Lean Manufacturing in a Semiconductor Environment: Production Leveling

by
Nima Subramanian

Bachelor of Engineering in Industrial Engineering, Anna University (1997)
Master of Science in Industrial Engineering, The Ohio State University (1999)

Submitted to the MIT Sloan School of Management
and the Department of Mechanical Engineering
In Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
Master of Science in Mechanical Engineering

In Conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology
June 2007

Signature of Author____
May 11, 2007
MIT Sloan School of Management and Department of Mechanical Engineering

Certified by____
Stephen Graves, Thesis Advisor
Abraham J. Siegel Professor of Management, MIT Sloan School of Management

Certified by____
Duane Boning, Thesis Advisor
Associate Dept. Head, Professor of Electrical Engineering & Computer Science

Certified by____
David E. Hardt, Thesis Reader
Professor of Mechanical Engineering

Accepted by____
Debbie Berechman
Executive Director, MIT-Sloan MBA Program

Accepted by____
Lallit Anand
Chairman, Graduate Committee, Department of Mechanical Engineering

This page is intentionally left blank.

Lean Manufacturing in a Semiconductor Environment: Production Leveling

by

Nima Subramanian

Submitted to the Sloan School of Management and the Department of Mechanical Engineering on May 11, 2007 in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration

and

Master of Science in Mechanical Engineering

Abstract

Intel Corporation's Fab17 located at Hudson, MA underwent a large scale manufacturing ramp-up, increasing its production volume by over 50%. As a result of this manufacturing ramp-up, the factory is faced with various capacity issues. These capacity issues along with current work-in-progress inventory (WIP) management strategies lead to an unbalanced inventory flow within the factory. The unbalanced WIP flow results in wafers accumulating in front of certain operations/areas. This WIP accumulation or "WIP bubbles" creates unexpected demand for the various resources on the shop floor, putting an undue strain on them. This strain is felt the most in the bottleneck area. The objective of this project is to develop a sustainable solution methodology to alleviate the strain on the bottleneck.

The scope of this project falls under Fab17's lean manufacturing organization, known as the manufacturing excellence (mX) group, and, the analysis used in this internship utilizes lean manufacturing concepts and principles. The solution methodology analyzes the wafer fabrication process in layers rather than in segments. This approach clarifies WIP movement and identifies problem areas that cause WIP bubbles. Further, the thesis applies the concept of production leveling to wafer fabrication in order to alleviate (and eliminate) the pressure on the bottleneck.

Thesis Advisors:

Stephen C. Graves
Abraham Siegel Professor of Management, Sloan School of Management

Duane Boning
Associate Dept. Head, Professor of Electrical Engineering & Computer Science

This page is intentionally left blank.

ACKNOWLEDGEMENTS

I would like to thank the Leaders for Manufacturing Program and Intel Corporation for providing the resources and support to make this project possible. It was a great learning experience that has had a decisive impact in my professional growth.

I would like to thank my project advisor, Melinda Manente, whose collaboration and enthusiasm made this project successful. Thank you to my thesis advisors, Steve Graves, Duane Boning, and Dave Hardt for making this thesis possible through their advice and guidance throughout the process. I would like to thank Bob Olson, Paul Cyr, Paul Cedrone, Gretchen Johnson, Kaine Gill, Eli Mirabal, and Charlie Vieth at Intel Corporation. All of them made a direct impact on this project by offering their valuable time, insights and knowledge.

I would like to thank my husband, Sanjay Subramanian, for his unconditional support to my career; and for encouraging me to apply to MIT, and my son Ashwin Sanjay for making me smile, no matter how hard the day.

Finally, I would like to dedicate this thesis to my parents, Savithri and T.S. Subramanian. They have been great role models for me and have encouraged me to follow my dreams.

This page is intentionally left blank.

Table of Contents

Table of Figures

Chapter 1: Introduction

This thesis is the culmination of a six-month internship at Intel Corporation's Fab 17 (F17) located at Hudson, Massachusetts. The Manufacturing Excellence Group (mX) at F17 sponsored this project as part of an ongoing effort to improve operations.

1.1 Project Overview

F17 is in the midst of a massive manufacturing ramp-up, increasing production volumes by almost 50% within a short period of 30 weeks. In April 2005, Intel made an investment of $17 M in order to make this ramp-up a reality. The upper management at F17 adopted such an aggressive ramp-up strategy due to an increased demand for the chipsets manufactured at F17. This increased business has provided a tremendous growth opportunity for F17; however, as with any manufacturing ramp-up, F17 is facing numerous operational and personnel challenges. During the planning stages, the ramp-up team identified some of the key challenges, which can be broken into four main categories (details in Figure 1):

1. Equipment and Facility related
2. Personnel / Manufacturing Workforce
3. Materials
4. Methods and processes

All the above-mentioned challenges are important for the manufacturing ramp-up to be successful. However, one of the most critical issues faced by F17 is that of limited equipment capacity in areas with constrained factory layout (no room for adding equipment). Material flow issues caused by unscheduled machine downtime further exacerbate this capacity issue. In an effort to solve this problem, the mX management team at Fab17 created a project proposal in October 2005 and sought LFM contributions as part of an off-cycle (spring and summer 2006) internship.

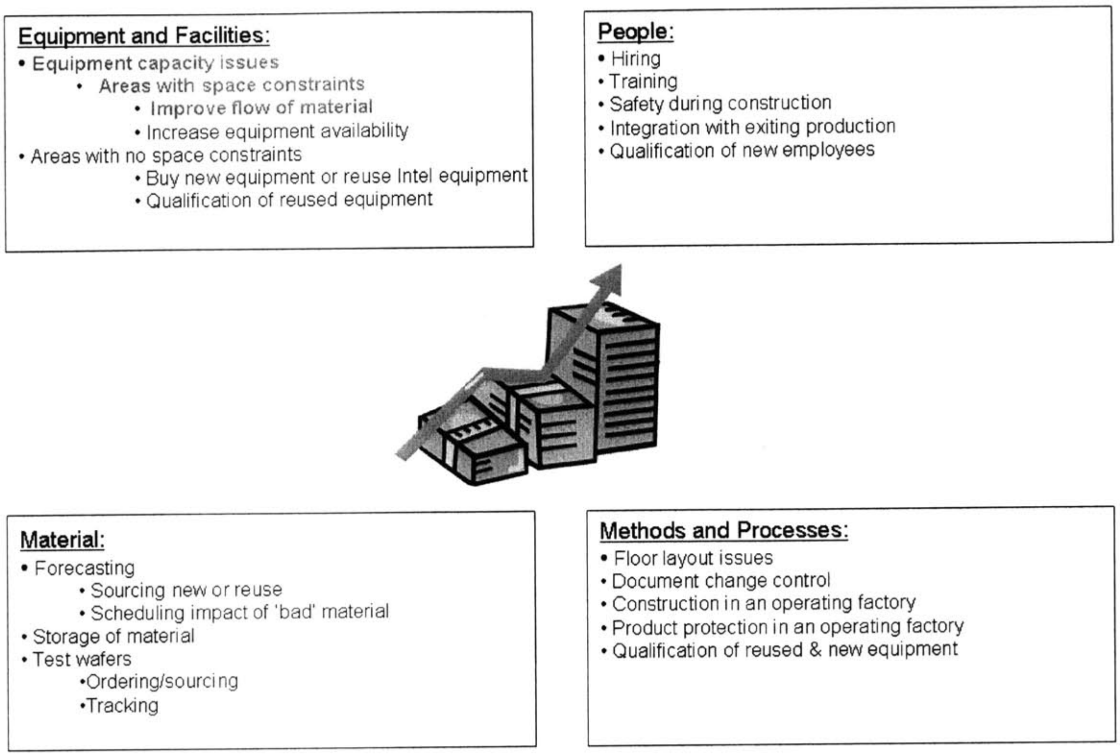


Figure 1: Challenges of Manufacturing Ramp-up at Fab17

1.2 Problem Statement

The following problem statement summarizes the focused improvement efforts this thesis would address:

- Fab17 is faced with capacity issues as a result of 50% increase in production volume
- Unbalanced WIP flow results in WIP bubbles that adversely affect equipment uptime and in turn equipment availability.
- WIP bubbles contribute to variability in finished product output.
- The current incentive system reinforces behaviors that create uneven distribution of WIP.

1.3 Thesis Overview

The remainder of this thesis covers the following topics:

Chapter 2 provides an overview of Intel Corporation and discusses some new trends in the semiconductor manufacturing industry.

Chapter 3 provides a history of F17, explains the existing organizational structure and outlines how the mX organization evolved. It also discusses the semiconductor manufacturing process and current WIP management strategy at Fab17.

Chapter 4 outlines how production leveling could be used at Fab17 to help with their capacity issues.

Chapter 5 describes the board game called New and Improved Manufacturing Approach (NIMA) that helped illustrate the benefits of production leveling to the manufacturing management team at Fab17.

Chapter 6 speaks about the Kaizen event that helped create run rules to achieve production leveling in the ion implantation area at Fab17.

Chapter 7 reviews the findings and provides additional recommendations for F17.

Chapter 2: Intel and Industry Background

This chapter provides a background of the Intel Corporation together with a brief overview of the semiconductor industry and Intel's position in this industry.

2.1 Intel Corporation

Founded in 1968 by two scientists, Gordon E. Moore and Robert Noyce, Intel Corporation (**INT**egrated **EL**ectronics) initially sought to design and build memory products. In 1971, Intel introduced the world's first microprocessor and since then it has continually produced leading edge microprocessors. Some of the most recognized brands of Intel Corporation's products include Pentium, Celeron, and Centrino. In 1965, Gordon Moore predicted the future of microprocessor technology. His prediction, known as *Moore's Law*, states that, "The number of transistors on a chip doubles about every two years" (Figure 2). Such an exponential increase in the number of transistors per chip also indicates an exponential decrease in cost. A focus on the ability to manufacture processors and chipsets at a lower cost, along with Intel's first to market strategy (Holly, 2006), have helped make Intel a Fortune 100 company.

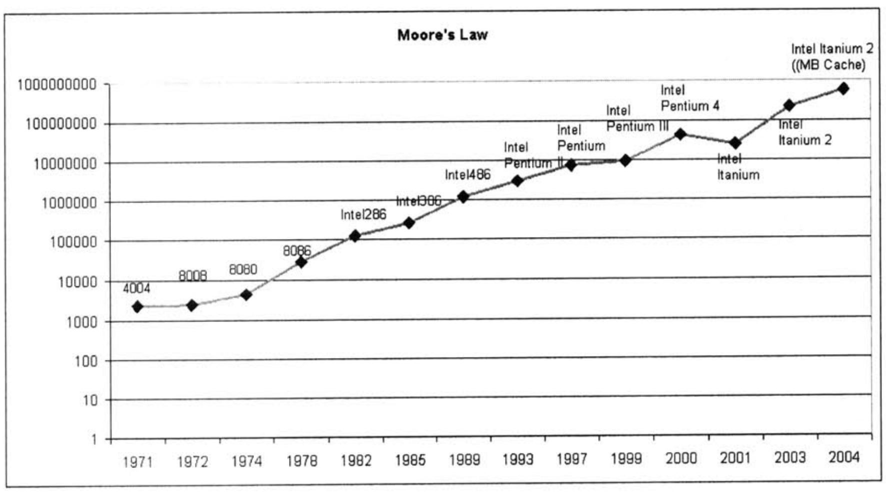


Figure 2: Moore's Law

Intel operates 11 fabrication facilities (Fabs) worldwide and has assembly/test facilities in six different countries. Various Fabs are moving towards a 300mm wafer configuration from a 200mm configuration due to the cost effectiveness of manufacturing 300mm wafers.

"This technology (300mm) greatly improves our capital efficiency by giving us more than twice the capacity at significantly lower costs. Additionally, by reusing an existing 200-mm factory we save additional capital and take advantage of the highly skilled workforce we already have in place" - Robert J. Baker, Senior Vice President, GM TMG, Intel Corporation (http://www.intel.com/pressroom/archive/releases/20030218corp_a.htm)

Further, due to the increased surface area (225%), the 300 mm wafer has a lower production cost per chip. In addition, the 300 mm facility utilizes 40% less energy and water per chip as compared to a 200mm fabrication facility. Given the advantages of converting to a 300mm Fab, it is imminent that in order to survive, conversion to a 300mm facility is critical for a fab (Fearing 2006).

2.2 New Trends in the Semiconductor Industry

Firms such as Intel, AMD, and IBM have dominated the semiconductor industry over the past decade. Such big firms that manage the design, manufacture, and marketing of their semiconductor devices are know as *Integrated Firms*. However, in recent years, the semiconductor industry has seen the emergence of "Fabless" and "Foundry" firms. Foundry firms are contract manufacturers for semiconductor companies, who retain the design and marketing functions and are hence "Fabless." Fabless firms have two choices when choosing a location to outsource their manufacturing: 1) A *pure-play foundry* that specialize only in manufacturing semiconductors designed by others, and 2) Integrated firms offering excess capacity for use. The later choice is less desirable by companies seeking outsourcing services since they would not like to share their designs with competitors, and due to the fear that the services would not be dedicated to them at all times (especially in time of need). Further, the foundry firms offer smaller firms with an advantage of quicker time to market, which is critical in a constantly evolving semiconductor industry.

It is widely believed that, in the future, foundry-fabless partnerships will dominate the industry, except for companies such as Intel and IBM, who rely on their superior process technology to maintain their leadership position in the industry. (R, C. Leachman et al., http://e-conomy.berkeley.edu/conferences/9-2000/EC-conference2000_papers/leachman.pdf).

Chapter 3: Semiconductor Manufacturing at Fab17

3.1 History of Fab 17

Intel's F17 at Hudson, MA was originally a part of Digital Equipment Corporation (DEC) and manufactured its Alpha microprocessors. DEC invested $450M to build a state of the art semiconductor manufacturing facility in 1994 at Hudson, MA. This facility won *Semiconductor International's* "Top Fab of the Year" award in 1997. In October 1997, Intel Corporation purchased DEC for $700M (Zhou, 2000, and F17 Communication Department).

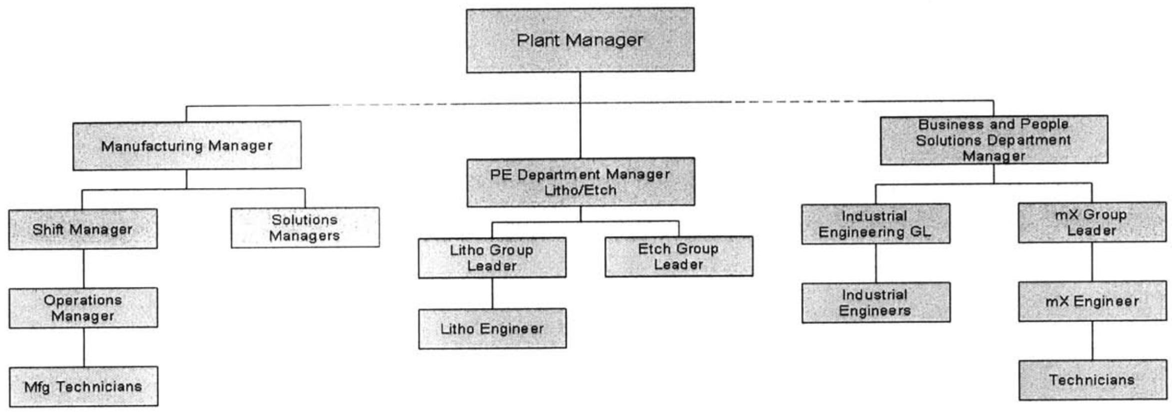
Since 2003, F17's product mix has started to shift increasingly towards manufacturing chipsets that support processors rather than solely manufacturing high-margin microprocessors. Because of the manufacturing ramp-up mentioned in Chapter 1, F17 is on course to increase its chipset manufacturing to make up 95% of its production volume.

F17 includes the 130nm process technology and the 200 mm wafer size. F17 produces two components of the mobile chipset used in Intel Centrino Duo Mobile Technology: Calistoga and ICH7. In addition, F17 produces Banias for Intel Pentium M and Intel Celeron M processors and Northwood for Intel Pentium 4 and Intel Pentium 4 HT processors. Further, F17 produces the Goldbridge and ISB2 for Xeon servers (Reference: F17 Communications Department). However, due to building constraints it is very unlikely that F17 would be converting to a 300mm facility anytime in the next few years.

3.2 Organization Structure at Fab 17

The head of operations at F17 is the Plant Manager. The factory staff report directly to the Plant Manager. The factory staff consists of Department Managers (DM) for each of the functional areas of the factory. For example, there is a department manager each for the Lithography and Implant Process Engineering groups. Group Leaders (GL) report to each of the Department Managers. Group Leaders are responsible for a group of individual contributors such as Process Engineers. Figure 3 gives an outline of the organization structure, separating the manufacturing organization from the rest since the titles used in the manufacturing organization can be different from the rest of the factory.

For example, DM for manufacturing organization is the Manufacturing Manager (MM), and Shift Managers (SM) report to the MM and are responsible for the operations of an entire shift. Operations Managers (OM) report to the SMs and coordinate the day-to-day production activities.



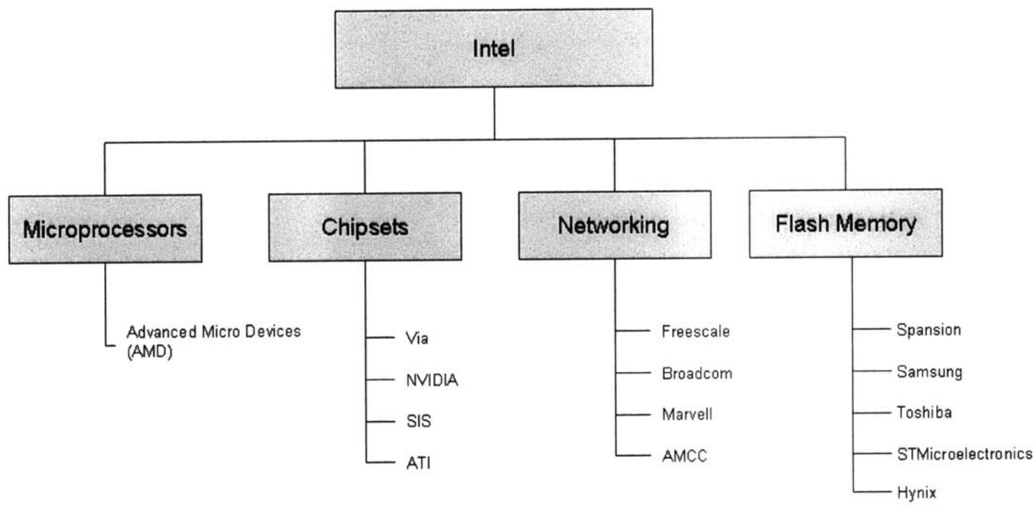
The dashed lines are used to indicate that the DMs are not restricted to the three shown on the figure. This is an incomplete organization chart used to give an outline of the organization.

Figure 3: Organizational Structure at Fab 17

Safety, Quality, Output, and Teamwork (SQOT) are the main manufacturing objectives of F17. To support these objectives, F17 has various cross-functional teams. The Group Leader-Operations Manager forum or GLOM is one such team that discusses tactical issues related to the F17 objectives. Module Teams (MOD Team) consisting of Process Engineers, Operations Managers, and Manufacturing Technicians report to the GLOM and support issues related to WIP flow, Equipment (will be interchangeably used with 'Tool' though out this thesis) Performance, and overall area operations (Holly, 2006).

3.3 Reason for Lean Manufacturing

Intel has been at the top of the semiconductor industry for over a decade. However, in the last few years it has seen increased competition from various domestic and international companies (Figure 4) and has lost some of its customers to its competitors due to cost benefits offered by them ("Not Paranoid Enough," *The Economist*, May 25th 2006). To regain this lost ground, it is critical for Intel to improve its operational efficiencies and reduce costs.



en.wikipedia.org/wiki/Intel

Figure 4: Competitors to Intel

This burden to improve operations is even heavier on F17 since it will not be converted to a 300m fab in the near future. In mid-2003, cognizant of this need to improve operations, F17 decided to implement lean manufacturing principles in the Fab. Due to the possible stigma attached to the word 'lean' (often interpreted as head count reduction), Intel chose the name manufacturing eXcellence (mX) for their lean manufacturing group. The main objective of this group is summarized as the 'mX House' (Figure 5), which has been adapted from the TPS House, developed by Taiichi Ohno and Eiji Toyoda. The mX House is a medium to get from the "Current Reality" to the "Ideal State."

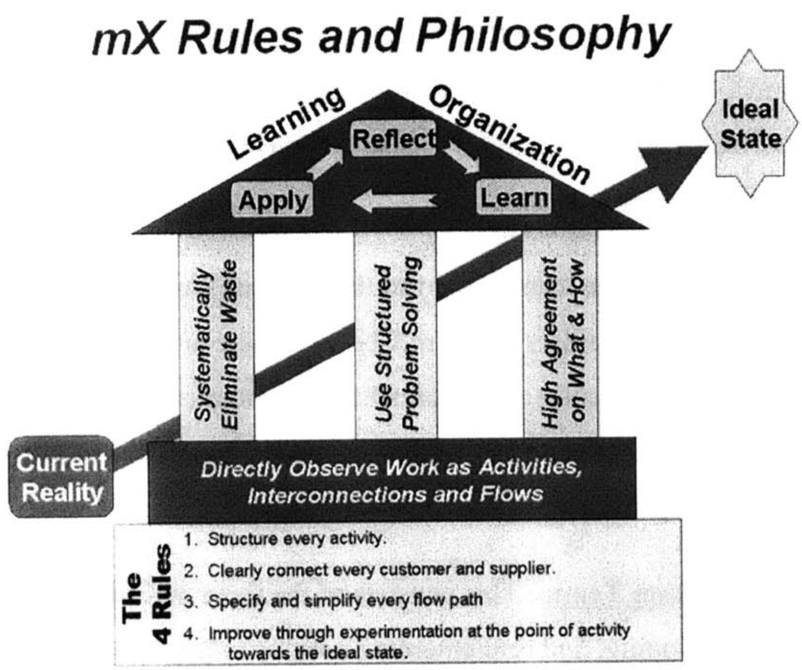


Figure 5: mX Lean Manufacturing House

At the bedrock of the mX House lies "The 4 Rules" that form the basis for the decision making process at F17. Over the bedrock lies the foundation that is at the core of all work performed, and the pillars are the operating philosophies supporting the roof of learning.

3.4 mX Organization

The mX organization can be broken down into three main sub-groups based on their function: the Steering Committee, Pillar Teams, and the mX Implementation Team (Figure 6).

Steering Committee: The Plant Manager and his staff meet weekly to discuss the status of various mX initiatives. Further, this committee serves as a forum to discuss future initiatives. The Steering Committee is also responsible for providing a strategic direction to various mX projects.

mX Pillar Teams: These teams report to the Steering Committee and were created based on the principles of the pillars of the mX house. There are five pillar teams in total, with each focusing on a particular aspect of mX: Autonomous Manufacturing (AM), Planned Maintenance (PM), Focused and Continuous Improvement (FI & CI), Waste Elimination (WE), and Training, Education and Communication.

mX Implementation Team: This consists of the three mX engineers who are involved in planning, developing, and implementing various mX solutions on the floor. The mX implementation team aligns itself with the GLOMs in order to leverage the cross-functional resources of the GLOM. The team facilitates initial activities for the mX teams and trains the people involved in the project. The goal of the implementation team is to provide mX solutions in the various areas in the Fab and later empower the various teams to create and implement similar solutions in other areas they work in.

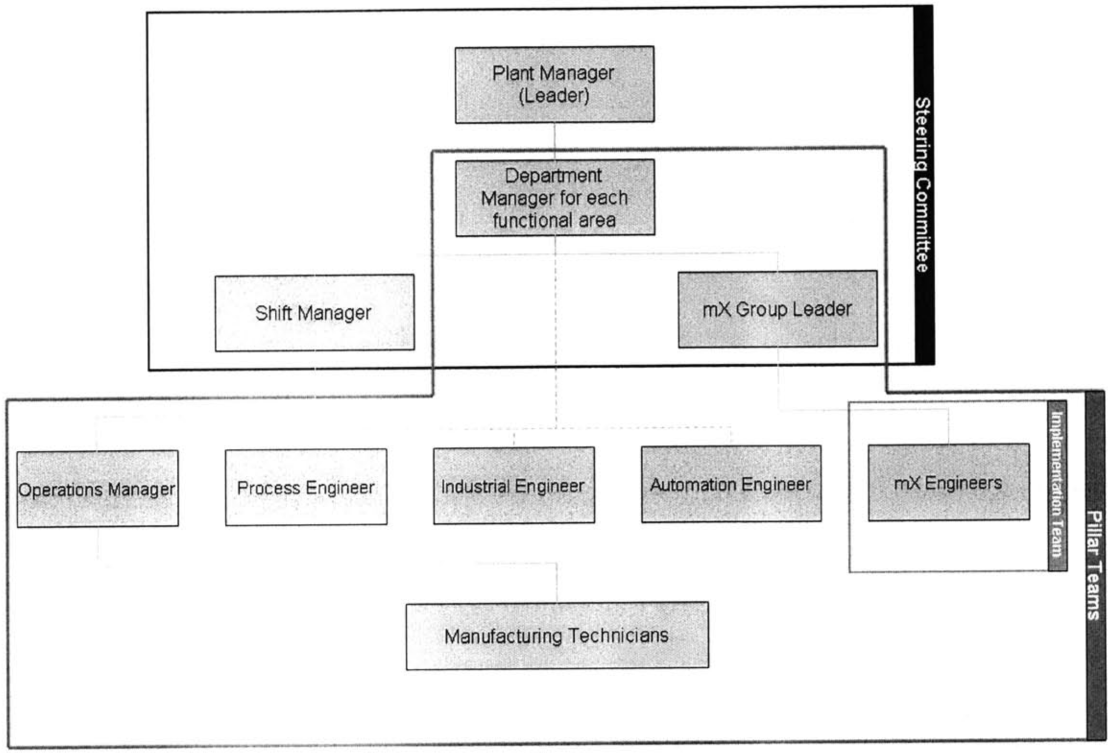


Figure 6: mX Organization Based on Function

3.5 An Overview of the Semiconductor Manufacturing Process

A semiconductor chip is a miniaturized electronic circuit manufactured on the surface of a thin substrate of semiconductor material. The semiconductor manufacturing process is described in Figure 7 (adapted from N. Zhou, 2000). The manufacturing process is described in detail in Appendix A.

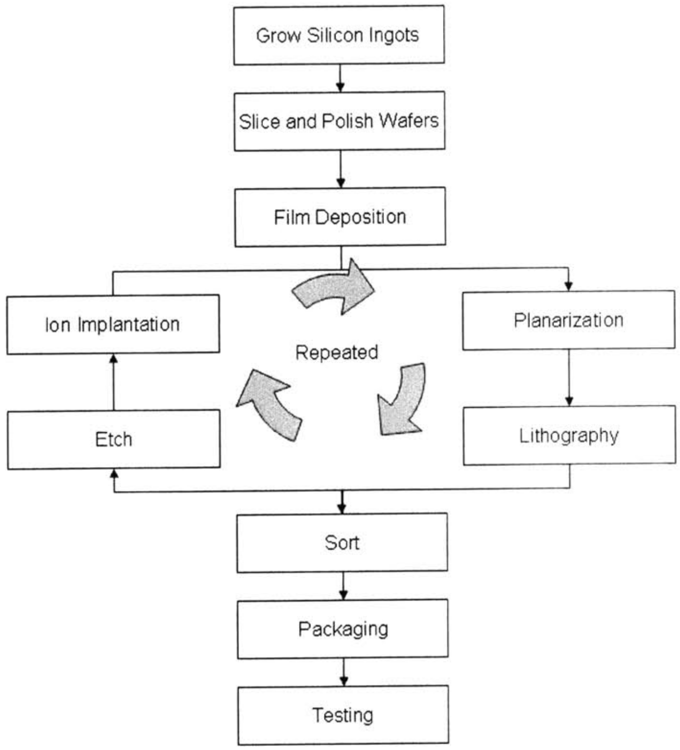


Figure 7: Semiconductor Fabrication Process

Semiconductor manufacturing/fabrication is a highly reentrant process with wafers returning to the same tools multiple times for various operations. Such a reentrant nature of the process means that at any given time (in steady state) when tools become available they can choose from wafers at various stages of manufacture. Hence, the operations employees need to prioritize the various lots of wafers (25 wafers per lot at Intel) on the tools. In addition to the priority decisions made at the technician level, the shift managers set operational goals (Connally, 2005).

3.6 WIP Management at Fab17

Fab17 employs a "Push" system with regard to their WIP (work in progress inventory) management. Based on the forecasted customer demand (Fab out schedule), the manufacturing management determines a production number known as "Wafer Starts per Week" (WSW). Wafers started in a particular week should be ready for customer shipment at the end of the theoretical fabrication cycle-time. For example, if the customer demand is equivalent of 5000 wafers in 10 weeks from now, and the fabrication cycle time is 8 weeks, than the factory will start 5000 wafers two weeks from now. Irrespective of amount of WIP in the system, the wafers are started according to the WSW plan. This plan to push wafers can cause a "WIP bubbles" to form in front of areas where tools are down for unscheduled maintenance. "WIP bubbles" refer to the accumulation of WIP in one area, either in front of a process or an operation. Such an accumulation occurs if a downstream process is unable to handle the volume of partially finished wafers pushed to it due to capacity issues or downtime (both scheduled and unscheduled).

To manage the level of WIP in the Fab, the Fab is divided into week-long segments. For instance, if the target cycle time for a wafer is 8 weeks, then the line is divided into 8 segments and each segment consists of the operations a wafer needs to complete within one week in order to be on target. Once the segments are established, the OMs and SMs continually "goal" the line by assigning priorities to segments, and setting targets for the number of wafers each operation should process in a shift. This "goaling" of the line is done at the start of each shift. An important step in goaling the line is to assign a drumbeat to each segment of the line. Drumbeats represent a goal for how many wafers per shift each operation in a segment should process. Wafers are primarily managed on a gross wafer basis throughout most of the line. The drumbeats and operation specific goals communicate how many wafers should move through an operation in one shift, but do not communicate how many of what product to process (Connally, 2005).

3.7 Customer-Supplier-Customer Relationships

The reentrant nature of the semiconductor manufacturing process makes the various areas in the fab both a customer and a supplier to each other. For example, a wafer that leaves the photolithography (Litho) area after completing a layer will go the Ion Implantation (Implant) area to be doped. However, after it leaves Implant the wafer would complete various steps and come back to Litho for one more layer and so on. Therefore, implant is both a customer and a supplier to the litho area/process. Further, litho is also a customer and a supplier to implant. Such a "customer-supplier-customer" relationship poses various challenges to effective WIP management. First, if a WIP bubble originates at Litho, it would leave Litho but can resurface, at Litho later. Second, when the issue does resurface the cause may not be easily apparent due to long lead-time of the process. For example, if the lead-time of the entire process is 10 weeks, and the time to finish a particular layer is 1 week, then the wafer would come back for the next layer to Litho only after a week (assuming there is no additional delay, in which case the time may be even longer). In this week, many lots would have passed through litho making it difficult to attribute cause.

Analyzing the WIP data over an extended period makes this customer-supplier-customer relationship between Litho and Implant evident (Figure 8 and Figure 9). The WIP data for both Litho and Implant was collected for 39 workweeks and normalized with respect to the mean. For example, if the mean WIP was 100 ($\bar{x}$), and the WIP for workweek 'a' was 165 (x_a), then the normalized value of WIP (x_n) using formula below is .65.

$$x_x = (x_a - \bar{x}) / \bar{x}$$

Normalized data is required in order to maintain data confidentiality and further to remove any bias from the analysis. This data is plotted to get a feel for the WIP movement (Figure 8 and Figure 9). Upon comparing the circled areas in the charts, it is clear that Litho has a tremendous impact on the way WIP moves through the fab. Further, one can make an argument that the effects of the WIP are amplified as it moves from Litho to Implant. The upward trend of the WIP line indicates an ever-increasing amount of WIP, which should be a cause for concern as it could result in unbounded queues if not checked.

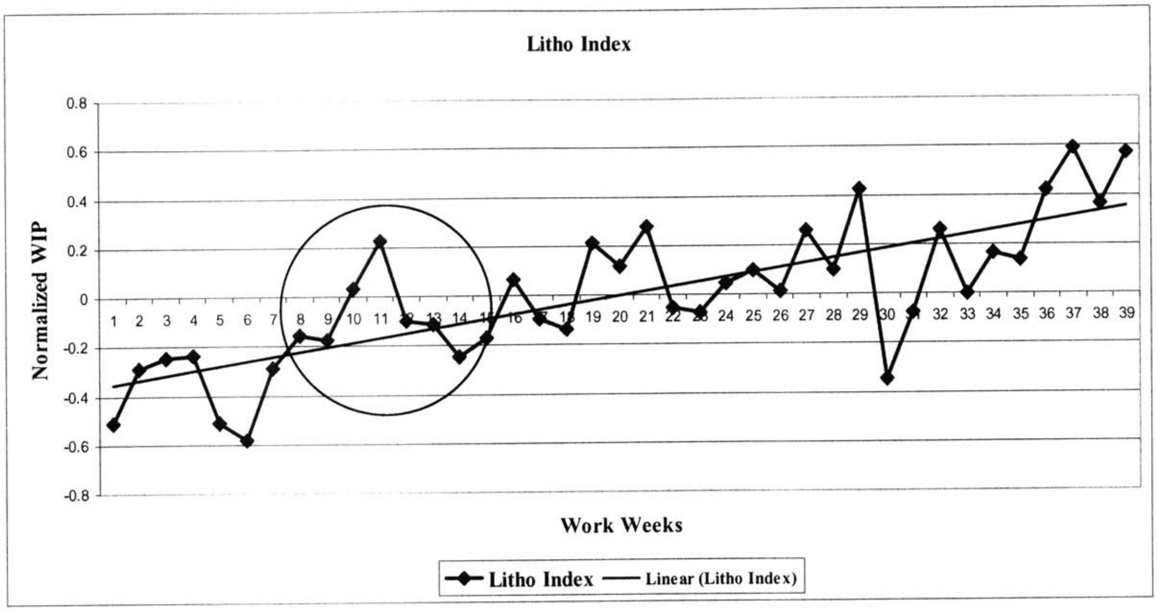


Figure 8: WIP Flow in the Lithography Area at Fab 17

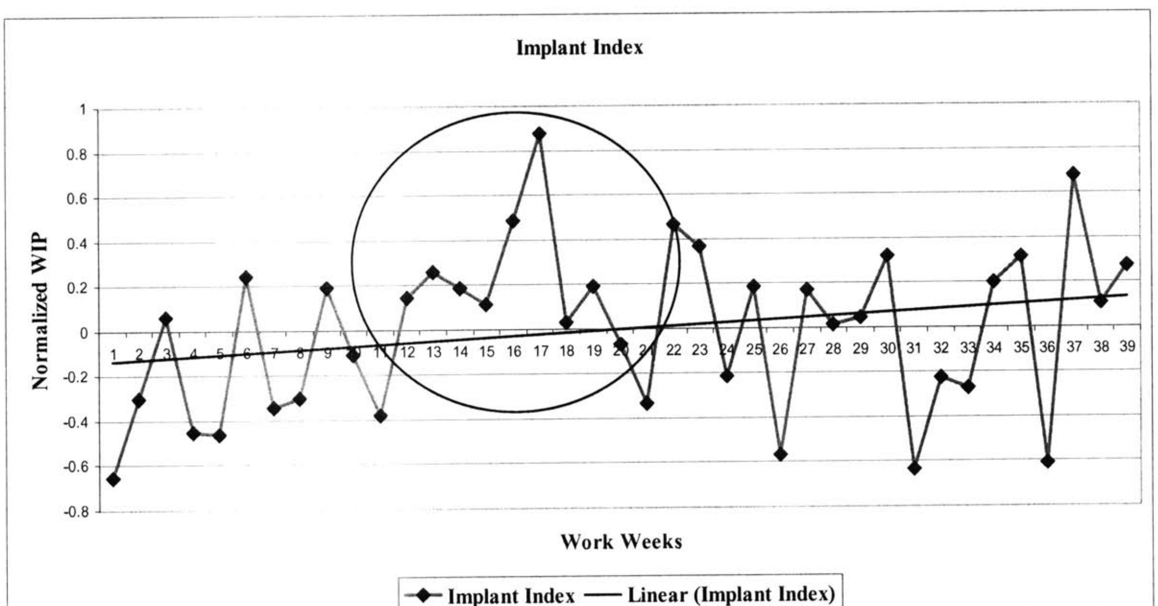


Figure 9: WIP Flow in the Ion Implantation Area at Fab 17

3.8. Capacity Issue at Fab17

As discussed earlier, one of the primary issues facing Fab17 is the shortage in capacity due to the manufacturing ramp-up. The Fab management has tried to alleviate this capacity issue by adding equipment/tools and hence increasing capacity. This approach has been unsuccessful in the Lithography area where even the addition of new equipment has not alleviated the capacity shortage. Even if the management was inclined to add more equipment, there is no room available in the Lithography area to add any new equipment. Further, within the Lithography area, Backend Lithography (BE Litho) is the most severely constrained area making this the bottleneck. Adding equipment is the long-term solution that will solve this capacity issue. However, in the meantime, it is essential to ensure that there are no unanticipated demands made on the bottleneck. Such unpredictable demands may be created due to one of the following:

1. Uneven production volume with sudden spikes in the volume created by WIP bubbles,
2. Uneven product mix,
3. Unscheduled downtimes of equipment in BE Litho.

This thesis primarily deals with an approach that aims at reducing unexpected demands on the bottleneck that will result in exacerbating the capacity issue. The next few chapters discuss the approach used to achieve this objective.

Chapter 4: Solution Approach: Production Leveling

4.1 Background and Motivation

Chapter 3 discussed the need to eliminate sudden spikes in demand for the bottleneck resource(s). This chapter aims at introducing the various methods by which this can be achieved. The solution methodology considered is a two-pronged one (Figure 10). First, perform regular preventive maintenance to ensure higher equipment availability and less downtime due to unscheduled maintenance. Second, achieve production leveling; which will smooth the variability in the process, ensure even product mix, and reduce/eliminate WIP bubbles.

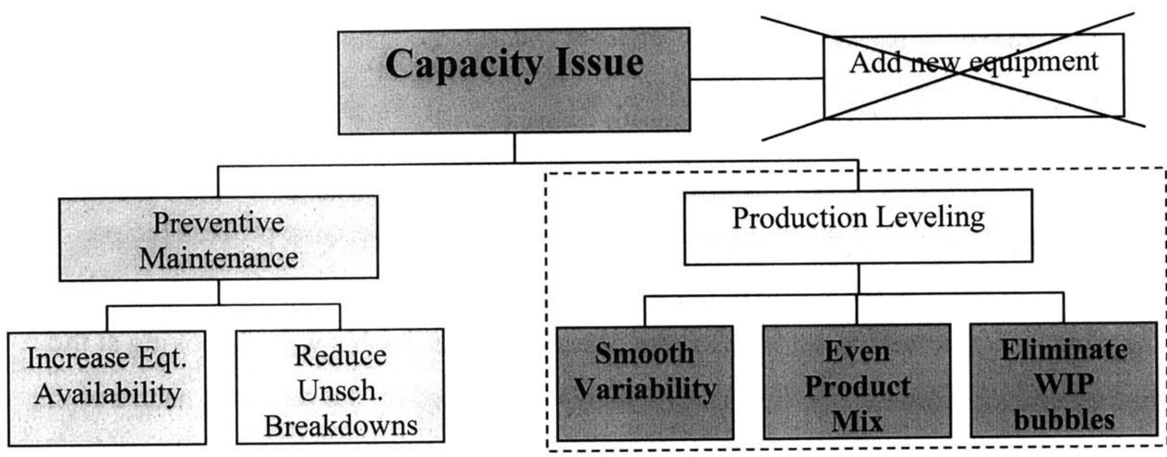


Figure 10: Proposed Solution Methodology

4.2 Production Leveling in Context of Fab17

The LFM project done by Sean Holly, LFM 2006, concentrated on the area of preventive maintenance. The main area of concentration for this project is production leveling. Production leveling or *heijunka* refers to the creation of a "level schedule" by sequencing orders in a repetitive pattern and smoothing the day-to-day variations in total orders to correspond to long-term demand (Womack, Jones, 306). In other words, can be thought of as a system of production smoothing designed to achieve a more even and consistent flow of work. In order to achieve production leveling, as a first step, it is important to identify the "product(s)" whose mix needs to be made even in order to achieve the result of predictable demand on the bottleneck. The next few sections aim at identifying this "product(s)".

4.2.1 Layers versus Segments

Microprocessors and chipsets are fabricated in layers on a silicon wafer through various chemical processes. On the wafer, the first layer of silicon dioxide is grown by exposing the wafer to extreme heat and gas. The wafer is then coated with a photoresist that becomes soluble when exposed to ultraviolet light. A mask or a stencil is used to cover the areas that should not be exposed to the ultraviolet light and a pattern is formed on the silicon dioxide. The exposed part of the silicon dioxide is etched away to leave the pattern behind. Layer upon layer is built in this fashion, with each layer using a different pattern to make the microprocessor or chip set.

Fab17 manufactures chipsets that primarily (more than 95% by production volume) use the same process technology and are made up of the exact same layers in the same order. This means that most of the chipsets made at Fab17 follow the same process recipe. Therefore, if the process were viewed as a series of steps making one layer upon the other, the process itself would look more like a flow line. Such a way of looking at the line takes away some of the complexity, that is introduced by the reentrant nature of the process, out of the analysis. Figure 11 gives an illustration of viewing the process in layers. Layer 1, Layer 2, up to Layer N are the various layers making up the wafer. Each of the layers consists of various operational steps like lithography, etch, and ion implantation.

For a wafer to be complete, it requires all the layers be performed in a particular order. Further, if the steady state output of the Fab were 500 wafers per day, then taking a snapshot of WIP in the fab should show 500 wafers in each stage of manufacture, i.e. having completed various layers. If the production were smoothed then there would be an even mix across all the layers in the Fab. Therefore, the "product" whose mix we are aiming at making smooth is in fact the various layers that make up the wafers manufactured at Fab17.

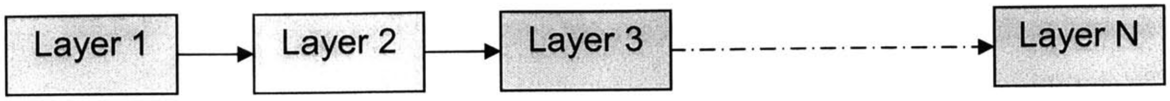


Figure 11: Illustration of Wafer Fabrication in Layers

4.2.2 Layers in Obtaining Optimal Granularity

At Fab17, the fabrication process is monitored in segments, where each segment may contain more than one layer. "Goaling of the line" ensures that there are a particular number of wafers (target) at the end of the segment. Further, it also dictates the number of wafers an operation should process in a week. This approach will give us an even mix of various layers if all the wafers were started at the same time and were in the same stage of fabrication. However, with multiple lots started each week, there is opportunity for an operation to pick one lot over the other that started at a different time. This kind of selective picking results in an uneven mix of layers in the fab. This uneven mix of layers will result in an uneven number of wafers finished per week, even if the wafers started per week (WSW) were the same. Therefore, in addition to monitoring the total segment output, it is essential to monitor the output for each layer. This will result in an even layer mix that will ensure an even output and therefore a predictable customer shipment.

The manufacturing process at Fab17 involves few hundred steps. It would be ideal to achieve production leveling in all the areas and operations. However, given the time and scope of the internship, this project will discuss a strategy of attaining an even layer-mix prior to the wafers entering the bottleneck (BE Litho). In other words, if an even layer mix is achieved prior to the wafers reaching the bottleneck (i.e. before starting layer BE1, see Figure 12), then there will be a predictable or "metered" input into the BE Litho. Therefore, the area that feeds into BE Litho, the implant area, should ensure that it creates a set of conditions that result in an even layer mix before the layers enter BE Litho. Figure 12 illustrates the various layers that make up the chipsets (in order to maintain confidentiality, only a condensed version of the various layers making the chipsets is shown); there are four different implant layers, and five different BE litho layers, required in this example chipset. The layers are depicted in the order they are added to the chipset.

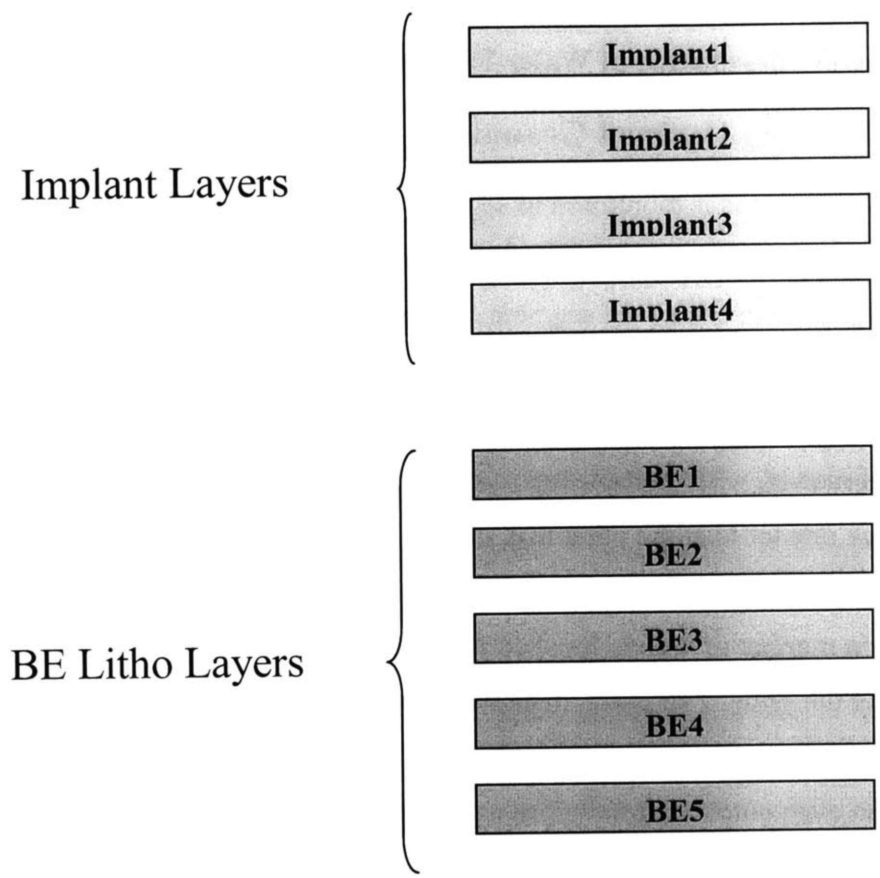


Figure 12: Implant Layers and BE Litho Layers

Since input into the BE Litho area needs to be "metered" the current layer mix in the Implant area needs to be understood. In order to do this, the variability in WIP over 14 consecutive shifts was analyzed (Figure 13). Visually it is apparent that the WIP levels for the various layers look disparate; however, a more accurate measure of variability would be the coefficient of variation. The coefficient of variation (CV) (http://www.statistics.com/resources/glossary/c/coeffvar.php) is the ratio of standard deviation to mean. CV for the implant area varies from a high of 0.54 (Implant5) to a low of 0.26 (Implant4). In order to level this variability and arrive at an even product mix, the cause for this variability needs to be understood. Additionally, any regularly occurring problems in the implant area that cause machine down time should be eliminated.

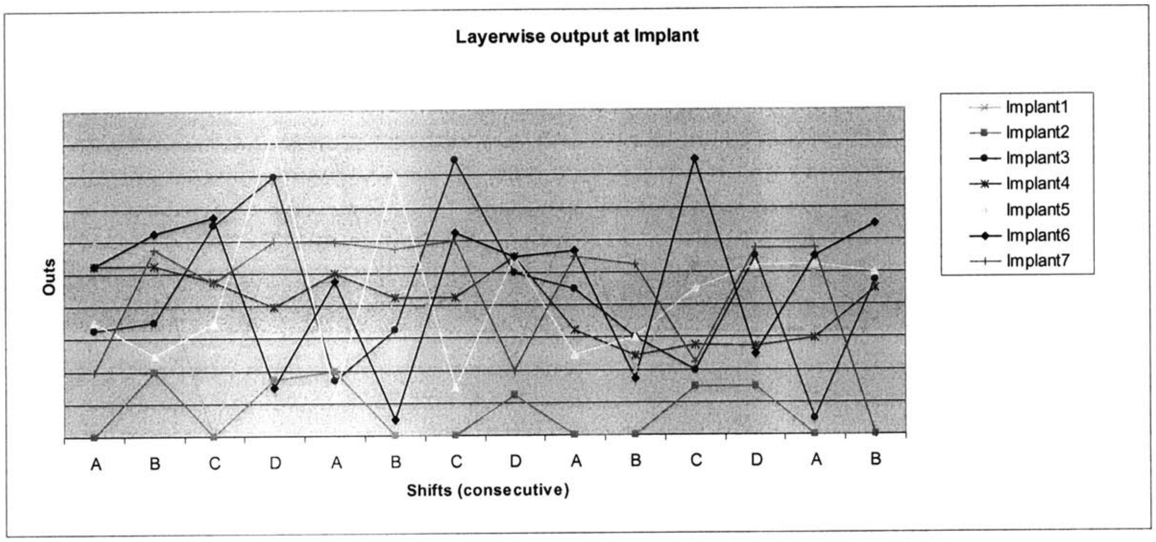


Figure 13: Current Data for Implant

4.3 Incentive for Wafer Outs

The first and major cause for the variability in the output from the implant area (and other areas too) is the current incentive system at Fab17. The areas in the Fab are measured on the "outs" (the number of wafer produced per unit of time, usually a week) the area makes irrespective of the type of outs. Therefore, if a given layer (referred to as *layer m*) had a cycle time of 10 min/lot versus another layer (referred to as *layer n*) that has a cycle time of 60 min/lot, the area would prefer to make *layer m* rather than *layer n* since they can make 6 lots (150 wafers) in an hour as opposed to just 1 lot (25 wafers). To maintain data confidentiality, Table 1 shows synthetic values for cycle time for some of the implant layers.

S.No	Layer	Cycle Time Range (min/lot)	Run Rate Range (wafers/hour)
1	Implant1	1.2 - 2.2	680 - 1250
2	Implant2	4.5 - 5.5	270 - 330
3	Implant3	2.0 – 3.0	500 - 750
4	Implant4	7.0 – 8.0	185 - 215
5	Implant5	0.5 - 1.5	1000 - 3000
6	Implant6	2.0 – 3.0	500 - 750
7	Implant7	2.0 – 3.0	500 - 750
. . .	. . .	. . .	. . .
n	Implantn	4.5 – 5.5	270 - 330
Weighted Average		3.5	700

Table 1: Cycle Time and Run Rate for Various Layers

A quick look at Table 1 shows that a shift incentivized purely on the number of outs would choose to produce the layer Implant5 over any other layer. Further, the shift is least likely to make the Implant4 layers (unless the shift manager directs the shift to make the Implant4). In effect, the run rates indicate that if a shift chooses to run only Implant5 it would make 18000 outs, instead of 2400 if it choose to make Implant4 – clearly a disproportionate advantage in processing Implant5. This also provides an insight into the variability in the data shown in Figure 13. For instance, a shift (say D) comes in and over processes the Implant5 layer, then the next shift, Shift A, over compensates by producing none of the Implant5 layer. However, it runs the layer that has the next shortest cycle time, Implant1, in order to maximize the outs for the shift. This yo-yo effect caused by the shifts trying to make their outs is responsible for the high degree of variability (Table 2). Variability is also created when the shift managers try to correct the over-production of certain layers by overcompensating production of other layers.

The second reason for the variability is also a result of the outs based incentive scheme at Fab17. Every time equipment in an area undergoes setup changeover, it loses some processing time. Since this lost processing time could have been used to make more outs, shifts put off setup changeover as long as possible. However, not having a regular setup changeover can have a negative effect on tool life. At Implant, the life of the ion implantation source is diminished if the setup is not changed periodically[1]. Some layers, like Layer4, consume the source more aggressively, therefore, if they are processed in large batches, it results in a diminished source life. However, general industry practice shows that alternating layers results in an extended source life.

The third reason is related to the fact that the process is still viewed in segments and not it layers. Therefore, the shifts are governed on an even WIP movement within segments and not necessarily between layers. As mentioned earlier, an even mix in the layer is critical to get wafers to the customer shipment stage.

Since the cause of variability for various layers is understood, the next step in metering the input into BE Litho is to understand issues at implant that cause machine breakdowns. Observation of the implant area over a period of one month showed that a recurrent problem at implant was the "crashing" of the implanters due to the over-processing of certain layers. For instance, as mentioned earlier, the Implant4 layers consume the ion implanter's source at a very aggressive rate compared to the other layers (almost twice as much as an Implant3 layer); this causes the source life to be greatly reduced causing the implanters to crash unexpectedly. This problem is significantly worsened when setup is not changed to run different layers on the implanters. Crashing of the implanter results in shutting off arrivals proceeding to BE Litho. The effect of this is the underutilization of the tools at BE Litho (and underutilization of the bottleneck is clearly not a good

---

[1] Ion implantation equipment typically consists of an ion source, where ions of the desired element are produced, an accelerator, where the ions are electrostatically accelerated to a high energy, and a target chamber, where the ions impinge on a target, which is the material to be implanted.

operations strategy). Further, when the implanters are operational again, there is a flood of WIP to BE Litho, which puts a lot of stress on the bottleneck.

The crashing of the implanters is currently viewed as an implant issue rather than an issue that affects the entire production line. The functional areas, such as Implant and Litho are currently isolated from one another mostly as functional areas; this makes it much harder to communicate a need for a common production strategy for the entire Fab. This thesis isolates one issue and proposes a solution to it. Though it is beyond the scope of this thesis/project, it is essential to make a system change to see an improvement to the WIP management strategy.

Chapter 5 New and Improved Manufacturing Approach (NIMA) Board Game

5.1 Background and Motivation

The issues related to the current incentive scheme and the effect of over-processing certain layers are discussed in the previous chapter. These issues are readily apparent when the entire production system is viewed in terms of wafers being made up of layers. However, since the current process is run in segments, it is difficult to visualize the problems. Further, it is difficult to generate buy-in for a concept that is not easily apparent. In order to overcome this issue, a board game was created (the game was named by the mX team as New and Improved Manufacturing Approach - NIMA). The board game was designed to be a simple representation of the Fab. The game incorporates the issues related to the current WIP management scheme along with the issues related to producing too much of specific layers.

5.2 The Board Game

5.2.1. Objectives of the Game

The main objective of the game is to illustrate in a simple way the problem and potential solutions to the problem. Therefore, the following are some of the requirements of the game:

- The game should illustrate the importance of viewing the process in layers.
- The game should show the importance of production leveling.
- The game should illustrate the value of setup changeover.
- The time taken to play the entire game should be short (under an hour).
- The time to understand the rules and operational intricacies of the game should be very short (less than five minutes).
- The set-up of the game should be easy.

- The components of the game should be made of easily available items and should be easily replicated.

With the above requirements in mind, the game selects the Implant3 and Implant4 layers as the two target layers. This choice is made since the Implant4 layer is the layer that feeds into BE Litho and the Implant3 layer is ideal to condition the implanters after the Implant4 layer is processed. The processes/operations that are used for the game are Lithography and Ion Implantation, since they are the two primary areas of concentration for this thesis.

5.2.2 Details of the Game

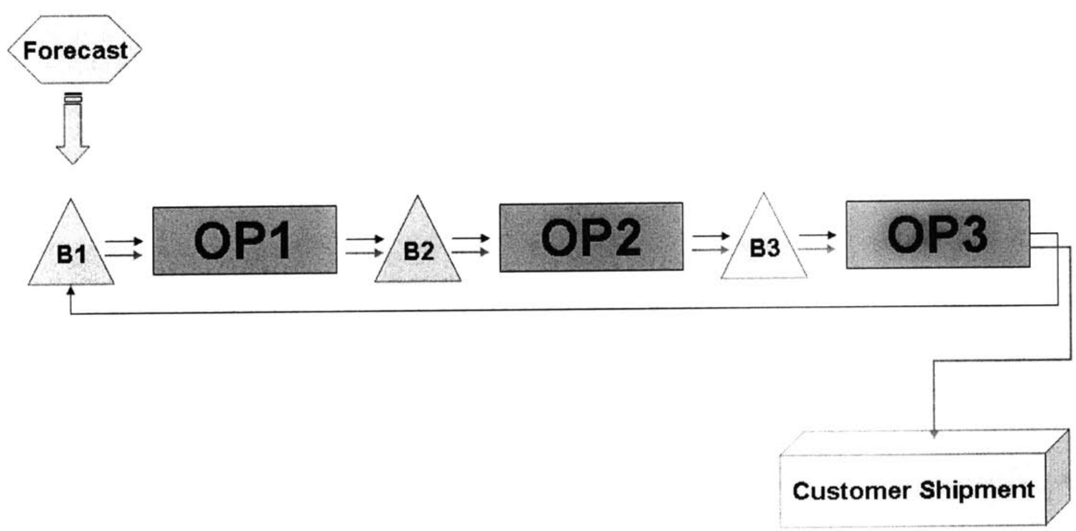


Figure 14: Flow diagram of the NIMA game

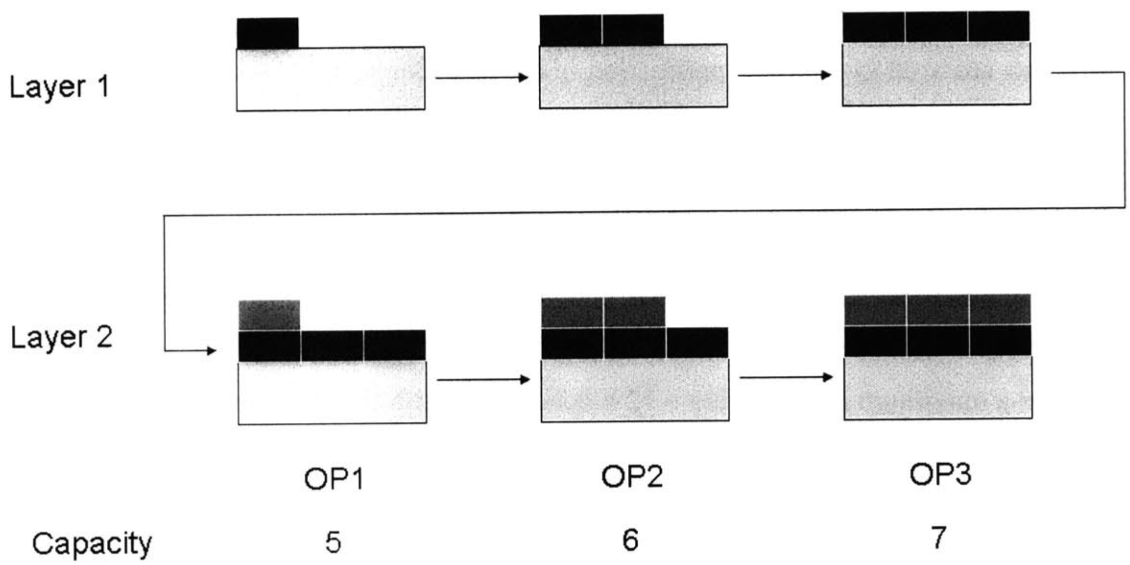


Figure 15: Illustration of the NIMA Game

Figure 14 and Figure 15 gives a pictorial representation of the game. The game will have three operations, OP1, OP2, and OP3. OP1 is used to represent Lithography, OP3 is designed to be Ion Implantation, and OP2 is designed to be a process like Etch, that is an intermediary between Litho and Implant (the players are not told what each of the operations represent prior to the start of the game). B1, B2, and B3 are the buffers located in between the various operations, and these buffers are used to hold WIP.

The final product (wafer) consists of two layers, the black layer and the red layer. The base is a blue colored Lego block (LB1) and is meant to represent a bare silicon wafer. The black layer is created in three steps – first, OP1 puts on a single black Lego block (that is $1/3^{rd}$ the size of the blue one, LB2); second, OP2 puts on the a black Lego block adjacent to the first one; and third, OP3 puts on a final Lego block to complete the black layer. Once the black layer is complete the wafer comes back to OP1 to start a red layer (this is to represent the reentrant nature of the wafer fabrication process). The red layer is completed in three steps just like the black layer. Once the two layers, the black and red, are completed, the wafer is ready for customer shipment.

Forecasts are used to drive the production in the Fab. Based on a forecast, the target production quantity is set at 25 wafers per shift. Each shift lasts for 6 minutes and each operation takes 30 seconds to complete. Therefore, in a single piece flow the time taken to complete a wafer for shipment would be 3 minutes. The batch capacity of OP1 is 5, OP2 is 6, and of OP3 is 7; i.e. OP1, OP2, and OP3 can process at the maximum 5, 6, and 7 wafers simultaneously. Therefore, OP1 is the bottleneck. Within a shift OP1 can produce a maximum of 60 wafers – 12 batches each with 5 wafers. Further, every time OP1 changes its setup, 20% productive capacity is taken away from the process. Therefore, whenever OP1 changes from making black layers to red layer or vice versa, it can only process 4 wafers instead of 5 wafers. When OP3 processes 8 wafers requiring red layers back to back, then it crashes and cannot come back for 1½ minutes. The game is played for two shifts, namely for 12 minutes. The following summarizes the rules of the game:

- OP1, OP2, OP3 are operations and B1, B2, & B3 are the buffers.
- Each operation takes 30 seconds to complete.
- Black is the first layer. Once units pass through OP3 they come back to OP1 for red layer. After the red layer is finished the units are shipped to customer.
- OP1 is the bottleneck with a capacity of 5 wafers in 30 seconds.
- OP3 crashes every time it processes 8 reds back to back and needs 90 seconds to complete repair and become operational again.
- Changing from rlack to red OR red to black is a setup change.
- Immediately following a setup change, OP1 can only process 4 units.

- OP2 and OP3 have no setup changeover impact.
- All stations are preloaded at the beginning of the game i.e. there are 5 units of the product at each buffer and they are in various stages of the black layer. In other words, B1 has 5 units with one black LB2, B2 has 5 units with two black LB2s, and B3 has 5 units with three black LB2s.
- The game is played for two shifts and the demand is 25 wafers per shift. Therefore, the target is to get 50 wafers at the end of two shifts.

The game has one player per operation and these players can change at the end of a shift (6 minutes). The game board and the location of the operations are included in Appendix B.

5.3 Playing the Game

Game Phase I

The rules of the game are explained to the players and the game starts. As expected, the players start by trying to make as many outs as possible for their particular operation. Typically, there is no communication between the operations even though they have not been told to refrain from communication. The production of 5 units at a time continues smoothly until OP3 has to process 8 red units back to back. Table 3 shows the results of the Phase I of game. At the start of the game, there are 5 units waiting in the buffers ready to be processed. When time starts, each operation (OP1, OP2, and OP3) processes the wafers and at the end of 30 seconds pass the wafers to the next operations. In Table 3, the numbers in black denote the number of wafers needing the black layer and the number in red denotes the number of wafer needing the red layer. For instance, at the end of 1:30 minutes, there are 5 wafers waiting for the black layer and 20 wafers waiting for red layer. Every time the product completes a black layer at OP3, its moved to B1 and wait for the red layer. Similarly, after completing each operation, the in-process wafers move to the next available buffer for the next process. The last column in blue denotes the number of finished wafers. At the 4 minute mark, OP3 has processed 8 wafers requiring red layers in a row, causing OP3 to crash. The shaded box under OP3 denotes the downtime of OP3 that lasts for 1 ½ minutes. At 5:30 there are no wafers available in B1 for OP1 to process; this results in OP1 being underutilized (the bottleneck is underutilized!). At 6:00, Shift 1 ends, Shift 2 begins, and the game is continued until the end of Shift 2.

The game was played many times and the results were the same. The bottleneck OP1 determined the flow of the entire process. When OP1 processed black layers back to back instead of alternating between black and red (since that would mean a setup changeover) it created a situation that would cause OP3 to "crash" (when it processed eight red units back to back). This crashing of OP3 causes OP1, the bottleneck, to be underutilized. The important things to notice in Table 3 are the "outs" (this is the sum of the columns for OP1, OP2, and OP3) for the different operations in the two shifts and the number of finished products (customer shipments) made. At the end of the Phase I, it is evident that the shifts did not make their goal of completing 25 units. In fact, they were 36% and 8% short in Shift 1 and Shift 2, respectively. The players were asked to discuss the cause for this and then come up with a different strategy to play the game.

	Capacity				5		6		7		
	Time	**Black Layer**	Waiting for Red	**B1**	**OP1**	**B2**	**OP2**	**B3**	**OP3**	Waiting for Red	Finished Product
SHIFT 1		**20**		**5**		**5**		**5**			
	0:00	**20**			**5**		**5**		**5**	5	
	0:30	**15**	5		**5**		**5**		**5**	10	0
	1:00	**10**	10		**5**		**5**		**5**	15	0
	1:30	**5**	15		**5**		**5**		**5**	20	0
	2:00	**0**	20		**5**		**5**		**5**	25	0
	2:30		25		4		**5**		**5**	26	0
	3:00		26		5		4		**5**	26	0
	3:30		26		5		5		4	21	5
	4:00		21		5		5	4	4	16	4
	4:30		16		5		5	6		11	0
	5:00		11		5		5	**11**		6	0
	5:30		6		5		5	**16**		1	0
	6:00		1		1		5	14	7	0	7
	EOS 1		0		**60**		**64**		**50**		16
		25		**25**		1		**19**			0
SHIFT 2	0:00	**20**		**20**	**5**		1	**12**	7		7
	0:30	**15**		**15**	**5**		**5**	**12**	1		1
	1:00	**10**		**10**	**5**		**5**	**5, 12**			0
	1:30	**5**		**5**	**5**		**5**	**10, 12**			0
	2:00	**0**		**0**	**5**		**5**	**15, 12**			0
	2:30						**5**	**13, 12**	**7**	7	0
	3:00		7	2	5			**6, 12**	**7**	7	0
	3:30		7	4	5		5	**6, 5**	7		7
	4:00				4		5	**10**	**6**	6	0
	4:30		6	1	5		4	7	7		7
	5:00				1		5	**10**	1		1
	5:30										0
	6:00										0
	EOS 2				**45**		**45**		**43**		23
					105		**109**		**93**		**39**

Legend

- (dark) Station is starved
- (light grey) Station is underutilized
- (hatched) Station is under maintenance
- EOS End of shift
- Black numbers indicate black layers
- Red numbers indicate red layers

Table 2: Results of the Game Phase I

Game Phase II

The main changes to the player strategies in the game that come about because of the discussion of the players are as follows:

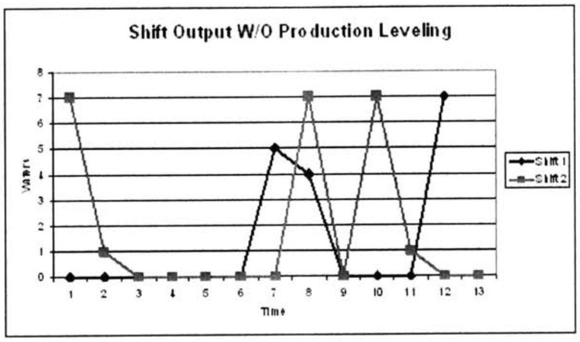
- The process should be viewed as a whole rather than as separate operations.
- Customer shipments are the most critical measure of the Fab's success.
- The different operations must communicate with each other in order to understand each other's constraints and to create an even work flow devoid of unscheduled downtimes.

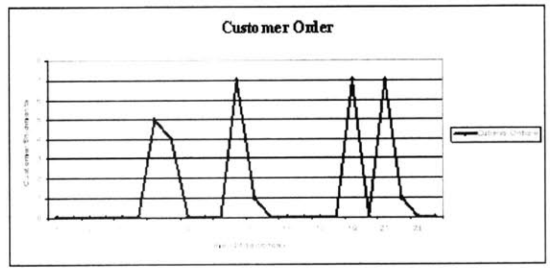
These suggestions are made clear to the players and the game is played again. This time the results are markedly different (Table 4 and Figure 16). Each of the shifts make 24 finished products each. Further, there are no unscheduled downtimes for OP3. The important thing to note is that the "outs" for the various operations are better in Phase II of the game. The customer shipments become more predictable (finished product every 4 minutes) and machine downtimes completely disappear. Though there is a cost associated with changing setups, overall, the outs for the two shifts improve.

SHIFT 1	**Capacity**				**5**		**6**		**7**		
	Time	**Black Layer**	Waiting for Red	**B1**	**OP1**	**B2**	**OP2**	**B3**	**OP3**	Waiting for Red	Finished Product
		25	5		**5**		**5**		**5**		0
	0:00	**25**	1		4		**5**		**5**		0
	0:30	**21**	1		**4**		4		**5**		0
	1:00	**21**	2		4		**4**		4		4
	1:30	**17**	2		**4**		4		**4**		0
	2:00	**17**	2		4		**4**		4		4
	2:30	**13**	2		**4**		4		**4**		0
	3:00	**13**	2		4		**4**		4		4
	3:30	**9**	2		**4**		4		**4**		0
	4:00	**9**	2		4		**4**		4		4
	4:30	**5**	2		**4**		4		**4**		0
	5:00	**5**	2		4		**4**		4		4
	5:30	**1**	2		**4**		4		**4**		0
	6:00	**1**	2		4		**4**		4		4
	EOS				57		**58**		59		24
		26	6		4		**4**		4		0
SHIFT 2	0:00	**22**	6		**4**		4		**4**		0
	0:30	**22**	2		4		**4**		4		4
	1:00	**18**	2		**4**		4		**4**		0
	1:30	**18**	2		4		**4**		4		4
	2:00	**14**	2		**4**		4		**4**		0
	2:30	**14**	2		4		**4**		4		4
	3:00	**10**	2		**4**		4		**4**		0
	3:30	**10**	2		4		**4**		4		4
	4:00	**6**	2		**4**		4		**4**		0
	4:30	**6**	2		4		**4**		4		4
	5:00	**2**	2		**4**		4		**4**		0
	5:30	**2**	2		4		**4**		4		4
	6:00										0
					52		**52**		52		24
					109		**110**		**111**		**48**

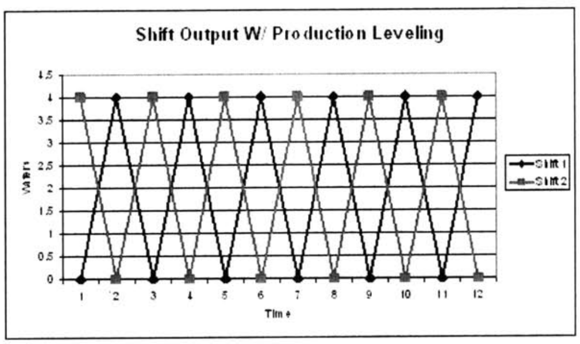
Table 4: Results of the Game Phase II

Phase I





Phase II



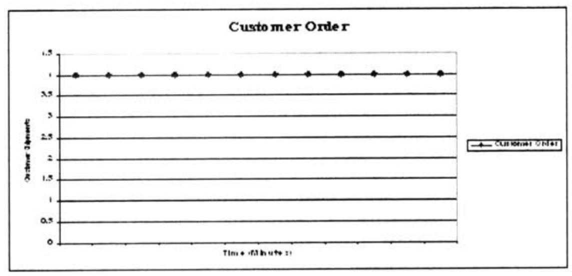


Figure 16: Comparison of the Results from Phase I and Phase II of Game

5.4 Lessons Learnt from the Game

The game was a powerful learning tool that provided insight into some of the WIP management issues that Fab17 is facing. In Phase I of the game there were no "winners" since the target number of customer shipments were not met (not even close!). The focus of the teams was on making as many "outs" as possible. In doing so, each of the operations (especially OP1 that was the driver of the WIP) was not comfortable changing setup from black to red or vice versa. This cost the "Fab" lost production due to unscheduled downtime, underutilization and starvation of bottleneck. Overall, the strategy of concentrating only on outs was not a good "winning" strategy. This scenario, though simplified, is a good representation of the "outs-focused" production strategy at

Fab17. Such a strategy at Fab17 has resulted in high variability in WIP and has placed undue demand on the bottleneck resource.

In Phase II of the game, the players moved away from this "outs-focused" approach. The teams communicated with each other to determine a coherent strategy that would result in maximum customer shipment, and high equipment availability (no unscheduled downtimes). The result of the strategy was to change setup on a regular basis to make an even mix of black and red layers, in other words to achieve production leveling by layers. The result of this production leveling was

1. Increased customer shipments (24 units per shift - close to the target of 25 units).

2. Predictable customer shipments – a shipment was made every minute.

Though outs per operation were reduced in the short term (OP1 went from 5 units in 30 sec to 4 units in 30 sec due to the setup changeover), in the long term, the total number of outs were more in Phase II than in Phase I.

Chapter 6: Kaizen Event and Run Rules for Ion Implantation

The results of the game discussed in the previous chapter are compelling and make an excellent case for a leveled production. However, the game is still an over-simplified version of the Fab and the results need to be proven in actual production settings. In order to do this, it is extremely critical to obtain buy-in from the various stakeholders. These stakeholders are the Manufacturing Technicians, Operations Managers, Process Engineers, Shift Managers, Department Heads and the Plant Manager. Such buy-in is important both to help understand the practicality of a solution, and to help implement the solution in the Fab.

6.1 The Kaizen Event

A three-day Kaizen event[2] was put together with the help of the mX team. Potential decision-makers within each job function were identified and invited to the event. To ensure representation from all the shifts, there were key stakeholders invited from the night shift. A facilitator with a lot of experience in conducting Kaizen events (who is a part of the mX organization) oversaw the event. The Kaizen event started as a forum to discuss the status of level production within the Implant area and a brainstorming session brought about the following hopes and concerns regarding operations at Implant:

Hopes

---

[2] Kaizen in Japanese means continuous improvement.
http://www.isixsigma.com/dictionary/Kaizen_Event-411.htm

Kaizen events are commonly referred to as a tool that:

1) Gathers operators, managers, and owners of a process in one place
2) Maps the existing process (using a deployment flowchart, in most cases)
3) Improves on the existing process
4) Solicits buy-in from all parties related to the process

Kaizen events are extremely efficient to quickly improve a process. The true intent of a kaizen event is to hold small events attended by the owners and operators of a process to make improvements to that process that is within the scope of the process participants.

- Four Shift Agreement – The Implant team hoped to garner support for their efforts from all the four shifts. This support was essential for any change to be implemented and be sustainable.

- Work Towards Ideal State – The team wanted to use the Kaizen event to learn how to steadily move towards an ideal production state (this was not a fixed state but a situation with minimal work stoppages).

- Operations Managers (OM) on same page with Manufacturing Technician and Process Engineers – There are different goals that the various members of the manufacturing team pursue at Fab 17. The OM's objective was to manage the day-to-day operations of the floor, and therefore they are in a fire-fighting mode during their shift. On the other hand, the Process Engineer's job is to improve the performance of the equipment under their care. Manufacturing Technicians are interested in meeting their daily goals. However, they are also interested in making the process more efficient in order to make their jobs easier (it is important to note that the work force at Fab17 is non-unionized and the culture encourages every member of the community to make process improvements).

- Consistent output while maintaining plant's ability to adjust – The manufacturing team was interested in meeting their daily goals for output. However, they found it essential to maintain a degree of flexibility to change their production strategy (this is at odds with the less setup changeover philosophy).

- Reduce tool downtime – This was a major concern for the team and they were working with the Preventive Maintenance (PM) and High Precision Maintenance (HPM) teams to meet this goal. However, according to our discussions in the previous chapters it is evident that this is not a stand-alone strategy, but rather one that will be affected by production leveling as well.

- All four shifts should work consistently towards meeting capacity challenges – In not just the capacity issues, but in all areas of manufacturing, Fab17 is faced with the challenge of implementing a consistent strategy through all the shifts. This inconsistency results in wasted effort and constant finger pointing.

- Learn how to facilitate – Some of the team members wanted to use the Kaizen event as an opportunity to learn how to facilitate other such "continuous improvement" activities.

- Productive event – Some team members were skeptical about the efficacy of the Kaizen event and did not want it to be a "waste of their time."

Reservations

- SCANDOs died out – SCANDO is the acronym used to describe the 5S initiative at Fab17. This was implemented in various parts of the Fab. However, after a few months the areas were not maintaining themselves as per SCANDO.

- Not on the same page down the road – This particular comment referred to the fact that the change may not be sustainable.

- Faith in past mX initiatives – There was no clear understanding of what mX or Lean Manufacturing were. Further, due to this, past mX initiatives had not been successful or the teams had prematurely aborted efforts. Therefore, a training program to educate manufacturing teams was critical to the success of mX initiatives like the Kaizen event.

The issues above are not restricted to the Implant area, therefore any 'behavioral' changes (between shifts, within shifts, incentive systems) can be applied to the entire Fab. Once general issues were identified, the discussion of the Kaizen event was narrowed down to the issues related to leveling of production within implant. The discussion identified the various machines in the implant area and the issues associated with having highly varied cycle times for the various layers. Further, the team felt the need to address the issue that caused lower source lives for the implanters (over processing Implant4 layers). The various options it proposed were:

1. Restrict the number of Implant4 layers processed between setups.
2. Change the equipment that processes Implant4s on a regular basis. For example, if *equipment m* processes 500 Implant4s, then *equipment n* should process the next 500 Implant4s.
3. Run layers alternately in order to improve the health of the implanters.
4. Dedicate an equipment to process Implant4 layers and do preventive maintenance on another machine so it can be ready in case the first machine is down due to shortened source life.

These inputs from the above discussions, along with the knowledge of the cycle times and run rates for the various XR80s (the implanters that are used to make implant layers including the Implant4 layers), gave rise to a set of considerations called run rules.

6.2 Implant Run Rules for Implant

The run rules were designed to eliminate the problems associated with the implant area having a very uneven layer mix. Further, the run rules were also designed to ensure that the implanters are not "tied-up" in processing the same layer, since this is not good for the equipment health, especially for the layers like Implant4 that consume the source very aggressively. The solution was as follows:

- In a given shift, particular equipment should be dedicated to process a certain layer. For instance, machine 1 should be dedicated to process Implant4 layers for the current shift.

- The number of machines dedicated to a certain layer should be proportional to the cycle time required to process the layer. In other words, the layers that required a longer time to process per lot would be dedicated to multiple (more) machines as compared to a layer with a lower processing time per lot. Further, if a layer required a much shorter time to process, it may share the equipment with another layer that has a short cycle time.

- The machines that process a certain layer (layer 1) in a shift will be changed to process a different layer in the next shift (layer 2). In case of layers dedicated to multiple equipment, this would mean that all the equipment processing layer 1 would move to layer 2. This would continue in every shift in order to maintain equipment health.

- Equipment will periodically undergo preventive maintenance. This would be according to a schedule maintained by the Operations Manager (and the Equipment Engineer).

- The run rules will be followed by all the shifts, and there would be a hand-off meeting between the shifts. Any inconsistencies to the run rule will be managed by the shift manager.

Table 4 gives a compiled look at a possible scenario using just seven layers (the values for cycle time are fictitious to maintain data confidentiality) that could be a result of the run rules that were created during the Implant Kaizen. Therefore, in Table 4; Shift 1: Implanter 1, Implanter 2, and Implanter 3 will process wafers requiring the Implant4 layer. Implanter 4 will process layers Implant5 and Implant1 for half a shift each. Implanter 5 will process Implant7. Implanter 6 will process Implant3 and so on. In Shift 2, the equipment will change the layers they are processing. Implanter 4, Implanter 5, and Implanter 6 will process Implant4; Implanter 7 will process Implant5 and Implant1 etc. By following these run rules, the output per layer looks more level (the # of wafers/shift are calculated based on the theoretical cycle time and the actual run rate). For instance, there are 7200 wafers with Implant4 and there are 7050 wafers with Implant5, and 7000 with Implant2.

Implanter	Layer	Cycle Time	# of Wafers / Shift
1	Implant4	*7.5*	*2400*
2	Implant4	*7.5*	*2400*
3	Implant4	*7.5*	*2400*
4	Implant5	*1.0*	*6000*
	Implant1	*1.7*	*7050*
5	Implant7	*2.5*	*7200*
6	Implant3	*2.5*	*7200*
7	Implant2	*5.0*	*3500*
8	Implant6	*2.5*	*7200*
9	Implant2	*5.0*	*3500*

Table 3: Run Rules Created as a Result of Implant Kaizen

The results in Table 4 assume that the output will be the same every shift. However, there is inherent variability in any system that will cause the output to vary from shift to shift. Therefore, the results were simulated under the leveling production scenario with a variability (of volume of wafers with a particular layer) of anywhere between -10% and +10%. This was achieved by randomizing the output in Excel (by using Rand ()). For instance, in Shift B, the output of Implant4 layer was randomized between 6480 wafers and 7920 wafers (+/-10%). In one simulation run, this resulted in an output of 7900 wafers in Shift B, 7180 wafers in Shift C etc. as shown in Table 5.

	Implant4	Implant5	Implant1	Implant7	Implant3	Implant2	Implant6
A	7200	6000	7050	7200	7200	7000	7200
B	6852	6718	7907	6792	6827	7270	7262
C	7816	6831	7656	7480	6990	6883	7838
D	7530	6415	7559	6787	6541	6610	7367
A	6605	8060	7728	6862	7506	7287	7229
B	7452	7737	7651	7641	7729	6305	7363
C	6643	6830	8097	6764	6927	7360	7169
D	7859	7536	6794	6815	6781	7025	6969
A	6724	6551	7314	6888	6638	7672	6750
B	7336	7594	6402	7344	7108	6314	7593
Std Dev	473	665	526	329	374	453	302
Mean	7202	7027	7416	7057	7025	6973	7274
CV	*0.065728*	*0.094623*	*0.070994*	*0.046684927*	*0.053246845*	*0.065033*	*0.041563*

Table 4: Output per Layer at Implant as a Result of New Run Rules

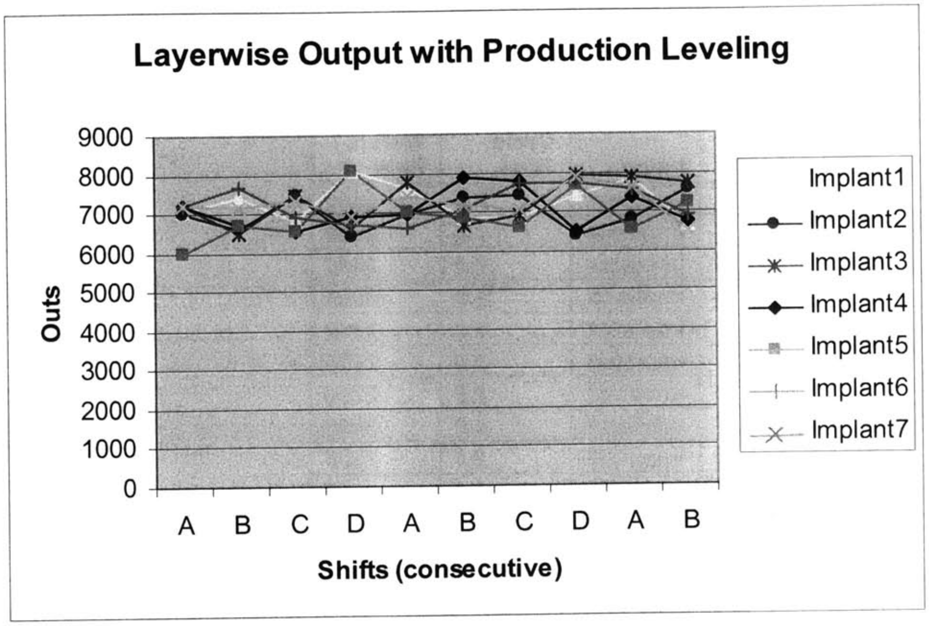


Figure 17: Simulated results for Production Leveling at Implant

It is impressive to note the significant decrease in the variability seen in Table 5 and Figure 17. For instance, the coefficient of variation for the Implant5 layer was 0.54 before the simulated production leveling, as opposed to 0.068 in the new scenario. These results are based on a simulation and are not the real results. Currently (as of 26 Feb 2007), Fab17 reports a 10% increase in machine availability, since short source life is no longer a critical issue:

"Run rules are followed and adhered to by every shift. Over all availability has gone up in excess of 10%. Primarily due to less source changes." – *Paul Cyr, mX Engineer, Intel Fab17*

Chapter 7: Observations and Areas of Future Work

Fab17 has come a long way since they began their lean implementation efforts. They have obtained support and commitment from the senior management for their lean implementation initiatives. They have trained all factory management on the basic concepts of lean manufacturing by collaborating with the Lean Learning Center (located in Novi, MI). They have led improvement activities that have realized tangible results in direct labor savings, decreased safety incidents, and decreased quality events. Their successes, though numerous, are still restricted to certain areas of the Fab. Efforts have been focused on implementing Lean Tools like 5S, 5Whys, etc. instead of using lean thinking to improve manufacturing processes. In other words, Fab17 still views Lean Manufacturing as a set of tools as opposed to a way of thinking. Moving forward, the biggest challenge facing Fab17 is making mX mainstream (i.e. leading process improvement initiatives using lean principles) and garnering buy-in from all levels of management. The following section discusses observations from the six-month LFM Internship. Recommendations based on those observations provide potential next steps for an mX implementation.

7.1 Observation 1: Establish the Customer-Supplier Relationship

The various functional areas within Fab17 operate as independent entities. However, in reality, due to the highly reentrant nature of the wafer fabrication process, each area is both a customer and a supplier to the other areas. Therefore, it is essential for the various areas to communicate with each other in order to understand and tackle any upstream or downstream issues that might affect them. Currently, the only formal forum for communication, within the Fab, is the daily operations meetings that are held by the shift managers to communicate the goals of the Fab for the day (there are also weakly operations meeting held on Wednesdays). Issues that occur in one area should be communicated immediately to other areas though a proper process. Such a communication path will ensure that other areas (both upstream and downstream since there is a customer-supplier-customer relationship) do not perform actions that could affect the flow of products downstream.

7.2 Observation 2: Alignment of Incentives and Monitoring in Layers

The current incentive system at Fab17 focuses heavily on the number of outs that each area makes. This methodology ensures that each area/operation is using their resources to the fullest. However, the inherent issue with such a methodology is that it does not tie the number of outs per area to the total number of complete wafers that are manufactured. Therefore, each area tries to maximize its own output by selectively picking certain layers to process over others. The incentive system at Fab17 should be focused on maintaining a balanced flow or output throughout the Fab. This output should tie into the overall goal of the Fab output and be monitored at the Operation Manager level and at the Shift Manager level. Such a balance is difficult to achieve in the current approach of monitoring segments. Segments are typically week-long and hence not granular enough to maintain a tighter control on the WIP movement. Fab17 should start monitoring their production process in layers as opposed to segments. However, there are difficulties to converting to a layer-based system now. Some of the potential issues are listed below:

- Current computer-based production systems do not support tracking the wafers according to the layers completed.
- Education and training of all operations personnel on the new methodology
- Complying with the new method.

As with any change, Fab17 management would have to be prepared for pushback from the operations personnel. There needs to be buy-in established at an early stage in order to make the transition smooth.

7.3 Areas of Future Work: Color-Coded Lots

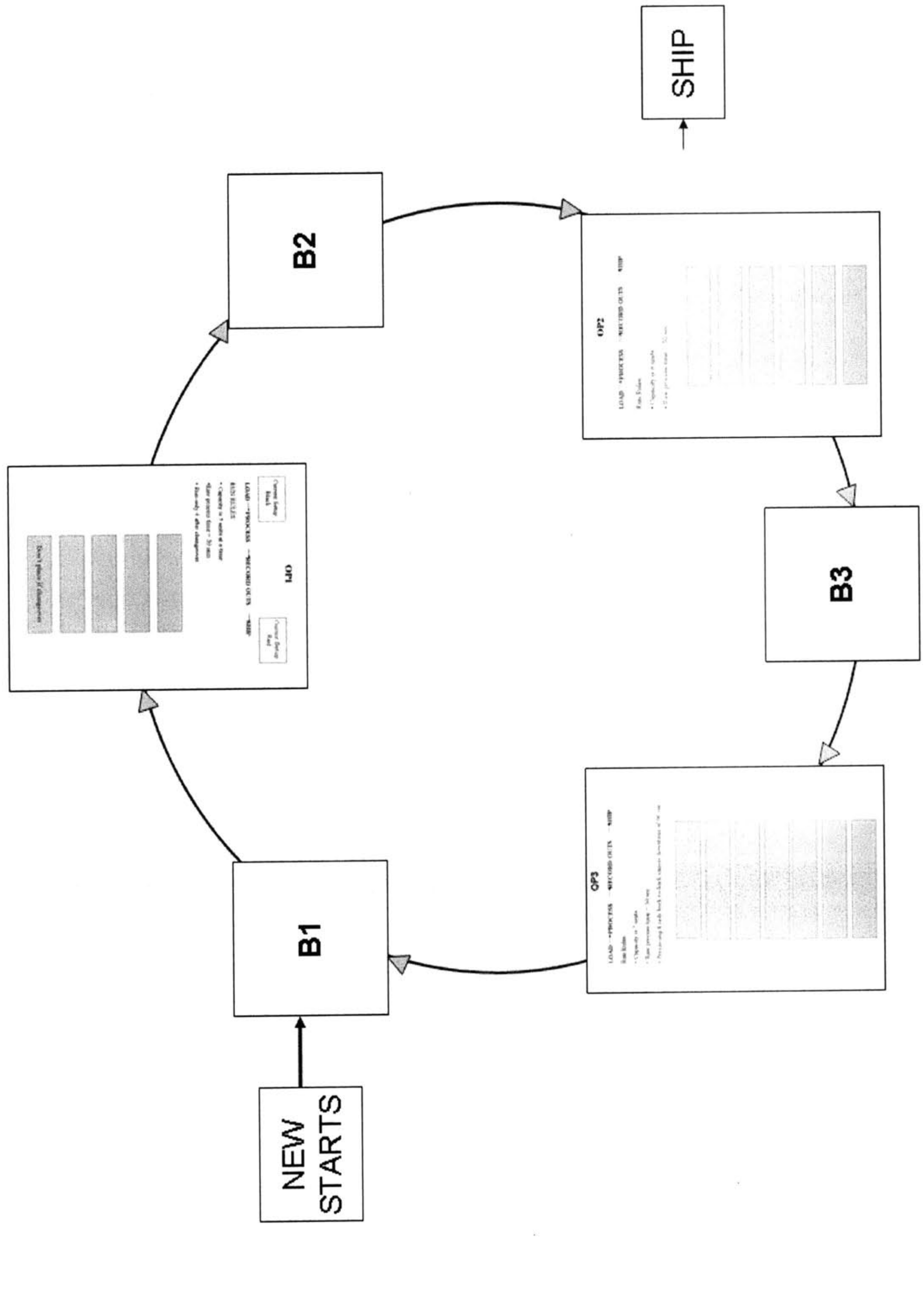
Lean manufacturing encourages the use of visual indicators and simple methods of error proofing. Visual indicators provide an "at-a-glance" indication of how well the system is performing. Computer based systems are effective; however, they do not provide quick feedback and the issues often remain hidden or buried within data. An area of future work at Fab17 could involve color-coding lots as they complete a layer. For instance, if a lot completed Layer 1, the lot will be marked by a red card and move on to complete the next layer, Layer 2. When it completes Layer 2, the red card would be removed and a blue card inserted and so on. With such a method of color-coding, a operations personnel could immediately "see" if there is a build-up of a certain type of layer (as this would cause an imbalance in the line) and identify the cause for this build-up immediately.

7.4 Conclusions

Level production was implemented in the ion implantation area of Fab17 after the Kaizen event, and the run rules are being followed. Because of the implementation, implant has seen an increase in equipment availability in excess of 10%. However, there are still some critical challenges ahead. First, the implementation needs to move beyond the implant area to drive level production in the entire Fab. This would require monitoring of the Fab to ensure that the benefits of an even layer mix that is gained at implant does not get lost down stream. Second, engagement of all levels of management is the key to the sustainability of the mX effort. Finally, management needs to put educational initiatives in place to inculcate lean thinking within the Fab.

Appendix A

GAME BOARD SETUP



OP1 SETUP SHEET

OP1

Current Setup
Black

Current Setup
Red

LOAD → PROCESS → RECORD OUTS → SHIP

RUN RULES:

- Capacity is 5 units at a time
- Raw process time = 30 secs
- Run only 4 after changeover

Don't place if changeover

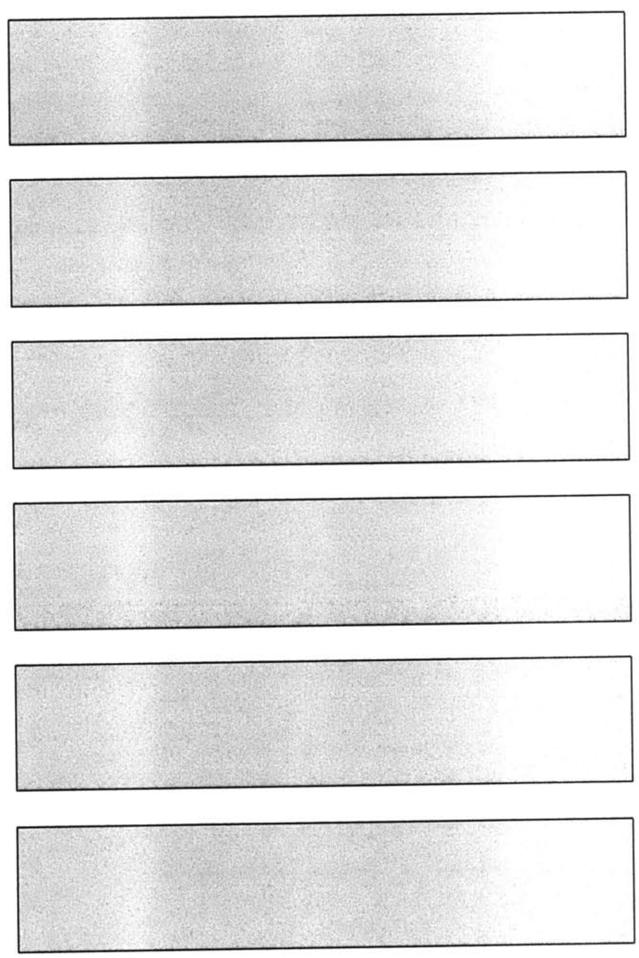
OP2 SETUP SHEET

OP2

LOAD → PROCESS → RECORD OUTS → SHIP

Run Rules:

- Capacity is 6 units
- Raw process time = 30 sec



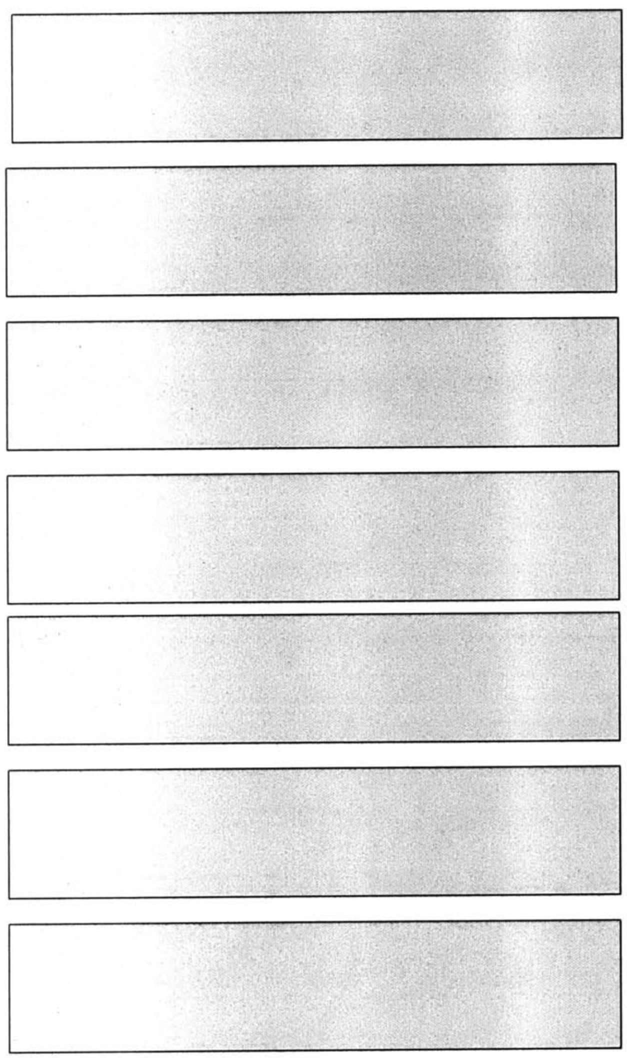
OP3 SETUP SHEET

OP3

LOAD → PROCESS → RECORD OUTS → SHIP

Run Rules:

- Capacity is 7 units
- Raw process time = 30 sec
- Processing 8 reds back-to-back causes downtime of 90 sec



References

Connally, Jason Walker. "Introducing Pull Methodologies in a Semiconductor Fab." Massachusetts Institute of Technology (MIT) Master's Thesis. June 2005.

Dennis, Pascal, Lean Production Simplified: A Plain Language Guide to the World's Most Powerful Production System. Productivity Press, 2002.

Fearing, Rebecca Cassler. "Managing Preventative Maintenance Activities at Intel Corporation." Massachusetts Institute of Technology (MIT) Master's Thesis. June 2006.

Flinchbaugh, Jamie and Carlino, Andy. The Hitchhiker's Guide to Lean. Society of Manufacturing Engineers, 2006.

Goldratt, Eliyahu and Cox, Jeff. The Goal, North River Press; 3 edition, July 2004.

Holly, Sean Michael. "Lean Manufacturing in a Semiconductor Environment: Use of Variation Analysis to Focus Continuous Improvement Efforts." Massachusetts Institute of Technology (MIT) Master's Thesis. June 2006.

F17 Communications Department. "External Communications Presentation." Intel, 2003. Kissel, Chris. "The Trendy Chipset for the x86 Processor." Reed Electronics Group. July 2005.

Intel Corporation Website, www.intel.com

Koch, Karl. "Pillar Summary." mX Group. 2005

Leachman, Robert and Leachman, Chien. "E Commerce and the Changing Terms of Competition in the Semiconductor Industry", June, 2006, <http://e-conomy.berkeley.edu/conferences/9-2000/EC-conference2000_papers/leachman.pdf>

Sonnet, Maria Claudia. "Introducing Cost of Stockouts in the Microprocessor Business and its Impact in Determining the Optimal Service Level." Massachusetts Institute of Technology (MIT) Master's Thesis. June 2005.

Spear, S. and Bowen, H. K., (1999) "Decoding the DNA of the Toyota Production System", Harvard Business Review, September/October 1999, Vol. 77, Issue 5, pp. 96-106.

Statistical Glossary, Retreived February 12, 2007 from: <http://www.statistics.com/resources/glossary/c/coeffvar.php>

The Economist, "Intel: Not Paranoid Enough", May 25, 2006.
Womack, James and Jones, Daniel. Lean Thinking : Banish Waste and Create Wealth in Your Corporation, 2nd Revision,. Free Press, 2003.

Zhou, Ning. "Introducing Improving Line Yield at Fab17." Massachusetts Institute of Technology (MIT) Master's Thesis. June 2000.