion implantation is the fundamental process to fabricate semiconductor devices by which dopants are introduced in the wafer. Ions from an ion source are electrostatically accelerated to high energy (10−500 keV) and implanted onto the wafer target. The ion implantation process is characterized by the dose and penetration of the dopant and these characteristics are dependent on the ion beam current. VSEA was incorporated in 1999 and acquired by Applied Materials, Inc. in 2011.

The four main product lines at VSEA are medium current, high current, high energy, and ultra-high dose. A detailed description of the different product lines can be found in previous research work [8, 9, 12]. These products are sold in low volume (not more than 300 tools a year) but are highly customizable. Thus, there are a high number of stock-keeping units (over 20,000 SKUs). There are three storage areas: Building 80, 70, and 5. Majority of the parts are stored in the warehouse (Building 80), bulky parts are stored in Building 70, and the main building (Building 5) has two primary storage locations- supermarket (SMKT), which stores parts required for subassemblies, and Module (MOD), which stores larger components and material kits required to build a module.

1.3. Current Operations of the Warehouse

Material flows through the warehouse in three main stages: receiving, picking, and consolidation for shipping. Parts are first received at the receiving area and put away into their respective storage locations. When an order is released to the warehouse, parts are picked for the order and then consolidated. There are three main types of orders: sales, production, and transfers. Sales orders are spare parts shipped to the customers [8, 10]. A production order can either be classified as a shop order or a z-pick. A shop order is a list of assemblies to be built, and details the parts required for each assembly. Parts required from the warehouse are pulled 24 hours in advance of laydown using z-pick kit codes. Production orders are typically delivered to the assembler on the production floor in the main building. Transfers are parts issued from the warehouse to replenish the supermarket inventory in the main building. The supermarket inventory is an inventory of parts for sub-assemblies that are assembled into sub-modules of the ion implantation equipment.

The focus in this project is to improve the flow of material from picking to consolidation and the two key aspects that affect the material flow are: where the parts are picked from i.e. part storage locations, and how they are picked i.e. picking methods.

1.3.1. Storage Locations

Parts in the warehouse are picked from three distinct storage locations: high racks, GL, and Vertical Lift Modules (VLMs) as illustrated in Figure 2. High racks contain bulky parts, which are picked with the aid of cherry pickers and forklifts. Parts in GL are less bulky and manually picked from shelves. VLMs are automated storage and order picking systems that mechanically eject trays for parts to be picked (see Figure 3a). A VLM is made up of vertically arranged trays for storage, a delivery lift platform mechanism, and a computerized control system. A tray is automatically brought to the access area (picking zone) upon request from the software, and a location indicator guides the picker about the part to be picked. The operator picks or replenishes the stock, and then the tray is returned to its position after confirmation. Figure 3b highlights the differences in picking from VLMs and GL.

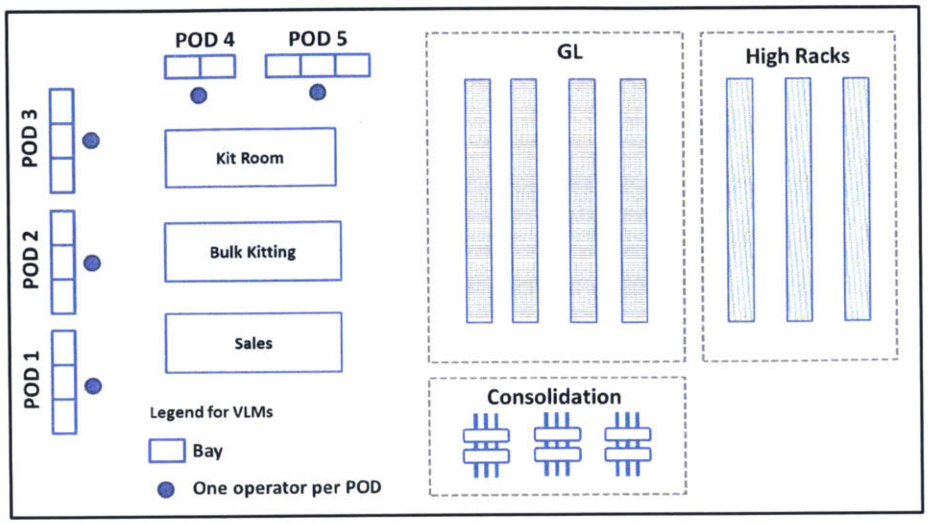
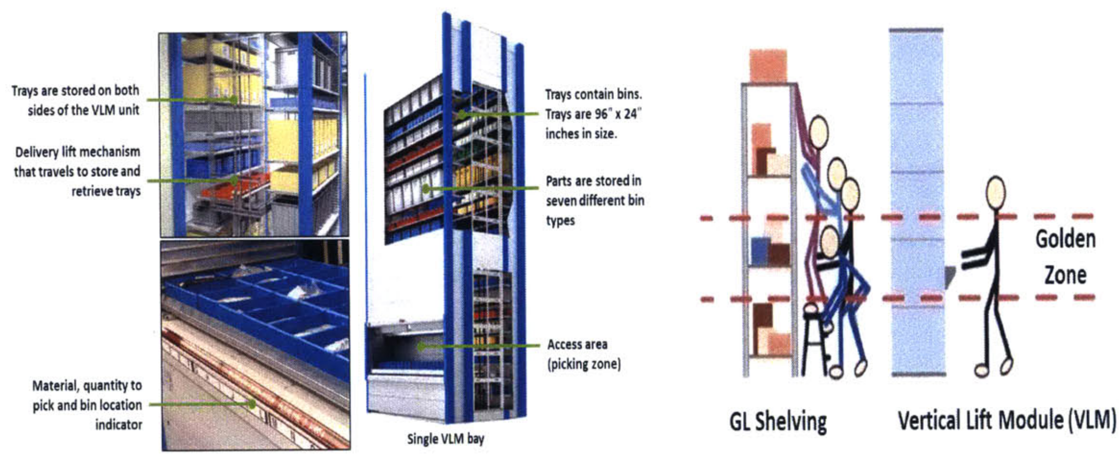


Figure 2: Warehouse floor plan shows various picking areas, kitting rooms, and consolidation areas



(a) (b)

Figure 3: a) VLM description breakdown; b) GL picking and VLM picking

At VSEA, a single VLM machine is known as a bay and a pod comprises of a set of bays grouped together and integrated as one entire setup (see Figure 4 and 5). While the term 'VLM' can be used to describe either a pod or a bay, 'VLM' refers to a pod in the remaining parts of this thesis in accordance with the terminology used at the company. There are five pods at VSEA and one operator is assigned to every pod. Four of the pods (Pods 1, 2, 3, and 5) comprise of three bays each and one pod (Pod 4) has two bays. There are 40-50 trays per bay and each tray (96" x 24") can fit 192 smallest sized (BA04) bins. There are 24 bins along the length and six bins along the width of one tray. There are seven distinct bin sizes (see Table 1) and 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 location identification number is VL01-02-03-A01 where the first two digits '01' represent the pod number, the next two digits '02' represent the bay number, the next two digits '03' represent the tray number, and the last three characters 'A01' represent the bin location.

Figure 4: A pod with three bays

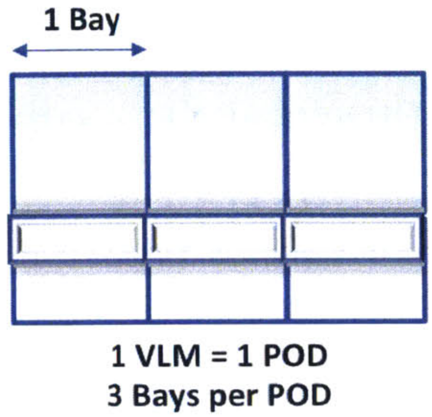


Figure 5: Graphic showing a pod and a bay

Table 1: Bin sizes used in the VLMs for part storage

Bin Type	Width (in)	Length (in)	Height (in)
BA04	4	5.5	4
BB04	4	11	4
BC04	8	11	4
BD04	8	16.5	4
BF04	8	33	4
BG04	12	16.5	4
BG08	12	16.5	8

There are a couple of advantages of VLMs as compared to the GL shelves, which have led to a push for increasing the number of VLM picks at the warehouse in recent months. The primary benefit of picking from VLMs is that the travel time associated with walking along aisles to pick the parts from shelves is eliminated since the VLMs bring the parts to the pickers. Typically each VLM setup has multiple bays so as to reduce the waiting time for the parts to be delivered to the access area [11]. While the picker is picking from one of the bays, the remaining bays search and retrieve the next part to be picked. This time saved translates into higher pick rates at the VLMs than GL. The pick rate from VLMs is expected to be 60 picks per hour while GL averages 20 picks per hour based on historical data. This is critical for delivering parts on time since the faster the parts can be picked, the sooner the orders can be consolidated to be ready for shipping.

In addition, unlike picking from the shelves of GL, which may require pickers to climb ladders, picking from VLMs is safer since the parts are brought to the picker at the access area. Last but not least, VLMs have a greater storage capacity than GL shelves.

1.3.2. Picking Methods

The warehouse picking process is summarized in Figure 6. The warehouse first receives orders either from the production floor or sales and a designated employee releases the parts periodically as pick waves to the warehouse pickers. The parts are then picked in parallel from three storage areas. VLM consolidation involves grouping the parts picked from the VLMs that belong to the same consolidation group. 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, they are packaged together in the consolidation area and finally staged for trucking.

There are two VLM picking methods: pick-and-consolidate and pick-and-pass. In the former method, parts of a consolidation group are picked in parallel from the pods separately as illustrated in Figure 7a and thus arrive to the VLM consolidation racks in up to five different totes (since there are five pods). In contrast in the latter case, parts are picked sequentially and a tote is passed along the pods as shown in Figure 7b. Parts for a consolidation group are put together in a single tote as the tote is passed along.

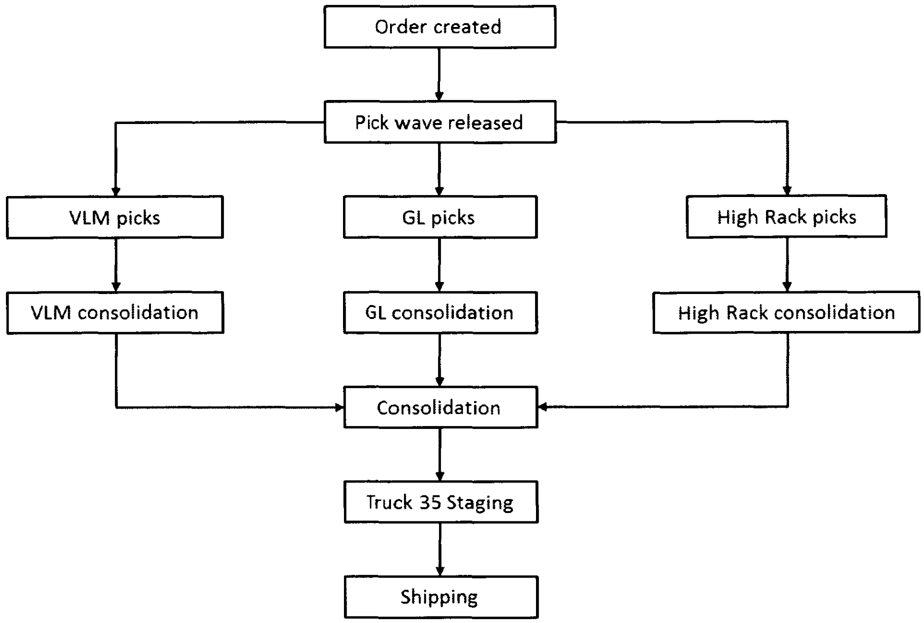
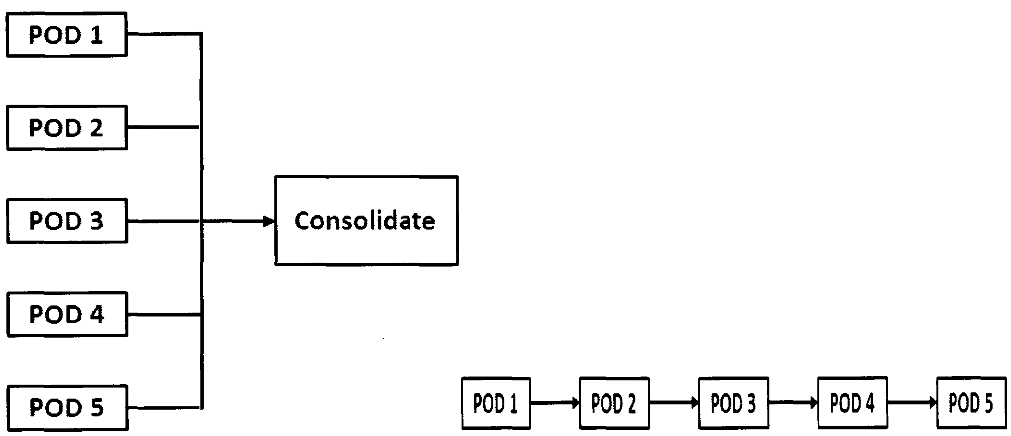


Figure 6: Picking process flow chart



(a) (b)

Figure 7: Picking strategies (a) Pick-and-consolidate; (b) Pick-and-pass

1.4. Motivation and Problem Statement

A crucial function of the warehouse is to deliver parts on time to the production floor. At VSEA, the target part delivery time from the warehouse to the production floor is 24 hours. However, during the busy periods, parts are not delivered on time to the production floor as the warehouse struggles to keep up with the high demand of parts required by the production floor and their customers. During the last busy period between November 2014 and January 2015, the time to deliver the parts to the production floor was more than three days. This project will focus on ensuring that the parts are delivered on time, especially during busy periods when timing is critical. Reducing the part delivery time would have several benefits.

First, reducing the part delivery time would reduce delays in machine laydown date and prevent missed or late shipments. Since the industry is primarily customer driven with a few key customers, missed or late shipments are a major concern for VSEA, as they stand apart in customer service. Furthermore, each shipment is worth at least a million dollars and a missed shipment may mean loss in potential sales revenue.

Second, to ensure high customer service level, VSEA allows its customers to change any configurations or 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 the company. Since customers routinely make configuration changes, VSEA must reduce risk by pushing the actual production to the last possible minute. In order to minimize production time and push it to the last possible moment, VSEA and other MIT teams have focused on reducing lead times and improving assembly times [8, 9, 12].

Third, for the lean manufacturing strategies to work, the parts must arrive on time consistently. This minimizes work-in process (WIP) inventory on the production floor. Having reliable delivery times help the production planners to develop a more accurate production plan and allocate labor resources more carefully. This would also prevent production planners from ordering parts way ahead of the laydown date and would reduce WIP inventory, which is not only costly, but also takes up valuable space on the production floor.

Fourth, oftentimes there is a problem of part shortage on the production floor that prevents important assemblies from getting completed on time and increases WIP inventory. Having a more streamlined and reliable part delivery system would allow handling part shortages and emergency orders more effectively.

1.4.1. Problem Finding

The problem finding process began by interviewing the department managers, production floor supervisors and assemblers, warehouse management, and material handlers. A detailed process flow map of the warehouse and the production floor was created to better understand the material flow and opportunities for improvement. The problem was defined and the delivery data time was obtained to quantify the problem. Root cause analysis was performed and three major problems were identified:

First, more than 16% of parts in the automated storage location (VLMs) were not picked in the past one year and were not expected to be picked during the next six months. On the contrary, there were a large number of very frequently picked parts in the manual storage location (GL).

Second, the five pods of VLMs in the warehouse had a skewed workload distribution, with one pod contributing to as high as 34% of total VLM picks and another pod contributing to as low as 11% VLM picks.

Third, pick-and-consolidate picking strategy was identified as a more effective picking strategy for VLMs, but pick-and-pass strategy was also being practiced along with pick-and-consolidate to pick parts from the VLMs.

To address these three problems three different projects were created:

1. Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Strategic Allocation of Parts: To increase the number of VLM picks and reduce the workload at the GL shelves.
2. Improving Operational Efficiency of a Semiconductor Equipment Manufacturing Warehouse through Effective Utilization of Vertical Lift Modules: To balance the workload across the five pods.
3. Improving and Maintaining the Efficiency of a Semiconductor Equipment Manufacturing Warehouse: To evaluate which picking method, pick-and-pass or pick-and-consolidate is more efficient.

The first thesis goes in greater depth explaining the problem identification process and root cause analysis. It focuses on reallocating parts to different warehouse picking locations to make the picking process faster and more efficient. This includes identifying and moving 2,052 FISH parts (parts that are not picked in past one year and are not expected to be picked in next six months) to the GL area in the most efficient manner possible. It also further identifies fast moving parts in the GL area that could be fit into VLMs and discusses the allocation process of these fast movers to specific VLM pods to balance the workload distribution of VLMs. It also identifies the fast moving GL parts that could not be fit into VLMs so that they could be put into easily accessible shelves in GL. Finally the thesis discusses how the actual implementation process was carried out.

The second thesis [14] concentrates on balancing the workload distribution of all five pods to maximize the throughput of the VLMs. The thesis goes into greater depth to evaluate three different part allocation strategies: random, snake, and order grouping. The random strategy as the name implies randomly distributes parts amongst pods, the snake strategy distributes parts in VLMs based on their pick frequencies, while the order grouping strategy focuses on arranging parts that get ordered together evenly among all five pods. The thesis also discusses, which strategy would be easiest to implement and sustain.

The third thesis [13] compares the picking performance of pick-and-pass, and pick-and-consolidate picking strategies. Pick-and-pass is a time consuming and less efficient strategy wherein a single tote is inducted for one order. The parts are picked sequentially and moved from one pod to another. For pick-and-consolidate, the order is picked from different pods in parallel and then consolidated. The thesis examines the pros and cons of both these strategies and identifies which strategy would be best for busy periods. The sustainability of the proposed solutions is also discussed in this thesis.

1.4.2. Thesis Organization

Chapter 2 presents an overview of the production floor and the warehouse, and discusses the problem identification process. Chapter 3 starts off by identifying the improvement areas in the picking process and then points out the root causes leading to poor part delivery performance. Chapter 4 discusses possible solutions that can be implemented to reduce part delivery times.

Chapter 5 discusses the implementation process and also presents the challenges faced during implementation. The conclusion and future work is provided in chapter 6.

This page left blank intentionally.

2. Problem Identification

2.1. Tour of the Production Floor

To understand the manufacturing operations of the company, the team first took a tour of the production floor, clean room, and the inbound and outbound shipping areas of the production floor. Figure 8 shows the layout of the production floor. The production floor has a receiving dock where the material from the warehouse and suppliers is received. The sub-assemblies that are later used for module assembly are built and tested in the supermarket area, and the module assembly and testing is done in the flow line and the universal end station area of the production floor. Upon customer's request all the four modules (source, analyzer, corrector, and universal end station) of the ion-implantation tool are assembled in the clean room for testing and are then disassembled, cleaned, and packaged in the air shower area. Finally, the tool is inspected and packaged at the shipping area and is shipped through the outbound shipping dock.

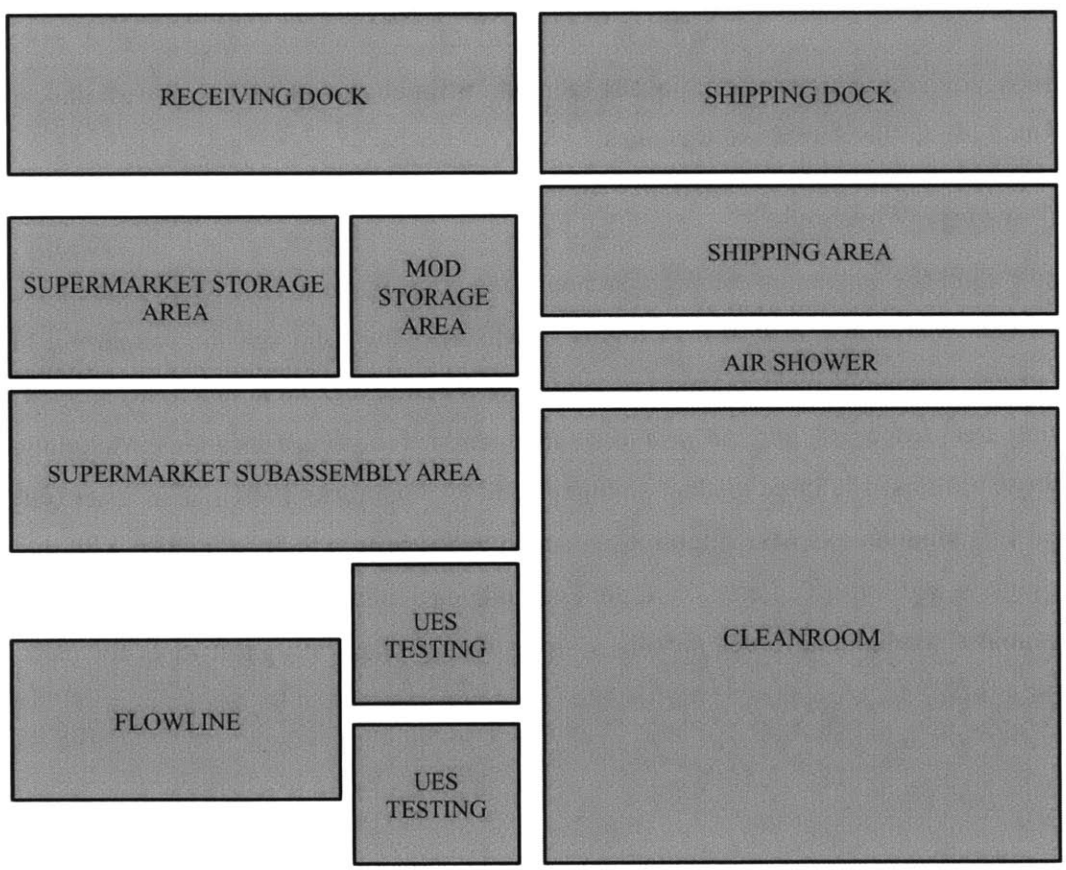


Figure 8: Production floor layout

During the next few days the department managers, and the production floor supervisors and assemblers were interviewed. The purpose of these interviews was to understand the key problem areas that make the production floor less efficient. The employees brought up a few key issues:

1. Shortages of subassemblies at the kanban squares; these subassemblies are used for the module assembly.
2. Allocated assembly times are longer than actually needed.
3. Missing parts are not labeled properly when they arrive from the warehouse.
4. Large WIP inventory at the universal end station assembly area.
5. Parts are often mistakenly delivered from the suppliers directly to the production floor. These parts have to be sent back to the warehouse for inspection.
6. Warehouse is unreliable and disorganized, and parts are often not delivered on time.
7. Trucking from the warehouse is unreliable, especially during morning hours.

The frequency with which the employees brought up warehouse unreliability was alarming and therefore a trip to the warehouse was made.

2.2. Tour of the Warehouse

To understand the operations of the warehouse the team made a visit to the warehouse. The layout of the warehouse is shown in Figure 9. Material flows through the warehouse in three main stages: receiving, picking, and consolidation for shipping. Parts are first received at the receiving area, inspected, and put away into their respective storage locations. Materials in the warehouse are stored in three distinct storage locations: high racks (WH), Gloucester (GL), and Vertical Lift Modules (VLMs). High racks contain bulky parts, which are picked with the aid of cherry pickers and forklifts. Parts in GL are less bulky and manually picked from shelves. VLMs are automated storage and order picking systems that mechanically eject trays for parts to be picked.

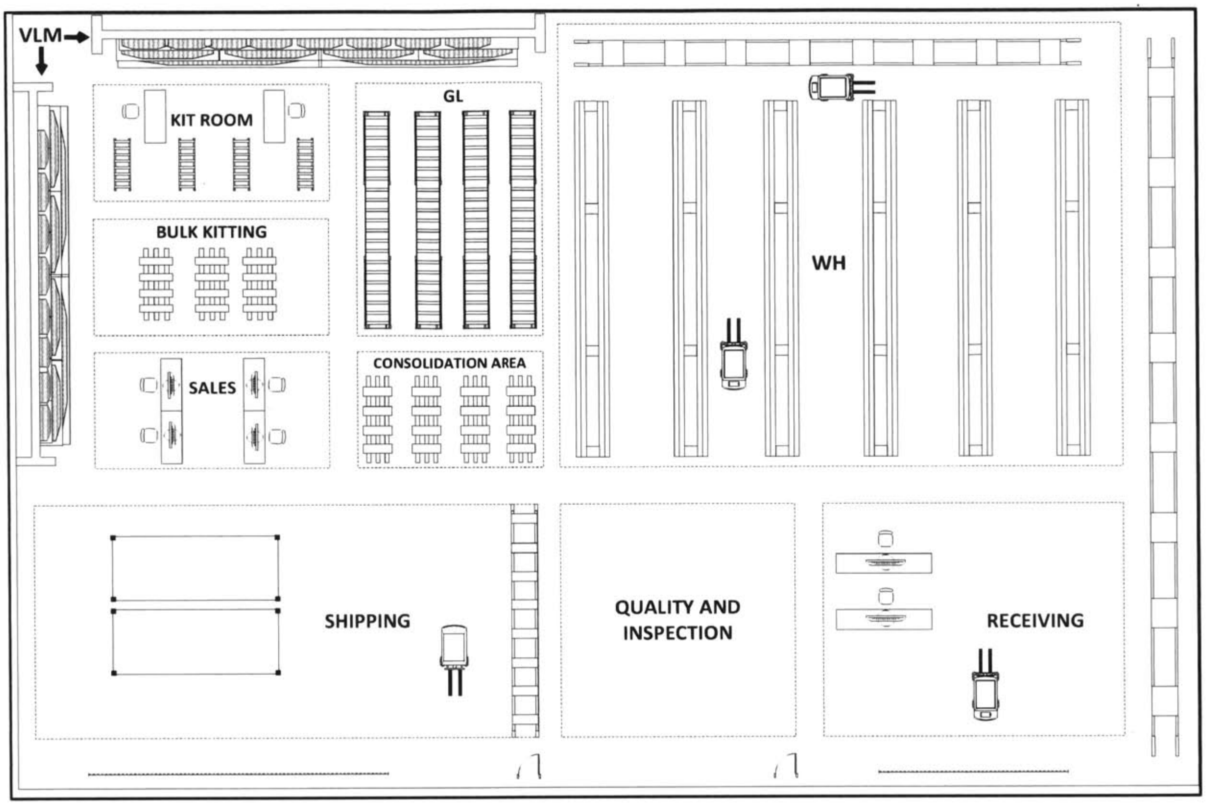


Figure 9: Warehouse layout [9]

The warehouse receives shop orders from the production floor and pick waves are released to the material handlers. Parts are then picked from the VLMs, GL, and high racks. The consolidation is first done separately for all three picking locations, and then final consolidation is done to prepare a completed order. At last, the parts are either shipped directly to the customers or sent to the production floor by a truck. The complete order picking process is shown in Figure 6.

The following key observations were made after the visit to the warehouse:

1. There was excessive scanning associated for parts that had to be sent for inspection.
2. There was decent amount of activity in the manual-picking zone (GL), where pickers were constantly climbing stairs to pick parts.
3. During the entire duration of the stay at the warehouse, the material picker at Pod 1 was picking parts continuously, while the pickers on Pod 4 and Pod 5 were standing idle, and often walked away from the VLMs.

4. It was observed that the picks from VLMs that took the maximum amount of time were picks for screws and other small parts that were ordered in large quantities. This was because pickers had to manually count these parts.
5. There was excessive amount of scanning involved in the consolidation area. Each set of parts had to be scanned thrice before it could be grouped into a single handling unit.

After taking the tour of the production floor and the warehouse, a process flow chart was made in which the problems and the improvement areas were identified at each stage of the process (see Figure 10).

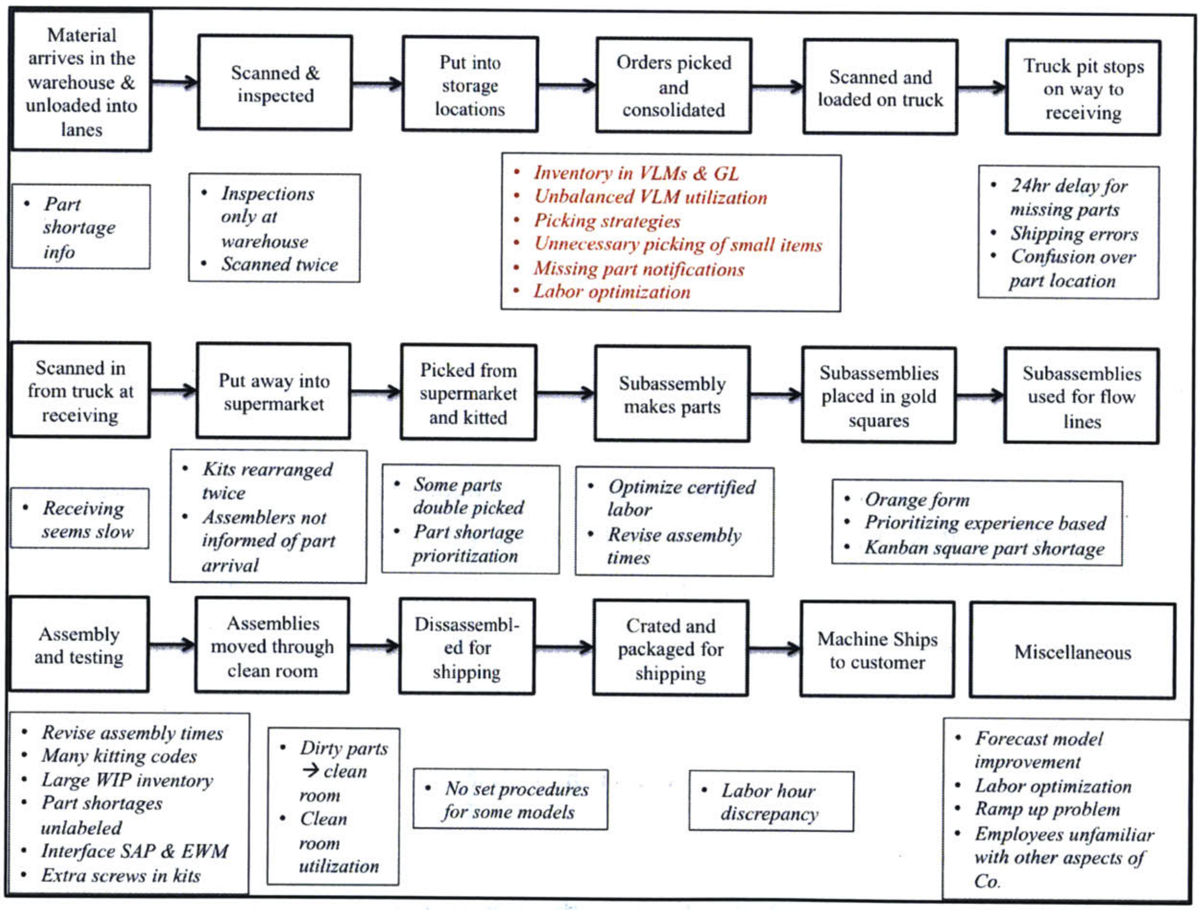


Figure 10: Process flow chart of the warehouse and the production floor

The process flow chart helped the team to trace the problems and inefficiencies in the production process. After analyzing the process flow chart, it was concluded that the warehouse needed the most attention in terms of reducing the part delivery times and making the picking process more efficient.

2.3. Current Part Delivery Performance

To track the performance of the warehouse, VSEA has set a part delivery target of 24 hours. The warehouse maintains an excel worksheet to track its performance. The graph (see Figure 11) shows the average time it takes for the warehouse to deliver parts to the production floor once the order has been placed. The horizontal line shows the target delivery time and we notice that there are several instances when the warehouse fails to meet its target. The actual situation is worse than what the graph shows. First, the 24-hour target set forth by the management requires that each order should arrive to the production floor within 24 hours, but the graph is plotted after taking an average of day's worth of orders. If the mean delivery time for a day is below 24 hours it doesn't give any information about the spread of the underlying distribution, and therefore the delays are underestimated. Second, the graph is missing a 3-month busy period between November 2014 and January 2015. The warehouse couldn't track its performance during this time and the management quoted an average delivery time during this period to be 36 hours.

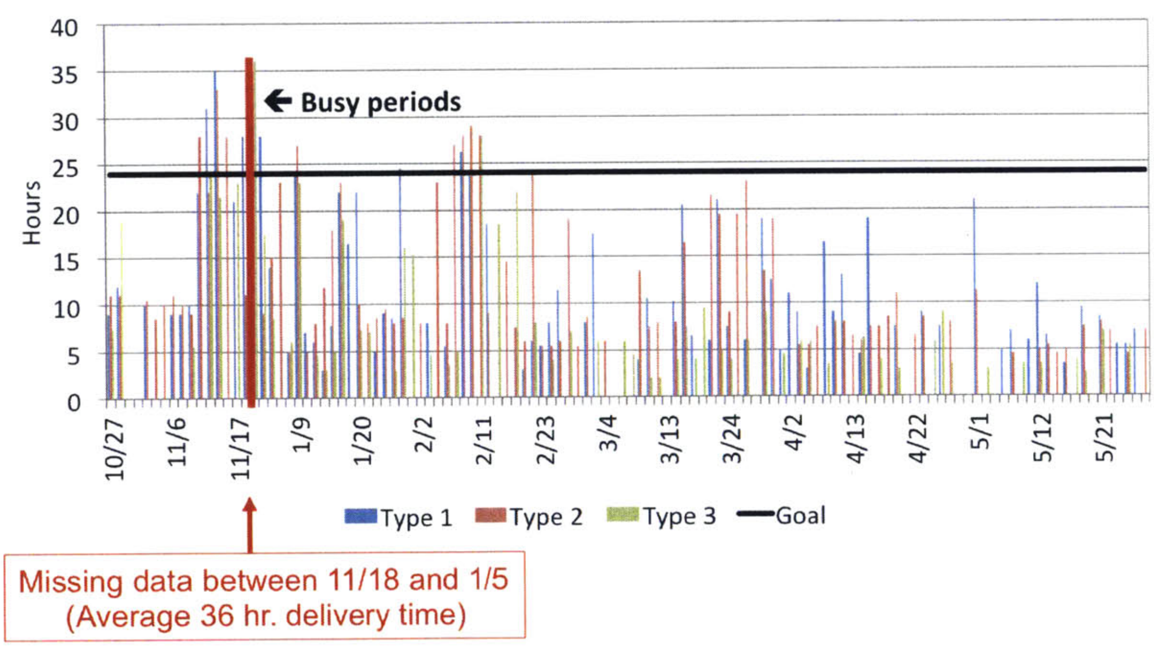


Figure 11: Part delivery times from the warehouse to the main building

To confirm that the actual part delivery time matched with the quoted time, part delivery data from Extended Warehouse Management (EWM) system was analyzed during this time duration and it was discovered that the average delivery time was in fact above 3-days. This was a shocking discovery and made the problem even more crucial to be solved.

Another interesting graph (see Figure 12) that was provided by the management shows the number of picks and picks per hour made at the VLM and the manual picking locations (high racks and GL). The number of picks per hour for the VLMs was steadily decreasing over past couple months and was approaching the number at manual picking locations. VLMs were installed in the warehouse to increase the picks per hour and to save space. Clearly, the benefit of faster picking rates at VLMs was diminishing and the picking rate were one-third of the expected value of 60 picks per hour. The picks per hour in the manual picking location were around 20, and for VLMs they were 35 but declined down to 22. Also, the number of picks made from the VLMs was consistently lower than the number of picks made at the manual picking locations.

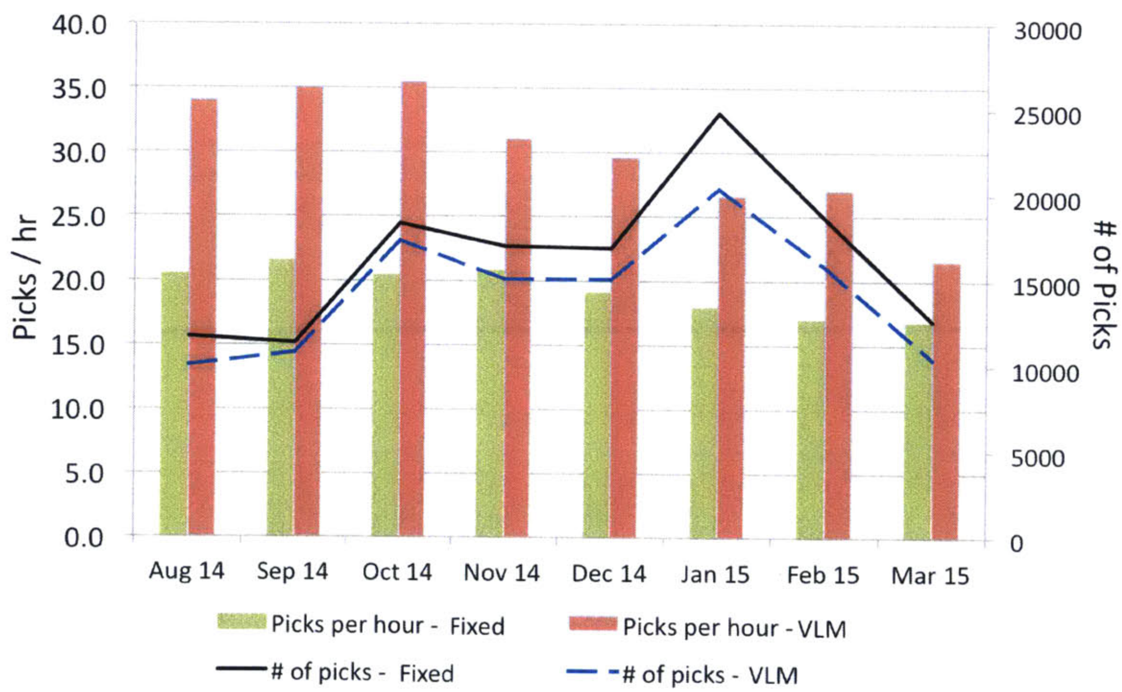


Figure 12: Picking rates and number of picks from VLMs and manual (fixed) picking locations

3. Root Cause Analysis

Once the current delivery times and performance of the warehouse were studied, the next step was to pinpoint the areas that required the greatest attention during busy periods. After identifying the improvement areas, the root cause analysis was performed.

3.1. Identifying the Improvement Areas

The order picking process at the warehouse is shown in Figure 6. When the warehouse receives an order, it is released to the part pickers and the picking process starts. To complete an order, parts have to be picked from up to three different locations (VLMs, GL, and RK). Once the parts have been picked from their respective locations, they are sent to the consolidation area. Consolidation could only be performed when all the parts from a specific order have been picked and arrive in the consolidation area. Parts can be picked from all three locations (VLMs, GL, and RK) in parallel. If an order is released at 9AM, then pickers at all three locations can begin to pick parts at 9AM. Say, pickers from VLMs were able to complete picking parts by 2PM, pickers from GL finished picking parts by 5PM, and pickers from RK finished picking parts by 3PM, then consolidation can be started earliest at 5PM. To pinpoint which area out of these three (VLMs, GL or RK) takes the majority of the time (during busy periods), data was obtained from EWM and processed.

Table 2 shows the average time it takes to complete picking an order at each picking location. The time at each location starts when an order is released and ends when the last part at each location corresponding to that order is picked. Based on the previous example if picking at VLMs, GL, and RK is completed by 2PM, 5PM, and 3PM, respectively, then it took 5 hours to pick an order from VLMs, 8 hours to pick an order from GL, and 6 hours to pick an order from RK. Then, the critical path in the picking process becomes GL as it takes the most time and therefore delays consolidation. The table also shows the consolidation time, which is the time it takes to consolidate the parts that are picked from VLMs, GL, and RK to prepare a completed order. The consolidation time starts when parts from all three picking areas arrive at the consolidation location and ends when they are consolidated together under a single handling unit.

Table 2: Time study results from busy period

	Total Average [*hr*]	**3 Day Cutoff** [*hr*]	**1 Day Cutoff** [*hr*]	**3 Day % Excluded**	**1 Day % Excluded**
VLM	41.9	11.8	7.0	13%	23%
GL	49.5	19.1	9.4	16%	40%
RK	27.9	12.7	8.2	8%	20%
CONSOLIDATION	24.3	3.1	1.1	11%	15%
Total Pick & Consolidation Time	73.8	22.2	10.5		

In Table 2, the numbers corresponding to the total average time at VLMs, GL, RK, and Consolidation are likely to be on the higher side because part shortages or weekends tend to exaggerate actual part picking and consolidation times. Therefore, 3-day and 1-day cutoff times were calculated. 3-day cutoff ignores all the orders that took more than three days to be picked from their respective locations. The table also shows how many orders were ignored because of a 3-day cutoff. We can see that 13% orders took more than three days (72 hours) to be picked from VLMs, 16% orders took more than three days to be picked from GL, 8% orders took more than three days to be picked from RK, and 11% orders took more than three days to be consolidated. Similarly, 1-day cutoff is calculated and ignores orders that took more than one day to be picked from their respective locations.

Finally, total pick and consolidation time is listed, which is the sum of the consolidation time and the maximum of VLM, GL, or RK picking time. We notice that the total average pick and consolidation time is 73.8 hours, which is the sum of consolidation time (24.3 hours), and GL picking time (49.5 hours). 1-day cutoff excludes a high percentage of total orders (23% for VLMs, 40% for GL, 20% for RK, and 15% for Consolidation); therefore 3-day cutoff can be taken as a more objective data set. Figure 13 presents the 3-day cutoff data in a graphical manner. The data corresponds to the busy period from November 2014 to January 2015, and based on a 3-day cutoff it takes around 19 hours to pick an order from GL and 12 hours to pick an order from the VLMs.

Based on the available data it is tough to breakdown what exactly happened during this 19-hour (GL) or 12-hour (VLM) time period. Therefore, the warehouse staff was consulted; it was mentioned that the majority of the picking time at the GL location was because of the formation of queues. At a single point in time, up to three or four orders are picked in GL (one order per picker), and the remaining orders are waiting in a queue. Another major component of the picking time at GL includes nighttime when not much activity happens at the warehouse. With VLMs, sometimes a part picker at say Pod 5 has finished picking a previous order and part picker at Pod 1 hasn't (happens often because of imbalance in VLM pod utilization), then the next order is split and a wave is released that contains parts that come from Pod 5 only. Therefore, not all parts in an order are released at once and the overall completion time for an order increases considerably because of being split into different waves. Also, parts are not picked from the VLMs during the night shift and therefore some orders have to wait in a queue to be completed. The queues during the busy period were long partly because of staff shortage in the warehouse. It should also be noted that even though there is a 3-day cutoff, the numbers could still be exaggerated due to part shortages, night times, and weekends.

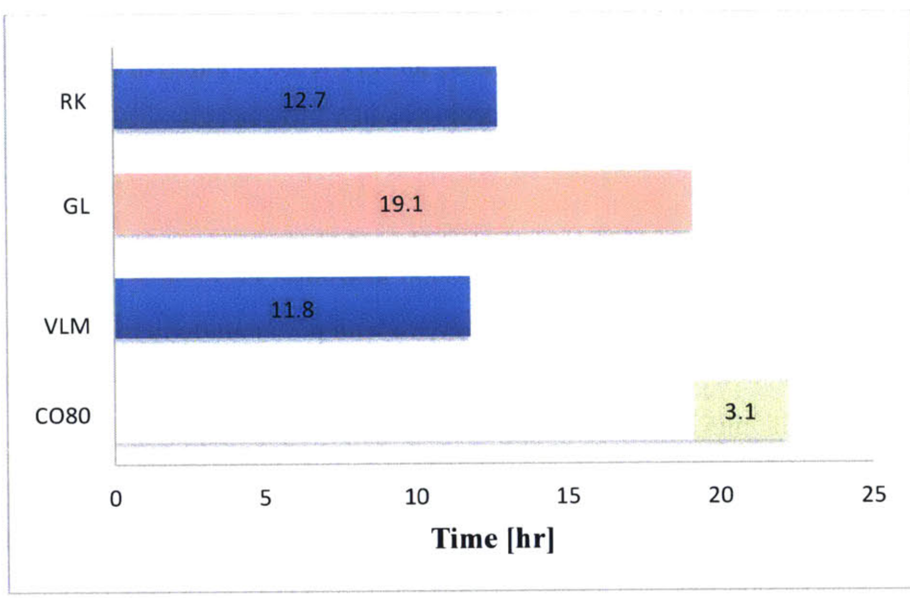


Figure 13: Imbalance in the picking process (3-day cutoff)

The picking process takes 19.1 hours and the consolidation process takes 3.1 hours to complete. Clearly, on average it takes longer to pick from GL, which causes an imbalance in the picking process. By reducing the GL picking time, the overall order picking time could be reduced. Although, the above numbers could still be exaggerated because of shortages, night times, and weekends, there is no reason to believe that GL is not a critical path. Currently, on average an order can be picked in 19.1 hours, but if the GL time is reduced down to 12.7 hours, then the order can be picked in 12.7 hours. But this is not valid for RK or VLMs; if RK or VLM picking times are reduced but GL picking time remains the same, then the order picking process is still going to take 19.1 hours on average to complete. Therefore, GL is said to be a critical path in the picking process. Now, the parts that are stored in GL and VLMs are very similar in nature. Parts from GL can be moved to VLMs and vice versa to shift the workload. Hence, these two picking locations became the primary focus of the project.

Ideally, it is preferred that the parts that are picked more often are stored in VLMs and the parts that are picked less often are stored in GL. As it was discussed earlier, VLMs offer an advantage for part pickers because pickers don't have to walk across aisles to pick parts, and this saves time.

Although there is certainly a need to make consolidation, RK, and trucking more efficient, keeping in mind the time constraints of the project the best areas to improve were identified to be VLMs and GL. The next step was to identify problems in VLMs and GL.

3.2. Fishbone Diagram

Within the domain of VLMs and GL, the fishbone analysis (see Figure 14) proved to be an effective method to identify the root causes associated with poor part delivery performance. Four main areas: machine, method, measurement, and policy were studied. The main problems under machine were high GL picks, low VLM picks, and skewed part distribution in VLMs, under method were part picking strategies, under measurement were declining pick rates, and under policy were lack of manpower during busy periods. Majority of these areas are addressed in detail.

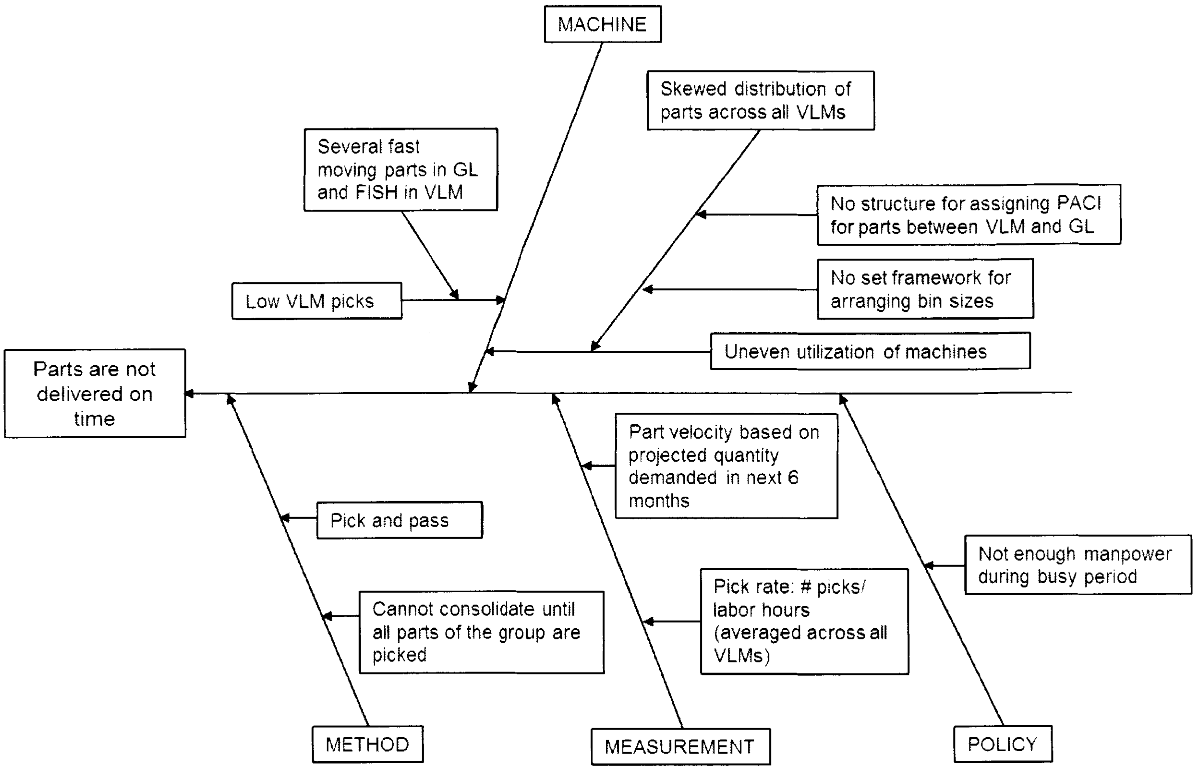


Figure 14: Fishbone analysis

3.3. Vertical Lift Modules (VLMs)

In 2012, the warehouse layout of VSEA had to be reorganized to make way for five new Vertical Lift Modules (VLMs). The introduction of the VLMs would provide for greater storage capacity in a smaller footprint and was expected to improve picking efficiency at the warehouse.

VLMs are inventory storage systems made up of vertically arranged trays for storage, a delivery lift platform mechanism, and a computerized control system. A tray is automatically brought to the access point upon request from the software. The operator picks or replenishes the stock and then the tray is returned to its position after confirmation.

VSEA bought a total of five pods. Four of the pods comprised of three bays each and one comprised of two. There is one operator for every pod. There are 40-50 trays per bay and each tray (96"X24") can fit 192 smallest sized (BA04) bins. There are seven distinct bin sizes (Table 1) and a total of approximately 25,000 bins across all five pods.

Figure 15 shows the distribution of different bin sizes in all five pods. The general trend is to have more smaller bin sizes and fewer larger bin sizes. The distribution of different bin types in VLMs is not uniform.

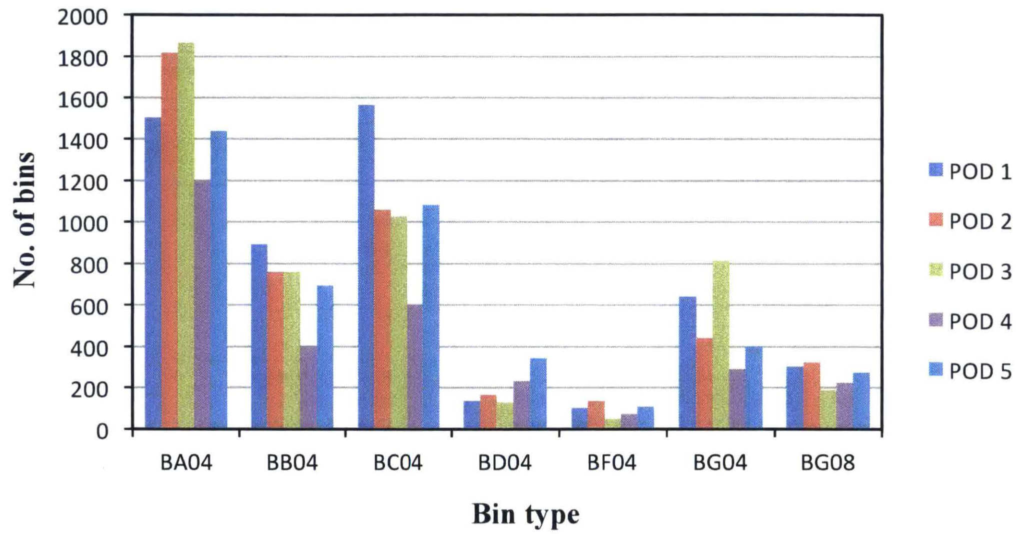


Figure 15: Number of different type bins in each pod

3.3.1. Low VLM Picks (FISH Parts)

When the VLMs first arrived, the team at the warehouse had to reallocate thousands of parts from the carousels, racks, and GL to these machines. Initially, the parts were said to be placed in the VLMs somewhat randomly but later on parts were classified as fast moving or slow moving. Fast moving parts were stored in the VLMs whereas slow moving parts were stored in the GL. While this approach has been continued till date, it has not been implemented effectively. Recent trends have revealed low picking rates, low utilization, and rising number of excess and obsolete parts in the VLMs.

When the management first invested in the Vertical Lift Module (VLM) machines, they expected high VLM pick rates (up to 60 picks per hour), and greater number of VLM picks than manual picks. However, from Figure 12, we can observe that the number of VLM picks was consistently lower than the number of manual picks between August 2014 and March 2015. The VLM picking rates are also significantly lower than expected and diminishing with time. The trend shows the declining picking performance at the warehouse.

To study the nature and the velocity of parts stored in the VLMs, we collected past twelve months and projected six months of data for 13,223 unique parts in the VLMs. The data showed how frequently each part was picked during this time duration. Table 3 shows the percent of FISH parts in the VLMs. FISH stands for First In Still Here, and is a generic term used by the warehouse management to define parts that are not picked over extended duration of time.

Table 3: % FISH parts in VLMs

	Past 6 months	**Past 12 months**	**Past 12 & next 6 months**
# of FISH Parts	4306	2879	2052
Total # of Parts	13223	13223	13223
% FISH	33%	22%	16%

Based on the data presented in Table 3, 33% of the parts in VLMs have never been picked over past six months, 22% parts have never been picked over the past year, and 16% parts were neither picked over the past year, nor projected to be picked over the next six months. FISH parts for the rest of the thesis are defined as the ones that have not been picked over past one year and are not expected to be picked over next six months and are highlighted in red in Table 3.

This was a shocking finding; the declining trend of total number of picks from the VLMs can be attributed to a large percentage of FISH parts located in the VLMs. The graph in Figure 16 shows the quantity of parts corresponding to their pick frequencies. FISH parts have been removed from this graph.

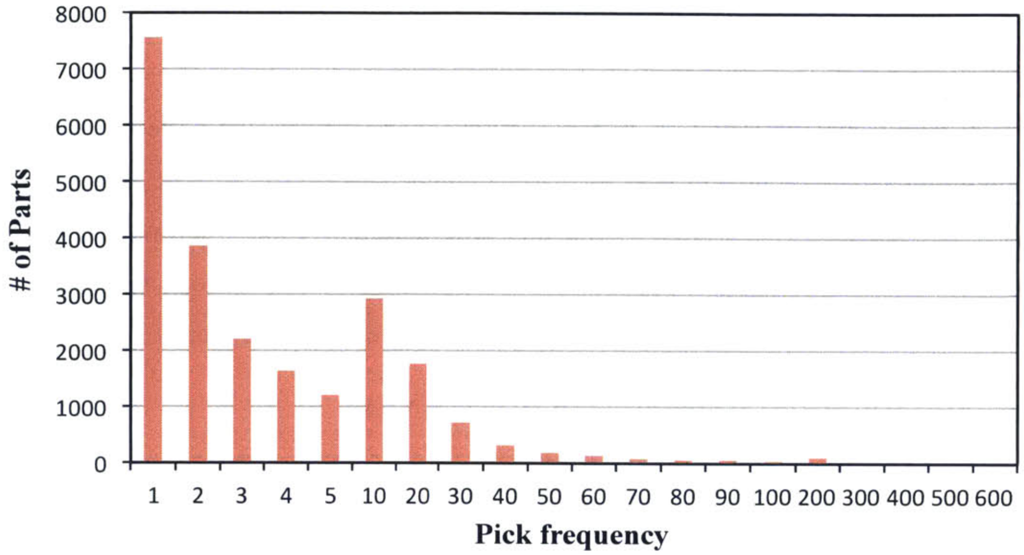


Figure 16: Breakdown of pick frequencies in VLMs (18 months)

From Figure 16, it can be noticed that despite removing FISH parts, a large majority of parts stored in the VLMs have pick frequencies lower than 30 picks per eighteen month time duration. 73% parts have pick frequencies less than 30, 70% parts have pick frequencies less than 20, 54% parts have pick frequencies less than 10, and 37% parts have pick frequencies less than 5.

There are certainly parts in the VLMs that have very high pick frequencies, but the number is small relative to low pick frequency parts. Figure 17 shows a higher resolution graph for parts with pick frequencies greater than 100. These are the set of parts that contribute to the most number of picks from the VLMs and VLMs are the most effective place for placement of such parts. 5% parts in VLMs have pick frequencies greater than 100 picks per eighteen month time period.

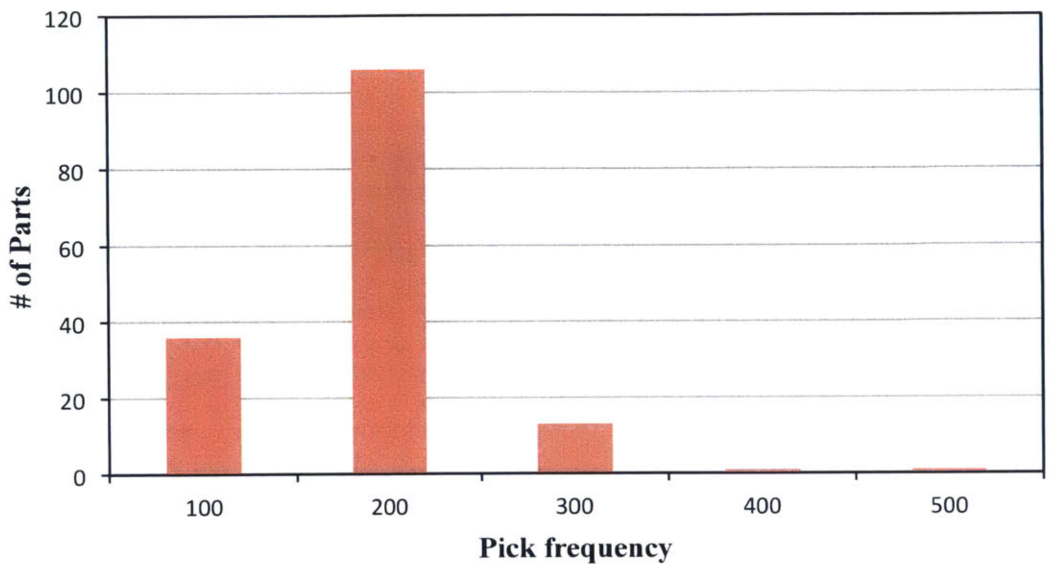


Figure 17: Higher resolution breakdown of pick frequencies in VLMs (18 months)

Approximately 23% parts in the VLMs contribute to 90% of VLM picks. The graph in Figure 18 shows the cumulative number of picks coming from VLMs corresponding to the number of parts. To plot the graph, parts were first sorted from the highest pick frequency to the lowest, next the cumulative frequency was calculated, and was plotted against the cumulative number of parts. Therefore the first, second, and n^{th} data points (x,y) on the graph would be (1, pick frequency of part 1), (2, pick frequency of part 1+2), and (n, pick frequency of part 1+2+...+n), respectively, where pick frequency of part 1 is the highest, and pick frequency of part n is the lowest. The graph is smooth because the pick frequency of an individual part is negligible compared to the overall cumulative pick frequency, and also because the scale of the graph is in thousands, therefore each additional part doesn't contribute enough to bring about a sudden jump in consecutive data points.

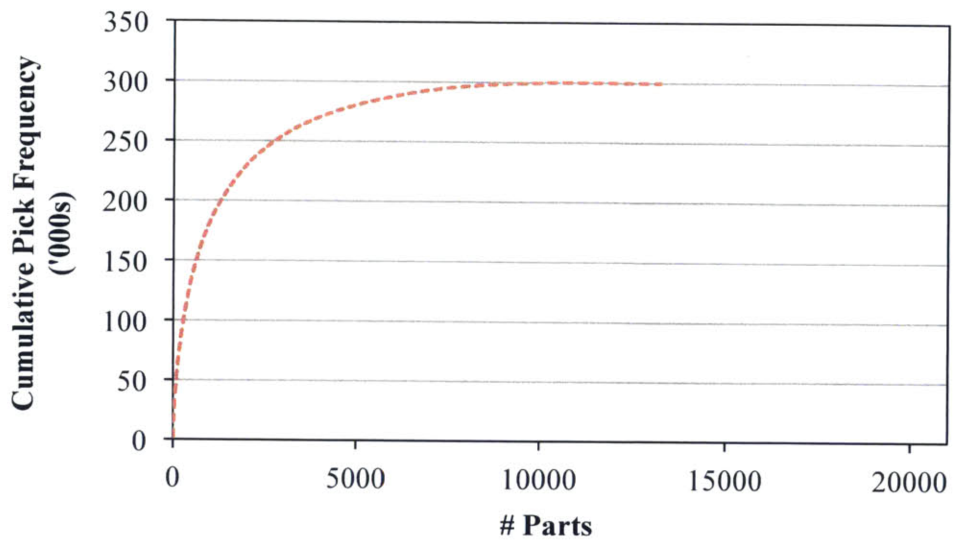


Figure 18: Cumulative number of picks coming from VLMs (18 months)

The underlying cause of many FISH parts in the VLMs is that when the VLMs first arrived, the team at the warehouse had to reallocate thousands of parts from the carousels, racks, and GL to these machines. Initially, the parts were said to be placed in the VLMs somewhat randomly without paying much attention if the parts were fast moving or not. As a result, a large number of slow moving and FISH parts were transferred into VLMs. Additionally, some of the fast moving parts overtime became slow moving parts as new ion-implantation tools were introduced and BOMs were updated. The reason for a large majority of slow moving parts present in the VLMs is that when a part is received into the warehouse, there is no set framework to classify the part as fast moving or slow moving. Therefore, parts end up being randomly allocated to GL or VLMs. Therefore VLMs end up getting slow movers along with some fast movers.

Such a large number of FISH parts and slow moving parts result in underperformance of VLMs because these parts end up taking valuable space in the machines, which could have been dedicated to fast moving parts.

3.3.2. Skewed Pick Distribution

The second major finding for low overall utilization of VLMs is uneven machine utilization. Figure 19 shows the number of picks in the past twelve months across five pods. Pod 1 has the most picks, roughly twice as many picks as Pod 3 and Pod 5. There is a strictly decreasing trend of picks from pod to pod except for Pod 5. This is because there are two bays in Pod 4 and three bays in remaining four pods. One person is assigned to each pod, therefore an imbalance not only implies unequal workload for the assigned staff but also implies lower overall throughput of five pods. Table 4 portrays the consequences of such an imbalance.

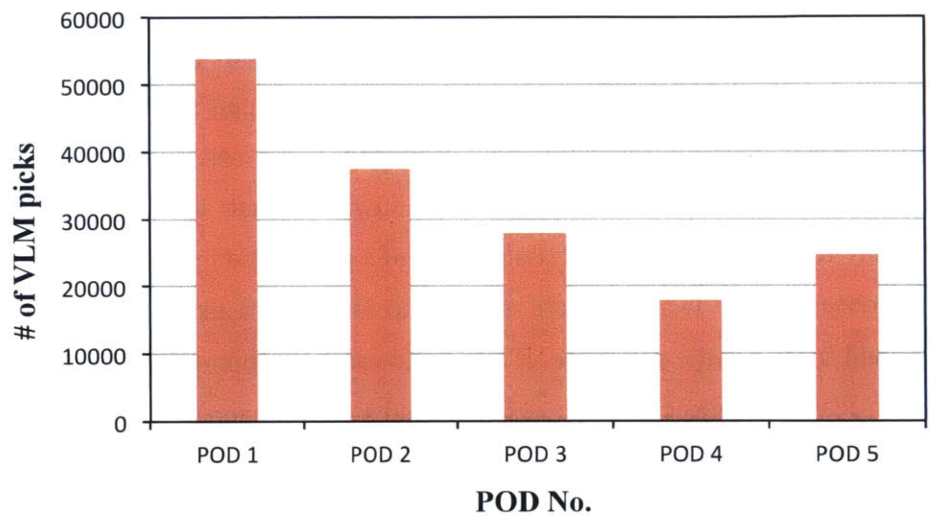


Figure 19: Breakdown of number of picks across five pods (past 12 months)

The skew in number of picks is such that on average 34% of parts that are picked from VLMs come from Pod 1, 23% from Pod 2, 17% from Pod 3, 11% from Pod 4, and 15% from Pod 5. As an illustration, if an order of 100 parts is placed, based on the current distribution of picks and assuming picks per hour to be 30, if parts are picked in parallel then it would take 68 minutes to pick from Pod 1 and only 11 minutes to finish picking from Pod 4. Therefore, Pod 1 behaves as bottleneck in the process. Ideally, it is preferred that each pod contributes to one-fifth (20%) of VLM picks. Pod 1 would remain a bottleneck even if parts are picked in series, where a tote is passed from Pod 1 to Pod 2 and so on, because Pods 2, 3, and 4 will starve (imagine an assembly line). This largely explains the observation made on the very first day of team's visit to the warehouse, where the material picker at Pod 1 was seen busy the entire time picking parts, while the pickers at both Pod 4 and 5 were idle. This not only limits the overall throughput of the VLMs but also leads to ineffective labor utilization.

Table 4: Pod 1 is a bottleneck

Order Size = 100 parts & Pick Rate = 30 picks per hour					
	Pod 1	**Pod 2**	**Pod 3**	**Pod 4**	**Pod 5**
# of Picks	34	23	17	11	15
Pick Time (minutes)	68	46	34	22	30

The reason for such a skewed part distribution is because of software put-away logic. The search sequence for put-away software starts with the pod to which a part's PACI is assigned. PACI stands for Put Away Control Indicator, and each and every part in the warehouse has a corresponding PACI assignment that instructs the software to start its search sequence from a specific pod. Say for example, the PACI assignment of a part is Pod 3; when this part is received into the warehouse, the put-away software would start looking for an appropriately sized empty bin in Pod 3, and try to put the part there. If in case, there is no appropriately sized empty bin in Pod 3, then the search sequence goes on to Pod 4, and so on. It happens to be that the PACI for majority of the parts that were assigned to VLMs was Pod 1 until unless a material handler manually changed it. Therefore, every time a part is received into the warehouse, the software search sequence starts from Pod 1. This has been a major cause of skewed part distribution in VLMs.

The reason why the utilization of pods is not uniform is because the number of parts stored in each pod is different and part demand is not accounted for when allocating parts to pods (see Figure 20- Pod 1 has the most number of parts). Figure 21 shows a frequency breakdown at the pod level for fast moving parts (pick frequencies greater than 100). We see that faster moving parts are stored in Pod 1 and therefore a large majority of overall VLM picks come from Pod 1, which leads to uneven utilization of pods.

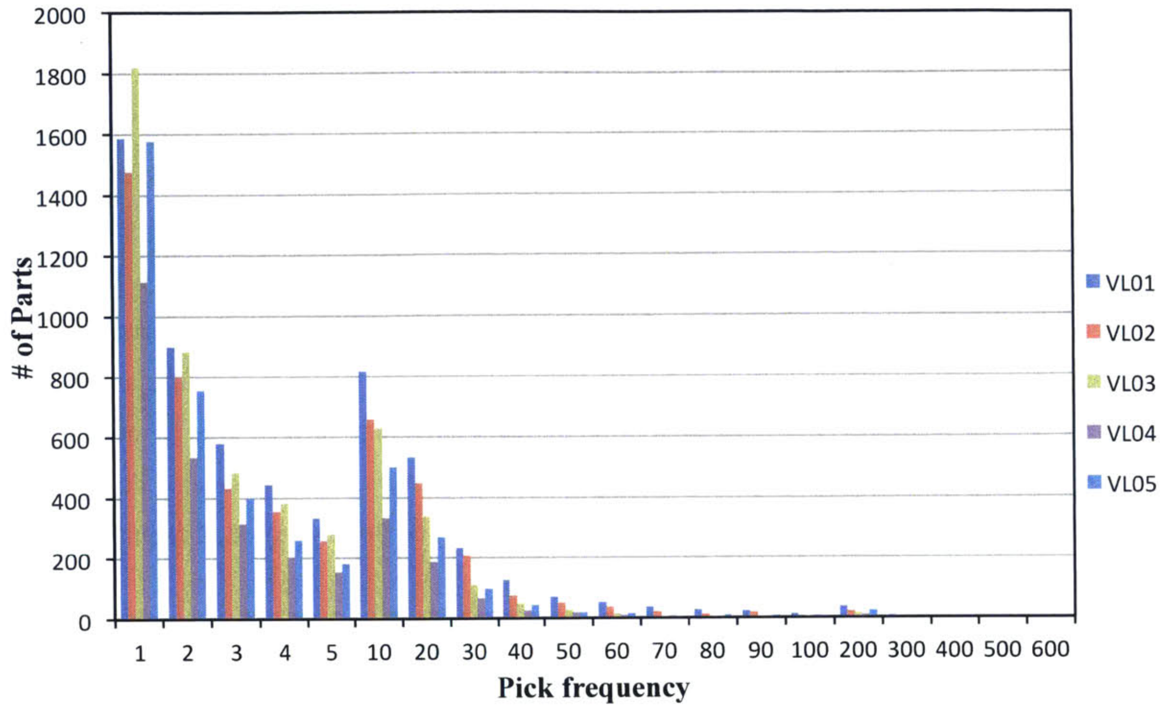


Figure 20: Breakdown of pick frequencies in VLMs at the pod level (18 months)

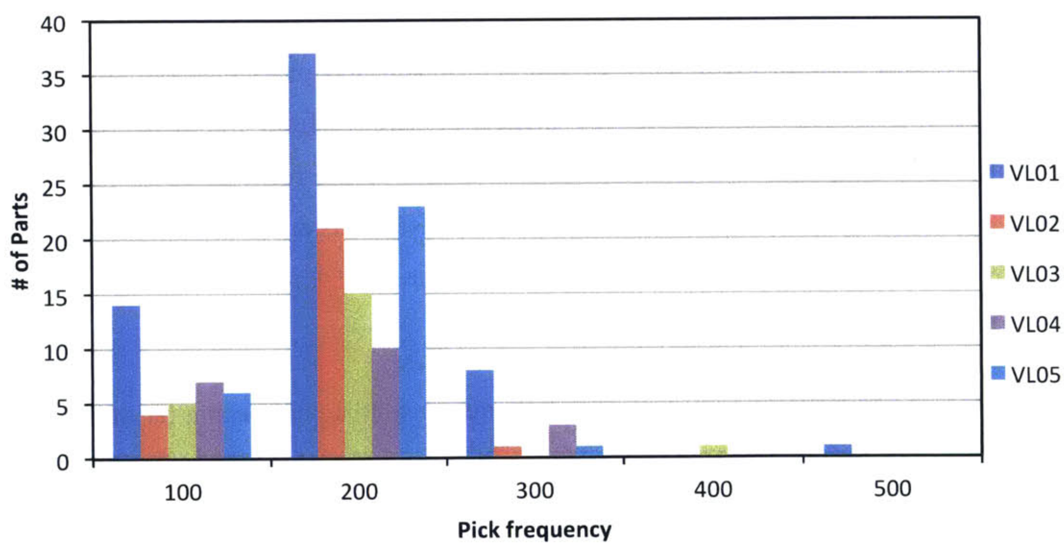


Figure 21: Higher resolution breakdown of pick frequencies in VLMs at the pod level (18 months)

A possible explanation for this distribution is that the first three pods were installed in August 2012 timeframe and the last two pods were installed in December 2012 timeframe. Pod 1 was built first, Pod 2 took additional four weeks to be built, and Pod 3 took four more weeks after Pod 2 was built. Even though the part assignment was somewhat random, Pod 1 was loaded completely as soon as it was built, and was loaded with the parts from manual carousels that the company had. The majority of the parts in manual carousels were fast movers. Next, Pod 2 was filled and finally Pod 3 was filled. Pods 4 and 5 were partially filled with the parts from GL (possibly slow movers). Also, most of the parts that are received into the warehouse are ordered because they have projected demand and are usually fast or medium movers. Due to the software put away logic, Pod 1 keeps getting these fast and medium movers. Other pods also do get these parts, but possibly at a lower rate since the slow movers that are already stored in them are rarely picked and less space is created for new parts.

Also, part of the reason why the number of picks per labor hour kept declining for VLMs over past few months (see Figure 12) is that the workers who stood idle at Pods 3, 4, and 5 often walked away and did other work in the warehouse but still logged their times under VLMs.

3.3.3. Picking Strategies

There are two VLM picking methods: pick-and-consolidate and pick-and-pass. In the former method, parts of a consolidation group are picked in parallel from the pods separately as illustrated in Figure 7a and thus arrive to the VLM consolidation racks in up to five different totes (since there are five pods). In contrast in the latter case, parts are picked sequentially and a tote is passed along the pods as shown in Figure 7b. Parts for a consolidation group are put together in a single tote as the tote is passed along.

It has been observed that during busy periods, pick-and-consolidate has been a far more effective picking strategy than pick-and-pass. The time taken to complete picking an order is faster using pick-and-consolidate method than pick-and-pass method. This is because pick-and-consolidate is a parallel picking process and there is little waiting time associated. In contrast, pick-and-pass is a sequential picking process and there is waiting time between passing the totes from one pod to another. Still, the warehouse doesn't consistently follow the right strategy at the right time. Table 5 shows the advantage of pick-and-consolidate method over pick-and-pass method. Picking time per order has been used as a metric instead of picks per labor hour because the former relates more closely to the main issue being addressed: picking a given order faster to meet delivery times.

Table 5: Results for picking strategies

	Pick-and-consolidate	**Pick-and-pass**
No. of trials	6517	3628
Avg. pick time per order	232 min	252 min

3.4. Gloucester (GL) Picking Area

GL is a manual picking area comprising of shelves containing storage boxes of different sizes. Part pickers walk along the aisles and often have to climb stairs to pick parts that are placed on higher shelves. There are a couple of disadvantages of GL as compared to VLMs, which have led to a push for decreasing the number of GL picks at the warehouse. At GL the travel time associated with walking along aisles to pick the parts from shelves is significant and leads to less picks per hour. In addition, picking from the GL shelves, which require pickers to climb ladders, is not as safe as picking from VLMs where parts are brought to the picker at the access area.

3.4.1. Fast Moving Parts in GL

Because of the benefits of VLMs, we expect that majority of the picking at the warehouse should be done from the VLMs; therefore it is desired that VLMs hold faster moving parts and GL holds slower moving or FISH parts.

Based on Figure 12, we find that more number of picks from the warehouse come from the manual picking areas (including GL). GL has about 8,000 parts and approximately 30% of the parts stored in GL contribute to 90% of the GL picks.

The graph in Figure 22 shows the quantity of parts corresponding to their pick frequencies in GL. The large majority of parts in the GL are slow moving (81% parts with pick frequencies less than 10 over eighteen months, and 27% FISH parts), but the graph also points out towards considerable number of fast moving parts. Surprisingly, there are large numbers of parts (~800) with pick frequencies above 30 over 18-month duration. Also, the number of parts with pick frequencies of above 100 is close to 280.

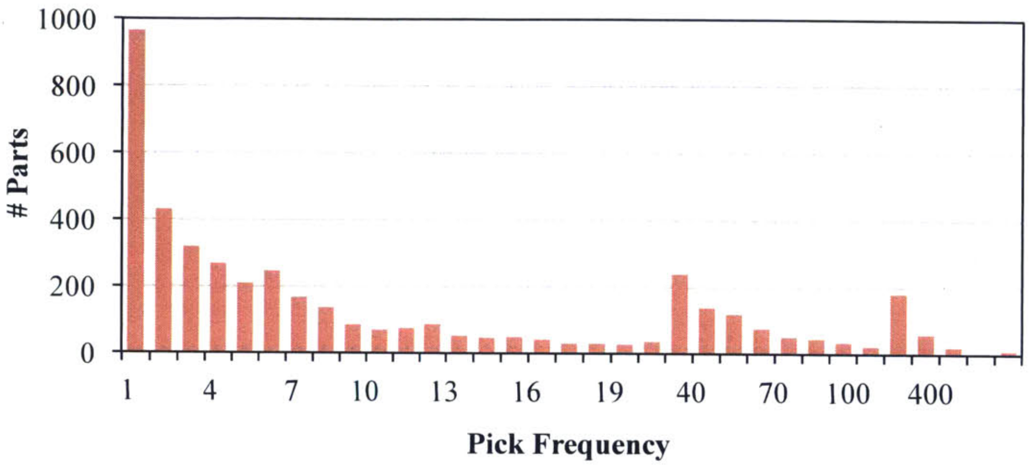


Figure 22: Breakdown of pick frequencies in GL (18 months)

We compare the frequency breakdown of GL relative to that of VLMs in Figure 23. We observe that in terms of part composition GL is very similar to VLMs, hinting a rather random distribution of parts between VLMs and GL, despite the fact that the management claims that slow moving parts are allocated to GL and fast moving parts are allocated to VLMs. Notice, there are large number of FISH parts both in GL and VLMs. We observe that (from right to left), first there is a drop of number of parts from 'no pick' to '1 pick' and then rise in number of parts from '1 pick' to '2<=x<5', and then a drop from '2<=x<5' to '5<=x<10', a further drop from '5<=x<10' to '10<=x<20', and finally a rise from '10<=x<20' to 'x>20'. Here, the focus is not on the drop and rise of the number of parts, the focus is that both GL and VLMs follow exactly the same trend; we conclude that in general the nature of parts placed in VLMs and GL appear to be very similar. Also, GL hosts roughly as many parts with pick frequencies greater than 20 as do the VLMs.

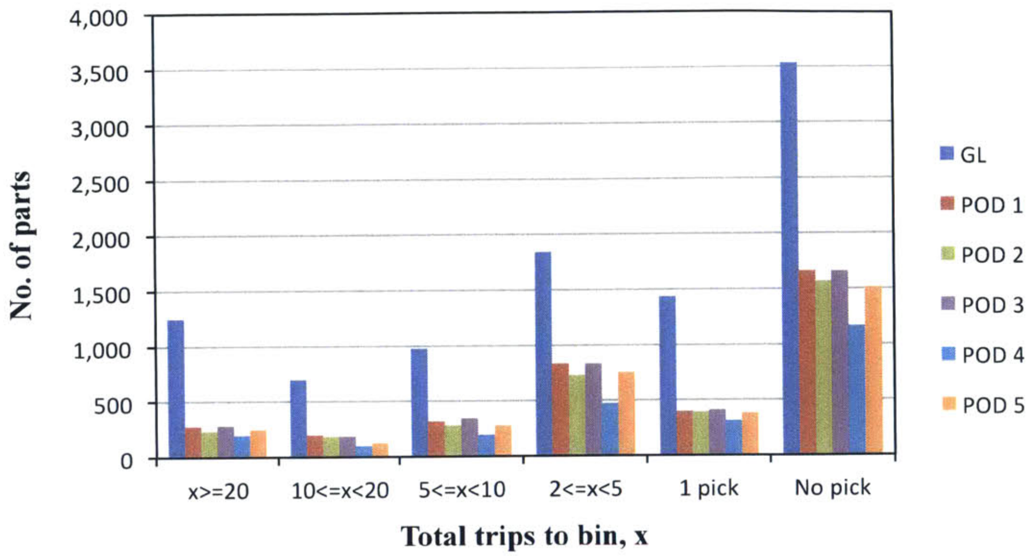


Figure 23: Breakdown of pick frequencies of parts in GL relative to those in VLMs (18 months)

Therefore, the warehouse relies heavily on the GL picking zone, despite having the privilege to utilize VLMs more and reduce the picking times. Hence there is an opportunity to increase the picking efficiency by swapping the high frequency parts in the GL with low frequency parts stored in the VLMs.

3.4.2. Slotting within GL

The problem of slotting was briefly touched upon. The GL area has shelves that are numbered A through K, where A is the bottom most shelf and K is the top most shelf.

Shelves A, B, and C are accessible without a ladder while D through K have to be reached with a ladder. We observed that the warehouse is not taking the most advantage out of shelves A, B, and C, as parts were allocated to these shelves randomly. Therefore there is a mix of fast moving, slow moving, and FISH parts in the 'Golden Zone', which is a term coined to identify the shelves that are accessible without a ladder (A, B, and C). Likewise, there are very fast moving parts stored on the shelves that have to be reached by a ladder. A better placement of parts within GL is possible.

This page left blank intentionally.

4. Solutions

Now that the root causes have been identified, the next step is to discuss possible solutions to the problems that were identified in the previous section.

To address the imbalance pointed out in Figure 13 (reproduced below as Figure 24) three major problems were identified:

1. Large number of FISH and slow moving parts in VLMs and large number of fast moving parts in GL.
2. Uneven workload distribution across the five pods.
3. Utilization of a less effective picking strategy during busy periods.

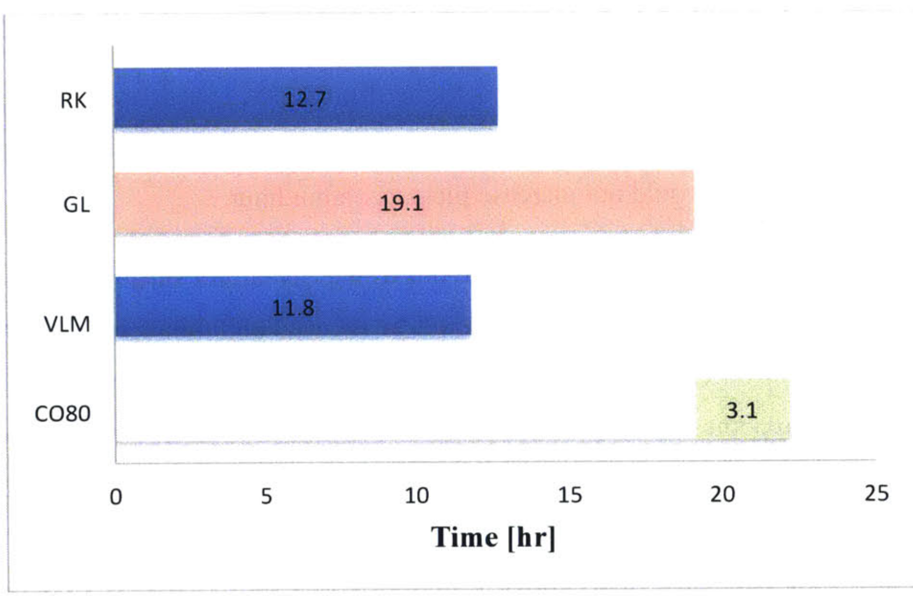


Figure 24: Imbalance in the picking process (3-day cutoff)

To find an effective solution, all three problems have to be tackled in unison. The idea is to reduce the overall picking time at the GL location. This can be done in three possible ways:

1. By arranging the parts already in GL in a more strategic manner (slotting) by utilizing golden pick zone to store the fastest moving parts and by using the higher shelves for FISH and slow moving parts.
2. By allocating more labor resources to the GL area.

3. By transferring GL workload to another area (possibly VLMs).

The first solution will be partially implemented, but by merely putting fast movers into golden picking zones, the number of picks per hour can't be drastically increased. The time it takes for a picker to walk around aisles and manually pick and scan parts is a major component of the total time, which is hard to reduce.

The second solution would require that more workers be delegated to GL picking area to reduce the time required to pick parts. First, there is a limit to the number of workers that can work in GL because of the width of the aisles. Each worker carries a rolling stair and only one rolling stair can fit in a single aisle. During busy periods there are already enough workers dedicated to GL and adding additional workers is going to hamper the progress of others. Second, moving workers from another area to GL would reduce the performance of the area that lost a worker. Third, hiring an additional worker would imply higher labor costs. Fourth, at the moment the GL picking is close to 20 picks per labor hour. Adding additional worker could increase the number of picks per time period but would not increase picks per labor hour.

The third solution seems to be the most effective but with a caveat. By shifting fast moving parts from GL to VLMs, the workload of GL area could be reduced and transferred to VLMs, and a new critical path could be created. The caveat is to come up with solutions that make VLMs more effective. The hypothesis is that since VLMs are drastically underperforming with a pick rate in the range of 20 to 35 picks per hour compared to their expected pick rate of 60, if the picks per hour are improved to an expected level, the VLMs can have more throughput and could sustain the extra work load that will be shifted from GL.

Three major problems that limit the throughput of the VLMs are FISH parts, skewed part distribution, and sub-optimal picking strategies. The overall idea is to swap FISH parts from VLMs with fast moving parts from GL, balance the utilization rate of VLMs by appropriately balancing all five pods, and by coming up with an effective picking strategy.

The rest of the section goes into the specifics of how many parts need to be swapped between VLMs and GL, what type of strategies need to be used to balance the VLM workload distribution, and what type of picking strategies (pick-and-pass or pick-and-consolidate) need to be used during busy periods. VLM workload balancing and picking strategies are covered in greater depth by Racca's thesis [14], and Fong's thesis [13], respectively.

4.1. Low VLM Picks and High GL Picks

To increase the number of picks in VLMs and reduce the number of picks in GL, FISH parts were identified in the VLMs and fast moving parts were identified in GL. FISH, as explained earlier, are the parts that have not been picked over past one year and are not expected to be picked in next six months. In total 16% of the parts (2,052 parts) were identified as FISH. Table 3 (reproduced below as Table 6) gives the percentage of FISH parts based on three different definitions.

Table 6: % FISH parts in VLMs

	Past 6 months	**Past 12 months**	**Past 12 & next 6 months**
# of FISH Parts	4306	2879	2052
Total # of Parts	13223	13223	13223
% FISH	33%	22%	16%

The idea is to fill the empty space that is created by moving FISH parts out with fast movers from GL. The overall target is to increase the number of picks from the VLMs and reduce the number of picks from GL to reduce the workload at GL. But this needs to be done by moving the least number of parts around. Analysis was done to study the right number of parts to move around, for which the frequency distribution breakdowns for GL and VLMs were very helpful (see Figure 16 and 22).

There are several parts in VLMs that are located in more than one bin location. Out of 2,052 FISH parts, 1,612 parts were located in a single location, and the remaining 440 parts were located in more than one location. These 440 parts occupied 1,058 locations in total. Therefore if all 2,052 parts were to be moved out, 2,670 empty spaces will be created in VLMs. Figure 25 shows the number of FISH parts corresponding to the number of locations they are stored in.

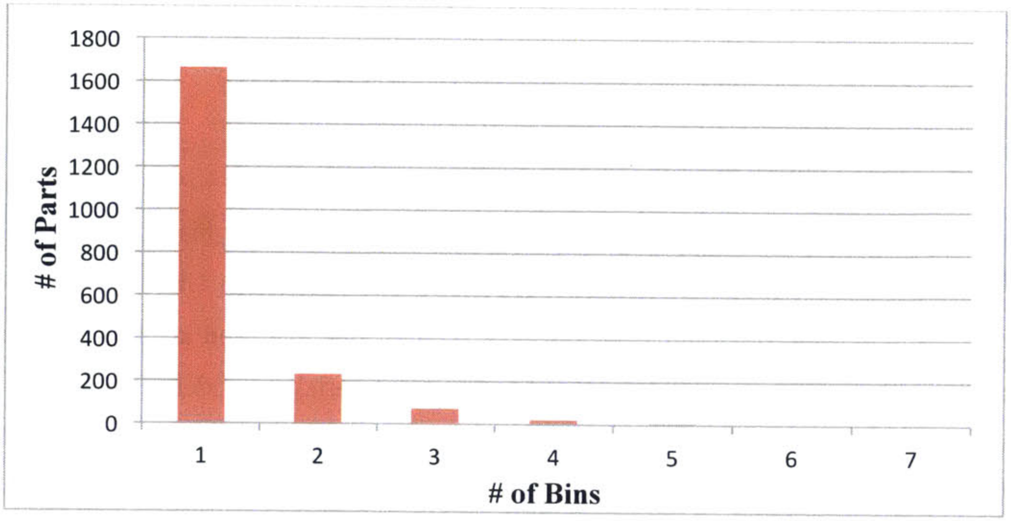


Figure 25: FISH parts and corresponding number of locations

Analysis was done to recognize fast moving parts to move from the GL area. The idea was to move enough fast moving parts from GL to approximately fill 2,670 empty spots that will be created in the VLMs. It was decided that any part in the GL with a pick frequency of more than or equal to 11 would be moved to GL. **Table** 7 shows the pick frequencies along with the corresponding number of GL parts.

Table 7: Pick frequency with corresponding number of parts

Pick Frequency	# of Parts	Cum. # of Parts
>500	9	9
500	4	13
400	21	34
300	58	92
200	179	271
100	23	294
90	34	328
80	46	374
70	51	425
60	75	500
50	118	618
40	138	756
30	239	995
20	37	1032
19	27	1059
18	31	1090
17	31	1121
16	41	1162
15	49	1211
14	47	1258
13	53	1311
12	86	1397
11	74	1471
10	68	1539
9	85	1624
8	135	1759
7	167	1926
6	245	2171
5	209	2380
4	268	2648
3	317	2965
2	430	3395
1	964	4359
0	3556	7915

A cutoff of 11 was chosen based on the corresponding cumulative number of parts (1,471) and based on an estimate of how many GL parts would be needed to fill 2,670 VLM slots. The quantity available of each individual part in GL was in general more than that of parts in VLMs. Also, 11 seemed to be a high enough number to justify the effort of moving these parts to VLMs.

4.2. Skewed Pick Distribution

In the root cause analysis section, non-strategic PACI assignment was found to be an underlying cause of skewed part distribution. As noted in Table 4, due to skewed distribution, Pod 1 acts as a bottleneck.

Ideally, all the pods and their corresponding bays should have equal workload utilization. If there is perfect balance of workload then a total time savings of 40% is possible relative to current levels.

Even though, a perfectly balanced pick distribution would bring about the best results, there is always variation, and therefore a perfect balance is not possible.

To get the best distribution possible, three different methods of allocating PACI assignments i.e. distributing parts amongst pods were studied. These three methods were named random, snake, and order grouping, and are briefly explained below.

Random: In this method all the parts will be randomly assigned to pods. Since, the magnitude of parts in VLMs is in the order of thousands, it is expected that the random assignment would bring about good results.

Snake: In this method pick frequency is used to allocate parts to different VLMs. Parts would first be arranged in descending order of their pick frequency, and then assigned to the pods as shown in Table 8.

Table 8: Snake assignment

Parts (arranged according to decreasing pick freq.)	**PACI Assignment**
A	VL01
B	VL02
C	VL03
D	VL04
E	VL05
F	VL05
G	VL04
H	VL03
I	VL02
J	VL01

Such an assignment is expected to ensure that the picking output of all pods is balanced. The motivation for this type of assignment came from childhood; when picking two teams for soccer from a group of players, the captains would pick players using this strategy. Therefore, the first captain would pick the first best player, the second captain would pick the second and the third best player, then the first captain would pick the fourth and the fifth best player and so on. As a result the captain who gets a chance to pick the player first, doesn't end up gaining an advantage, and the teams were usually balanced.

Order grouping: In this method, parts that are ordered together i.e. parts from a same subassembly are evenly distributed amongst all the five pods. This would ensure that when an order has to be picked for such subassembly, all the five pods end up delivering about the same number of parts.

Analysis was done to find out which PACI assignment strategy is the most effective. Historic data was pulled and the performance of each assignment strategy was studied based on historic orders picked from the VLMs.

Figure 26 shows the relative percent variance in number of picks per pod of the original distribution, random distribution, snake distribution, and order grouping distribution. We observe that all three proposed strategies significantly outperform the current strategy.

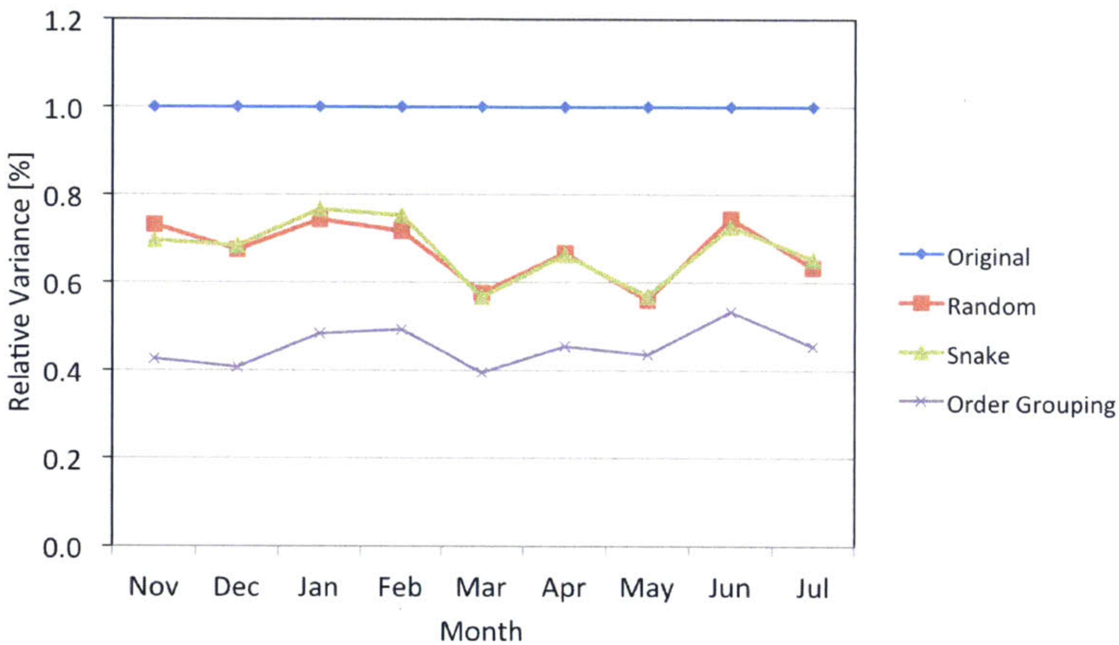


Figure 26: Relative percent variance across pods [14]

Figure 27 shows the relative percent variance in number of picks per bay of the original distribution, random distribution, snake distribution, and order grouping distribution. We again observe that all three proposed strategies significantly outperform the current strategy.

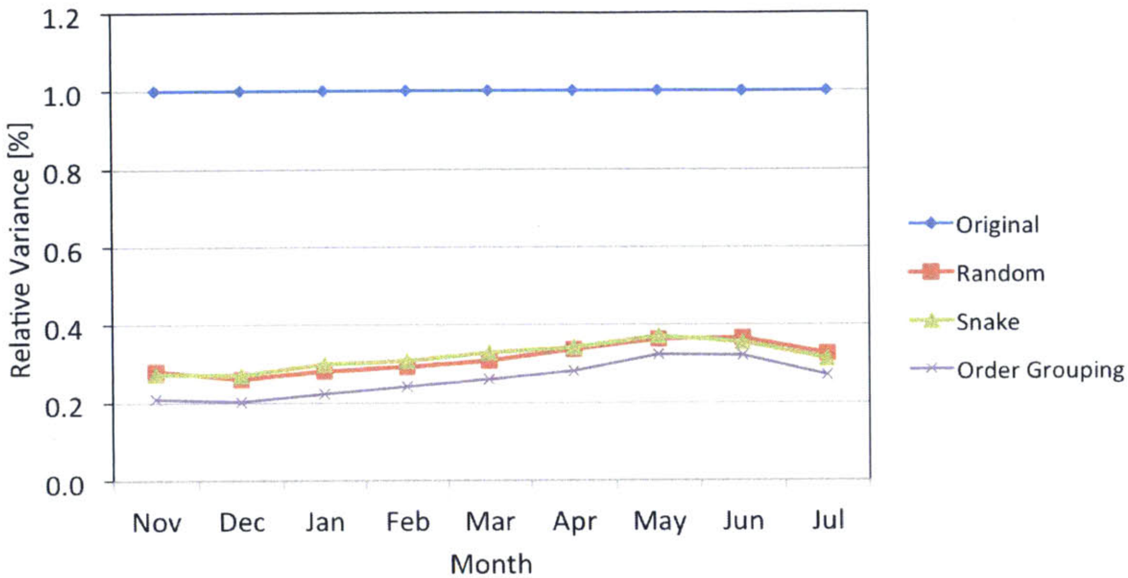


Figure 27: Relative percent variance across bays [14]

If there is an ideally balanced utilization of all the five pods, expected time savings is 40%, but due to variance, the expected time savings is going to be around 35% if any of the random or snake strategies are implemented. Racca's thesis [14] goes in greater depth and explains the performance of the proposed strategies in detail.

4.3. Picking Strategies

On average, the time required to complete picking an order by pick-and-consolidate is shorter than pick-and-pass by 8%. The primary reason is that pick-and-consolidate is a parallel picking process whereas pick-and-pass is a serial process. The results further reveal that despite the busier workload, pick-and-consolidate has a better average picking performance than pick-and-pass. The additional consolidation time for pick-and-pass is estimated to be 15 seconds per tote and it has a minimal effect on the overall time required to complete picking an order from the VLMs. The resultant average pick time per order for pick-and-consolidate is 232 minutes as compared to pick-and-pass, which is 252 minutes. In summary, pick-and-consolidate is the better picking method and should be used during busy periods so as to deliver parts on time from the warehouse to the production floor. Fong's thesis [13] goes in further depth to compare the two strategies.

4.4. Slotting within GL

We found that not all the fast movers from the GL that were identified to be shifted to VLMs could be moved because of size or quantity constraints. These fast movers will be moved to the golden picking locations. Also, the FISH parts that will be moved out of the VLMs will be assigned to higher shelves in the GL. This would make picking in GL faster, more convenient, and safer.

5. Implementation

We discuss here the actual implementation process of moving in GL fast movers to VLMs and moving out VLM FISH parts into GL. The implementation process for balancing the workload of VLMs is discussed in detail in Racca's thesis [14], and the implementation process for utilizing a more effective picking strategy is discussed in detail in Fong's thesis [13]. The implementation process was planned in a manner to make it efficient and less labor intensive.

5.1. Swapping VLM FISH parts with GL Fast Movers

Moving parts out of VLMs and out of GL had to be a parallel process because there are not enough empty bins available in GL to store all VLM FISH parts and vice versa.

A detailed step-by-step flow chart for moving out GL fast movers is shown in Figure 28. The parts that couldn't be moved out of manual picking zones such as vendor-managed inventory were excluded from the analysis. The remaining parts were sorted based on their pick frequencies and 1,471 parts were selected to be moved into the VLMs.

Moving out 1,471 parts manually was an extremely labor intensive task. The warehouse is currently in a stage of a busy period and therefore a method was developed to reduce the workload. These 1,471 parts are considered to be fast movers and if the corresponding bins for these parts are not replenished, then over time the bins will empty. The replenishments for these parts could directly be sent to VLMs by changing the PACI assignment for these parts to VLMs. To get a sense of time duration in which all these 1,471 parts would empty, a bin-depletion metric (equation 1) was created that took into account the current quantity of a certain part in hand and its corresponding rate of usage.

$$Expected\ Bin\ Depletion\ Ratio = \frac{Total\ Available\ Quantity}{Quantity\ Picked\ in\ a\ Given\ Time\ Duration} \qquad (1)$$

The expected bin depletion ratio for all 1,471 parts was calculated and was converted into the number of months it would take to empty these parts. We found that out of 1,471 parts, 771 parts would empty within the next three months and the remaining 701 parts would take more than three months. We decided that the PACI for all 1,471 parts would be assigned to VLMs, and only the 701 parts that would take more than three months to empty would be manually transferred to VLMs.

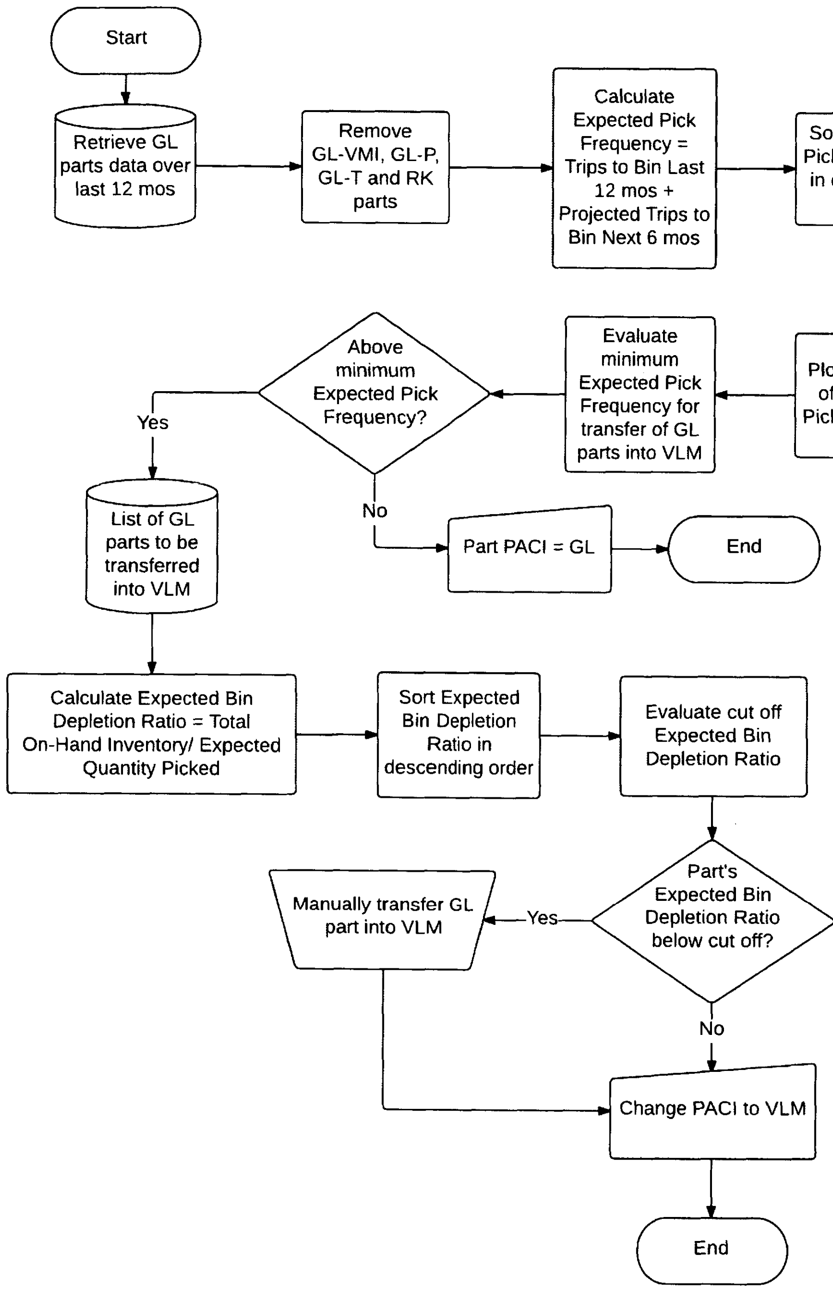


Figure 28: Flow chart for moving fast movers out of GL

Of the selected 1,471 parts it was known that not all parts would fit into VLMs because of size or quantity constraints. Some parts in GL are large and wouldn't fit in VLM bins, while other parts that could fit might have very large quantities and therefore would take up a large number of boxes in VLMs. Also, we needed to know which part would fit in which bin type in VLMs. The volume (cm^3) data was available for all the parts but it was not enough to conclude which parts would fit in VLMs and which wouldn't. Therefore, the team manually inspected all 1,471 parts and made a list of parts that could be moved into VLMs with their corresponding bin sizes.

It was surprising to find out that 736 parts out of 1,471 parts couldn't be moved into VLMs. VLMs had a lot of small bins (BA04, BB04, and BC04) that were empty, therefore the team decided to move 266 small sized GL parts with pick frequencies lower than 11 but more than 5, to fill small sized bins in VLMs. Note that not all GL parts with frequencies between 11 and 5 were to be moved, only small sized GL parts with these frequencies were to be moved. Of these 266 parts, 177 had to be manually moved.

As mentioned earlier, moving out FISH parts and moving in fast movers needed to be a parallel process. A detailed step-by-step flow chart for moving out FISH parts is shown in Figure 29. FISH parts were first identified and then sorted based on a priority list. FISH parts that were stored in multiple locations (440 parts in 1,058 locations) were to be moved out first. By moving parts out with multiple locations first, 1,058 bins in VLMs could be emptied, but on the GL side, only 440 locations will be occupied.

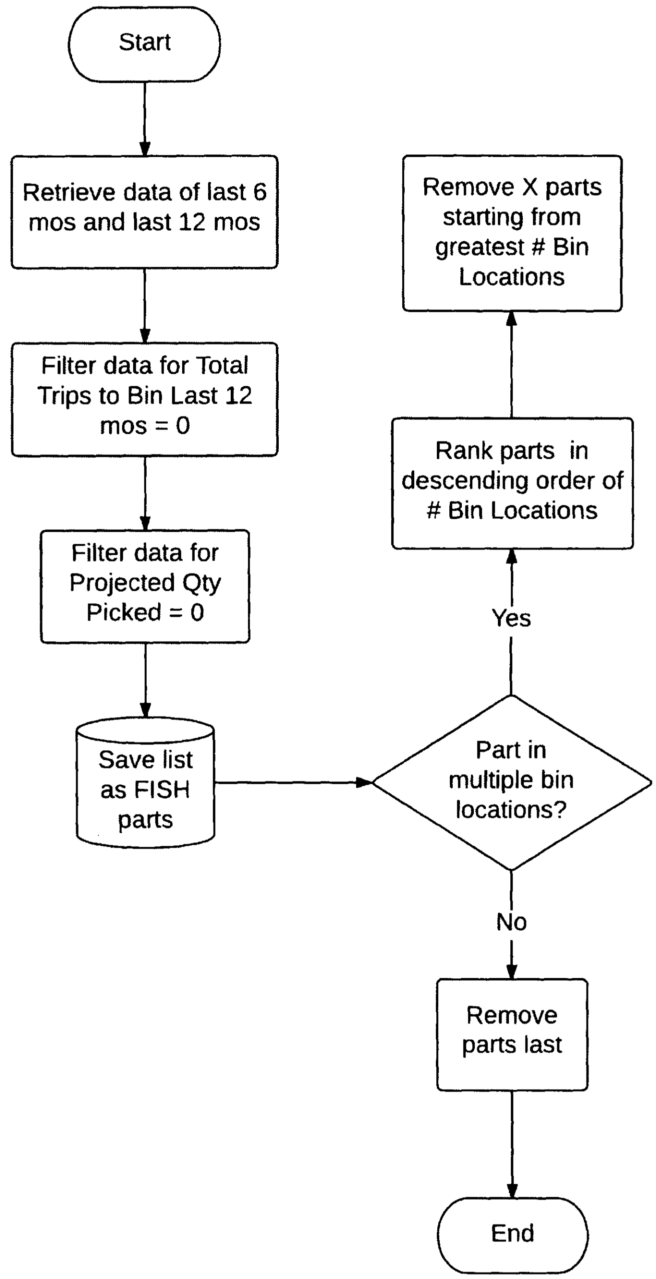


Figure 29: Flow chart for moving FISH parts out of VLMs

VLMs are automated machines that pull out parts automatically when an order is released. But moving out all the FISH parts had to be a manual process, where a part picker had to manually put in part number to retrieve the part. This was impractical; therefore other options were being explored to figure out a way to move these parts out automatically.

It was not practical to stop the implementation process during the time a solution is being found to move FISH parts out. The team decided to start moving in fast moving parts, but before these parts could be moved in, PACI assignments needed to be made to allocate these parts to a specific pod. Care also had to be taken to check if the right number of appropriately sized bins is available in the assigned pod.

Another objective was to balance out the workload distribution of the five pods as much as possible by assigning these fast movers strategically. To accomplish this the assignments were made based on the available bins in the VLMs. There were not many BG04 and BG08 bins available and therefore the projected number of bins that would be emptied by moving FISH parts out was also taken into account. Since, the number of picks coming out of first two pods was already high, no parts were assigned to these two pods.

Table 9 and Figure 30 show the percentage of empty bins available in each pod. We notice that Pod 2 has a large percent of empty bins available; yet no part was assigned to Pod 2.

Table 9: % Empty and filled bins

	Total	**VL01**	**VL02**	**VL03**	**VL04**	**VL05**
Filled Bins	22410	5257	4676	4965	3021	4491
Empty Bins	4172	280	1329	427	156	1980
Total Bins	26582	5537	6005	5392	3177	6471
% Filled	84%	95%	78%	92%	95%	69%
% Empty	16%	5%	22%	8%	5%	31%

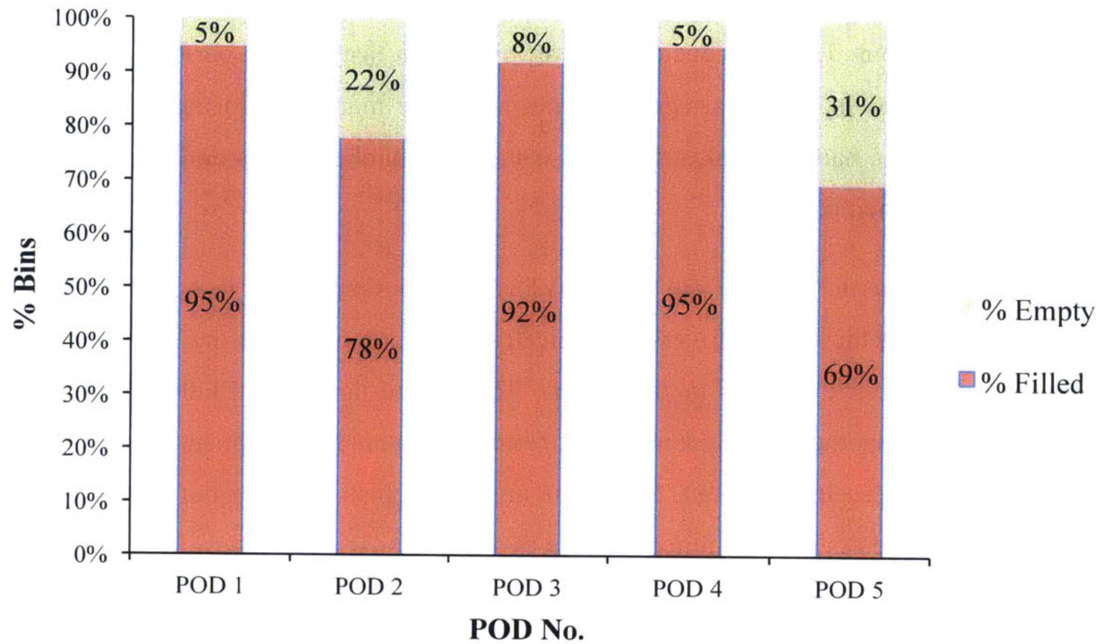


Figure 30: % Empty and filled bins

To assign locations to 735 parts of initial 1,471 parts, the number of empty bins of different sizes was studied to assign PACIs to parts to ensure that the parts find a space in a pod they are assigned to. It can be noticed from Figure 31 that there was less number of bins for BG04 and BG08 bin sizes. Therefore, Figure 32 was studied to find out how many bins will empty after FISH parts are taken out.

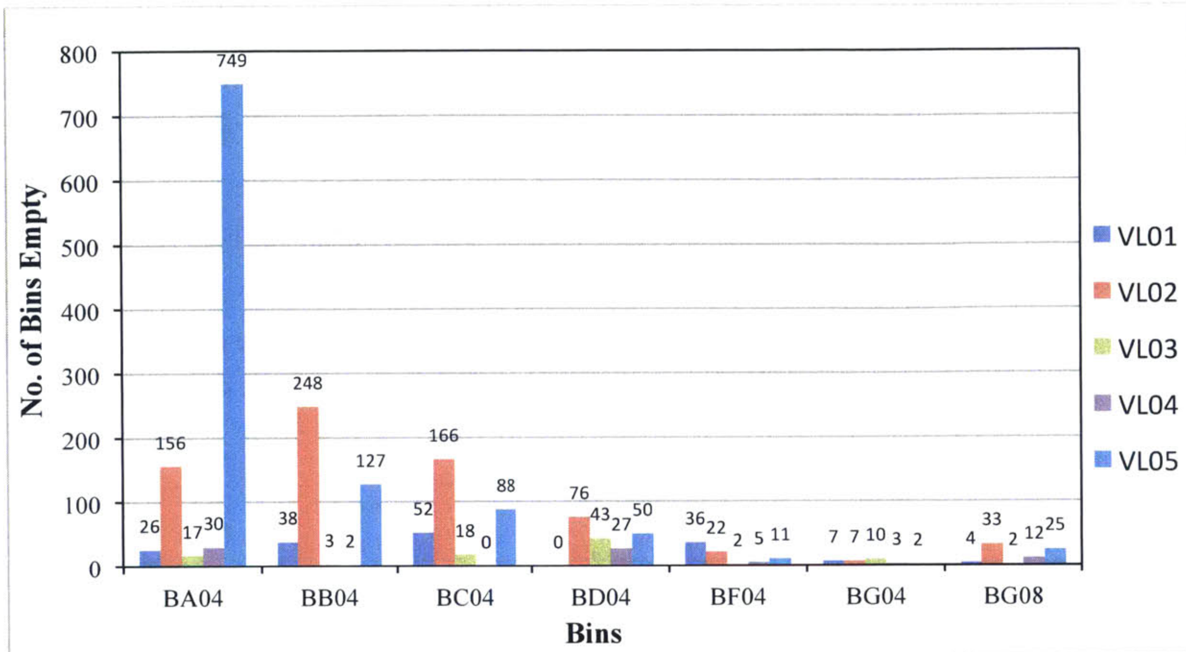


Figure 31: Current number of empty bins for each bin size

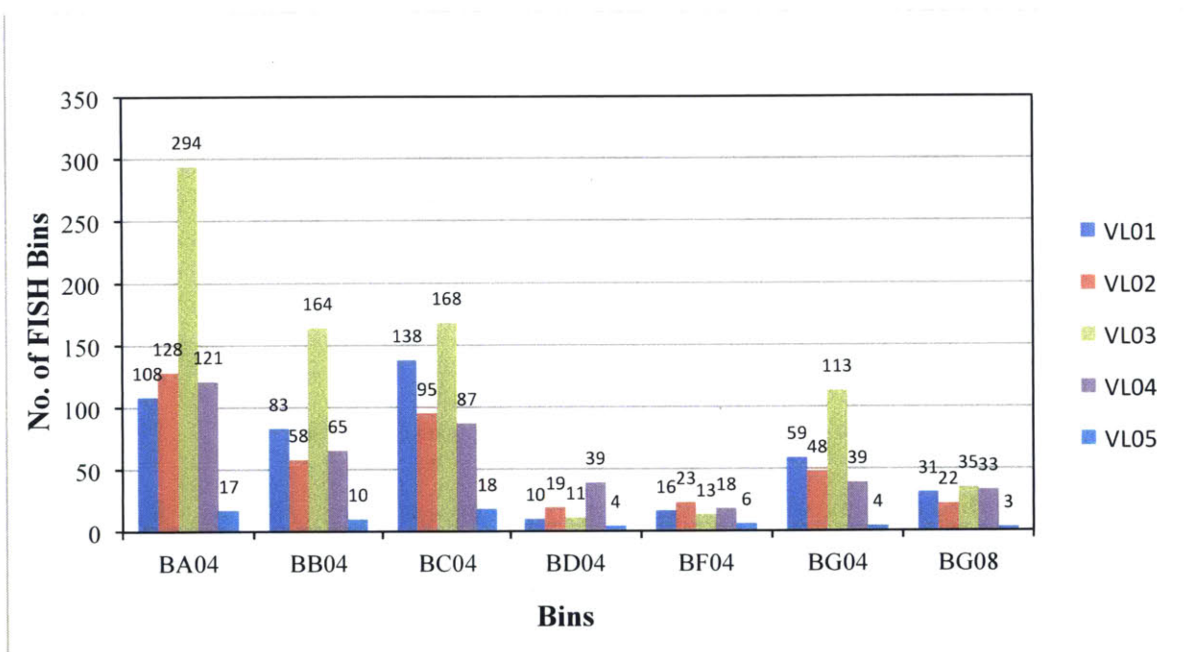


Figure 32: Number of FISH bins that will empty out for each bin size

This allocation ensured the best possible balance that could be achieved by correctly assigning these GL parts to VLMs. Figure 33 shows the improvement in the previously skewed pick distribution. The bars in red are the original distribution and the bars in blue show the change in distribution. The original distribution is based on how many picks were made from each individual pod over the duration of past twelve months. The new addition (blue bars) adds the pick frequencies (over past twelve months) of the GL parts that have now been moved to a specific pod. Therefore, if these GL parts were present in the pod that they are currently assigned to since past twelve months, then the pick distribution would have looked like Figure 33.

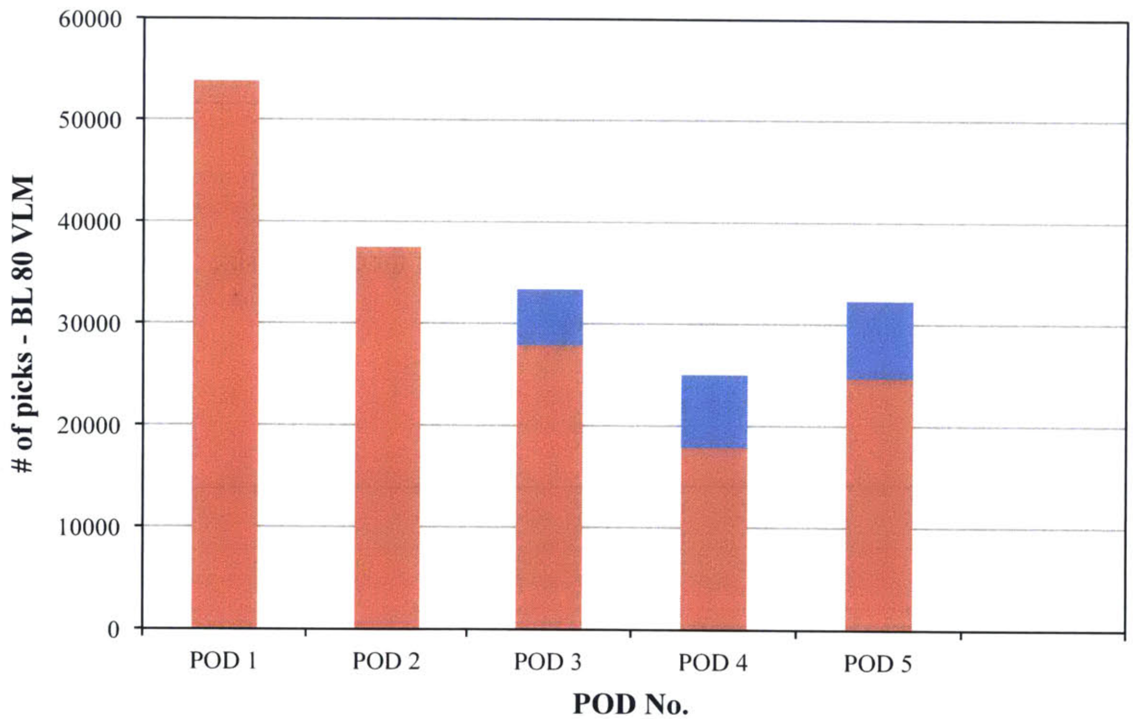


Figure 33: Improvement in skewed number of picks

All the parts that had to be manually moved were moved except the parts that had to be put into BG04 and BG08 sized bins.

In the mean time a solution was found to move the FISH parts out. A new BOM was created with a list of all the 2,052 FISH parts and the BOM was split into several kit codes of 50 to 70 parts each. Each kit code had parts corresponding to a specific bin size. Each day one kit code is released to move FISH parts out; priority was given to move out parts that were located in BG04 and BG08 bin sizes to create space for GL fast movers that had to be moved into these bins. The implementation process is still under way and is expected to be finished by the end of August 2015.

Of the initial list of 1,471 fast moving parts, 736 parts could not be moved in because of size and quantity constraints. To some extent this was seen as a positive because this would prevent driving down GL picks extremely low, and transferring the entire workload to VLMs.

5.2. Skewed Pick Distribution

Based on Figure 26 and 27, all three proposed strategies (random, snake, and order grouping) gave similar results in terms of balancing the workload of the pods and the bays. Random strategy is the easiest and least costly to implement. Snake strategy is a little more expensive to implement, as it requires more coding hours. In snake strategy, new parts will be assigned to the pod with the lowest utilization. Order grouping is a complex strategy, requires many more coding hours and needs continuous updates as BOMs are updated and is therefore difficult to implement.

It is proposed that the snake strategy be implemented despite higher costs as it can resolve the current VLM workload imbalance in a short period of time. This is because new parts are assigned to the pod with the lowest utilization. Racca's thesis [14] goes in further depth and talks about the cost benefit analysis and the implementation process.

5.3. Picking Strategies

On average, the time required to complete picking an order by pick-and-consolidate is shorter than pick-and-pass by 8%. Pick-and-consolidate is a better picking method and should be used during busy periods so as to deliver parts on time from the warehouse to the production floor. Fong's thesis [13] goes in further depth to talk about the implementation process.

5.4. Slotting within GL

Out of the initial list of 1,471 fast moving parts in GL, 736 parts could not be moved to the VLMs either because of size or because of quantity constraints. These 736 parts are being moved to the golden picking zones (A through C), and the FISH parts that are being removed from VLMs are being placed onto the shelves that are only accessible through ladder (D through K).

As of now, the priority is to move out FISH parts and move in fast movers from GL to VLMs. Once these parts are moved, the process of moving GL fast movers into golden picking zones will be accelerated.

6. Conclusion and Future Work

Our work is focused on improving the operational efficiency of the warehouse so as to consistently deliver parts on time from the warehouse 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 shipment is costly to the company, it is critical for the warehouse to meet the target delivery time of 24 hours. However, during the busy periods, the warehouse could not keep up with the target delivery time. The long delivery times are addressed by the simultaneous reduction in workload on GL and improvement in picking efficiency at the VLMs. There are two ways to improve the picking efficiency: i) balancing the workload across the VLMs and ii) using an efficient picking method. Our proposed solutions and their potential benefits are summarized below.

Strategy 1: The workload in GL is reduced by increasing the number of VLM picks. This is accomplished by moving fast moving parts from the GL to the VLMs. The primary motivation for increasing the number of VLM picks is that it is more efficient to pick from the VLMs than GL. Additionally, FISH parts have to be removed from the VLMs in order to make room for fast moving parts.

Benefit: The GL workload is expected to decrease by 25% because fast moving GL parts are moved to the VLMs, and the VLM workload is expected to increase by 13%. This would reduce the imbalance in the picking process pointed out in Figure 13.

Strategy 2: We recommend balancing the workload across the VLMs by using a dynamic slotting technique- the snake frequency balancing method. The dynamic slotting framework is a preventive approach that constantly seeks to minimize the workload imbalance across the pods and bays. This eliminates the need for periodically re-slotting the VLMs, which is labor intensive. Both the snake frequency balancing and randomization methods show comparable results in terms of achieving a balanced workload distribution across the pods and the bays. However, we recommend the snake frequency balancing method. The primary advantage of the snake frequency balancing method is that the current VLM workload imbalance can be resolved in a shorter period of time.

Benefit: 35% time savings can be obtained by using the snake frequency balancing method. Taking into account the 13% increase in VLM picks, the resultant time saving is 23%. This additional time savings will also provide additional time for workers to perform cycle counts and put aways.

Strategy 3: We recommend employing the pick-and-consolidate method during busy periods. Pick-and-consolidate is more efficient than pick-and-pass. Because each pod is working independently of each other for pick-and-consolidate, there is no starvation and blocking. In contrast, because the pickers downstream the picking line has to wait for picks to be completed upstream, there is the problem of starvation and blocking for pick-and-pass.

Benefit: The expected makespan time savings per order for pick-and-consolidate is 20 minutes (or 8%). Furthermore, once the workload across the VLMs is balanced in the near future, an order is expected to be picked from four to five pods, rather than from one to three pods. The time savings gained from using pick-and-consolidate is thus expected to increase further.

Additional recommendations to maintain the operational efficiency of the warehouse and monitor the picking efficiency at the VLMs are also provided below.

1. We propose three new additional performance metrics to effectively evaluate the performance of the warehouse and VLMs. In order to effectively measure the service level of the warehouse, the management should report the percentage of completed orders within the 24 hours goal instead of the average delivery times. The flow time of an order should also be measured to evaluate the picking efficiency of the warehouse. This metric closely relates to how fast an order can be completely picked. Finally, in order to monitor the performance of the VLMs, the workload distributions of the VLMs should be measured as well.
2. In order to maintain a high number of VLM picks, we also recommend reviewing the list of FISH parts and the list of fast moving GL parts every six months. This is a reasonable estimate because the pick frequencies of parts require time to be accumulated. Whereas slower moving parts should be retained in the GL, fast moving parts should be kept in the VLMs. This ensures a higher picking efficiency since picking from VLMs is more efficient than picking from GL.

In our work, we find that about 50% of the fast moving GL parts cannot be moved into the VLMs because they are either too large in size or the lot size is huge. In the latter case, the size of the parts is relatively small and able to fit into the bins of the VLMs. However, because the order quantities of these parts are large, the consequence is that large sized VLM bins or multiple medium sized VLM bins are required to store the parts. A proposed solution is to reduce the lot sizes of these parts. Future work could examine what is the optimal lot size of these parts and also analyze the impact of reducing the lot sizes of these parts on inventory management of the VLMs.

During our analysis, a noticeable amount of part shortages resulted in late deliveries. The ordering methodologies such as the lot sizes and review periods of the inventory in the warehouse should be examined.

The present evaluation, on which slotting method is the most effective in balancing the workload, is based purely on historical data. Future work could focus on evaluating the actual performance of the slotting method once the method is implemented so as to determine the effectiveness of the dynamic slotting method.

While at present the goal of the warehouse is to ensure that orders are delivered on time within 24 hours to the production floor, a possible longer-term goal could be to reduce the target delivery time from 24 hours to 8 hours. By reducing the time it takes to deliver parts from the warehouse to the production floor, the lead-time of the ion implantation tool can be reduced.

This page left blank intentionally.

References

[1] S. Benerjea. (2015, January) Semiconductor stock outlook. [Online]. Available: http://www.zacks.com/commentary/37436/semiconductor-stock-outlook

[2] (2005, May) How to respond to changes in the semiconductor value chain. [Online]. Available: https://www.gartner.com/doc/480913

[3] L. Spencer. (2015, April) Global semiconductor market hit $340 billion in 2014: Gartner. [Online]. Available: http://www.zdnet.com/article/global-semiconductor-market-hit-340-billion-in-2014-gartner/

[4] (2015, April) Worldwide semiconductor foundry market grew 16.1 percent in 2014, according to final results by gartner. [Online]. Available: http://www.gartner.com/newsroom/id/3027717

[5] (2014, April) Worldwide semiconductor manufacturing equipment spending declined 11.5 percent in 2013, according to final results by gartner. [Online]. Available: http://www.gartner.com/newsroom/id/2701117

[6] (2015, April) Gartner says worldwide semiconductor sales expected to reach $354 billion in 2015, a 4 percent increase from 2014. [Online]. Available: http://www.gartner.com/newsroom/id/3034518

[7] (2015, January) Gartner says worldwide semiconductor capital spending to increase 0.8 percent in 2015. [Online]. Available: http://www.gartner.com/ newsroom/id/2961017

[8] A. S. Bhadauria, "Production lead time reduction in a semiconductor capital equipment manufacturing plant through optimized testing protocols," Master's thesis, Massachusetts Institute of Technology, September 2014.

[9] S. Jain, "Assembly lead time reduction in a semiconductor capital equipment plant through improved material kitting," Master's thesis, Massachusetts Institute of Technology, September 2014.

[10] Y. Wu, "Operations improvement in a semiconductor capital equipment manufacturing plant: Component level and assembly level inventory management," Master's thesis, Massachusetts Institute of Technology, September 2012.

[11] R. D. Meller and J. F. Klote, "A throughput model for carousel/vlm pods," IIE Transactions, vol. 36, no. 8, pp. 725–741, August 2010.

[12] B. W. C. Sedore, "Assembly lead time reduction in a semiconductor capital equipment plant through improved constraint based scheduling," Master's thesis, Massachusetts Institute of Technology, September 2014.

[13] G. H. N. Fong, "Improving and maintaining the operational efficiency of a semiconductor equipment manufacturing warehouse," Master's thesis, Massachusetts Institute of Technology, September 2015.

[14] S. D. Racca, "Improving operational efficiency of a semiconductor equipment manufacturing warehouse through effective utilization of vertical lift modules," Master's thesis, Massachusetts Institute of Technology, September 2015.