A Constraint-based Systems Approach to Line Yield Improvement in Semiconductor Wafer Fabrication

by

Viju S. Menon

M.S.E., Computer Science Engineering
University of Michigan, Ann Arbor, 1990

Submitted to the Department of Electrical Engineering & Computer Science
and the Sloan School of Management
in Partial Fulfillment of the requirements for the degrees of

Master of Science in Electrical Engineering and Computer Science

and

Master of Science in Management

at the
Massachusetts Institute of Technology
May 1994

Signature of Author_____
Department of Electrical Engineering & Computer Science
MIT Sloan School of Management
May 1994

Certified by_____
Associate Professor Charles H. Fine, Thesis Advisor
MIT Sloan School of Management

Certified by_____
Professor Lionel C. Kimerling, Thesis Advisor
Department of Material Science Engineering

Accepted by_____
Jeffrey A. Barks, Associate Dean
Sloan Master's and Bachelor's Programs

Accepted by_____
Frederick R. Morgenthaler, Chairman
Department Committee on Graduate Students

A Constraint-based Systems Approach to Line Yield Improvement in Semiconductor Wafer Fabrication

by

Viju S. Menon

Submitted to
the MIT Sloan School of Management and
the Department of Electrical Engineering & Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management

and

Master of Science in Electrical Engineering & Computer Science

ABSTRACT

Intel Corporation is a leading semiconductor manufacturer, with 1993 sales exceeding $8 billion. Since market demand is not the constraint for Intel, the key to maximizing profits lies in maximizing die shipments by improving capacity, die yield and line yield. This thesis adopts a "systems thinking" approach, in confluence with Eli Goldratt's "theory of constraints", to develop line yield improvement tools for semiconductor fabrication.

Intel's historical approach of responding reactively to major wafer loss events provides very little power to alter the course of those events. This thesis uses Systems Thinking to emphasize the underlying structure driving line yield behavior rather than focusing on discrete wafer loss events, thus uncovering high leverage strategies to improve line yield.

Intel has typically prioritized line yield problems and improvement projects based on the number of wafers being scrapped due to each problem. This assumes that the significance of a line yield problem depends only on the number of wafers lost, not on the extent of processing done to the wafers before they were lost. My thesis uses Theory of Constraints to prove that the importance of a line yield loss incident should depend not only on the number of wafers lost, but also on where in the process flow, the loss occurred. This research develops a constraint-based wafer scrap value model to systematically assign relative values to wafers depending on where in the process flow the loss occurred. This approach helps to focus a factory's limited line yield improvement resources on those line yield problems having the most revenue impact.

After identifying fundamental issues that contribute to (or limit) line yield performance, this thesis recommends key changes to management systems, structure, policies, and processes, to attain higher line yields in semiconductor wafer fabrication.

Thesis Advisors

Dr. Charles H. Fine, Associate Professor, MIT Sloan School of Management
Dr. Lionel C. Kimerling, Professor, Material Science Engineering Department

The author gratefully acknowledges the support and resources made available to him through the MIT Leaders for Manufacturing Program, a partnership between MIT and major U.S. manufacturing companies.

Acknowledgments

This thesis was far from an individual effort. Without the assistance of numerous people at Intel, MIT professors, students, and professionals in the field, the data and analysis would have suffered. It would be impossible to list everyone who influenced my thinking, both before and during the thesis, but I would like to thank the most significant contributors.

I am deeply grateful to my company supervisor, Al Stone, for his valuable guidance, support, and timely feedback before, during, and after my internship. Al regularly provided direction for focusing my efforts on solving a problem which was at the same time manageable and beneficial to Intel.

Special thanks go to David Marsing and Dan Goranson for initiating and supporting this internship. Many thanks go to Bob Baker and the entire Fab 9 organization, especially the Factory Support Team for their support and hospitality.

I am thankful to my Steering Committee members Brian Harrison, Brian Krzanich, Bruce Sohn, Don Myers, Jerry Marcyk and Tim Hendry for providing valuable insights and cross-factory support. Special thanks goes to Jacques Vuye and Mike Stark for giving me the opportunity to co-chair and drive the cross-site line yield benchmarking team. I would like to thank all the team members for their valuable help in data collection.

I am deeply thankful to the Intel Manufacturing Excellence Conference organizers for giving me the opportunity to present my findings and recommendations at the January '94 conference. In this context, I also gratefully acknowledge the excellent feedback I received from the conference's four hundred attendees.

I would also like to thank my thesis advisors, Professor Charlie Fine and Professor Lionel Kimerling, for providing me with very valuable support and guidance.

Finally, and most importantly, I thank my parents, and dedicate this thesis to my mother and to the inspiring memory of my father. All of my successes and achievements are a direct reflection of their support, guidance and confidence.

Bibliographical Note on Author

Viju S. Menon graduated from the University of Michigan, Ann Arbor, MI, in December 1990 with a Master of Science degree in Computer Science Engineering. He graduated from the University of Kerala, India, in 1988 with a Bachelor of Technology degree in Computer Science Engineering.

Viju joined Intel Corporation in January 1991 and has worked in various design engineering roles for Intel in Arizona and Oregon. Upon his acceptance to MIT's Leaders for Manufacturing program, Intel Corporation decided to sponsor him, and he will be joining Intel's Component's Manufacturing Division (Fab 11, New Mexico) in June 1994. He is currently working towards masters degrees in management and electrical engineering and expects to graduate in May 1994.

Table of Contents

List of Figures

List of Tables

Reader's Guide

One of the issues I faced at the end of my internship was how to document my research. Since I had used the Seven Step Method for a systematic diagnosis of the fundamental causes of line yield problems, I decided to use the "QI (Quality Improvement) Story" format, which is an efficient sequential format used by seven-step researchers to present to management, the findings of quality improvement studies.

The Seven Step Method is a step by step process to systematically solve problems.	
Step 1: Theme Selection	Step 5: Evaluation of Effects
Step 2: Data Collection & Analysis	Step 6: Standardization
Step 3: Causal Analysis	Step 7: Reflection on Process &
Step 4: Solution Planning & Implementation	Identification of next problem

Chapter 1 (Theme Selection) discusses the line yield improvement theme, and the motivation for this project. It also clearly defines the research objectives and deliverables.

Chapter 2 (Data Collection & Analysis) discusses the framework I used to collect and analyze line yield related data at each Intel site.

Chapter 3 (Causal Analysis) introduces the tool (i.e., Systems Thinking) I used to analyze the root causes of line yield problems, and also discusses how I applied systems thinking to uncover underlying causes of slow line yield learning rates at Intel factories.

Chapter 4 (Solution Planning & Implementation) first talks about the solution planning tools I used (e.g., Theory of Constraints (TOC), Systems Thinking), and then introduces my solutions including the TOC based wafer scrap value model and other facets of my line yield improvement tool-kit.

Chapter 5 (Evaluation of Effects) discusses the tangible results of my research and also the intangible results in terms of satisfaction level of customers with this tool, skill level, and motivation of people involved in implementing my recommendations. This chapter also discusses communication of my key findings across Intel.

Chapter 6 (Standardization) defines how the solution can be permeated in the company.

Chapter 7 (Reflection on the process/Next Problem) discusses key recommendations and findings of my research and also outlines tentative research areas based on this thesis.

Chapter 1. Theme Selection

Purpose: The purpose of theme selection is to recognize the importance of the problem and to define the problem clearly. This chapter discusses the line yield problem, motivation for this research, the research objectives and deliverables.

1.1. Introduction

Intel Corporation is the world's leading microprocessor manufacturing firm. Much of its recent success is based upon its dominance in providing the microprocessors that run IBM compatible personal computers. In 1981, IBM selected the Intel x86 architecture as the microprocessor that would be used in its initial foray into the personal computer market. Successive generations of this architecture (286, 386, 486, and Pentium™) have each spurred the rapid deployment of new product lines by PC manufacturers.

The ability of Intel, as well as the other semiconductor manufacturers, to ramp new products and processes has become critical in today's business environment [1]. In fact, competition in many industries has moved beyond quality and performance to reaching the target market in the required time window. Intel Corporation has now committed itself to more rapid introductions of new generations of microprocessors. In order to achieve this goal, design teams for different microprocessor generations are now working in parallel on their circuit designs. This approach as well as the use of sophisticated software tools will enable more rapid releases of product designs. The important challenge facing Intel and other semiconductor firms is whether their manufacturing facilities can match the accelerating pace of the designers when they ramp production of new microprocessors.

1.2. Project Motivation & Problem Definition

Factory output is a function of installed equipment base, utilization of that equipment, and yields. Yields include both line yield (the % of wafers that make it through the entire process) and die yield (the % of good die in each wafer). Output is ramped through the simultaneous installation of capacity and improvement in yields. Various studies within Intel have shown that of the three output levers (capacity, line yield, and die yield), right from factory startup through ramp, capacity and die yield

improvements have higher die output leverage than line yield improvements. This has led Intel to focus its efforts on learning how to ramp up capacity and die yields, with the result that it now has a better understanding of driving rapid capacity ramps and die yield improvements [1].

While Intel has historically made dramatic progress in capacity and die yield improvements, it still has to better understand the complex issue of line yield improvement, in terms of understanding the controllable factors causing line yield losses. This would require reviewing all contributors of line yield losses, including equipment, process, and procedural (human) elements to see if there are intrinsic, systematic causes to line yield losses.

Each of the 5 evils **(waste, delay, mistakes, defects, accidents)** could describe the problems of having line yield losses. A line yield loss results in wasting human and capital equipment resources that were used in processing the wafer before it was scrapped. Depending on where the loss occurred in the process flow, it may result in delaying shipments to customers. It could be the result of a mistake in terms of misprocessing, it could be the result of defective equipment, or it could be the result of an accident. Each of these evils provided strong motivation for initiating this project. Once the theme of line yield improvement was selected, the next step was to set challenging, but realistic, measurable goals, keeping in perspective the 7 month internship time frame.

Key Project Objectives

To develop and provide line yield management models to enable Intel to attain

(a) faster line yield learning rates for new processes

(b) step function improvements in line yield for mature processes

Key Project Deliverables

- Document the line yield methodologies at each Intel wafer fabrication plant.
  - Collect line yield related data on process, equipment & human elements.
- Compare line yield definitions and trends at each Intel wafer fabrication plant.
- Analyze the line yield management systems at each Intel wafer fabrication plant.
  - Deliver a survey and analysis of the line yield methodology used at each site.
  - Deliver a Pareto of line yield loss causes at various stages of process maturity.
- Develop, refine, and apply conceptual models for line yield improvement.
- Recommend key changes to management systems, structure, policies, and processes, to accomplish world class line yield performance at Intel.

Chapter 2. Data Collection

Purpose: The purpose of data collection is to investigate the specific features of the problem from a wide range of different viewpoints. In other words, this phase identifies factors that are the most important things to investigate for root causes in Step 3 (causal analysis). This chapter discusses the framework I used to collect and analyze line yield related data at each Intel site.

2.1 Introduction

The line yield improvement recommendations and findings discussed in this thesis are based on research performed at Intel's Technology & Manufacturing Group's various technology development and wafer fabrication facilities in New Mexico, Arizona, California, Oregon, and Israel. The nature of this research makes it important to understand the organizational context in which my model was developed and applied. For this reason, this chapter starts with an overview of Intel's Technology & Manufacturing Group. This overview is followed by a discussion of the silicon wafer fabrication process [7,19]. This chapter then discusses the benchmarking [2] framework I used to collect, document, and analyze line yield related data at each Intel site.

2.2 Intel's Technology & Manufacturing Group

At a time when silicon manufacturing is not viewed as strategic by companies that design chips and then have them manufactured outside, Intel's perspective is that manufacturing and process technologies are areas of strategic competitive advantage. According to Intel, having the most advanced process technology is key to being able to design and manufacture leading edge microprocessors. Intel's expertise and continuing investment in process and manufacturing technology indeed help to give it a significant advantage over competitors. Over the years, Intel's manufacturing processes have been refined to allow for chips with smaller circuit sizes, fabricated on larger wafers. The company's first products were made on 2-inch wafers; today, Intel's most advanced products are made on the new 0.6-micron technology process using 8-inch wafers.

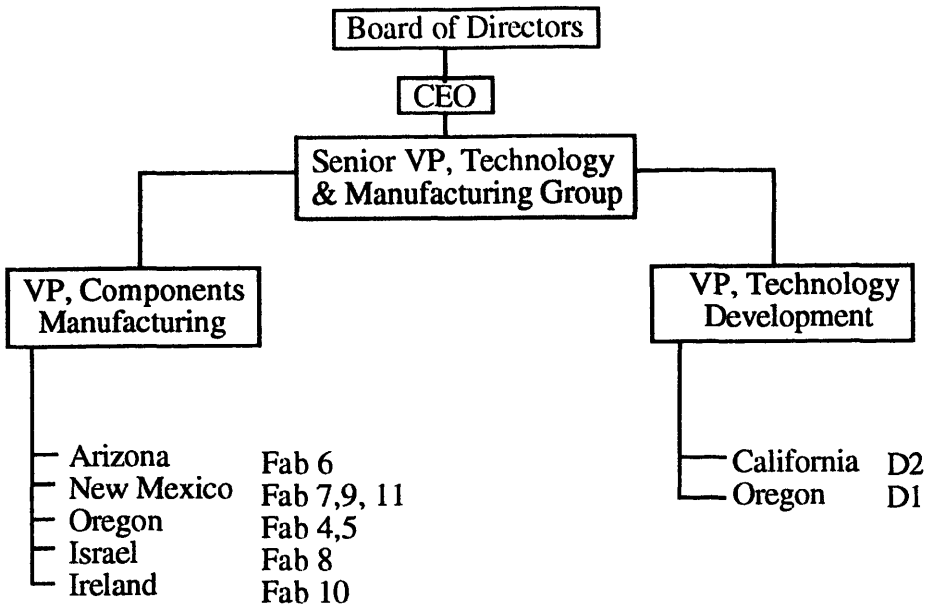


Figure 2.1 TMG (Technology & Manufacturing Group)'s Wafer Fabrication and Process Development Sites & Organizational Structure

The Technology Development (**TD**) sites at Oregon and California (Figure 2.1) have the charter to develop and deliver new processes that yield at "manufacturable levels". In other words, their charter is to bring the new processes they develop, to a point where the die per wafer yields are high enough to make high volume fabrication economically feasible. Historically, the emphasis has been on developing process technology in a stable environment (e.g., at low volumes). When they got the yields up on a new process, the high volume manufacturing sites (**Fabs**) had to replicate the process exactly. The Fabs then had to (1) copy exactly, the processes developed in the technology development sites, and (2) ramp the production volumes and yields.

The overriding goal of meeting yield targets in a TD site directly competes with another activity: surfacing potential problems [1]. Surfacing potential manufacturing obstacles requires stressing a system that is closely aligned with the Fab environment. However, this works against getting the process up and running smoothly in the TD site. Moreover, some problems arise only in a full scale manufacturing environment, which differs from development in many areas including personnel policies, equipment capacity and utilization, and operational scale and procedures. These differences result in Fabs and TDs adopting differing line yield management practices, which are discussed in Chapters 3 & 4 of my thesis. My research was based out of Fab 9 in Rio Rancho, New Mexico.

2.2.1 Fab 9

Although my research and data collection efforts led to significant travel to several Intel sites, my home base was Intel's highest volume wafer fabrication plant, Fab 9, in Rio Rancho, New Mexico. Fab 9 is currently Intel's largest and most profitable logic factory. Fab 9's charter is to manufacture a variety of microprocessors (primarily the 486 and Pentium™ processors on 1micron and sub-micron 6-inch Silicon wafer technologies.

2.3 An Overview of Microelectronic Device Fabrication

Although semiconductor materials have been used in electronics since the early decades of this century, it was the invention of the transistor in 1948 that set the stage for what would become one of the greatest technological advancements in all of history. Semiconductors have become the foundation for electronic devices because their electrical properties can be altered by adding controlled amounts of selected impurity (dopant) atoms into their crystal structures. The electrical operation of semiconductor devices is controlled by creating regions of different doping types and concentrations. Although the earliest electronics devices were fabricated on germanium, silicon has without a doubt become the industry standard.

2.3.1 Crystal Growth and Wafer Preparation

Silicon occurs naturally in the forms of silicon dioxide and various silicates. They must undergo a series of purification steps in order to become the high-quality, defect-free, and single crystal material required for semiconductor device fabrication. Single crystal silicon is almost always obtained by using the Czochralski process [19], utilizing a seed crystal that is dipped into a silicon melt and then slowly pulled out while being rotated. At this point, controlled amounts of impurities can be added to the system to obtain a uniformly doped crystal. The result of this growing technique is a cylindrical single-crystal ingot, typically 50-200 mm (2-8 inches) in diameter and over 1 m in length. Next, the crystal is sliced into individual wafers about 0.5 mm thick. Finally, the wafers must be polished and cleaned to remove surface damage caused by the cutting process. *These silicon wafers provide the starting material at Intel.* Device fabrication takes place over the entire wafer surface, and hundreds of identical circuits are generated on each wafer at the same time. Once processing is completed, the wafer is sliced into individual "chips", each containing one complete microprocessor.

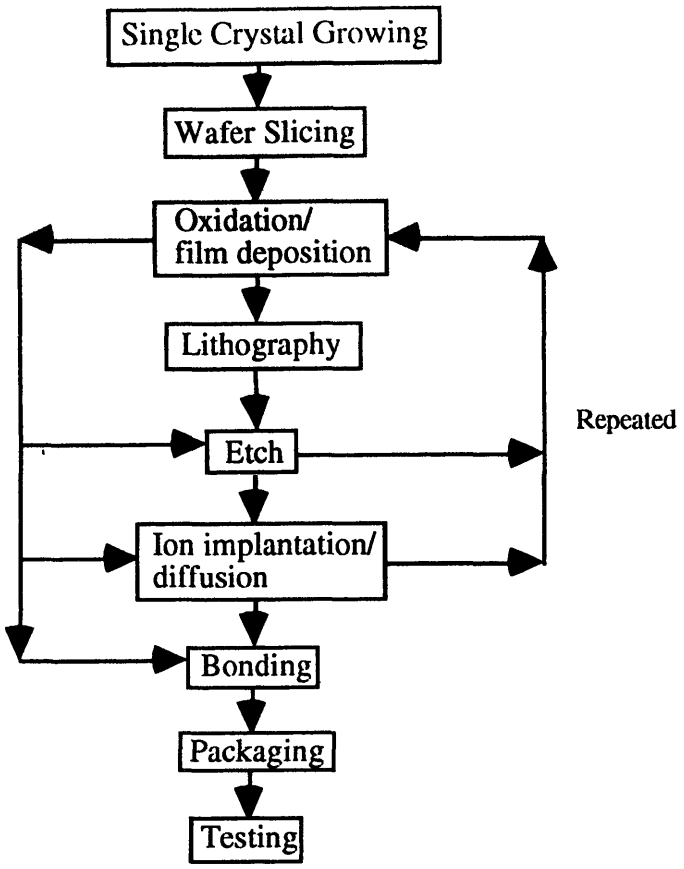


Figure 2.2. General fabrication sequence for integrated circuits.

Figure 2.2 outlines the processes used in fabricating microelectronic devices and integrated circuits. This section also briefly discusses each of the major fabrication steps [7,9, 19].

2.3.2 Film Deposition

Films of many types, particularly insulating and conducting, are used extensively in microelectronic device processing. Examples include polysilicon, silicon nitride, and silicon dioxide. Some of the major functions of deposited film include masking for diffusion or implants, and protection of the silicon surface. In masking applications, the film must effectively inhibit the passage of dopants while also displaying an ability to be etched into patterns of high resolution. Upon completion of device fabrication, films are applied to protect the underlying circuitry. Conductive films are used primarily for device

interconnection. Film deposition techniques involve a variety of temperatures, pressures, and vacuums such as evaporation, sputtering, and chemical vapor deposition (CVD).

2.3.3 Oxidation

Oxidation refers to the growth of an oxide layer by the reaction of oxygen with the substrate material. Silicon dioxide is the most widely used oxide in IC technology today. Typical functions are dopant masking, device isolation, and surface passivation. Modern IC technologies require oxide thicknesses in the hundreds and thousands of angstroms. Dry and wet oxidation techniques are used for coating the entire silicon surface. Selective oxidation is used to oxidize only certain portions of the substrate.

2.3.4 Lithography

Lithography is the process by which the geometric patterns that define the devices are transferred from a mask to the substrate surface. In current practice, the lithographic (or "litho") process is applied to the wafer surface several times, each time using a different mask to define the different areas of the working devices. Once the film deposition process is completed and the desired masking patterns have been generated, the wafer is cleaned and coated with an emulsion, called photoresist (PR), which is sensitive to ultraviolet light. The wafer is then aligned under the desired mask in a mask aligner. This crucial step is called "registration". Once the wafer and mask are aligned, they are subjected to UV radiation. Upon development and removal of the exposed PR, a duplicate mask pattern will appear in the PR layer.

One of the major issues in lithography is **line width**, which refers to the width of the smallest feature obtainable on the silicon surface. As circuit densities have escalated over the years, device sizes and features have become smaller and smaller. Today, minimum commercially feasible line widths are 0.8 to 0.6 microns.

2.3.5. Etching

Etching is the process by which entire films or particular sections of films are removed, and it plays an important role in the fabrication sequence. An important criterion in selecting an etchant is its selectivity, which refers to its ability to etch one material without etching another. An etching process must effectively etch the silicon

dioxide layer with minimal removal of the underlying silicon or the resist material. In addition, polysilicon and metals must be etched into high resolution line with vertical wall profiles and minimal removal of the underlying insulating film.

2.3.6 Diffusion and Ion Implantation

The electrical operation of microelectronic devices depends on regions of different doping types and concentrations. The electrical characteristics of these regions is altered by introducing dopants into the substrate by the *diffusion* or *ion implantation* processes. This step in the fabrication sequence is repeated several times, since many different regions of microelectronic devices must be defined.

In the diffusion process, the movement of atoms results from thermal excitation. Dopants can be introduced to the substrate surface in the form of a deposited film, or the substrate can be placed in a vapor containing the dopant source. The process takes place at elevated temperatures, usually 800-1200 degree centigrade. Dopant movement within the substrate is strictly a function of temperature, time, and the diffusion coefficient of the dopant species, as well as the type and quality of the substrate material.

Ion implantation is a much more extensive process and requires specialized equipment. Implantation is achieved by accelerating ions through a high voltage field of as much as one million electron-volts and then choosing the desired dopant by means of a mass separator. The major advantages of ion implantation including a high degree of anisotropy and wider range of dopant species has made it the dominant technique in modern IC technology.

2.3.7 Metallization and Testing

Generating a complete and functional integrated circuit requires the individual devices fabricated on the wafer surface, to be interconnected. Interconnections are made by metals that exhibit low electrical resistance and good adhesion to dielectric insulator surfaces. Aluminum, aluminum-silicon-copper alloys, and tungsten are the materials commonly used for this purpose in VLSI technology today. The metal is deposited by standard deposition techniques, and interconnection patterns are generated by lithographic and etching processes. Modern ICs can typically have 1-4 layers of metallization, in which case, each layer of metal is insulated by a dielectric. Layers of

metal are connected together by vias and access to the devices on the substrate is achieved through contacts.

Wafer fabrication is complete upon application of a passivation layer, and the next step is to test each of the individual circuits on the wafer. Each chip (or die) is tested with a computer controlled probe platform that contains needle-like probes to access the aluminum pads on the die. The platform steps across the wafer, testing whether each circuit functions properly with computer generated simulations. If a defective chip is encountered, it is marked with a drop of ink. After completion of this primary testing, each die is separated from the wafer using diamond sawing. The chips are then sorted.

2.3.8 Bonding, Packaging, and Final Testing

The working dies must be attached to a more rugged foundation to ensure reliability, and they must be accessible to electrical connections to the outside world. Bonding pads are located around the perimeter of the die. The most popular method of attaching these pads is by "wire bonding", which utilizes very thin gold or aluminum wire. The pads are attached by thermocompression, ultrasonic, or thermosonic techniques.

The connected circuit is now ready for final packaging. Packages are available in a variety of styles. Consideration of a circuit's package includes chip size, number of external leads, operating environment, heat dissipation, and power requirements.

After the chip has been sealed in the package, it undergoes final testing. Because one of the main purposes of packaging is isolation from the environment, testing at this stage usually encompasses heat, humidity, mechanical shock, corrosion, and vibration. Destructive tests are also performed to investigate the effectiveness of sealing.

2.4 "Systems thinking" as a framework for data collection

Data collection can be approached from various perspectives as shown in Figure 2.3.

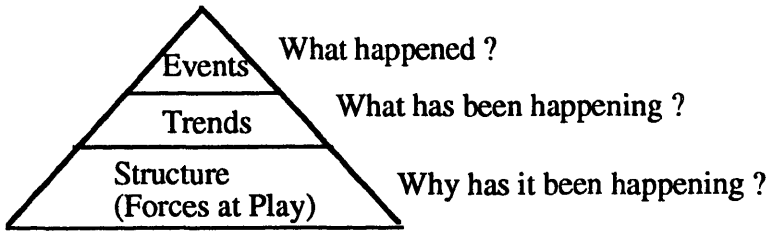


Figure 2.3. Events, Patterns, Structure

Event level: Data collection at this level helps answer the questions: "What happened? What crisis are we facing?". An example would be line yield loss events. Much of the attention is on daily events, responding and reacting to something that has happened. This level of data collection and action provides very little power to alter the course of these events. **The best hope is for an optimal reaction.**

Trend or pattern level: Data collection at this level answers the questions: "What has been happening ?". An example would be a deteriorating line yield trend in a Fab. This mode helps us to take a step back and look at trends, hoping to understand why a recent event has happened based on a number of events in the past. This level of data collection provides us very little power to alter the course of patterns and trends. **The best hope is that the trends continue and the manager optimally responds to them.**

Structural level: Data collection at this level leads us to answer the questions: "Why has it been happening?". An example would be policies affecting line yield improvement. By considering the relationships that exist among the components of the entire organization, this level of data collection fosters a more holistic approach to developing coherent solutions. Structural explanations address the underlying causes of behavior at a level that patterns of behavior can be changed. **Structure produces behavior, and changing underlying structures can produce different patterns of behavior [16].**

In this thesis, I have taken a systems thinking approach to data collection and analysis, which has revealed that underlying structure [16] does drive line yield behavior, and that

maximum leverage comes from changes at the structural level. Aspects of structural level data collection are illustrated in Figure 2.4.

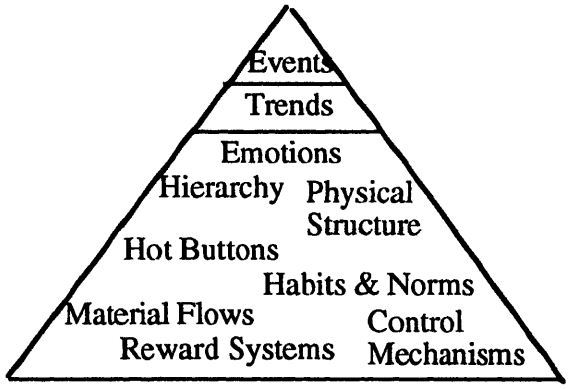


Figure 2.4. A Systems Framework for Data Collection

Applying this structural level approach to collecting data specifically for line yield improvement revealed several underlying structural facets driving line yield behavior. The major structural forces shaping line yield behavior are shown in Figure 2.5. The associated *causal loop diagrams* are shown and discussed in chapter 3.

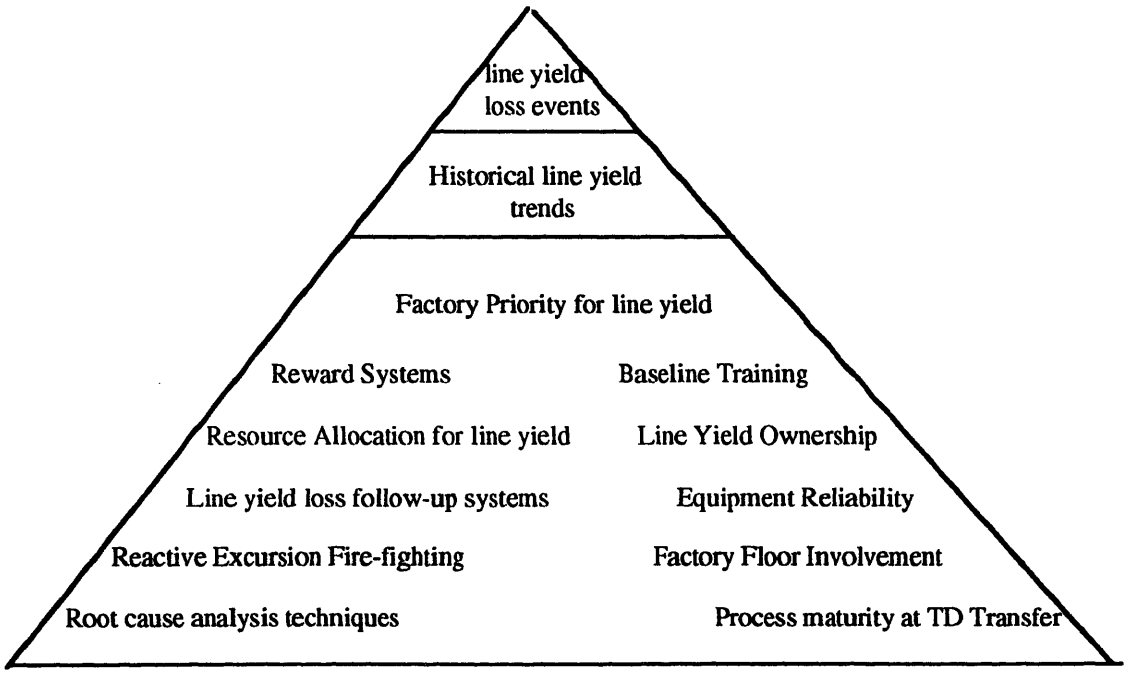


Figure 2.5. A Systems Perspective to Line yield data collection

A number of approaches were used to arrive at the list of structural forces described above: face-to-face interviews (both structured and unstructured), remote conversations (via telephone and e-mail), attendance of meetings, site visits, and plant tours. I conducted interviews on a 1-1 basis, and interviewed plant managers, manufacturing and engineering managers, team-leaders, engineers, shift managers, shift supervisors, operators, and technicians on the factory floor. I attended numerous meetings that enabled me to understand several operating policies in the factories. I also co-chaired a ten member inter-site line yield benchmarking team.

2.5 Line Yield Methodology Data Collection Questionnaire

Based on the above knowledge sources, I designed a line yield methodology questionnaire, which I then used to collect data and document the line yield improvement methodology of each Intel Fab and TD site. One of my deliverables for Intel was a comparison of its factories along several dimensions related to line yield. **Since the data is proprietary to Intel, it is not included in this thesis document.** However, the questionnaire, and the format I used for comparing the various factories is described in this chapter. Both the questionnaire and the comparison format can be adopted by any semiconductor manufacturing firm, to analyze the line yield methodologies followed in their factories. The questionnaire and the comparison format are as follows:

Line Yield Methodology
Questionnaire

Viju Menon

General Fab Information

Name:
Location:
Age of Fab:
Process(s):

Process Maturity & LY Historical Trends for each process

process name:

Startup date:

Please provide the following graphs

Wafer Starts Per Week (WSPW) versus Time graph:

Line Yield versus Time graph:

Line Yield Loss cause Pareto from startup by quarter

- by process, people, equipment
  - a) For the whole factory
  - b) For each functional area
    - Lithography
    - Etch
    - Thin Films
    - Diffusion/Implant

Die Yield (ISO) versus Time graph

1. Please comment on the above Paretos to explain excursions/abnormal behavior, if any.

2. Please describe in a paragraph how the loss cause paretos have changed since ramping up from a minimum volume threshold (say 1000wspw). Both, the loss cause contribution by a) process, people & equipment and b) by cluster. What improvements led to these changes in the paretos historically.

3. Please list historical key losses in LY, what methods were adopted to deal with these, and the success of these methods, so as to provide learnings to other factories.

Existing Line Yield Analysis System

a. Definitions

1. How is Line Yield defined and calculated? (please provide equations).
   What are the factory floor line yield metrics? e.g. %LY, Losses per 10,000 activities?

2. Do line yield calculations include
   a. monitor, short loop and engineering losses?
   b. sort zero yield wafers ? low yielding wafers?
   c. ETEST losses?
   d. losses due to natural disasters (e.g. power loss)?
   e. miscellaneous/other?

3. Do die yield calculations include
   a. all wafers that make it to sort?
   b. only the wafers that are shipped out?
   c. miscellaneous/other?

4. Where are the sort limits set? How are these established?

b. Data collection, tracking, and analysis

1. What is the database system used to track wafer losses? (e.g. LCS)

2. Please describe the documentation process while scrapping wafers. Does it depend on number of wafers scrapped? Who receives this documentation? Please include samples of document(s) generated when a wafer is scrapped.

***c. Disposition methods** (to get a flavor of the operating philosophy)*

1. Who is allowed to scrap material in-line? Who actually does the scrap? Technicians, Operators or Engineers? (To get a flavor of whether it is more objective or judgemental) Are spec limits for scrap well defined?

2. Are spec limits for scrap well defined? Are spec limits established using statistical bin limits, or are they true spec limits?

3. Do you use statistics in setting disposition limits in the fab, etest, and sort areas? Are wafers out of SPC control always scrapped?

3. If a wafer fails spec, is it always scrapped? Do you challenge specs with each scrap? If not, how often is it done?

4. Are wafers out of spec passed and out into quarantine? How many wafers are currently (or typically) in quarantine?

5. Do you tend to pass wafers on and screen them at e_test & sort?

6. When production wafers get scrapped, do they get downgraded to engineering wafers? If so, how does line yield get affected? Is this a line yield loss?

7. Do you have a procedure for bonusing or upgrading wafers back into production?

***c. Follow-up or Postmortem System** for the documented losses*

1. Who tracks the documented losses (e.g. first order scrap reports)?

2. What is the forum where the documented losses get addressed? Who owns this postmortem co-ordination? What are the roles & responsibilities of the members of this forum? How often are they conducted?

3. Who presents at the postmortem about the losses? How do the presenters view the forum? (intimidatory ?) Typical Agenda? Do they discuss both line yield & die yield in this forum?

4. What is the threshold (number of wafers lost) for a loss to be investigated?

5. What is the follow-up system for action items generated at the postmortem? Please include a flowchart of the documentation, postmortem, and follow-up processes?

6. Do the people doing the scrap documentation get feedback on the analysis of the scrap & root cause fixes? Is there a feedback loop?

7. What in the system ensures a closed loop, between generating documents for losses, and implementing fixes for the losses documented?

e. Root Cause Analysis

1. What are the conditions under which root cause analysis is done for scrap?

2. What is used to prioritize what line yield problems are worked on? How do you distinguish between an excursion mode issue and a baseline issue? How do resourcing and root cause analysis systems work for excursion and baseline issues?

3. What is the forum at which root cause analysis gets done?

4. Is there a documented structured procedure for root cause analysis ? How do you establish causality ? Is it empirical evidence, statistics, FMEA, seven step method, or is the root cause analysis an adhoc procedure?

5. Do e_test & sort losses get billed back to the appropriate clusters ? How is it done ? Who does the causal analysis for e_test & sort losses?

6. Please describe a typical real life scrap incident and provide a walk through of how the system operates (from the scrap decision, documentation, through root cause analysis and fix implementation)?

f. Implementing Root Cause Fixes

1. What is the forum where implementing root cause fixes gets addressed?

2. What is used to prioritize what line yield fixes are implemented?

3. Please address ownership, responsibility & accountability for the question below.
   What actions are taken when the fix is identified to be
   a) process related
   b) equipment related
   c) procedural related

4. Please give specific examples for each of the above in a paragraph each.

g. Line Yield Improvement Efforts

1. How would you define roles and responsibilities of each of the functional departments in line yield improvement?

2. What *reactive & proactive* measures are taken by these departments?

3. Is each of these departments in a reactive or pro-active mode of operation?

4. Are there formal line yield improvement programs? Is there a dedicated individual/group that owns driving line yield improvements?

Departments:

Manufacturing	Industrial Engineering	Yield Engineering
Process/Eqpt Eng:	Training	Automation
Factory Support	Functional Area Teams	Other

5. Please comment on any suggestions you may have on improving line yield?

h. Line Yield related Success Indicators

For each of the departments listed above,

1. Is Line Yield a CSI (Critical Success Indicator) for that department?
2. If it is a CSI, How is it communicated ? (e.g., dept meetings, passdown)?
3. In what forums does line yield get talked about in each department?
4. What are the top three priorities for each department?
5. What is the budget ($) for line yield improvement? What are the quarterly goals?
6. What are the positive incentives (reward systems) for line yield performance?
7. What if any are the negative incentives (disciplinary systems) for not meeting line yield goals?

i. Communication Strategy for Line Yield

1. Is overall line yield well known on the factory floor?
2. Does each equipment cluster know what their contribution to line yield loss is?
3. What is the communication mechanism for people on the factory floor? Is it part of the normal visual experience e.g. (display boards, shift passdowns) or is it just on the computer (e.g., vax, bulletin boards) that people have to look up?
4. Does each cluster have a line yield goal? Do the people know what their goal is?
5. Who establishes line yield goals for each cluster? Who drives this goal?
6. Do the factory personnel generating scrap reports feel comfortable about entering the correct loss cause? Has the message been conveyed to them that they wont get disciplined, and that the report will be used for identifying the correct root causes?
7. Are there any incentives for people to be frank & enter the correct loss code?

Miscellaneous

1. What systems are in place for fabs to learn from each other to avoid problems faced and solved by the other fabs?

2.6 Inter-Fab Line Yield Benchmarking Framework

Once the answers to the questionnaire were documented for every Fab, the next step was to compare the line yield methodologies adopted in the various factories, and to document the comparison framework in a concise format. The format I used to compare various aspects of line yield, such as definitions, management methods, feedback systems, communication strategies, etc. are outlined below:

2.6.1. Inter-Fab Comparison of Line yield definitions:

A fundamental issue in benchmarking is to make sure we are "comparing apples to apples". Thus, when comparing line yield performances of various factories, it was crucial to ensure that each Fab was calculating line yield in the same way.

Table 2.1. Inter-Fab Comparison of Line Yield Definitions:

Factory Line Yield calculation includes	Fab A	Fab B	Fab C	Fab D	Fab E	Fab F	Fab G	Fab H	Fab I
Monitor wafers									
Shortloop wafers									
Engineering wafers									
Production wafers									
Electrical-test wafers									
Sort low yield wafers									
Sort Zero yield wafers									
Wafer losses due to facilities losses									
Wafer losses due to Natural disasters									

When presented with a trend graph showing the line yield % achieved at several factories over a period of time, it is important to consider how each factory arrived at their line yield numbers. The first column in Table 2.1 shows the various types of wafers that may or may not get included in each Fab's line yield calculations. Each entry in Table 2.1 can contain either a 'Yes' or 'No', indicating what kinds of wafers are included in line yield calculations at each factory. The specific answers from each site are not included due to confidentiality reasons. However, this format would be useful to compare line yield performances in any bench marking effort, either within another company, or among other companies.

2.6.2. Inter-Fab Comparison of Line Yield Loss Follow-up Methodologies:

Line yield management methodologies varied on several aspects across the various Fabs. A critical component of any line yield methodology is the mechanism used to keep track of wafer losses, following up on analyzing causes of major losses, and implementing root cause fixes. In other words, ensuring a closed loop follow-up system where root cause fixes will get implemented at least for major problems selected for investigation. Table 2.2 represents the format I used for comparing the follow-up systems at the various Fabs.

2.6.3. Inter-Fab Comparison of Line Yield Communication strategy & incentives:

Another critical component of any line yield methodology is the communication strategy and incentive structure used to increase awareness, involvement, ownership, and accountability of factory floor personnel in line yield improvement. Table 2.3 represents the format used to compare such facets of line yield methodologies at the various Fabs.

Table 2.2. Inter-Fab Comparison of Line Yield Loss Follow-up Mechanisms:

LY Loss tracking & Follow-up systems	Fab A	Fab B	Fab C	Fab D	Fab E	Fab F	Fab G	Fab H	Fab I
What Databases are used to track losses?									
Who receives the loss notification?									
Wafer loss threshold over which losses are documented?									
Who is responsible for tracking these documents?									
What are the forums where these losses are addressed?									
Are Postmortems conducted for losses?									
Who coordinates the Postmortems?									
Who attends the Postmortems?									
How frequently are Postmortems held?									
Who presents losses at the Postmortems?									
Loss threshold above which Postmortems are conducted?									
Who follows up on tasks assigned at Postmortems?									
Is the Follow-up loop closed & effective?									

Table 2.3. Inter-Fab Comparison of Line Yield Communication Strategy & Incentive Structure:

Facets of Strategy	Fab A	Fab B	Fab C	Fab D	Fab E	Fab F	Fab G	Fab H	Fab I
Is line yield well known on the Floor?									
Does each cluster know their contribution to LY?									
What is the mode of communicating LY to the Factory floor?									
Does each equipment cluster have a line yield goal?									
Do people in each cluster know their line yield goal?									
What is the metric for measuring LY? (%LY, Loss/10K)?									
Who drives cluster line yield goals?									
Are there any positive line yield rewards?									
Are there negative incentives for poor LY performance?									
Are there any formal LY programs?									

Chapter 3. Causal Analysis

Purpose: The purpose of causal analysis is to find out what the main causes of the problem are. This chapter identifies fundamental issues that contribute to (or limit) Intel's line yield performance.

3.1 Introduction

This chapter analyzes the data collected via the line yield questionnaire, and arrives at root causes of line yield losses that cause slow line yield learning rates. The two tools I used were Ishikawa (Fishbone) diagrams and Systems Thinking[16]. I used Ishikawa diagrams to arrive at the first order causes. Then, I used causal loop diagrams and systems thinking principles to arrive at the fundamental issues underlying line yield losses.

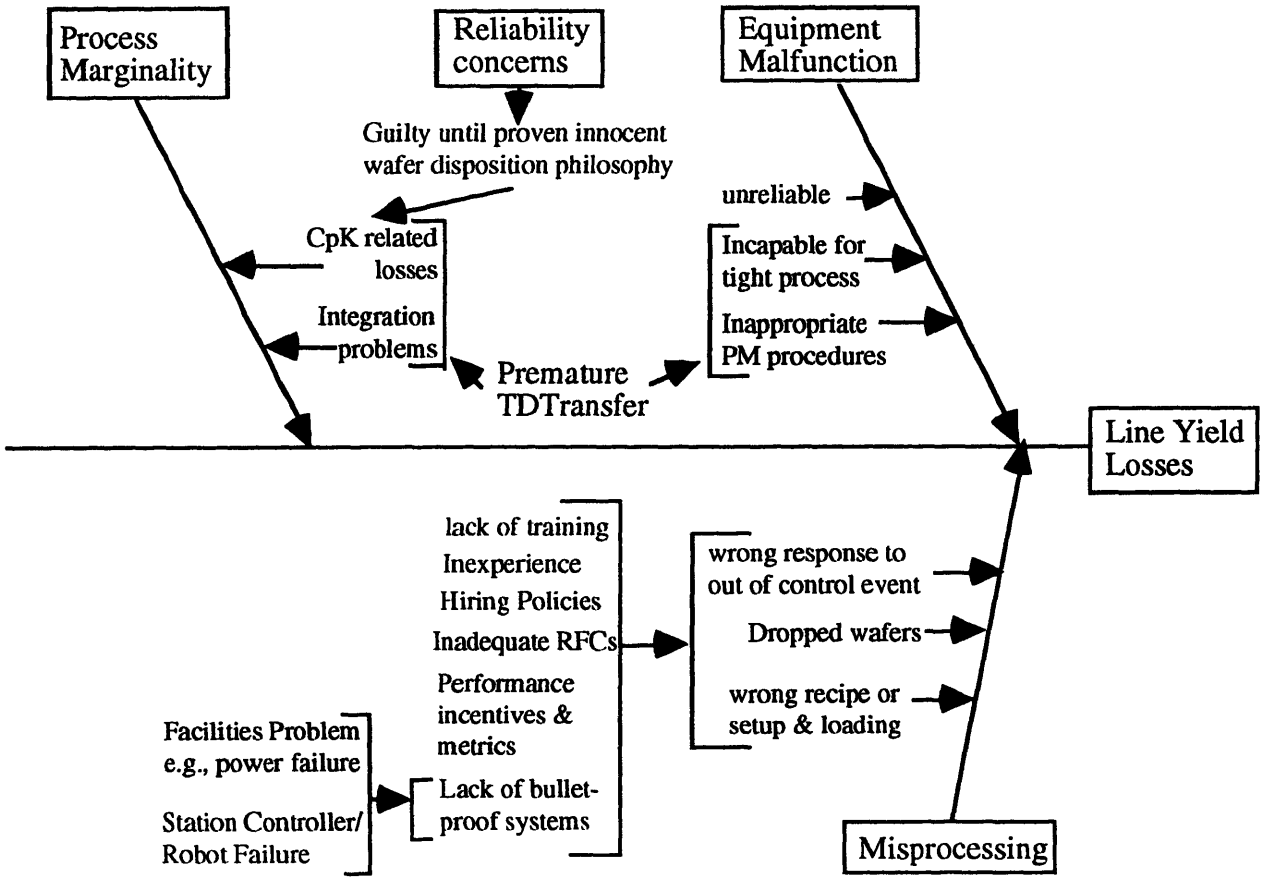


Figure 3.1. Major Causes of Line Yield Losses

3.2 Fishbone Diagram Analysis of line yield loss causes

Figure 3.1 shows my first order Fishbone diagram analysis of the major factors causing line yield losses. As this diagram indicates, line yield is indeed a very complex issue with multiple interactions among loss causes. This Fishbone diagram does not provide, and is not intended to provide, a comprehensive representation of all the underlying causes of line yield losses. Instead, it is intended to give the reader a flavor of the major line yield loss causes, the complexity of the problem, and to set the stage for the in-depth discussion following in the next section.

3.3. Systems Thinking analysis of line yield loss causes

All line yield losses are typically attributed to one of three major causes (a) equipment malfunction (b) process marginality (c) misprocessing, or in other words, human error. However, when we dig deeper, we find that these categories are not really causes in themselves, but are only symptoms of underlying problems. For instance, consider a line yield loss incident caused by an operator running the wrong recipe for a process step. Our typical response would be to attribute the loss to irresponsible operator behavior, and to categorize the loss as "human misprocessing". A typical remedial action would be to counsel the operator. However, the fundamental reasons may not lie with the operator at all. Maybe the operator had to juggle various lots of wafers each of which required a different recipe, in which case, it would be very easy to mix up lots and run the wrong recipe. The long term fix maybe lies in installing a station controller at that work-station, not in counseling the operator.

In fact, I would hypothesize that the vast majority of line yield problems are caused by systemic issues that transcend simple explanations such as "equipment failure", "process immaturity", and "human screw up". The following section highlights several underlying reasons for line yield losses and slow line yield improvement rates exhibited at wafer fabrication plants.

How do factories approach line yield improvement ? Has it been a high factory priority ?

The aim of a factory is to maximize revenue and profits. At Intel, where market demand is not the constraint, this aim is achieved by maximizing die shipments. The three levers for maximizing shipments are (a) improving capacity by installing new equipment and by

improving equipment availability and utilization (b) improving die yields (c) improving line yields. The simple causal loop diagram (Figure 3.2) illustrates the expected behavior of this policy. An arrow labeled '+' connecting 2 variables indicates that a change in one direction of the first variable results in a change in the same direction of the other variable. For example, Increasing the EDO goal in the diagram below will increase the EDO Gap. An arrow labeled '-' indicates that the variables change in opposite directions. For example, Increasing 'Current actual Die Output' decreases the 'EDO Gap'.

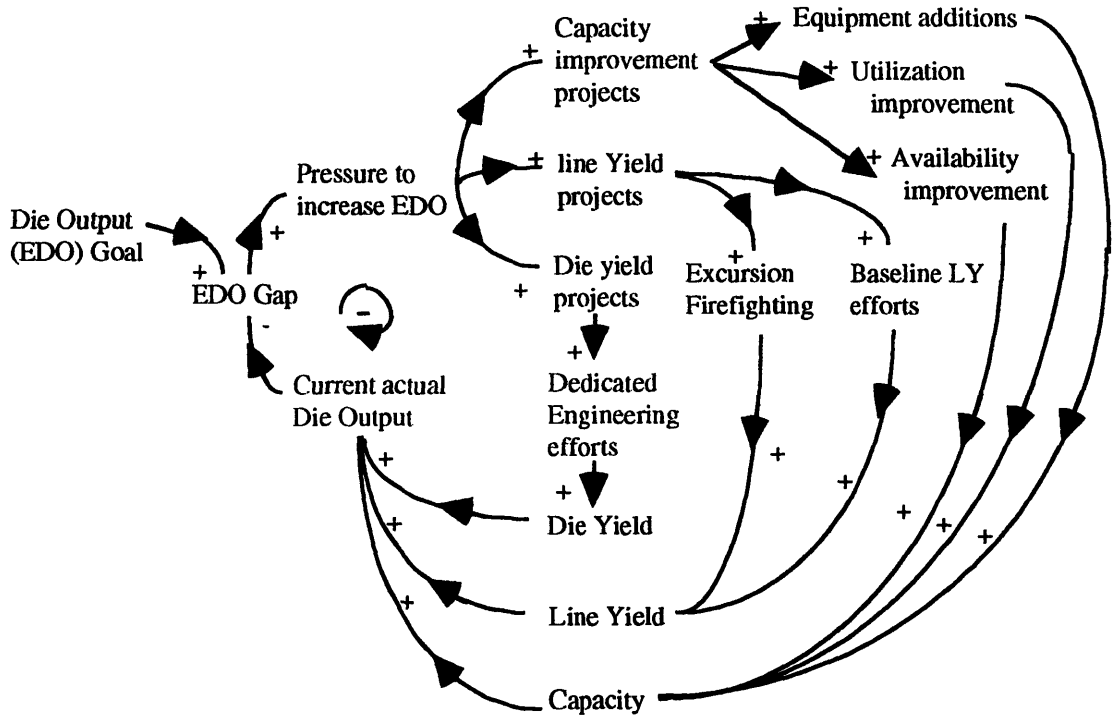


Figure 3.2. Expected Behavior of EDO Policy

However, since improvement opportunities may exist in all three areas (capacity, line yield, die yield), and since a Fab has only limited resources to work on improvement projects, the plant management prioritizes improvement projects according to their impact on maximizing die output. Typically, every quarter, the senior management in a factory prioritizes potential capacity, die yield, and line yield improvement projects based on quantifiable incremental die output resulting from the completion of each project. EDO is "equivalent die out", defined as the total weekly output of the factory measured by normalizing all good die to the equivalence of a single product. The goal of the factory is to maximize EDO.

EDO = Min (WSi) x DPW x LY x DY

WSi: Output of the ith station in WSPW (Wafer Starts Per Week),
Min(WSi) is the output of the bottleneck station which limits the factory.
DPW: Number of potential good die per wafer.
LY: Factory Line Yield in % = product of line yields at each station in the process flow.
DY: Process (or Product) Die Yield in %.

The inputs to the evaluation process are (a) Line Yield loss Pareto for the previous quarter, which identifies the potential line yield improvement projects. Only quantifiable and realizable line yield projects are chosen for consideration. e.g., projects to increase line yield at a particular functional area such as lithography or diffusion get chosen, while projects to reduce procedural errors across the entire factory do not get chosen, because of its wide scope and difficulty of implementation (b) Die Yield loss Pareto for the previous quarter (c) Capacity limiter Pareto, which identifies the existing potential capacity improvement projects.

The evaluation process is as follows: DPW is held constant throughout the process. For determining the EDO opportunities of each capacity project, the other variables in the EDO equation, LY and DY, are held constant (say, at 100%). Similarly, while calculating the EDO impact of LY projects, DY and Capacity are held constant, and while determining the EDO impact of DY projects, LY and Capacity are held constant.

The output of the evaluation process is a Pareto, ranking each potential improvement project in terms of its EDO impact. This Pareto is called the EDO Pareto. The limited improvement resources available (in terms of human and capital resources) are applied to the biggest Pareto items. Typically this results in several low EDO leverage improvement projects not being addressed due to lack of resources.

As part of my research, I studied the EDO Paretos at several Fabs, at several stages of process and factory maturity. Historically, right from factory start-up, all the way until after ramp, the EDO Pareto (Figure 3.3) indicates that capacity and die yield improvement projects have higher die output leverage than most line yield improvement projects.

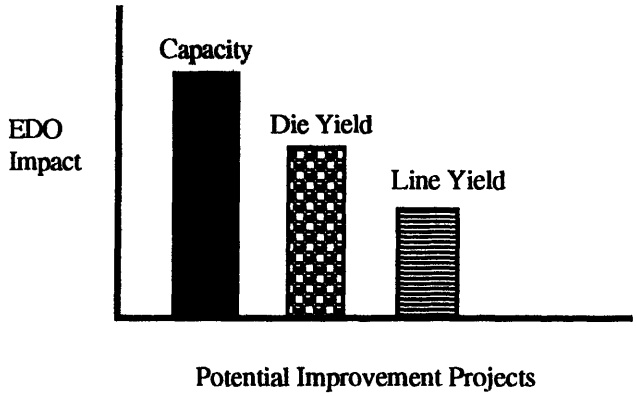


Figure 3.3 A Typical Factory EDO Pareto

Possible reasons why Line Yield is behind capacity and die yield on the EDO Pareto:

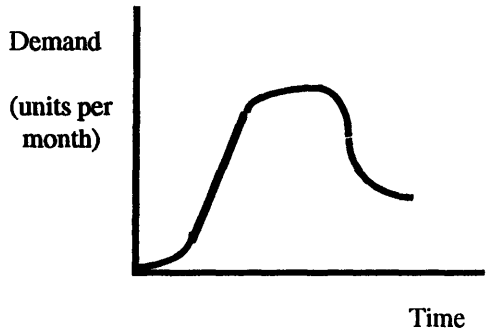


Figure 3.4

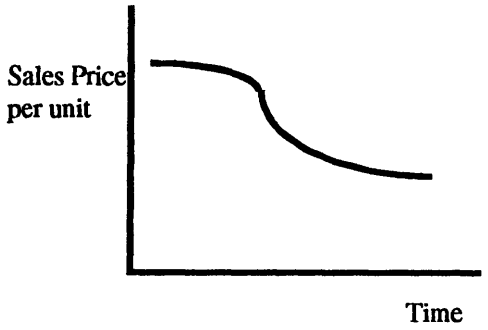


Figure 3.5

Intel's bread-and-butter lies in its leading edge microprocessors. Figure 3.4 shows a typical x86 microprocessor demand life cycle. Figure 3.5 traces the average sales price of a microprocessor at various stages in its life cycle. Most of the unit revenues generated for a new product occur soon after its introduction while demand for the product is booming. Due to the absence of clones at this stage, Intel can command high prices. Operating margins are highest in that time period. Demand typically outstrips supply at this stage, and there is tremendous pressure to increase factory output since revenue and profits are directly proportional to factory die shipments.

This leads the factory management to search for EDO improvement opportunities that quickly increase EDO with a high level of certainty. Note that "quick *returns*", *"high*

probability of success" and *"tangible effort requirement"*, are critical in selecting potential EDO improvement opportunities. These may not correspond to projects that fetch the "most bang for the buck", especially in cases where cost-effective projects do not fetch quick returns, or when cost-effective projects require wide-spread and not focused efforts. This explains the bias towards capacity and die yield improvement projects, and the bias against line yield improvement projects.

Although adding a new machine to increase die shipments may not be as cost-effective as improving yields, the resulting increase in capacity will fetch a faster and greater increment in EDO. Especially for state-of-the-art processes, die yield improvement entails focused engineering efforts, and although such projects have a high probability of success, they are time consuming, and not as easy to achieve as adding another machine. However, die yield efforts are still preferred to line yield efforts by factory management because line yield improvement is even more difficult to achieve in a short time frame than die yield. This is due to the fact that line yield is a complex function of procedural (human) errors, operating philosophies, equipment reliability, and process marginality, while die yield is primarily an engineering intensive problem with very little procedural error intricacies.

The bias towards identifying EDO improvement projects that (a) have quick returns (b) require focused effort (c) have high probability of success have a big influence on what kind of line yield projects are even considered to be included on the EDO Pareto. The Factory Wafer Scrap Pareto for the previous quarter (3 months) provides the list of potential line yield projects to be considered as inputs to the EDO evaluation process. At first glance, this appears okay, because it provides a list of the top line yield problems existing in the factory. Only those line yield projects requiring focused effort (mostly excursions), that are *perceived as realizable* within a relatively short time horizon (1-2 quarters) are considered. This Pareto contains individual station (or functional area) line yield losses, and does not include high leverage (baseline) line yield projects of global scope, such as reduction of procedural errors across the factory floor. For the most part, line yield excursions make it higher in the EDO Pareto than baseline losses that are more spread out across functional areas.

Implications of the EDO "low priority" policy on line yield:

The causal loop shown in Figure 3.2 shows the expected behavior of the EDO policy. However, in reality, there are various interactions that are not reflected in that diagram. The low EDO line yield priority causes various dynamic interactions, some of which are reflected in Figure 3.6.

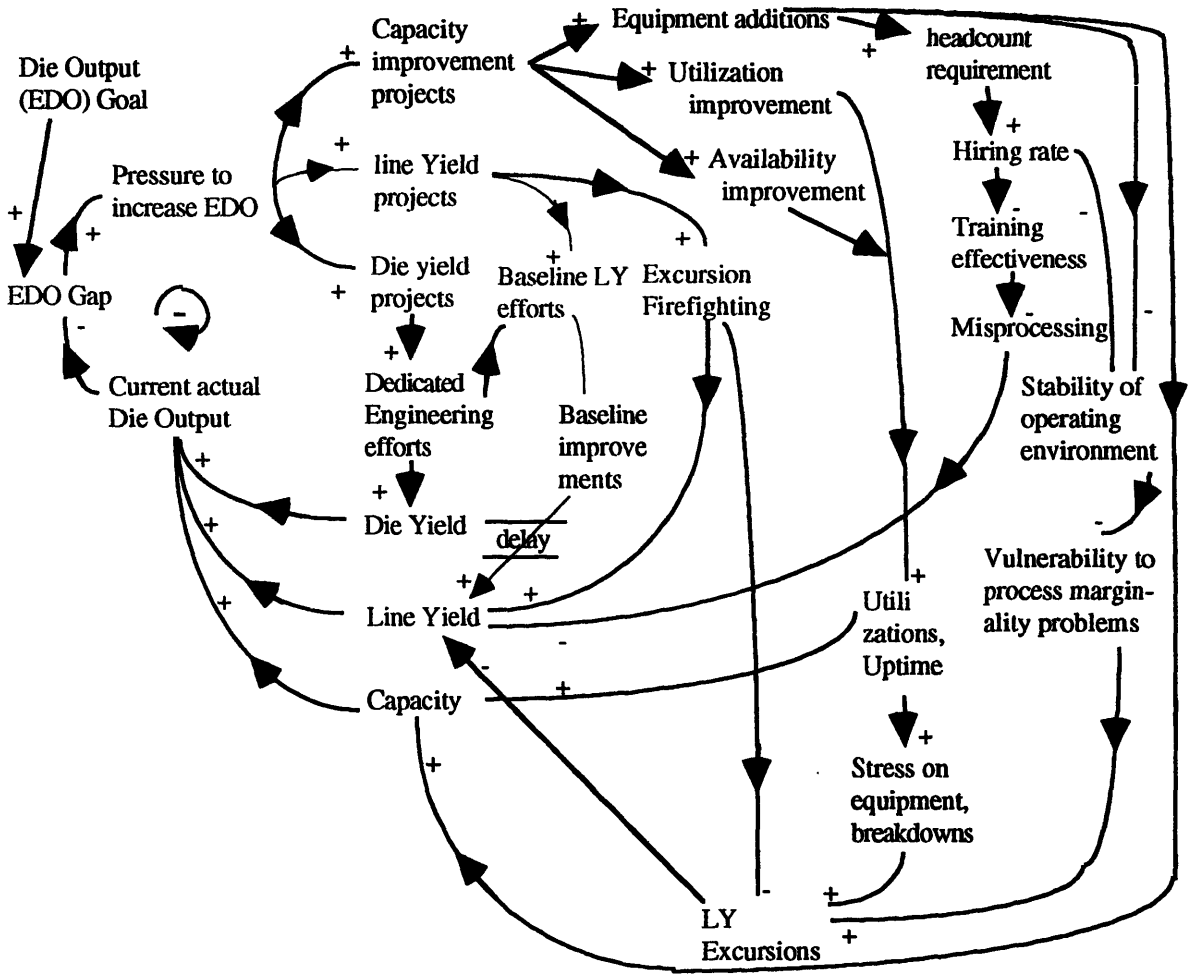


Figure 3.6. Actual Behavior due to low line yield EDO Prioritization

Due to the lower EDO line yield priority, initially more emphasis is given to improving die yield and capacity. As the above diagram indicates, this emphasis on die yield and capacity negatively impacts line yield performance in several ways. Rapid capacity expansion results in lower stability of operations in the ramp phase, which increases the

occurrence of line yield excursions. High levels of equipment uptime and utilization often leads to irregularities in carrying out preventive maintenance procedures, and increased stress on the machinery, which in turn leads to line yield losses. Emphasis on die yield improvement occupies significant engineering resources, leaving very few resources for baseline line yield improvement. All these reasons induce reactive fire-fighting. Another factor is the long delay involved between initiating baseline efforts and seeing tangible improvement in line yield as a result of these baseline efforts. This delay promotes even more emphasis on quick "band-aid" fixes, promoting excursion fire-fighting.

Since factory priorities are driven by the EDO Pareto, the apparently low line yield leverage influences management's mindset towards line yield improvement in the start-up and ramp stages of the process, and results in lower factory priority for line yield. Figure 3.7 illustrates the "success to the successful" system dynamics archetype where two activities compete for limited resources. The more successful one becomes, the more support it gains, thereby starving the other. The low priority for line yield results in most of the resources being given to capacity and die yield, which increases the success of die yield and capacity improvement activities. This low priority for line yield also translates into:

- fewer resources to work on line yield improvement (Process & Yield Engineering resources are mostly dedicated to die yield; Production personnel and Industrial Engineering concentrate on capacity improvements).
- reactive excursion mode fire-fighting approach to line yield improvement.
- lack of ownership, accountability, awareness, and involvement of factory personnel in line yield improvement.
- very low emphasis on baseline line yield improvements until after ramp.

Baseline improvements receive low priority during the startup and ramp stages. However, systematic fixes to line yield problems come from baseline improvements and not from fighting excursions. Meanwhile as the ramp stage nears completion, sizable improvements would have been accomplished in capacity and die yield, and further efforts in these sectors would only fetch diminishing marginal returns. It is at this stage that line yield projects climb to the top of the EDO Pareto. However, when line yield does become a higher EDO lever after the ramp phase, there are no systems in place to deal with it. Hence a low learning rate for line yield performance is experienced. This phenomenon is represented by the balancing loop in Figure 3.7.

Food for thought: what kind of priority should be given to line yield improvement ? Is there a way of improving line yield learning rates without adversely affecting die yield and capacity improvement rates ?

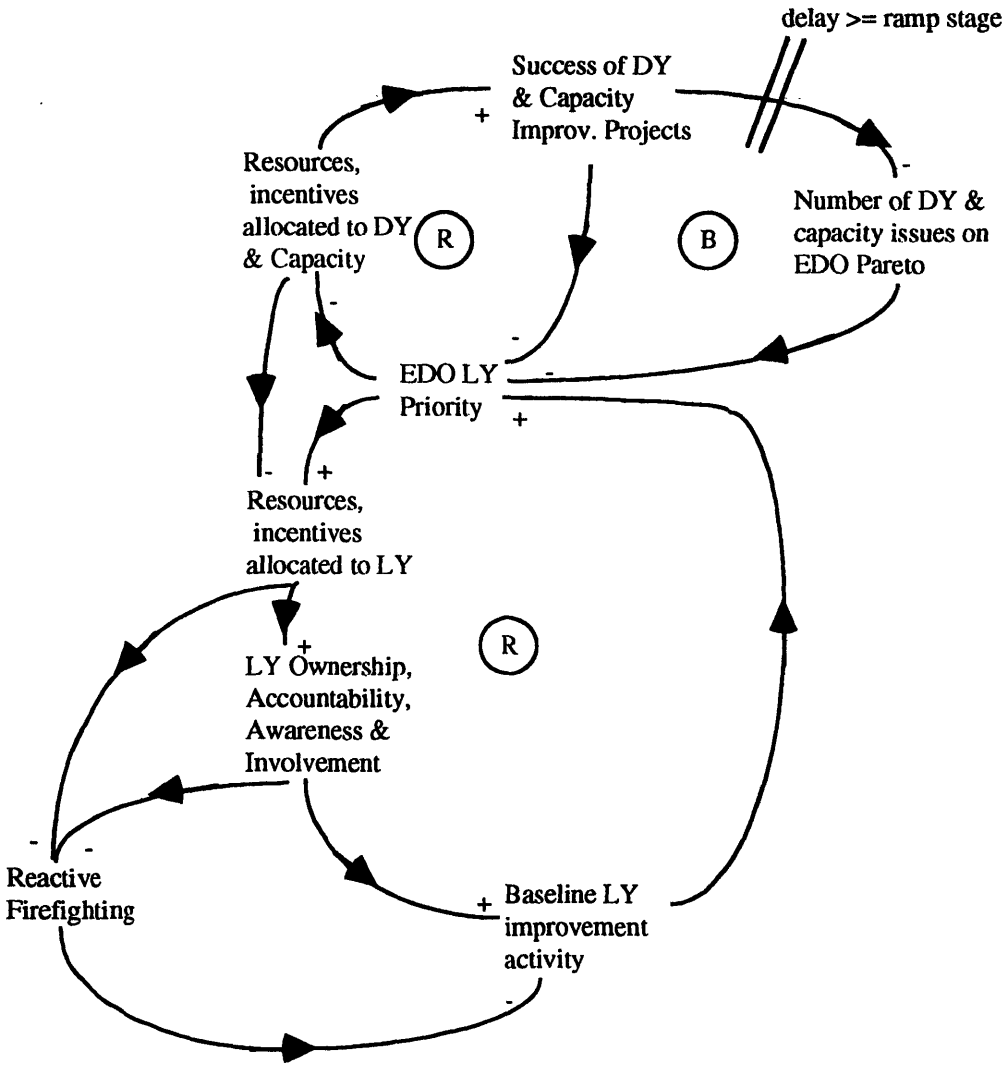


Figure 3.7. Reasons for slow rate of line yield improvement

Line Yield Ownership in a factory

One of the most popular answers to the question "Who owns line yield in a factory?" is "Everybody owns line yield". And they indeed should, because one person alone, or one

department alone cannot deliver high line yields, as is obvious from the Ishikawa diagram (Figure 3.1). However, the answer to the question "Who coordinates line yield improvements in a factory?", often is "Nobody". Herein lies the disconnect. The lack of a driving force, combined with a low EDO Pareto factory prioritization leads to a lack of awareness, involvement, and accountability for line yield improvements (Figure 3.7). A typical wafer Fab is organized into functional departments such as Manufacturing, Process Engineering, Yield Engineering, Industrial Engineering, Automation, Planning, Finance, etc. Due to the fact that line yield has multiple contributors such as process stability, equipment reliability, procedural factors, facilities and automation issues, low die yielding wafers, etc., it is difficult for any of the traditional departments to own line yield improvement.

Food for thought: who should own and drive line yield improvement ?

Line Yield Improvement Structure in a Factory

Realizing that it is difficult for any traditional department to own and drive line yield improvement, some factories have introduced the concept of a line yield coordinator whose typical task involves following up on line yield losses, driving root cause analysis, and also root cause fixes. However, the fixes to line yield problems could be as diverse as installing a station controller at a work station, designing a new process step, or changing preventive maintenance procedures, thus requiring diverse resources from multiple departments. Since each department's priorities are derived from the factory's EDO Pareto driven priorities which dictate that line yield is a lower priority, each department's resources will be busy on die yield and capacity improvements. Thus the line yield co-ordinator typically finds it very difficult to get resources from each department to participate in line yield improvement projects. Without sufficient top management support, the line yield coordinator's job becomes very frustrating, and line yield performance does not improve either.

In some factories, improvements are driven in a top down fashion by the Factory Improvement Steering Committee (FiSC) to the A-teams. There is one Area team (A-team) for each equipment cluster. E.g., Diffusion A-team, Litho A-team, Etch A-team, etc. Each A-team consists of representatives from the various departments. e.g., the Litho A-team has representatives from Litho Engineering, Litho Operations, Litho Industrial Engineering, Litho Automation, etc. Thus, each A-team has cross-department

representation. Based on the factory priorities set by the FiSC, each A-team establishes priorities for that area. Thus this structure helps establish a consistent improvement vision for the factory, keeping in perspective all aspects of maximizing good die output (die yield, line yield, and capacity). Each A-team typically initiates multiple improvement teams (I-teams) to work on specific EDO improvement projects relevant to that area. Typically, there would only be a couple line yield projects that are high enough on the factory EDO Pareto, to be allocated resources. Although the improvement efforts are focused on the key EDO limiters, this approach is not helpful in increasing awareness, involvement, and ownership of line yield on the factory floor.

Some other factories have "Bay Improvement Teams", or "Station Improvement Teams" that are focused on line yield improvement issues. This approach provides a very good grass-roots level line yield involvement across all shifts. The question though, is whether it makes sense for all areas to concentrate equally on line yield improvement. Yet another approach that exists in certain factories is to have the operations (manufacturing) group own and drive line yield improvements. Definitely, the extent of factory floor involvement is a highlight of this system. On the positive side, they looked at all the line yield losses in all the functional areas, thus heightening line yield awareness and involvement throughout the factory floor. Again, the question is whether it makes sense for all areas to concentrate equally on line yield improvement.

Some of the key organizational factors determining the effectiveness of the line yield improvement structure are (a) plant management's commitment and support to line yield improvement (b) line yield improvement role assigned to each individual in the factory, and (c) buy-in from factory floor personnel and commitment to line yield improvement.

***Food for thought**: what would be a good line yield improvement structure ? What tasks should be done to systematically improve line yield ? Who is best qualified to do these required tasks ? How do we get these right resources plugged into line yield improvement ?*

Factory-floor awareness & involvement in LY improvement

The importance of factory floor buy-in cannot be over-emphasized, and is absolutely critical to the success of any LY improvement effort. In this light, it is worth examining some of the hidden reasons behind procedural errors. Most people at work do

most of what they are supposed to do most of the time. They are co-operative, hard-working and dependable. However, there are times when even the good performers do something wrong.

Some reasons for non-performance leading to procedural errors

They don't know why line yield is important

The information on why a specific task is important falls into 2 categories. The first category includes the benefit to the organization for doing a task right, as well as harm to the organization for doing it wrong. The second category includes the benefits to the employees for doing a task right, as well as the harm to the employees for doing it wrong. The first category deals with why the task should be done; the second category deals with why that employee should do it. If they don't know the consequences of both of these categories, the result might be non-performance. Some of the reasons why an employee should do a task right would be:

To get paid
To get a good performance rating, raise or promotion
To get an award or recognition
To learn something
To avoid embarrassment
To prevent the pay from stopping
To avoid a reprimand, demotion, or termination

Food for thought: ***How can we motivate the factory floor to work on line yield ? How can we correct the problem of employees not knowing why line yield is important ?***

There is no positive consequence to them for line yield performance

There is a school of thought that getting paid for doing the job is the reward. Unfortunately, it does not always work that way. According to human behavior research, the weekly paycheck does not qualify as a reward that influences people's productivity. "People don't come to work to get paid, they come to work so that pay does not stop". Moreover, in most business organizations, an employee can perform badly for a long

period of time before the pay stops. An interesting analogy would be to breathing. Getting a weekly paycheck becomes important only when it stops.

Currently, in most factories, there are no positive incentives to factory floor personnel for line yield performance. The incentives are negative, if any, because disciplinary measures do exist for line yield losses due to chronic misprocessing.

Food for thought: ***How can we motivate the factory floor to work on line yield ?***

They think they are doing the right thing

"How do they know when they are doing a good job" ? The problem has to do with getting feedback. Feedback has a big influence on performance. All psychologists agree that feedback is one of the most critical requirements for sustained, high-level performance of any human act. Without feedback, you could be doing something much worse or much better than you think. If employees think they are performing okay, they have no reason to change. Typically, in factories, although line workers get to know the factory line yield performance as a percentage figure, say 90% factory line yield, it is not tangible to them in the sense that they find it hard to relate how their day-to-day tasks resulted in the overall line yield percentage figure. In other words, it is difficult to see the impact of their actions on the overall line yield percentage number.

Food for thought: ***How do we provide tangible feedback about line yield performance to operators and technicians so as to increase their awareness of LY ?***

There are obstacles beyond their control

This reason for non-performance includes all those situations, conditions, and influences outside the individual that are preventing performance from occurring. Examples include unavailability of resources, poor quality of resources, conflicting instructions, etc. An example of an obstacle beyond the control of floor personnel would be wafer scrap due to a poorly designed process step, where the remedy would be to redesign the process step.

Food for thought: ***How can we address line yield problems whose solutions lie beyond the realm of procedural errors ?***

They don't know how to do it

In many instances, procedural errors may be induced by not knowing how to perform a certain operation step or task. This could have its root causes in many factors, including hiring, training and re-deployment. It was observed that most factories, while they are being built and brought into production, adopt a just-in-time hiring policy. i.e., hiring of operators and technicians often gets delayed to reduce up front labor costs, with the result that training time is low. The training also could be termed "just-in-time" in this respect. A lack of training, and inadequate training practices, often manifest themselves as procedural error related line yield losses.

Food for thought: ***How do we impart proper training to the workforce so as to minimize line yield losses due to a lack of knowledge in wafer processing practices.***

They think something else is more important

This problem occurs when the operators and technicians on the factory floor are given conflicting signals about relative priorities of various tasks. The most obvious one being quality versus quantity. An important evaluation metric presently used to measure the performance of each shift is "number of wafers processed per shift". On the other hand, operators are also expected to do root cause analysis on wafers being scrapped and fill out a scrap report. Taking the time to analyze for root cause and fill out a scrap report would mean less time for processing wafers, leading to less wafers processed per shift. This would lead to a tendency to (a) move on to process the next wafer without filling out a scrap report (b) filling out a scrap report without really investigating for root causes, etc. Although this leads to the short term gain of being able to process more wafers per shift, by not arriving at the root cause of the problem, there is the increased risk of losing more wafers in future due to the same cause.

Another example would be as follows: When the particle count at a machine increases beyond a threshold, it would be advisable to shut down the machine, perform maintenance on it to bring down the particle count and then bring the machine back on line. This is because the presence of particles may spoil the die on the wafers, causing wafers to be scrapped. Shutting down a machine reduces number of wafers processed in that shift, while running the machine would cause more wafers per shift to be processed, especially if the wafers that became defective due to the presence of those particles are not detected until they reach a downstream test station, maybe during another shift.

Food for thought: ***How can we establish relative priorities on the factory floor , for key metrics such as capacity (number of wafers processed), die yield, and line yield ?***

Personal limits prevent them from doing it

Personal limits mean individual human physical limits that are unchangeable and are in fact preventing job performance. Such deficiency is not a function of learning. Personal limits could be temporary or permanent. Typically, on the factory floor, this problem would lead to chronic misprocessing, leading to repeated line yield losses, which would cause disciplinary action.

Food for thought : ***How can we deal with chronic misprocessing by individuals who are impaired by physical limits ?***

Personal problems prevent them from doing it

Such problems refer to those happenings in a person's private life outside the work environment that appear to be reasons for non-performance on the job. Seemingly trivial, in reality, this problem often manifests itself as line yield losses on the factory floor.

Food for thought: ***How can we address the issue of non performance at work due to personal problems outside the work environment ?***

Is there a closed loop follow-up system for investigating line yield losses ?

Ideally, one would want to document (i.e., fill out a scrap report), understand the root cause, and implement fixes for each wafer loss. However, realistically, given the volume of wafer starts in a factory (in the order of tens of thousands per week) and critical few investigation resources, even at a hypothetical line yield of 90%, the sheer number of wafer losses every week would make it impractical to close the loop on every wafer loss.

The typical approach adopted in most factories is to establish a wafer threshold for losses to be investigated. In a low volume fab, all losses involving say, five, or more wafers will be targeted for investigation. In higher volume fabs, this threshold may vary between ten

and twenty wafers. In any case, for each line yield loss incident where the number of wafers lost exceeds the threshold, a scrap report is supposed to get filled out. This scrap report is then supposed to be investigated for root causes and root cause fixes are supposed to be determined and implemented. However, even with a wafer threshold, these tasks are easier said than done, especially since there are no designated line yield coordinators at most factories to see that these tasks get executed.

Typically, in factories where there are no designated line yield coordinators, routine follow-ups do not happen. In factories where there are designated line yield coordinators, follow-ups do happen in the form of postmortems. To give the reader a flavor of how the follow-up system operates, I will describe the method used at a particular factory:

A particular Fab has established a flowchart that clearly depicts their line yield loss management policy. This flowchart is actually a specification that is adhered to, by all the individuals in the factory. The key players include all factory-floor personnel, area improvement teams, group leaders, engineering managers & the manufacturing manager. They have a threshold based on the number of wafers lost, which determines which losses get investigated. For those losses that get investigated, they have a closed loop system for root-cause analysis & implementation of root-cause fixes. Any wafer loss that is greater than or equal to ten wafers is selected for investigation. If it is a misprocessing (human procedural error) incident, it is investigated by the shift supervisor and the bay improvement team using a 7-step problem solving format. This results in an incident report which is presented by the supervisor at a weekly line yield meeting. If it is not a misprocessing incident, the station (equipment) owner and the bay improvement team will investigate the incident using a 7-step method. This also results in an incident report that is reviewed by the engineering group leader, and presented by the group leader in the weekly line yield meeting (postmortem). The general action items resulting from the line yield meeting are followed up by the Fab's line yield coordinator. The specific action items are followed up by the station owner or the bay improvement team. The station owner or supervisor that presents at the line yield meeting can take the incident review back to the improvement teams, to complete the feedback loop.

At some other Fabs, postmortem sessions are viewed as intimidatory. They feel that the postmortem method would tend to drive mistakes underground (i.e., passing faulty wafers on to the next production step, instead of scrapping them) for fear of disciplinary action,

because scrapping wafers would lead to filling out scrap reports, presenting at postmortems, and maybe receiving disciplinary action.

While there is the concern of facing disciplinary action, operators and technicians would like to see fixes implemented for the scrap report suggestions they turn in. The importance of having a closed loop system to follow up on wafer losses being documented cannot be over-emphasized. To an operator or technician on the line filling out a scrap report, if the report is not reviewed by anyone, or if no action is taken either for root-cause analysis or for implementing improvement suggestions, there is no motivation for attention to detail, let alone for filling out the scrap report.

Food for thought: ***Thus, the issue comes down to designing a follow-up system that ensures that key losses are followed up, analyzed for root causes, and fixed, without instilling fear of disciplinary action. An associated issue would be how to decide what the key losses are. Clearly, all losses cannot be followed up for investigation, as described above. So how do we decide what losses to select for investigation. Should the wafer threshold for investigation be the same in all functional areas, or should they differ based on whether the functional area processes wafers in batches (as in diffusion) or one wafer at a time (as in lithography). Should the threshold also vary depending on whether the losses occurred in the front -end of the process flow versus the back-end?***

Structured root cause analysis methods for losses

For losses that are selected to be investigated, what methods are used to arrive at root causes ? Arriving at root causes is critical to enable root cause fixes. A wide range of problem solving techniques exist, from structured methods such as Seven-Steps, or Fault Tree Analysis, or Failure Modes Effects Analysis, to unstructured, adhoc methods. Although structured techniques will maximize the changes of arriving at the root causes, their use is only slowly catching on in line yield loss cause analysis in the factories.

Food for thought: ***what would be a good strategy for causal analysis of line yield losses ?***

Implementation & Communication of root-cause fixes to line yield problems

The whole objective of tracking losses and performing root cause analysis is to arrive at root-cause fixes to line yield issues. However, the fixes could be as diverse as putting in place a station controller to prevent operators from running wrong recipes, better training procedures, designing a new metal etch process, or changing preventive maintenance procedures. This implies that the resources needed to implement the fixes would also be quite diverse, which increases the difficulty of implementation. This is where a line yield co-ordinator can play a key role in securing resources for driving root-cause fixes. It is for this reason that many factories have adopted the concept of a line yield coordinator. However, the job could be very frustrating because the line yield coordinator has to secure resources from various departments to work on line yield, when line yield may not be a top priority for that department. A line yield issue while being a top priority for the line yield coordinator may not be a top priority for the departments which have to lend resources to work on the issue. Thus, although the initial loop of investigating a line yield loss incident may be closed, the far more important loop of implementing root cause fixes often remains open.

In several factories that have dedicated line yield coordinators, weekly or bi-weekly line yield meetings are held. These meetings are supported and attended by senior Fab management. The Fab manager attends, in addition to engineering managers, manufacturing manager, group leaders, etc. Some of the positive aspects of this structure are (1) Universal ownership of the problem rather than only "production" (2) visibility and participation at the highest levels (3) clear expectations and standard formats (4) attendance by all the departments facilitates the implementation of cross-functional fixes to line yield problems (5) the line yield coordinator as a dedicated resource helps to keep the program on track. On the negative side however, the fab manager attending the meeting may lend a flavor of intimidation.

Food for thought: How can we design a follow-up system to ensure that root cause fixes get implemented ?

Line Yield losses due to process equipment failures

Equipment related losses account for a major percentage of line yield losses in any semiconductor factory. Typical reasons attributed are (1) equipment being incapable to

run very tight processes (2) inappropriate preventive maintenance frequencies and procedures (3) equipment unreliability, resulting in breakdowns while processing wafers. I will discuss each of these categories:

(1) Equipment being incapable to run very tight processes (Machine/Process Potential):

New semiconductor fabrication processes often require the use of new and more capable production equipment. U.S. firms have begun to rely to a great extent on foreign suppliers for several state-of-the-art processing equipment and technologies. There is a fierce controversy raging about whether or not U.S. semiconductor firms are being denied advanced equipment or technologies from foreign suppliers. Although this topic is outside the scope of my thesis, the intent of the above comment is to make the reader aware of the relationship between access to state-of-the-art equipment and making advances in process technology. Because of the rapid advancement of semiconductor technology, even a brief delay in obtaining a piece of equipment can cause a company to fall a generation behind in its technological capabilities, resulting in lost market share.

The fact that each new generation of Intel's microprocessor requires more transistors on the chip causes Intel to strive to be at the cutting edge of process technology. However, these new processes often have to utilize equipment meant for a previous generation process technology leading Intel to push process equipment beyond its original design capabilities. This results in low machine or process capability (Cpk). Cpk is a machine or process capability index that measures the ability of a machine or process to produce product within specification. Cpk is the ratio of the distance between the actual process average and the closest specification limit over three times the standard deviation of the actual process. Cpk measures the degree of centering of the actual process spread with respect to the allowable spread. A machine or process is referred to as being capable when its Cpk has a minimum value of one, and process stability has been proven.

Food for thought: ***What are some of the methods of increasing Cpks of process steps so that line yield scrap due to equipment incapability can be reduced ?***

(2) inappropriate preventive maintenance frequencies and procedures:

How frequently should we do a PM (Preventive Maintenance, or calibration) procedure? PM procedures are often developed subjectively without the benefit of sufficient data to

support the frequency. There are several metrics used today to determine PM frequencies, the most common being MTBF (Mean time between failures) and MWBF (Mean wafers between failures). In many cases, the PM frequencies actually used in the field would be the ones recommended by the Equipment supplier for the first production system. Due to the interaction of the equipment with the manufacturing processes, suppliers are not always in a position to optimize PM frequency.

How does PM frequency and procedure affect line yield performance? Unless you believe something is deteriorating rapidly, by bring down a piece of equipment to do a PM (calibration), and starting it back up, you are probably introducing more chance of failure by introducing a new source of variability into the system. An example is a case where certain valves were not turned on by technician oversight, after doing a PM, which resulted in losing wafers processed at the machine after the PM.

Food for thought: ***How do we address line yield losses induced by variability caused by either inappropriate PM frequencies or improper PM procedures ?***

(3) equipment unreliability, resulting in breakdowns while processing wafers.

Equipment breakdowns are bad due to several reasons including (1) losing the wafers that were being processed by the equipment when it broke down (2) downtime of the machine resulting in lost throughput for the factory. For example, if a wafer breakage happens in an etcher (a single wafer processing machine), the wafer loss may be low on the line yield loss Pareto, but since a time consuming cleaning operation is required after every break, the impact on throughput is significant.

Ideally, one would like to proactively detect potential failures and rectify them before the equipment actually fails and causes line yield losses. FMEA (Failure Modes & Effects Analysis) provides a structured, team approach for identifying and analyzing potential or existing failure modes. The purpose of conducting an FMEA is to understand all failure modes and their effects, so the higher risk items can be prioritized and eliminated. FMEA is common practice in several Japanese semiconductor companies such as NEC.

Although one would like to use FMEAs to proactively detect potential equipment failures and rectify them prior to an actual failure, in reality, given the number of diverse pieces

of equipment in a Fab, and the limited human resources available, it may not be practical to track down ALL failure modes of ALL pieces of equipment in a Fab.

Food for thought: ***How can we proactively use FMEAs on key pieces of equipment in a Fab ? How do we decide what the key pieces are, on which to do FMEA?***

Process Immaturity at transfer from a technology development site to a high volume Fab

The traditional role of process development has been to deliver a new process that yields at "manufacturable levels", i.e., to bring the process to a point where the die per wafer yields are high enough to make manufacturing the product economically viable. The emphasis is on developing process technology in a stable environment (e.g., at low volumes). Under this development paradigm, planning and designing of operational methods such as maintenance, automation, and training are issues that remain for manufacturing to resolve and develop.

The goal of meeting yield targets for process transfer to high volume facilities directly competes with surfacing potential problems, which requires stressing a system closely aligned with the factory environment. This works against getting the process up and running smoothly in the development center. This lack of congruence between the actual high volume manufacturing facilities and the environment in which the process is developed leads to low probability of exposing and solving potential production problems in the development stage.

Some problems occur only in a full scale manufacturing environment., which differs from development in many areas including personnel policies, equipment utilization, and operational procedures. The important issue is the state at which the process is transferred. Instead of "thrown over the wall" transfers, if there is a seamless evolution from development to manufacturing, it will reduce the chances that significant changes will be required to the process once it is running in a manufacturing plant. Increasing the congruence also leads to fewer "yield loss excursions" in the post-transfer stage.

Food for thought: ***How can we get a more mature process at the TD transfer stage ?***

The Environment in Ramping Wafer Factories (at the post process transfer stage)

The operating environment in ramping factories is worth examining since it will shed light on the slow line yield learning behavior (Figure 3.6) exhibited during the ramp stage. Some of the characteristics of a ramping facility are:

Dominance of Output Goals: Reaching the output goals negotiated for the plant is a critical success indicator for the factory management. Output goals take precedence over other indicators such as throughput time, yields, or cost targets. Although the importance of meeting market requirements is fully justifiable, this sense of urgency could lead to a reliance on short-term fixes, weakening the ability of the factory to focus on more fundamental improvements (such as line yield, or throughput time).

New incremental equipment installation: Adding new capacity while ramping a factory is a common occurrence. The timing and increment of the installation of new processing equipment adversely affects the ability of the factory to focus on key sustaining activities. Some adverse effects of adding capacity in small increments are (a) continued disruption of the line as equipment is installed (b) added process variation due to slight differences among sophisticated equipment sets acquired in different time periods.

New, in-training workforce: Ramping of a new facility invariably requires the hiring of personnel at various levels (engineers, technicians/operators, managers). Dedicated training is absolutely necessary, to develop the skill set of the workforce, and this training can be done either while the ramp is going on, or beforehand, while the site is being prepared for ramping. Tradeoffs between the two approaches include costs associated with early hiring and training, the impact of each role on overall factory performance, and the existing skill set of the workforce. Needless to say, hiring and training have a big impact on line yield performance, because they directly affect the extent of wafer losses due to misprocessing or procedural errors.

New and immature process: As mentioned in the previous section, processes are often transferred from development sites to high volume factories in an "immature" stage and require further changes as they are being ramped at the manufacturing site. This immaturity often results in yield loss excursions, hurting the ability of the factory to meet its die output goals.

New equipment in the works: With the introduction of a new process comes new and more capable production equipment. Older equipment with well known characteristics are often replaced by newer machines that have unknown idiosyncrasies. The introduction of this new generation of equipment implies that even the experienced personnel will feel the requirement of a learning curve, to understand the new machinery. This also increases the probability of equipment related yield loss excursions at the ramping stage of a factory or process.

Reactive excursion fire-fighting: All the above mentioned issues contribute to a ramping operating environment best characterized by two words: complexity and confusion. The result is the frequent occurrence of major wafer loss events, otherwise known as "yield excursions", which directly impact the factory's key performance metric: die shipment. Typically, during the ramp stage, a factory's resources are mostly allocated to capacity improvement projects and not yield improvement. However, since excursions severely reduce die output, they have the immediate effect of focusing attention on the problem area. Unfortunately, responding to excursions are "reactive", and because of the intense pressure to get yields back to normal, the solution found is usually more of a "quick fix" (e.g., replacing the defective part, or recalibrating the equipment) rather than a root cause fix.

Most of the yield improvement activities that exist during the ramping stage can be classified as excursion fire-fighting as opposed to longer term baseline improvements. However, the only result of excursion firefighting is the restoration of line yield performance back to where it was prior to the excursion. Hence, fundamentally, it does not lead to improving line yield, rather, all it does is prevent the deterioration of line yield performance. This would imply that for a factory to come up the line yield learning curve, excursion firefighting would not suffice. Rather, baseline improvements are needed, in order to fundamentally improve a factory's line yield performance. However, during the ramp phase, since line yield is only a tertiary output lever (behind capacity and die yield), resources do not get allocated to work on baseline activities until after ramp, when line yield has more leverage. However, due to this reason, when the ramp gets over, and line yield does become a high factory priority, there are no systems in place to deal with it.

Food for thought: ***What would be the best way to address line yield improvements during the ramp stage, and later on, during the maturity stage of a factory or process ?***

3.4. Summary

This chapter has identified the underlying reasons of line yield losses. This section will summarize key issues that have to be addressed to improve line yield performance.

1. **LY Factory priority:** what kind of priority should be given to line yield improvement? Is there a way of improving line yield learning rates without adversely affecting die yield and capacity improvement rates?

2. **LY ownership:** Who should own and drive line yield improvement?

3. **LY improvement structure:** What would be a good line yield improvement structure? What tasks should be done to systematically improve line yield? Who is best qualified to do these required tasks? How do we get these right resources plugged into line yield improvement?

4. **LY Reward Systems:** How can we motivate the factory floor to work on line yield?

5 **LY Awareness:** How can we communicate to the floor why line yield is important?

6. **LY Feedback:** How do we provide tangible feedback about line yield performance to operators and technicians so as to increase their awareness of line yield?

7. **LY losses due to systemic issues:** How can we address line yield problems whose solutions lie beyond the realm of procedural errors?

8. **LY training:** How do we impart proper training to the workforce so as to minimize line yield losses due to a lack of knowledge in wafer processing practices?

9. **LY Communication strategy:** How can we establish relative priorities on the factory floor , for key metrics such as capacity/output, die yield, and line yield?

10. **LY procedural errors:** How can we deal with chronic misprocessing by individuals who are impaired by physical limits?

11. **LY procedural errors:** How can we address the issue of non performance at work due to personal problems outside the work environment?

12. **LY Follow-up systems:** How do we design a follow-up system that ensures that key losses are followed up, analyzed for root causes, and fixed, without instilling fear of disciplinary action?

13. **LY Pareto prioritization:** How do we decide what the key losses are. Clearly, all losses cannot be followed up for investigation, as described above. So how do we decide what losses to select for investigation.

14. **LY loss investigation:** Should the wafer threshold for investigation be the same in all functional areas, or should they differ based on whether the functional area processes wafers in batches (as in diffusion) or one wafer at a time (as in lithography). Should the threshold also vary depending on whether the losses occurred in the front end of the process flow versus the back end?

15. **LY root cause analysis:** What would be a good strategy for causal analysis of line yield losses?

16. **LY solution implementation:** How can we design a follow-up system to ensure that root cause fixes get implemented?

17. **LY process instability issues:** What are some of the methods of increasing Cpks of process steps so that line yield scrap due to equipment incapability can be reduced?

18. **LY equipment unreliability issues:** How do we address line yield losses induced by variability caused by inappropriate PM frequencies or improper PM procedures?

19. **LY Proactive risk analysis for process equipment:** How can we proactively use FMEAs on key pieces of equipment in a Fab ? How do we decide what the key pieces are, on which to do FMEA?

20. **LY process maturity issues:** How can we get a more mature process at the TD transfer stage ? What would be the best way to address line yield improvements during the ramp stage, and later on, during the maturity stage of a factory or process?

Chapter 4. Solution Planning & Implementation

Purpose: The purpose of solution planning and implementation is to take action that will eliminate the causes of the problem. This chapter identifies key changes to line yield management systems, policies, and practices, in order to improve line yield performance.

4.1 Introduction

In this chapter, I adopt a systems thinking approach, in confluence with Eli Goldratt's theory of constraints [3,4,5,6], to develop line yield improvement tools for a semiconductor fabrication plant. Each of these tools can be incorporated into a wafer Fab's existing line yield strategy [10,11, 15], as appropriate.

4.1.1 Systems Thinking

--> How to develop a coherent line yield management system.

Intel's historical approach of responding reactively to major wafer loss excursions provides very little power to alter the course of those events. In the last chapter, I used systems thinking as the tool to determine the true drivers of line yield performance. By considering the relationships that exist among the components of the entire organization, systems thinking fosters a more holistic approach to developing coherent operating policies. My research [12], by taking a systems approach, reveals that underlying structure drives line yield behavior, and that maximum leverage in improving line yield performance comes from changes at the structural level. This chapter will describe my solutions [12], based on systems thinking, to improve line yield performance.

4.1.1 Theory of Constraints

--> How to focus a factory's limited line yield improvement resources on those line yield problems having the most revenue impact.

Line yield being a lower leverage die output tool (compared to capacity & die yield) during the startup & ramp stages of the process life cycle, Intel has historically focused its line yield improvement efforts on the top few line yield hitters (or problem areas) as indicated by a Wafer Scrap Pareto. In other words, line yield problems would be

prioritized based on the number of wafers being scrapped due to each problem, regardless of the process step at which they were scrapped. The underlying assumption here is that the significance of a line yield problem is dependent only on the number of wafers that are lost, and is not dependent on where in the process flow, the scrap occurred. My research suggests that this approach can be misleading in that it treats every wafer loss the same, regardless of the extent of processing done on the wafer before it was scrapped.

I use the theory of constraints to analyze whether wafers lost at different points in the process flow are worth the same. A constraint based approach clearly indicates that every wafer loss is not worth the same. e.g., a 100 wafer loss at a front-end buffer oxide diffusion step should not be considered as equally significant to a 100 wafer loss at the back-end polyimide cure diffusion step. Since market demand is not the constraint for Intel, my thesis proves that wafer value should depend on the amount of time it has spent at the Fab constraint (e.g., Lithography). *The importance of a line yield loss incident should depend not only on the number of wafers lost, but also on where in the production line the loss occurred.*

My research develops and applies a constraint-based wafer scrap value model to systematically assign relative values to wafers depending on where they were scrapped in the production line. The more processing done to a wafer at the Fab constraint, the more valuable it becomes. A wafer loss at an operation before the constraint is valuable only if that loss causes the constraint to be idle. If lithography (litho for short) is the constraint, a wafer that has been through litho 10 times should be twice as valuable as a wafer that has been through litho 5 times. This principle of constraint-based wafer value modeling can be applied in several ways to prioritize the various line yield improvement projects, based on the revenue impact of each project. *Thus, my research helps to focus a factory's limited line yield improvement resources on those line yield problems having the most revenue impact.*

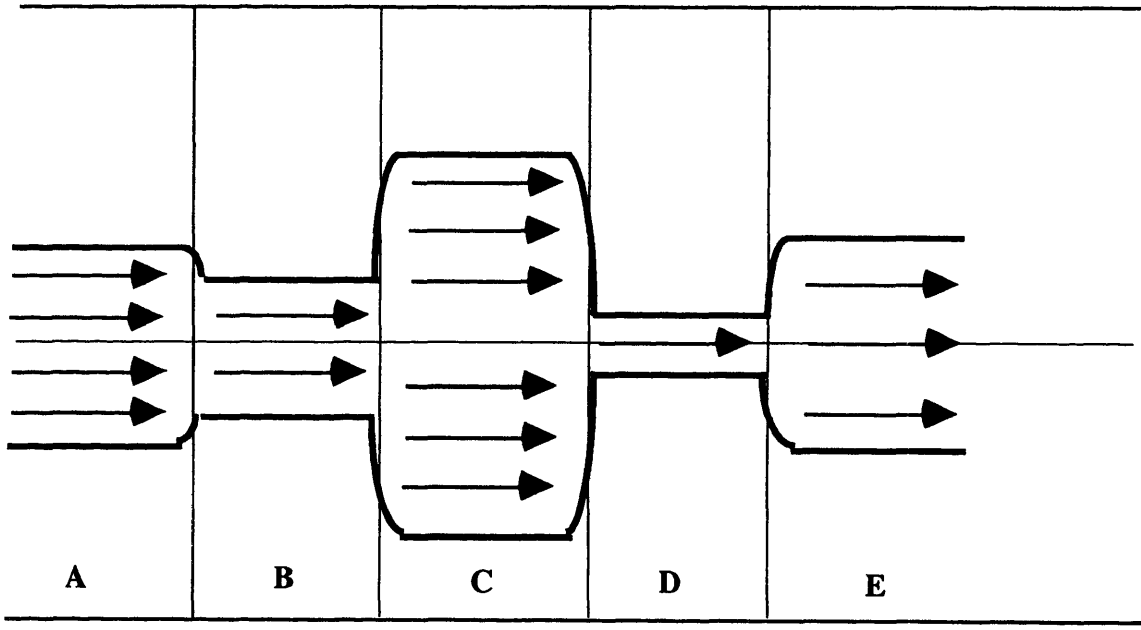
4.2 What is "Theory of Constraints" ?

In 1979, Creative Output of Milford, Connecticut, introduced a new production planning system called OPT (for Optimized Production Technology). Eliyahu Goldratt, an Israeli Physicist, developed the algorithm used in the system. At the heart of the OPT philosophy is the notion of *bottlenecks (constraints).* If we recognize that every system was built for a purpose or goal, a system's constraint is anything that limits a system from

achieving higher performance versus its goal. In reality, any system has very few constraints, and any system must have at least one constraint [5].

One can most easily understand a bottleneck by considering an analogy between the flow of material through a factory and the flow of liquid through a pipe [13]. Consider the pipe pictured in Figure 4.1. Clearly, the rate at which the liquid can flow through the pipe is the rate at which the liquid can flow through the most narrow portion of the pipe, at D. If we were to change the diameter of the pipe at A, B, C, or E (assuming the change did not decrease the diameter below that at D), there would be no change in the total throughput rate. However, if we were to change the diameter of the pipe at D, the total throughput rate would change accordingly.

Figure 4.1. A pipe of varying thicknesses to illustrate bottlenecks



If we think of A through E as steps in the production process with the diameters of the pipes representing the respective production rates, then the bottleneck in the process occurs at D. This means that any delay at D results in a decrease in the flow of material through the system, while a delay at another step will probably not result in a delay in the system (unless the delay causes the diameter to fall below that at D). Identifying bottlenecks and managing production resources and material flow, so that there is no loss in production time at the bottleneck facilities is at the heart of "theory of constraints".

4.2.1 Principles underlying "Theory of Constraints" (TOC)

The general TOC process of continuous improvement [3,4] can be stated as follows:

1. **Identify the system's constraints:** What are the factors that most constrain the output of the factory? Identification of the constraints also implies prioritizing them according to their impact on the goal.

2. **Decide How to exploit the system's constraints:** Once we have identified the constraints, the next step is to make sure we do not waste these few things that are in short supply. i.e., design a schedule that keeps the bottleneck areas busy. An hour lost at a bottleneck is an hour lost for the factory. The lost time can never be recovered and the production flow will decrease.

3. **Subordinate everything else to the above decision:** Activating a non-constraint resource does not correspond to intelligent utilization of that resource. Non constraint resources should be managed so that they will supply everything that the constraints are going to consume. Referring to Figure 4.1, there is no benefit to running C at full capacity if D cannot absorb its output, since the system's performance is dictated by D.

4. **Elevate the system's constraints:** It is now obvious that any improvement we achieve at a non-constraint is a mirage, since the system's performance is dictated by the constraint. Hence, to improve the system's performance, we have to elevate the constraint's performance. Whatever the constraints are, there must be a way to reduce their limiting impact, and so the way to do it is by focusing improvement efforts at the constraint operation, and elevating the constraint's performance.

5. **If the Constraint has been broken as a result of Step 4, return to Step 1:** If we elevate and continue to elevate a constraint, then there must come a time when we break it. This thing that we have elevated will no longer be limiting the system. This does not imply that the system's performance will go to infinity because another constraint will limit its performance. So once we have broken a constraint, we have to go to step 1, to identify the new constraint.

Thus, while a typical improvement program would spread efforts across each and every piece of equipment (constraints and non-constraints), TOC helps to focus improvement efforts on the constraints (i.e., areas actually limiting total system performance [8,17,18].

4.3 Applying "Theory of Constraints" to Line Yield at Intel

This section outlines how we can apply TOC to prioritize line yield improvement projects in a semiconductor Fab environment. Intel's fabrication processes can be modeled as a chain of process steps, each step dependent on the preceding steps. According to TOC, strengthening a chain requires strengthening the chain's weakest link. Intel's goal is to make money now and in the future. Moneymaking is measured with two simple parameters: profit and return on investment (ROI). Profit is an absolute measurement while ROI is a relative measurement, and both are important[17,18].

A problem with profits and ROI as measurement parameters is that they are difficult to use for making daily decisions. Throughput (T), Inventory (I), and Operating expenses (OE), can be applied more effectively. T is all the money that is generated through the system through sales. I is all the money that is inside the system at any point in time. OE is all the money required to turn inventory into sales. Profit is now defined as T - OE and ROI is defined as (T - OE) / I. From these simple equations, the desired results are clearly to increase T, and to decrease I and OE.

If T is money generated through sales, then T requires sales to occur. If Fab output is limiting chip sales, then T is directly proportional to wafer output. The fact is, **Market Demand is NOT the constraint for Intel.** Since Intel is in the enviable position of being able to sell all the high margin microprocessors its factories can produce, market demand is not the constraint for Intel. Hence, the company's revenues and profits directly depend on, and are proportional to, die shipments from its factories. Increasing Fab wafer shipments increases Intel's T. Die shipment from Intel's factories are dependent on factory throughput. i.e., the number of die the factories can ship out every day. Factory throughput in turn is directly dependent on the Fab constraint. In other words, **it is the constraint within the factory that is limiting Intel's throughput (T), die shipments, revenue, and profits.**

Let's examine the three variables T, I, and OE, in greater depth. Within a Fab, inventory dollars (I), can be classified into non-depreciated equipment costs, unused materials, and incompletely processed wafers. Equipment costs, while significant in size, are difficult to trim since the equipment is already purchased, and is unlikely to be sold, especially at book value. Unused materials and wafer inventories are required to produce wafers and hence, T. Note that wafer inventory is directly proportional to Fab cycle time. Since

defect levels are inversely proportional to cycle times, then T is improved by minimizing wafer inventory.

OE sums equipment depreciation costs, personnel expenses, major fixed expenses, cost of wafer scrap, etc. Like equipment inventories, depreciation is pre-allocated at the time of purchase and is difficult to alter. Personnel expenses (head count) is also not a very viable alternative. Wafers that are scrapped in the line contribute directly to OE, since the Fab had already invested equipment, materials, and personnel resources into processing the wafer that ended up being scrapped. However, as mentioned above, we can consider equipment and personnel as fixed costs. Material costs are negligible relative to the high price margin commanded by Intel's microprocessor products.

Thus, T appears to be the most critical of the three money making measurements. Certainly, a Fab would lower neither inventories nor operating expenses if doing so has an adverse impact on T. On the other hand, for Intel, there is no upper limit on T. Given the high profit margins in the semiconductor microprocessor business, if the goal is to make more money, stockholders would prefer an increase in sales (T) to a decrease in inventory (I) or a decrease in operating expenses (OE). The following argument assumes that **T is most critical,** and that **increasing wafer shipments increases T.**

How do line yield losses impact T ?

- **A pre-constraint wafer scrap has ZERO impact on T,** as long as it does not starve the constraint.
  (a pre-constraint scrap refers to a wafer that was scrapped before being processed by the constraint).
- **A post-constraint wafer scrap reduces T.**
  (a post-constraint scrap refers to a wafer that was scrapped after being processed by the constraint).

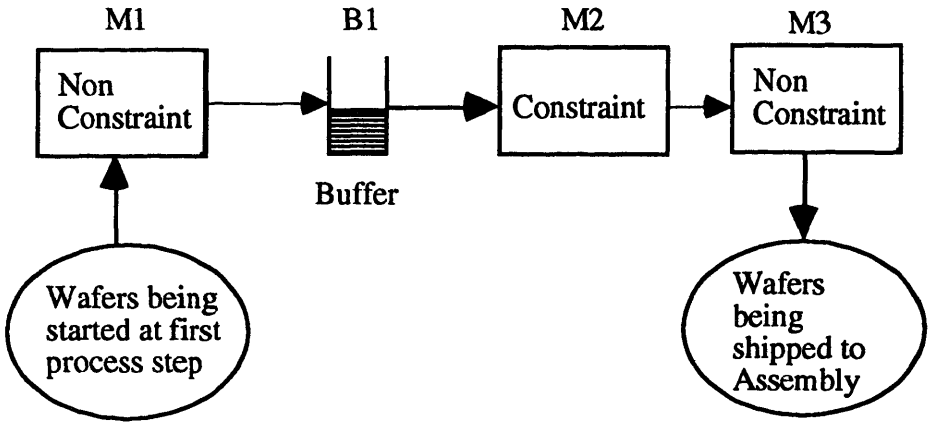
Note:

(1) **Line yield losses impact OE adversely,** regardless of whether the losses occurred at pre-constraint or post-constraint steps. However, as the above discussion shows, when the operating margins (Revenue - variable cost of production) are very high, as is the case with Intel, we are most concerned with the impact of line yield scrap on T.

(2) **Line yield losses may or may not impact I:** Typically, a pre-determined number of wafers enter the production line every week (WSPW - Wafer Starts Per Week).i.e., in-line losses are not compensated by more starts. In this scenario, line yield losses cause I to decrease, although at the risk of starving the constraint. However, if the losses are at post-constraint steps, this decrease in I is not desirable since it is accompanied by a much bigger decrease in T. If losses are at pre-constraint steps, a decrease in I does not impact T as long as it does not starve the constraint although it does impact OE. T is adversely impacted if the constraint is starved.

Figure 4.2 is simplified version of the Fab process flow, primarily intended to give the reader a better feel for the difference in throughput impact for pre-constraint versus post-constraint line yield losses. The factory's throughput can at best be equal to the throughput of the constraint M2, if no line yield losses occur at the post-constraint operations (M2 or M3). Any loss at M2 or M3 causes the Fab to have a sub-optimal throughput, causing sub-optimal revenues. However, line yield losses at pre-constraint steps (M1) do not affect the constraint's (M2's) throughput unless the loss at M1 causes the buffer B1 to be empty, which in turn causes M2 to be idle. Thus, a pre-constraint scrap incident does not affect the Fab throughput unless it starves the constraint.

Figure 4.2. A Simple model of Fab Processing Steps



Thus TOC points out that the importance of a line yield loss incident should depend not only on the number of wafers lost but also on **where** in the process flow the loss occurred.

4.3.1 A Constraint-based Wafer Scrap Value Model

Time is Money: Any time lost at the constraint is lost throughput for the factory. Since it is constraint time that is limiting Fab throughput, and hence Fab revenue, the value of a wafer depends on the amount of time it has spent at the Fab's constraint. Assuming that a pre-constraint loss is not going to starve the constraint (due to the presence of a large enough buffer), we can say that a pre-constraint scrap' s worth is negligible, but that a post-constraint scrap's worth is equal to the revenue the wafer would have brought in, had it reached the end of line without being scrapped. Since the raw silicon cost is around 1% of the potential revenue from a fully processed wafer, it is indeed appropriate to assign values to scrap based on potential revenue lost as opposed to assigning values based on raw material+processing expenses incurred until the wafer got scrapped.

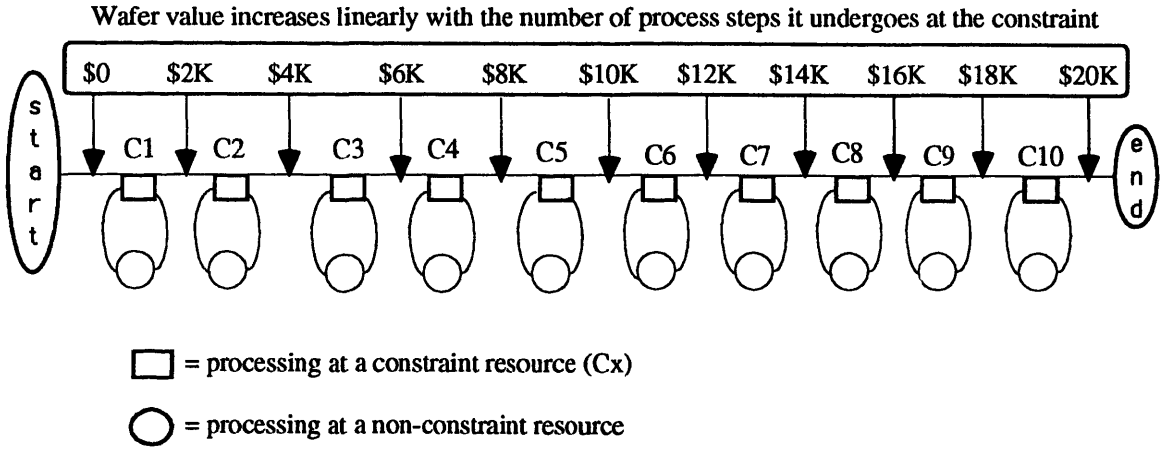
The above discussion assumes that it is easy to distinguish between pre-constraint and post-constraint wafers. However, the nature of semiconductor processing is such that a wafer may have to go through the same piece of equipment at multiple steps in the process flow. Due to the layering nature of semiconductor fabrication, and the fact that each layer resembles other layers in many ways, each wafer goes through lithography steppers, ion implanters, etchers, deposition systems, and resist strip equipment multiple times.

Given the above scenario, supposing lithography is the constraint, and there are 10 litho steps in the process flow. How can we distinguish a post-constraint wafer from a pre-constraint one? Clearly, a wafer that has not yet reached the first litho step is a "pre-constraint" wafer, and a wafer that has been through all ten litho steps is a "post-constraint" wafer. What about all the wafers that have been through one or more, but less than ten litho steps ? Assume that each litho operation takes roughly the same amount of processing time. Since the value of a wafer is determined by the amount of time it has spent at the Fab's constraint, intuitively, we could say that a wafer that has been through litho ten times should be twice as valuable as a wafer that has been through litho only five times. *Thus, theory of constraints can be applied in a systematic, scientific way, to assign relative values to a wafer, depending on its position in the process flow.*

To illustrate this point further, consider a semiconductor process with say, Lithography, as the constraint, and with 10 litho steps in the process flow. Assume that a fully processed wafer would fetch $20,000 in revenue. This process is modeled in Figure 4.3.

In this figure, C1, C2...C10 represent each of the 10 litho constraint steps, with the non-constraint steps indicated in circles...The dollar values indicate the value of the wafer at the completion of each constraint processing step. We know that the fully processed wafer, at the end of C10, is worth $20K. Since at that stage, the wafer has spent twice as much time at the constraint than a wafer that is currently between C5 and C6, it should be worth twice as much. In other words, a wafer that is currently between C5 and C6 should be worth $10K. The same logic is applied to arrive at the value of a wafer at each operation step in the entire process flow.

Figure 4.3



Clearly, from a potential revenue loss perspective, losing 100 wafers at C1 (or any step between C1 & C2) has the same impact ($200K) as losing 50 wafers at C2, or losing 10 wafers at C10. In other words, you can afford to loss 10 times as many wafers at C1 than you can at C10. This has several implications on how you prioritize line yield problems or issues to work on.

Let's assume that we are trying to look at the major line yield losses over the last quarter - 3 months (or any time frame, for that matter), in order to prioritize and select the top few, most critical line yield problems to work on. I will now describe a prioritization model that will look at the number of wafers that were lost at each operation step of the entire process flow, consider the revenue impact of a wafer loss at each of these steps, and come up with the revenue impact of the total losses at each operation step. Once we have the revenue impact broken up by operation step, we could consider categorizing the losses by equipment type, or loss cause type, or however we choose to look at it. The point here is that this model will enable us to prioritize and focus a factory's critical few line yield

resources on those line yield improvement projects having the most die-output or revenue impact for the factory.

Let N = Number of operation steps in the entire fabrication process flow: Let 'i' = 1 to N.
Let M = Total number of process steps at the Constraint (e.g., 10 in the above case)
Let D = Dollar value of a fully processed wafer (e.g., $20K in the above case)
Let W_i = Weight assigned to step 'i', = number of constraint steps upto & including step i
Let d_i = Dollar value of a single wafer loss at step 'i'.
Let L_i = Number of wafers lost at operation step 'i'.
Let R_i = Revenue impact in Dollars, of all line yield losses at step 'i'.
= Number of wafers lost at step 'i' * Dollar value of a wafer lost at step 'i'
Let R = Total revenue lost for the factory due to line yield losses during the time frame.

The following 3 equations summarize the constraint-based scrap value model.

$$di = \frac{D}{M} Wi$$

$$Ri = Li.di$$

$$R = \sum_{1}^{N} Ri$$

4.3.2 Constraint-based Wafer Scrap Value Modeling
- A Fab 9 Case Study on process P6x

Note: For confidentiality reasons, the actual names of process steps, wafer scrap values, and dollar figures stated in this section have been changed, without impeding the validity of the reasoning.

Nikon stepper systems in Lithography are the constraint at Fab 9. There are 200 process steps in the P6x process flow, which includes 14 litho steps at the Fab constraint (Nikons). The average revenue a typical product wafer would fetch is $25K, assuming 100 die per wafer, and $250 as the average sales price of each die.
Since Litho is the constraint, I assigned a weight (Wi) to each operation step in the p6x process flow, and this weight was equal to the number of litho steps upto and including that operation. For example, if "litho1", "litho2", "litho3" are the first, second, and third

litho operations, $W_{litho1} = 1$, $W_{litho2} = 2$; $W_{Fieldox} = 2$ (Fieldox is a diffusion step which occurs between the second and third litho operations), $W_{litho3} = 3$, and so on. The "wafer sort" operation which occurs at the end of the entire process flow (step 200) was given a weight of 14, since it occurs after all 14 litho operations. i.e., $W_{sort} = 14$. Table 4.1 applies the constraint based model to assess revenue impact of line yield losses.

Table 4.1

Operation Step	Number of Wafers lost at this step	Value of each wafer lost at this step	Revenue impact of all losses at this step
step 1	L_1	$d_1 = (W_1 * \$25000) / 14$	$L_1 * d_1$
.....			
step i	L_i	$d_i = (W_i * \$25000) / 14$	$L_i * d_i$
.....			
step 200	L_{200}	$d_{200} = (W_i * \$25000) / 14$ since $W_{200} = 14$, $d_{200} = \$25,000$	$L_{200} * d_{200}$

Fab 9 typically categorizes losses according to equipment cluster. For example, all the wafer scrap happening in each step of the litho area is summed together to get the total litho scrap for the quarter, and similarly for each of the other areas such as diffusion, etch, thin films, etc. This is because the Fab is organized into cluster specific improvement teams and groups, and it is logical to do a functional area scrap breakdown to figure out improvement projects relevant to each functional area.

I applied this constraint-based wafer scrap value model to Fab 9's 1993 third quarter P6x wafer losses to arrive at a Pareto of line yield losses having the most revenue impact. Losses in the Diffusion functional area, which were the #1 on the wafer scrap Pareto (based on number of wafers lost) slipped to #5 on the new constraint-based Pareto, behind Litho. Several other differences were also observed, proving that the wafer scrap Pareto, based purely on the number of wafers scrapped, does not truly reflect the significance of the resulting revenue impact to the factory. Suppose that the Factory's policy was to fund the top 3 line yield improvement projects. Then, taking a constraint-based approach instead of a pure wafer scrap Pareto approach would result in a different set of line yield improvement projects being funded.

A case study within the Diffusion area:

Figure 4.4 shows the revenue impact of losing a wafer at the various P6x diffusion steps. The Polycure step comes after all the 14 litho steps, and so a polycure wafer loss is worth $25K. However, the Nitride Dep & BufferOx1 steps are before the first litho step, and a wafer loss at these steps have no revenue impact at all. Even if one makes the argument that we have to consider the raw Si wafer cost, it comes to only less than $100. In Figure 4.5, I have grouped together several diffusion steps (e.g., Edge/Spacer/Poly2) because all these steps are between the same litho constraint steps, and hence wafer losses at any of these operations have the same revenue impact.

Figure 4.4 Wafer values at each diffusion step

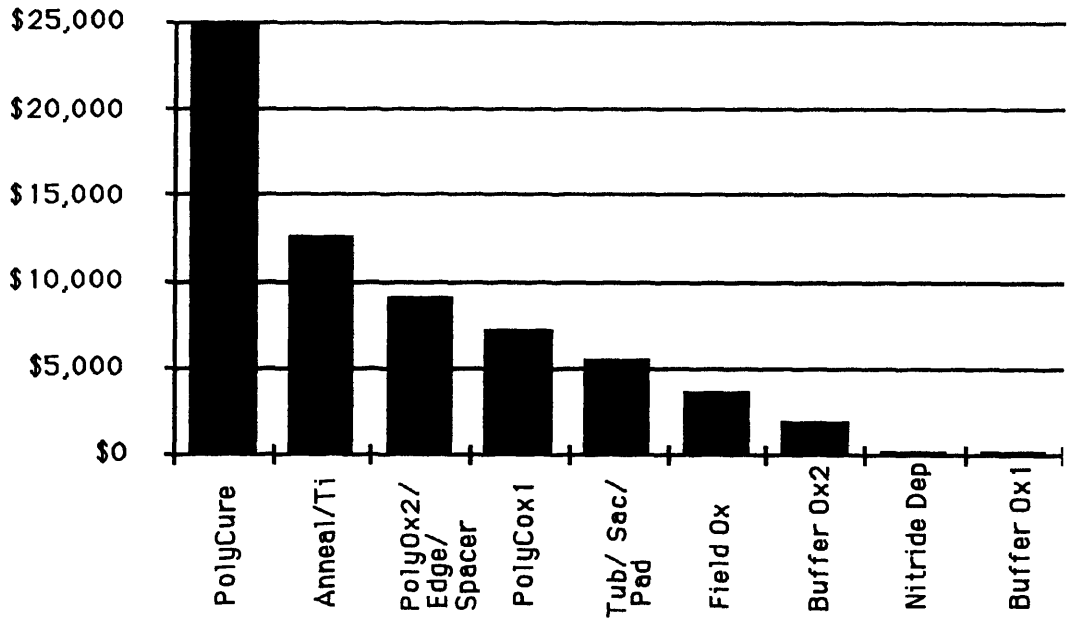


Figure 4.5 shows the Fab 9 Wafer Scrap Pareto for Diffusion, in Q3 '93. Typically such wafer scrap paretos are used to decide which line yield projects to initiate. Topping the list in terms of number of wafers lost is the "Nitride Dep/ BufferOx" step. Hence the wafer scrap Pareto indicates that the maximum leverage in line yield performance within Diffusion would come from improving line yield performance at this step.

However, if we factor in the constraint-based revenue impact of losing a wafer at each step, and arrive at a Pareto based on the revenue impact of the total losses occurring at each Diffusion step (following the method indicated in Table 4.1), we get different

results, as indicated in Figure 4.6. The contrast between Figures 4.5 & 4.6 clearly points out the value of the Constraint-based approach. Although the largest number of wafers were lost at the front-end BufferOx/Nitride step, since those wafers had not yet been processed by the constraint, each loss had very little revenue impact.

Figure 4.5 Fab 9 Q3 '93 Diffusion Wafer Scrap Pareto

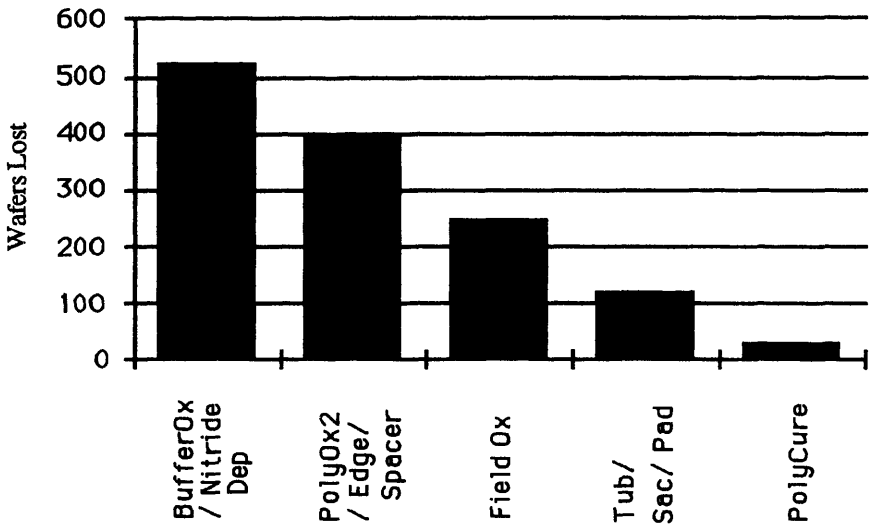
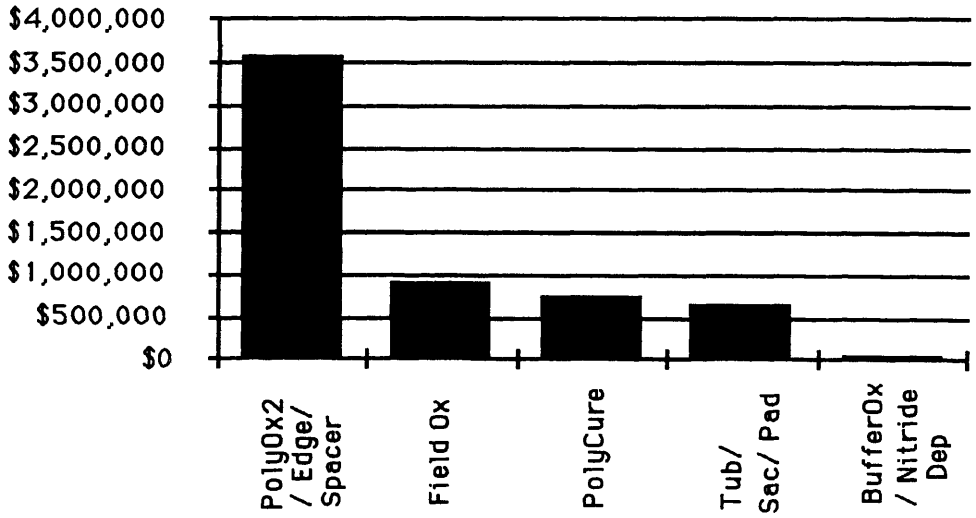


Figure 4.6: Diffusion Scrap Revenue Loss Pareto Fab 9 Q3 '93



4.4 A Line yield management tool-kit based on Systems Thinking and Theory of Constraints:

This section adopts a Systems Thinking approach, in confluence with Theory of Constraints, to develop both proactive and reactive line yield improvement tools for a semiconductor Fab [12]. Each of these tools may be incorporated into a Fab's line yield management strategy as appropriate. Specifically, these tools are intended to provide answers to the key issues and questions raised in Chapter 3.

Pro-active mechanisms

- Appropriate factory priority for line yield
- Line yield ownership in the factory
- Line yield improvement structure in the factory
- Constraint-based line yield goal setting for each equipment cluster & operation step.
- Factory-floor awareness & involvement in LY improvement via goals & **rewards.**
- Analysis of critical behaviors leading to procedural errors (BAPP™ approach).
- Risk analysis (FTA, FMEA) on constraints, to maximize constraint effectiveness.

Reactive mechanisms

- Track performance against LY goals at each cluster to identify areas of improvement.
- Establish accountability & ownership for line yield losses
- Track line yield losses, to identify trends
- Establish a closed loop system for following up on key line yield losses
- Establish structured root cause analysis methods for selected line yield losses
- Implement & communicate root-cause fixes to other sites.

LY Factory priority: *what priority should be given to line yield improvement? Is there away of improving line yield learning rates without adversely affecting die yield and capacity improvement rates ?*

The goal of a Factory is to make money now and in the future. We have seen that maximizing **T,** the throughput of the factory, is achieved by increasing die shipments. Thus, appropriate priority should be given to the three die output levers (capacity, die yield DY, and line yield LY) by ranking improvement projects based on their contribution to increasing the Factory's die shipments. As discussed in chapter 3, the Fab 9 EDO method of prioritizing improvement projects is a good step in this direction. However, I would like to suggest two changes to the EDO method:

(1) Line yield problems are currently prioritized in the EDO Pareto based on the number of wafers being lost due to each problem. However, increasing line yield from 85% to 90% does not automatically mean more die shipments from the factory. If the line yield improvements are achieved at pre-constraint steps, it does not increase Factory **T** at all. All it does is reduce **OE,** which has lesser leverage than increasing T. On the other hand, if line yield is improved at post-constraint steps, it directly translates into increased **T.** Hence, my recommendation would be to rank the line yield projects using the constraint-based wafer value model (discussed in section 4.2) instead of wafer scrap paretos.

(2) While the EDO Pareto keeps in perspective the various levers (capacity, DY, LY), it does not consider what kind of resources are needed to improve capacity, DY, and LY, while deciding what projects to fund, and what not to fund. Let's examine what kind of resources are needed to improve EDO. Most of the die yield improvements come from focused Engineering efforts. Both Engineering and Manufacturing play a role in capacity installs. However, around 25-40% of line yield losses typically are due to procedural errors which implies that **production personnel can be better leveraged to improve line yield.** The EDO Pareto also does not consider the interactions between die yield and line yield. How does one improve line yield ? By improving equipment reliability, preventive maintenance and wafer handling procedures, baseline training, etc. each of which also has a positive impact on die yield. So by working on line yield, the factory's die yield improves, and vice versa. One way to supplement the pure EDO approach would be to recognize the longer term payback of allocating resources up-front to line yield, and establish clear ownership for line yield. This also implies putting in place a LY improvement structure, with clearly defined roles, responsibilities, and empowerment.

LY ownership: *Who should own and drive line yield improvement ?*

The lack of a driving force, combined with a low EDO Pareto factory prioritization leads to a lack of accountability for line yield improvements. Due to the fact that line yield has multiple contributors such as process stability, equipment reliability, procedural factors, etc., it is difficult for any of the traditional departments to own line yield improvement. This introduces the concept of a **line yield coordinator** to integrate key players influencing line yield, and to focus their efforts in improving line yield.

Since each department's priorities are derived from the factory's EDO Pareto driven priorities which dictate that line yield is a lower priority, each department's resources will be busy on die yield and capacity improvements. Thus the line yield co-ordinator typically finds it very difficult to get resources from each department to participate in line yield improvement projects. My recommendations to alleviate this problem would be:

(1) Instead of trying to follow up on all line yield issues, if the line yield coordinator picks a critical few issues based on the constraint-weighted wafer value model, it will be easier to secure cross-functional resources to improve line yield in these few areas.

(2) Without sufficient top management support, the line yield coordinator's job becomes very frustrating, and line yield does not improve either. One way to get around this problem is for top management to realize the longer term payback of allocating resources up-front to line yield, and to support the line yield coordinator role. The existence of a dedicated resource empowered by the plant manager to co-ordinate LY improvement efforts within the plant would stimulate better line yield ownership.

(3) Factory floor ownership is vital for procedural error reduction. Accountability should also exist for all categories of line yield losses. Although the exact approach would vary from Fab to Fab, the key would be to define owners for each category of loss. Individual responsibilities for technicians, operators, engineers, and supervisors with respect to line yield have to be defined.

LY improvement structure: *What would be a good line yield improvement structure? What tasks should be done to systematically improve line yield? Who is best qualified to do these required tasks? How do we get these right resources plugged into line yield improvement ?*

The idea of having a LY co-ordinator as mentioned above, is to integrate all the players influencing LY to focus their efforts in improving the LY performance. Key organizational factors determining the effectiveness of the LY improvement structure are

- Plant management's continued support of the improvement program.
- Improvement role assigned to each individual in the fab.
- Buy-in from fab personnel & commitment to the effort.

Table 4.2 outlines the tasks that have to be done to improve line yield, and also the right players to do these tasks.

Table 4.2

Improvement area	Causal Factors	Solution	Right Players
Procedural errors (Misprocessing)	Lack of Skills Lack of experience Human behavior workstation issues PM procedures/specs	Training & Hiring practices On-the job manufacturing experience Awareness, involvement, incentives Automation (bullet-proof systems) Better oper spec/PM methods	Training Manufacturing Manufacturing Automation Manufacturing, Engineering
Equipment failures	Eqpt. Unreliability Eqpt. Incapability Robot/SC failures	PM procedures etc. Match process & equipment capabilities Increase automation reliability	Engineering, Mfg. Engineering Automation
Process related losses	Integration problems Cpk related problems	Match process & equipment capabilities	Integration (yield) Engineering
Low Yield @ sort	Die Yield related issues	Bill back to the correct operation step(s) Process/Equipment issues	Yield Engineering Engineering
E-test failures	process, equipment, operations	Billback to the correct operation step(s) Process/Equipment issues Misprocessing issues	Yield Engineering Engineering Manufacturing
Facilities losses	power loss/glitch natural disasters	system immunization fixes Pray to God & hope for the best	Site Services GOD

Now that we have identified the right resources to work on each line yield issue, the next step is to plug these resources in the right way, to work on the appropriate issues.

At Fab 9, improvements are driven in a top-down fashion by the Factory Improvement Steering Committee (FiSC) to the Functional area teams (A-teams), which initiate multiple improvement teams (i-teams) within each area to work on area-specific EDO limiters (could be capacity, die yield, or line yield depending on the area). This structure helps to establish a consistent improvement vision for the factory, keeping in perspective all aspects of maximizing good die out (DY, LY, Capacity). Each A-team analyzes the issues that are most important for that area, and constitutes i-teams (improvement subteams) appropriately. In other words, the improvement efforts are focused on key limiters, provided the EDO paretos are accurate. However, this approach is not helpful to increase awareness, involvement, and ownership on line yield related issues from the shop floor. Historical data indicates that procedural error related losses account for 25%-40% of total line yield losses in a factory. Hence, supplementing Fab 9's top-down improvement strategy with a grass-roots level factory floor involvement would result in a more effective improvement structure.

Certain other Fabs have Bay Improvement Teams that are focused on line yield issues, and hence provide a very good grass-roots level involvement across all shifts in LY improvement. The question though, is whether it makes sense for all areas to concentrate equally on line yield improvement. Constraint management theory, when applied to continuous improvement would argue for a "focused" improvement effort at the key factory limiters/constraints and just sustaining current levels, at the non-constraints.

Constraint-based goal setting for each cluster & operation step

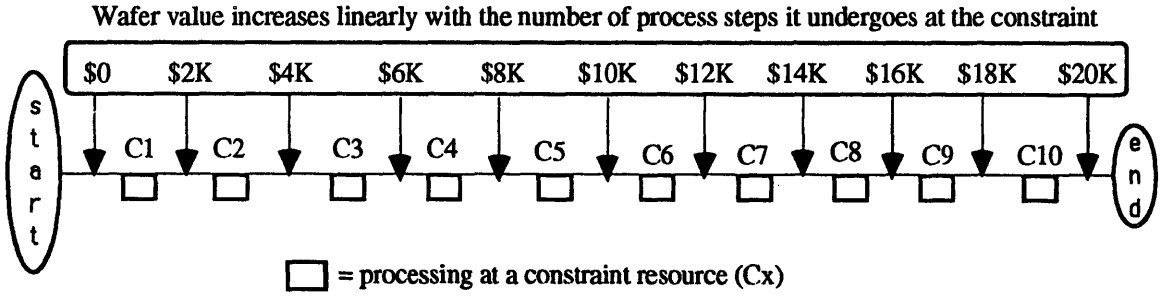
It is common practice to establish line yield goals in Factories. Setting a Factory LY% goal is useful because that it gives an overall sense of direction and progress. However, this level of granularity is not helpful in conveying to operators on the factory floor about how they are performing, relative to the goal. Tangible goals are very important for the shop floor. I studied line yield systems at several Intel Fabs and received feedback from several factory-floor personnel. My studies indicate that a weekly scrap goal by equipment cluster seems to be preferred to other goals such as losses per 10,000 activities or overall %LY. The weekly scrap goal is very tangible because people can relate to the number of wafers lost at their work-station or cluster on a daily or week-to-date basis.

So the logical next step would be to translate the overall Fab LY% goal into individual cluster goals. For a certain Fab wafer start level (say, 7000 wafer starts per week) and a goal of 90% factory LY, the scrap budget is 700 wafers per week. These 700 wafers have to be allotted to individual operations or clusters. One way of doing this is by the number of moves per operation or cluster. Although this method takes into consideration the extent of wafer handling done at each step, the fundamental problem is that it considers every loss as equally valuable, irrespective of where it happened in the production line.

The constraint-based wafer costing model discussed in section 4.2.3 gives us an idea of how to allocate the scrap budget among the various clusters/operation steps. It gives a good feel for the ideal number of wafers we can afford to lose at each operation step for a specified LY% goal. **The constraint-weighted model also helps to identify the most critical areas that really need to improve their LY performance, and those non-critical areas that only need to sustain their current level of LY performance.**

I will now describe how the constraint-based model can be applied to allocate scrap targets to operating steps at different locations in the process flow. For the sake of simplicity, lets assume a simple process flow as shown in Figure 4.7

Figure 4.7



From a revenue impact perspective, losing 10 wafers at C1 is the same as losing 2 wafers at C5 or a single wafer at C10. Assume that you are running at 7000 wafer starts per week, and your goal is to achieve 90% line yield. i.e., you will tolerate scrapping 700 wafers a week, and you are trying to decide how many wafers you can afford to lose at each of the 10 operating steps. Clearly, losing 700 wafers at C1 would have a different revenue impact than losing all 700 wafers at C10. So the key is how to allocate the scrap target of 700 wafers among all the operating steps such that the revenue impact felt at each step is the same. The method I used to arrive at constraint-based scrap targets for each operation step of this simple process flow of 10 steps is outlined in Table 4.3.

Table 4.3

Operation Step	Value of a wafer at this step = Number of steps at the constraint upto this step = (Weight W)	wafer scrap affordable at this step, relative to step1 = reciprocal of the step weight, (the more valuable the step is, less scrap you can afford)	Target wafer scrap at this step with a goal of 90%LY@7000WSPW i.e. scrap target = 700 How many wafers can we lose at each step ?
C1	1	1	$\frac{700 * 1}{2.93} = 240$
C2	2	1/2 = 0.50	$\frac{700 * .5}{2.93} = 120$
C3	3	1/3 = 0.33	$\frac{700 * .33}{2.93} = 80$
C4	4	1/4 = 0.25	$\frac{700 * .25}{2.93} = 60$
C5	5	1/5 = 0.20	$\frac{700 * .2}{2.93} = 48$
C6	6	1/6 = 0.167	$\frac{700 * .167}{2.93} = 40$
C7	7	1/7 = 0.143	$\frac{700 * .143}{2.93} = 34$
C8	8	1/8 = 0.125	$\frac{700 * .125}{2.93} = 29$
C9	9	1/9 = 0.111	$\frac{700 * .111}{2.93} = 26$
C10	10	1/10 = 0.10	$\frac{700 * .10}{2.93} = 23$
		Total = 2.93	Total = 700

Once the above method is understood, it is very easy to apply it to a more complicated process flow, such as the P6x process, with 200 process steps, including 14 constraint steps at Litho. Table 4.4 shows how you would apply the method to P6x.

Table 4.4. Methodology for Calculating weekly Cluster Scrap Targets for P6x

Operation Step	Value of a wafer at this step = Number of steps at the constraint upto this step = (Weight W)	wafer scrap affordable at this step, relative to step1 = reciprocal of the step weight, (the more valuable the step is, less scrap you can afford)	Target wafer scrap at this step with a goal of 90%LY@7000WSPW i.e. scrap target = 700 How many wafers can we lose at each step ?
step 1	W_1	$r_1 = 1/W_1$ if $W_1 > 0$ $r_1 = 0$ if $W_1 = 0$	$\frac{700 * r_1}{R}$
.....			
step i	W_i	$r_i = 1/W_i$	$\frac{700 * r_i}{R}$
.....			
step 200	$W_{200} = 14$ (14 constraint steps)	$r_{200} = 1/W_{200}$	$\frac{700 * r_{200}}{R}$
		Sum of reciprocals (R) $= r_1 + ... + r_i + ... + r_{200}$	Total = 700

Table 4.5 Weekly Cluster Scrap Targets for P6x

Cluster Toolset	Average cluster weight	Cluster Scrap Goal	Current scrap performance	Delta from the goal	Focus Areas
Metal Etch	11	27	250	(223)	Metal Etch
Planar	10	9	52	(43)	Planar
Anelva	9.6	31	64	(33)	Anelva
DIF frontend	3.6	186	210	(24)	Diffusion
Gold	14	3	10	(7)	Gold
WAMO	9.8	10	9	1	
AMAT	10.5	24	23	1	
DIF-backend	14	3	0	3	
WJ	7	6	2	4	
LRC	6.7	70	54	16	
Litho	7.5	130	101	29	
Implant	4.4	56	26	30	
AWS	5.1	145	66	79	
Total		700	867	167	

I applied the methodology outlined in Table 4.4 to translate Fab 9's overall goal of say, 90% LY, into individual operation step scrap goals. Then I grouped operation steps within each functional area to get wafer scrap goals for each equipment cluster as shown in Table 4.5.

The scrap targets in Table 4.5 could be used to select a few focal areas to concentrate line yield improvement efforts. Causal analysis could be done on line yield losses in these areas to pinpoint root causes. **However, one has to be careful in directly translating these into targets on which performance incentives for the factory floor are based** because this technique does not consider the extent of human wafer handling involved at each operation.

The constraint based wafer costing model also helps to do a sanity check on how the overall Factory 90% LY goal is achieved. In the absence of such a model, one would be happy if the Fab hits the target of 90% LY regardless of where in the production line the losses (10%) were taken. In other words, there would have been no distinction as to whether the losses were taken at the back-end (post-constraint) or at the front-end (pre-constraint). The constraint-based model also highlights how each area (and operation step) is doing relative to its allotted wafer budget, in addition to indicating whether or not the overall fab/factory LY% goal was met.

Factory-floor awareness & involvement in LY improvement

The importance of shop-floor buy-in cannot be over-emphasized, and is absolutely critical to the success of any LY improvement effort. Chapter 3 examined some of the hidden reasons behind procedural errors. Most people at work do most of what they are supposed to do most of the time. They are cooperative, hard working and dependable. However, there are times when even the good performers do something wrong. The key is in preventive management using pro-active interventions to prevent problems from occurring versus reacting to solve problems after they occur.

Problem: They don't know why line yield is important

Solution: Mandatory training classes to impart the importance of line yield would be very helpful in increasing awareness of line yield on the factory floor.

These training classes can (1) Explain the benefits to the organization, and to them, for doing a task right, as well as the harm to the organization and to themselves (via lost opportunity to obtain performance incentives) for doing it wrong. Personal rewards could be financial incentives, awards, etc. (2) Describe the relationship between the individual tasks people perform, and the mission of the organization.

Problem: They don't know how to do it

Solution: In cases where procedural errors are induced by a lack of knowledge on how to perform a certain operation step or task, *the solution lies in better training curriculum and hiring practices.* As mentioned in Chapter 3, hiring policies may have to be re-examined for appropriateness of timing. It was observed that most factories, while they are being built and brought into production, adopt a just-in-time hiring policy. i.e., hiring of operators and technicians often gets delayed to reduce upfront labor costs, with the result that training time is low. The training also could be termed "just-in-time" in this respect. A lack of training, and inadequate training practices, often manifest themselves as procedural error related line yield losses.

Problem: They think something else is more important. There will be a lack of relative priorities, or a conflict of priorities. This applies to the line yield problem, especially when there are seemingly conflicting tasks such as increasing wafer outs versus say, taking a machine down because of particles. Another example would be taking the time to do root cause analysis on a wafer being scrapped, as opposed to just scrapping the wafer and moving on to process the next wafer.

Solution: Clearly communicate the relative priorities to the factory floor. Obviously, priorities would be different for different functional areas, at any point of time, and would change within the same functional area, at different time periods. Hence if the work involves multiple priorities or if the priorities change frequently, explain to the floor why one task is of higher priority than the other. If you frequently change what you identify as "hot priorities", you should devote sufficient communication time and effort at the time of change to make certain everyone understands which project is now the "hot project", and why it is so. You should take care not to label everything as a "hot project". An example would be NOT to confuse employees by allocating line yield, die yield, capacity, safety, etc. as equal priorities to work on. The focus should be different for different areas, with

the exception of safety. Constraint-based wafer scrap goals, as described in Table 4.8 can be used to determine the few areas that should improve line yield performance. The message to the other areas should be to sustain their current line yield level, and to focus on other output levers as appropriate.

Problem: There is no positive consequence to them for doing it. Currently, there are no line yield related incentives for factory-floor personnel in most factories.

Solution: Reward LY performance. An excellent example is an incentive system adopted at a certain Fab where a reward of $10 was given to each technician or operator in every functional area that met it's weekly scrap goals. Human behavior research indicates that people do things for which they are rewarded, and don't do things for which they are not rewarded. In other words, performance that is rewarded will increase in its frequency. Research also suggests that these positive consequences or rewards are most effective when they occur immediately following the specific action being rewarded, and when the reward occurs at a high frequency. In other words, small awards received immediately and frequently seem to have more effect on performance than larger rewards delivered long after performance and infrequently.

Problem: They think they are doing the right thing. "How do they know when they are doing a good job" ? The problem has to do with getting feedback. Feedback has a big influence on performance. All psychologists agree that feedback is one of the most critical requirements for sustained, high level performance of any human act.

Solution: An easy way to give feedback on the performance of each cluster with respect to line yield goals would be via DISPLAY BOARDS at Fab entrances. This technique has been proven to be very effective in certain Factories. Week-to-date cluster-specific display boards at Fab entrances are more effective than bulletins or Fab newsletters that indicate overall Fab LY% performance. Bulletins & newsletters do not give cluster-specific tangible feedback to the factory floor.

Problem: Obstacles beyond their control. This reason for non-performance includes all those situations, conditions, and influences outside the individual that are preventing performance from occurring.

Solution: This is an opportunity for looking at systematic fixes. e.g., station controllers, training procedures on how to respond to out of control events, such as station controller failures etc.

Problem: Personal limits prevent them from doing it. Personal limits mean individual physical limits that are unchangeable and are in fact preventing job performance. Such deficiency is not a function of learning. Personal limits could be temporary or permanent.

Solution: If there are critical personal limits that relate to success or failure in a job, be sure to evaluate the applicant before hiring or promotion to determine whether or not those personal limits exist. Also, don't be too quick to label the non-performer as failing because of personal limits. If a permanent personal limit is causing failure, there is nothing much that can be done other than replace the person or live with the problem.

Problem: Personal problems prevent them from doing it. Such problems refer to those happenings in a person's private life outside the work environment that appear to be reasons for non-performance on the job.

Solution: Often times, the best way to address this would be to work it out with the person. The more systematic approach would be to put in place bullet-proof systems which would not be impacted by a person's emotional state of mind.

Analysis of critical behaviors leading to procedural errors

The above section describes some of the underlying reasons for procedural errors, and some solutions to these reasons. These hidden reasons induce certain behaviors in people that lead to actions which in turn cause misprocessing. The best way is to address each of the above-mentioned reasons so as to avoid undesirable behaviors. However, at this stage, it would also help to identify what behaviors are desirable, and what behaviors are undesirable. The more agreement there is on the right behaviors, the easier it is to make improvements.

The Behavior Accident Prevention Process (BAPP™ approach) adopted at Fab 9 in an effort to reduce safety-related incidents, can be applied to reducing line yield losses due to procedural errors. Since procedural line yield losses and safety incidents are both

caused by employee actions and behaviors, the behavior based accident prevention plan can potentially be applied to reduce procedural errors. The approach would provide a long-term, effective, and measurable plan that reduced accidents, with an emphasis on employee behavior and actions. The company that developed the concept "Behavioral Accident Prevention Plan (BAPP™) was Behavior Science Technology (BST) Inc.

The basic ideas behind BAPP™ are

- Accidents involve people's behavior.
- If you want to decrease accidents, you must improve the safety of their behavior.
- Since behavior is measurable, it can be changed.
- The way to manage safe behavior is to identify the behaviors that are critical to accident occurrence, train people to measure them and use the results of this measurement to provide feedback to employees to guide their future behaviors, and feedback (data-reports) to management to guide its decision-making.

Two important side benefits of this approach are

- Since you are measuring behavior (as opposed to accidents/procedural losses), you do not have to wait until an accident/loss happens to know you are in trouble. Unsafe-behavior is an early-warning system for accidents.

- Developing a list of behaviors that are critical to safety (error avoidance), and training people to measure these behaviors often entails shop-floor involvement. This leads to more commitment to desirable behaviors from the shop floor.

Risk analysis (FTA, FMEA) on constraints, to maximize constraint effectiveness.

Equipment related failures account for around 40%-50% of the total line yield losses in a typical wafer Fab. Ideally, one would like to proactively detect potential failures and rectify them before the equipment actually fails and causes line yield losses. However, in reality, given the number of diverse pieces of equipment, and the limited resources available, it may not be practical to track down all the failure modes of all pieces of equipment in a fab. In this situation, **a "theory of constraints" approach would come in very useful in determining what pieces of equipment to focus on.**

As discussed in the "constraint-based goal setting" section earlier, once goals/targets have been established for each equipment cluster, and performance against goals have been tracked, we can identify clusters where operations have to be improved in order to meet the goals. It would then be useful to proactively analyze key pieces of equipment in these clusters.

FMEA (Failure Modes & Effects Analysis) provides a structured, team approach for identifying and analyzing potential or existing failure modes [14]. The purpose of conducting an FMEA is to understand all failure modes and their effects, so the higher risk items can be prioritized and eliminated. A recent "Equipment Reliability Benchmarking" effort by Intel found that FMEA is common practice in several Japanese semiconductor companies such as NEC.

Training classes for FMEA are available at Intel. However, what needs to happen is the active use of FMEAs on the factory floor & by engineering, to analyze key pieces of equipment in the factory, and to proactively prevent equipment failures.

Track performance against goals at each cluster, to identify areas of improvement.

As mentioned earlier, constraint-based goal-setting at each cluster and operation step helps to track performance, and to focus improvement efforts on those clusters that are not meeting their goals. The message to be given to clusters that are meeting their goals is to sustain their current level of performance. Having established goals for each cluster, tracking performance and communicating it to the shop-floor are also crucial. Display boards at fab entrances, as described earlier provide an excellent mechanism for this. Rewards for meeting the goals also go hand in hand with tracking performance.

LY loss tracking and identifying trends

A real-time notification system is necessary to alert the appropriate factory personnel about wafer losses in the line. In addition to notification, the automated tracking system should also be helpful in identifying loss trends in the line. Almost all Intel fabs have instituted automated wafer loss tracking systems (LCS stands for "Loss Code System"). Losses are categorized into procedural, equipment & process related loss codes, and automatic mailing/distribution lists exist for each loss code. Such automatic reports

ensure that the station owner, the technicians on the floor, and the management structure knew whenever wafers were scrapped. It is also very important to bill-back end of line losses (e-test) to the station where the defect was introduced into the wafer. The loss tracking system can be used to aid this bill-back process. However, the follow-up system should ensure that the billed-back losses are investigated by the improvement team associated with the station where the losses were billed back to. This is very crucial since the defective wafers were using up precious resources all the way through the end-of-line in spite of being defective.

A closed loop system for following up on key line yield losses

The importance of having a closed loop system to follow up on wafer losses being documented cannot be over-emphasized. To an operator or technician on the line filling out a scrap report, if the report is not reviewed by anyone, or if no action is taken either for root-cause analysis or for implementing improvement suggestions, there is no motivation for attention to detail, let alone for filling out the scrap report.

Ideally, one would want to document, understand root causes, and implement fixes for each wafer loss. However, realistically, given the volume of wafer starts in a factory, and the available resources, it may be impractical to close the loop on every wafer loss.

Another argument against following up on every loss incident is the necessity of focused improvement efforts based on Goldratt's theory of constraints. Constraint management theory argues that efforts to improve operations at the non-bottlenecks serve only to make the people feel good that they are doing some improvements in the production line. However, such improvements at non-constraints do not have any impact on the overall output of the factory. The only improvements that really affect factory performance are those that improve the output at the constraint(s). At the non-constraints, the key focus should be on sustaining current levels of performance. This would argue for different wafer thresholds and follow-up policies for investigation, for different operations/stations, based on their impact on factory performance. In other words, more focus should be given to losses at the constraints.

Clear expectations on the follow-up process should be conveyed to the factory floor so that a person filling in a scrap report would get appropriate feedback, and not get the

feeling that the scrap report he/she is filling out is being ignored. For instance, a loss at a station where line yield improvement is not a high priority, should still be documented because this will come in useful in the future, for analyzing performance trends and root causes of failure at that station, if line yield at that station becomes a priority.

Structured root cause analysis methods for losses

For losses that are selected to be investigated, what methods are used to arrive at root causes? Arriving at root causes is critical to enable root cause fixes. A wide range of problem solving techniques exist, from structured methods such as Seven-Steps, or Fault Tree Analysis, or Failure Modes Effects Analysis, to unstructured, adhoc methods. Although structured techniques will maximize the changes of arriving at the root causes, their use is only slowly catching on in line yield loss cause analysis in the factories. In addition to training personnel in these techniques, it would help to reward improvement teams that actually use these structured techniques in their root cause analysis of line yield losses. In Factories where the seven-step method is now regularly being used, top management strongly drove the training of all Fab personnel in the use of these techniques.

Implementation & Communication of root-cause fixes to line yield problems

Line yield fixes could be as diverse as installing a station controller, designing a new process step, or changing preventive maintenance procedures, thus requiring diverse resources. This is where the line yield co-ordinator, with management support can play a key role in teaming the various resources and driving root cause fixes.

Is overall Factory Line Yield an appropriate performance metric ?

Should we be aiming for world class overall factory line yield **(95% factory LY)**, or for zero scrap **(100% step LY)** only at the post-constraint operations, with less attention to line yields at pre-constraint operations? Line yield improvements at operations before the constraint do not result in increased factory die output, but result only in reduced operating expenses. However, if a down-stream, post-constraint operation scraps wafers, it implies that constraint time is wasted recreating material that was previously

processed. The question to ask ourselves based on theory of constraints is whether overall Factory line yield is an appropriate factory performance metric. This could have several implications in how we approach line yield improvements.

Inter-fab shared learning for Line Yield

To date, there has been very little sharing of line yield learnings even among Fabs having significant module synergy. Also, there are well-known failure modes that are being allowed to occur in every Fab even though they can be minimized with better designs & known solutions. What is needed is a forum or method for regular real-time sharing of line yield learnings across sites within Intel, to prevent each site from re-inventing the wheel.

Higher process maturity at TD transfer to high volume Fabs

In order for Fabs to focus on long-term yield improvement, frequent exposure to diverting yield crashes (excursions) must be minimized. Process development and manufacturing performance are tightly coupled: an uncharacterized and immature semiconductor process creates great obstacles to meeting manufacturing objectives. However, as mentioned in Chapter 3, the emphasis at a process development (TD) site is on developing process technology in a stable environment (e.g., low volumes). Full scale manufacturing differs from development in many areas including personnel policies, equipment utilization, and operational procedures. Some problems arise only in a full scale manufacturing environment, thus requiring several modifications to the process once it has been transferred.

This lack of congruence between the operating environments has recently led Intel to change the process development sites to take on the dual charter of (1) developing the technology and (2) the operations, for the next generation process. This policy change is expected to enable the transfer of a process better optimized for manufacturing. By better aligning the process to the manufacturing environment, the process can ramp with fewer "unexpected" problems. This reduction in firefighting situations will allow manufacturing resources to concentrate on baseline long-term line yield improvements, as opposed to fighting yield excursions.

The additional responsibilities proposed for development centers would enable full-scale manufacturing to rapidly ramp and focus on fundamental improvements. These objectives include:

- Developing operational standards such as PM schedules, and automation.
- Uncovering "hidden" problems driven by intensity factors, and volume factors.
- Increasing fidelity over time with the manufacturing organization.

Dynamically changing wafer starts per week, to make up for line yield losses

Currently, the number of wafers started per week (WSPW) is pre-determined before the beginning of the week, and kept constant at the pre-determined level (e.g., 5000 WSPW). Even if wafers that have been started are lost in the line, they are not compensated by starting more wafers. This may lead to a situation where a pre-constraint loss may cause the constraint to be idle if the wafer loss causes the pre-constraint buffer to be empty. To avoid idling the constraint, it may be prudent to start more wafers to compensate for the wafers that were lost. This is easy and meaningful when the loss is near the front end of the line (e.g., at front-end diffusion). However, when the loss is further down the line, and beyond the constraint, there is no point in restarting wafers since the constraint time has already been lost (in processing the wafer that was subsequently scrapped). Thus, it would be good to have a conditional wafer restart policy of backfilling wafers that were lost before the constraint.

Chapter 5. Evaluation of Effects

Purpose: The purpose of Step 5 is to evaluate the tangible and intangible results of my solution. This would include satisfaction, skill level, and motivation of the people involved in the solution.

5.1 Introduction

This chapter discusses the tangible results of my research and also the intangible results in terms of satisfaction level of customers with this tool, skill level, and motivation of people involved in implementing my recommendations. This chapter also discusses communication of my key findings across Intel.

5.2 Tangible Results of my Research

The tangible results of my research are reflected in the key project deliverables outlined below.

- *Documentation of line yield methodologies at each Intel wafer fabrication plant.*

  Tangible Results: I've documented the line yield methodologies of each Intel site in my report titled **"Line Yield Methodology Questionnaire Results"**. This report documents line yield methods, line yield trends, loss codes, and wafer loss cause paretos at each of the Intel sites. Hence it serves as a 1-stop-shop for understanding the line yield methodology at any Intel site.

- *Compare line yield definitions and trends at each Intel wafer fabrication plant.*

  Tangible Results: I've compared line yield definitions and trends in the report titled **"Line Yield Benchmarking team Face-to-Face meeting Presentation Materials"**. As part of my internship, I have been co-chairing an Inter-site line yield benchmarking team, with representatives from each Intel site. I arranged a face-to-face meeting of the team members on October 26, 1993, in Santa Clara, CA. The purpose of the meeting was for each site to present their line yield management system to the rest of the team-members to extract, share, and learn line yield methods across sites. I've compiled a report that has a comparison of line yield definitions, metrics, and trends at each site.

• *Analyze the line yield management systems at each Intel wafer fabrication plant.*

- *Deliver a survey and analysis of the line yield methodology used at each site.*
- *Deliver a Pareto of line yield loss causes at various stages of process maturity.*

Tangible Results: The above-mentioned report, **"Line Yield Benchmarking team Face-to-Face meeting Presentation Materials"** includes an analysis of line yield management systems used at each site, and wafer loss paretos.

• *Develop, refine, and apply conceptual models for line yield improvement.*

Tangible Results: My thesis in itself is the tangible result, as far as this deliverable is concerned, since it describes how I developed my Constraint-based systems model of line yield improvement.

• *Recommend key changes to management systems, structure, policies, and processes, to accomplish world class line yield performance at Intel.*

Tangible Results: In addition to my thesis, I've written a concise five page paper entitled "A Constraint-based Systems Approach to Line Yield Improvement". I presented this paper at the 2nd **Intel Manufacturing Excellence Conference (IMEC)**, in San Diego, in January 1994. This paper is also included in the IMEC '94 Digest of Technical Papers. The same paper has also been accepted for presentation at the **International Symposium on Semiconductor Manufacturing '94 (ISSM '94)** to be held at Tokyo, Japan, in June 1994.

5.2 Intangible Results of my Research

This section will describe the intangible results of my research in terms of satisfaction level of customers with this tool, and their motivation in implementing the recommendations outlined in this thesis.

Customer Satisfaction Level:

Key customers of this project's deliverables included both high volume manufacturing Fabs, as well as process technology development (TD) sites at Intel. I've received extremely positive feedback from several sources at all levels, from Plant Managers to Technicians, from both Fabs and TDs about the line yield improvement ideas based on Systems Thinking.

Although I received very complimentary feedback for having come up with a "thought-provoking" and "innovative" approach in applying constraint management to identify and focus on key line yield losses, there were several doubts about its applicability in a TD environment. In other words, the Theory of Constraints based line yield project prioritization approach seems to be more applicable to Fabs than TDs. The reasoning is as follows: The goal of a high volume facility is to make money by maximizing throughput. Hence, it makes sense to discriminate between pre-constraint and post-constraint line yield losses. However, the goal of a TD facility is to deliver a robust process to a high volume manufacturing site. Just as throughput is the means to achieving the goal in a Fab, the means to achieving the TD goal of a robust process is "Information turns". There is some learning about the process to be obtained from each and every wafer loss, regardless of the location of the loss. i.e., regardless of whether the loss was before or after the constraint in the TD site. Moreover, even for the same process, the location of the constraint may be different in a TD site from that in the Fab to which the process is to be transferred. Hence, a pre-constraint step that is losing a lot of wafers at the TD site may end up being a post-constraint step in the Fab. In this case, if the TD site had followed the TOC approach to focusing line yield improvement, it would result in the Fab encountering huge post-constraint losses due to a problem that was neglected in TD.

I agree with the above reasoning in that if the TD sites currently have enough resources to examine every line yield loss incident, and even near-misses, then they clearly do not face

the resource allocation problem that is prevalent in high volume factories. Hence they can afford to look at every line yield loss incident without resorting to a TOC prioritization. However, in a TD facility that has a dual charter of not only developing a new process technology, but also ramping up the new technology to high volumes, the further along they are on the volume ramp curve, the more problems they typically have with improving line yield performance. Sufficient data exists within Intel that to prove that in a TD site that is ramping up volumes to a level that is comparable to wafer start levels at a high volume Fab, there will be a need to prioritize line yield projects, due to the lack of resources to examine each and every line yield loss incident. In such a scenario, TOC based line yield prioritization approaches are expected to come in useful.

The other concern that has been raised about a TOC based prioritization is that it may cause us to turn a blind eye to a problem currently impacting a pre-constraint step, but which could later on potentially impact a post-constraint operation. This is because the underlying issues or causes may be the same, and by ignoring the problem when it occurs at a pre-constraint operation, we are taking a "reactive" mode in the sense that we are waiting for the problem to come back and bite us at a post-constraint step later on. I agree that this risk does exist. However, I would point out that due to a lack of enough resources to tackle every problem, some kind of prioritization has to be done, and as was proven in chapter 4, a TOC based approach is better than the current wafer scrap Pareto prioritization approach.

Following along the same lines, due to the fact that multiple line yield loss incidents have the same root cause, by examining the problems at the post-constraint steps, we could arrive at root cause fixes, which could then be applied to solve problems occurring at pre-constraint operations as well.

On the whole, I would say that customer satisfaction about the ideas proposed in this thesis is high. What remains to happen is the translation of these ideas into actual operating policies at the manufacturing sites.

Chapter 6. Standardization

Purpose: The purpose of Step 6 in the Seven Step Problem Solving Method is to standardize the solution by permeating it throughout the organization.

6.1 Introduction

This chapter describes methods and ideas for communicating the key findings of this research throughout the Components Manufacturing organization. Specifically, I will describe how the "steering committee" approach to managing the internship has facilitated cross-site communication from its inception. I will also talk about a line yield benchmarking activity I was closely involved in, the Intel Manufacturing Excellence Conference where I presented my findings, and also the Road show of Intel sites that I went on, to communicate my findings to each site. I will also discuss the current status of implementation of key ideas, as well as sketch out a tentative roadmap for spreading and implementing my findings at the various Intel sites.

6.2 The Line yield Steering Committee

My project was directed by a Steering Committee with representatives from 4 High volume Fabs, and 2 TD sites. These representatives typically were, or had been Manufacturing Managers or Engineering Managers at their respective sites, which implied that they had significant contacts, influence, and communication channels at their sites. The deliverables of the project were arrived upon by this steering committee at the beginning of the internship, and the committee members have greatly facilitated my data collection efforts at the various sites. After reviewing the key findings of the project, this committee recommended a road-show to communicate my findings to all the sites, and also set up target audiences and meeting schedules at each site. The road show is discussed more in section 6.4.

6.3 The Line yield Benchmarking Team

A Line Yield Benchmarking Team was initiated in July 1993 by Intel's Competitive Analysis Group. Due to the close alignment of its goals with those of my internship, I got actively involved in this effort, and eventually became the co-chair of this team. This team would provide another good resource to help in communicating the key findings in this thesis to the various sites, since it has representatives from each site.

6.4 Intel Manufacturing for Excellence Conference (IMEC) '94

The IMEC provides an ideal opportunity for spreading new ideas within Intel, across all the Intel manufacturing and process development sites. I had the privilege of presenting my paper on line yield improvement at the 1994 IMEC in San Diego, California. The IMEC had more than 400 attendees, including Manufacturing Vice Presidents, Plant Managers, Engineering Managers, Manufacturing Managers, Engineers, Technicians, and Operators. It was a great forum for permeating the findings within Intel. As a result of the IMEC presentation, a wide range of people within the Manufacturing division are now exposed to the key findings of the project.

6.5 Line Yield Road show '94

The line yield steering committee for my project recommended a road show at the conclusion of my internship, which entailed traveling to each of the domestic sites and conducting a presentation of my key findings. This was conducted in January 1994. The schedules are shown below.

Site	Location	Presentation Date
Fab 9	New Mexico	January 10, 1994
D1, Fab 5	Oregon	January 11, 1994
D2	California	January 12, 1994
Fab 6	Arizona	January 13, 1994
Fab 11	New Mexico	January 14, 1994

The audience at each site included Plant Managers, Engineering & Manufacturing Managers, Shift Managers, Shift Supervisors, Engineering Team Leaders, and Engineers.

6.6 Next Steps .....

Via all the communication methods described in sections 6.2. to 6.4 above, there is a wide spectrum of people within Intel's Manufacturing Organization that are familiar with the key findings of this research. What needs to happen now is the actual adoption of the findings in the line yield management policies of individual sites.

Several sites have expressed interest in implementing at least some of the recommendations. Since what my thesis has provided is a set of tools, each of which can be incorporated into a Plant's existing line yield management strategy, it would be easy to implement the recommendations incrementally, without totally revamping the existing line yield management system in a Factory. This would lead to a gradual, incremental adoption of the findings into Intel's factories.

Due to the typical challenges involved in instituting change across several sites simultaneously, one strategy may be to first select a pilot site that is committed to implementing the changes. Considering the natural change-inertia of existing organizations, it may be best if the pilot site is a new start-up Fab that has not yet formalized its line yield management system. The hope is that a successful implementation in the pilot site will induce other sites to adopt the pilot site's line yield management strategy, at a later date. It is currently expected that Intel's new upcoming wafer fabrication facility, Fab 11, in New Mexico, will act as the pilot site for implementing the key findings of this research. This expectation is based on several talks I've had with Fab 11's senior management. Since I will be part of the team responsible for developing improvement strategies in Fab 11, I will be able to follow through with implementing key ideas from this research.

To enhance communication and share line yield learnings across various sites, it may be good to institute a cross-site forum or team, with representatives from the various sites, with the charter of ensuring a robust line yield communication channel across sites. This would help overcome inter-fab parochial boundaries, if any. It would also help to standardize at least critical facets of a line yield management system across all Intel sites.

Chapter 7. Summary of Key Recommendations

Purpose: The purpose of Step 7 in the Seven Step Problem Solving Process is to reflect on the learnings from the improvement project and to identify tentative research topics (or problem areas) based on the current work.

7.1 Introduction

This chapter will summarize the key recommendations of the project, and identify tentative themes for further research on line yield improvement in semiconductor wafer fabrication.

7.2 Summary of key learnings

The key findings of this project can be divided into two categories:
(a) Core Recommendations that provide "must have" features in a LY system.
(b) Desirable Recommendations that provide "nice to have" features in a LY system.

Core recommendations form the heart of any line yield system. They provide a set of tools whose presence is absolutely critical to an effective line yield management strategy. It would be nice to have all the Intel Fabs implement at least the core recommendations. This would lead to all the Fabs having a standardized approach to line yield improvement, which would not only ensure that each Fab has an effective LY system, but also facilitate inter-Fab shared learning.

Desirable recommendations provide features that would be "nice to have" in a line yield system. However, these recommendations would require a significant effort to implement, and a long time frame to observe tangible results. For this reason, given the varying nature of relative priorities for line yield in various factories, it may not be practical to implement these recommendations in every Fab.

I will now list the "core" and "desirable" recommendations separately. Each of these recommendations have been described in detail earlier in this document, and I would refer the reader to Chapter 4, for an in-depth discussion on any recommendation.

7.2.1 Core Recommendations - A Summary

1. **LY Factory priority:**
   The overall goal of a Factory should be to maximize die shipments. Hence, appropriate priority could be given to line yield improvements, with respect to other die output levers such as die yield & capacity. The **EDO Pareto** (Equivalent Die Out) Approach is a good way to assign a Factory's relative priorities and resources.

2. **LY ownership:**
   It would be good to have a "**Line Yield Coordinator**" in each Factory, to integrate key players influencing line yield, and to focus their efforts in improving line yield performance. However, strong management support is necessary for this person to be effective in the co-ordinator role.

3. **LY related incentives:**
   Line yield performance could be rewarded, to motivate the factory floor to improve line yield. Although the exact **LY reward mechanism** would vary from Fab to Fab, the key is to have positive incentives for line yield performance.

4. **LY Awareness & Feedback:**
   The factory floor could be given feedback on their performance. Display boards at Fab entrances that show equipment cluster-specific week-to-date number of wafers scrapped, along with scrap goals, would provide very tangible feedback to the floor.

5. **LY Project prioritization:**
   It would be good to prioritize all potential line yield improvement projects based on the incremental die output contribution of each project. However, instead of using the current wafer scrap paretos to prioritize projects, it would be better to use the **Constraint-based Wafer Scrap Value model** described in this thesis, because it accurately reflects the revenue impact of each line yield improvement project. Scrap goals can be established for each equipment cluster based on the Constraint weighted model, and performance of each cluster relative to the goals can be tracked. This helps to focus our limited line yield resources on those projects having the most die output impact.

6. **LY Follow-up systems:**

   At a minimum, line yield losses in **Key areas** (as identified by the Constraint-based wafer scrap model) should be followed up, analyzed for root causes, and fixed, without instilling fear of disciplinary action. A **closed loop** is absolutely necessary.

7. **LY root cause analysis:**

   Systematic root cause analysis techniques are necessary to be able to arrive at the root of the problem. Examples are the **Seven-Step** method and **Fishbone** diagrams.

8. **LY solution implementation:**

   The line yield **co-ordinator** should be responsible to ensure that solutions are found and implemented for at least **key line yield problems.**

7.2.2 Desirable Recommendations - A Summary

1. **LY proactive risk analysis for process equipment:**
   It would be good to proactively use **FMEAs** (Failure Modes & Effects Analysis) and **FTAs** (Fault Tree Analysis) on key pieces of equipment. The constraint-based wafer scrap model can be used to decide what the key pieces of equipment are.

2. **Analysis of critical behaviors leading to procedural errors:**
   The **BAPP™ approach** currently used at several Intel factories for improving safety could potentially be applied to decrease the occurrence of procedural errors. This would, in turn, lead to better line yield performance, since procedural related errors account for a major portion of current line yield losses in any Factory.

3. **Inter-Fab Shared Learning For Line yield:**

   To date, there has been very little sharing of line yield learnings even among factories that have significant module synergy. Also, there are well known failure modes that occur in every Fab. It would be good to establish a forum for real-time communication and sharing of line yield related learnings across sites. This would not only prevent the same problem from occurring in other sites, but would also enable sharing of solutions, preventing the "re-inventing the wheel" phenomenon.

4. **Changing how line yield is used as a Factory performance indicator:**

   Which would we prefer, of the 2 following options:
   1. A factory that runs at 85% line yield, with most of the losses occurring at pre-constraint operations, OR
   2. A factory that runs at 95% line yield, with most of the losses occurring at post-constraint operations ?

   In other words, should we be aiming for world-class overall factory line yield, (e.g., 95% Factory LY) or for zero scrap only at post-constraint operations (100% post-constraint step line yields) ? If a factory improves its line yield from 85% to 95% by reducing line yield losses at pre-constraint operations, it does not increase the factory's die shipments at all. All it does is reduce operating expenses. However, if a post-constraint operation improves line yield, it increases the factory's die shipments.

Bibliography

1. Benfer, Rich, *"Learning during Ramping: Policy Choices for Semiconductor Manufacturing Firms"*, S.M. Thesis, MIT Sloan School of Management, 1993.

2. Camp, Robert C., *"Benchmarking: The search for Industry's best practices that lead to Superior Performance"*, Quality Press, 1989.

3. Goldratt, Eliyahu and Jeff Cox, *"The Goal"*, North River Press, Inc. 1986.

4. Goldratt, Eliyahu and Bob Fox, *"The Race"*, North River Press, Inc. 1987.

5. Goldratt, Eliyahu, *"Theory of Constraints"*, North River Press, Inc. 1990.

6. Goldratt, Eliyahu, *"The Haystack Syndrome"*, North River Press, Inc. 1990.

7. Kalpakjian, Serope, "Manufacturing *Engineering and Technology"*, Addison Wesley, 1992.

8. Kempf, Karl and Dub Dierke, *"The concepts of Constraint Management applied to Fab"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

9. Kozicki, S.Hoenig, and P. Robinson, *"Cleanrooms - Facilities & Practices"* , Van Nostrand Reinhold Books, 1991.

10. Marcyk, Gerald, et al, *"Line Yield Improvement in a Development Environment"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

11. McMichael, Dave, *"Operator Driven Scrap Teams"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

12. Menon, Viju, *"A Constraint-based Systems Approach to Line Yield Improvement"*, Proceedings of the Intel Manufacturing for Excellence Conference, San Diego, 1994.

13. Nahmias, Steven, *"Production and Operations Analysis"*, Richard R. Irvin, Inc. 1989.

14. Nakajima, Seiichi, *"Introduction to TPM"*, Productivity Press, 1988.

15. Rodgers, Bob et al, *"Fab 6 Line Yield System"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

16. Senge, Peter, *"The Fifth Discipline: The Art & Practice of the Learning Organization"*, Doubleday Publications, 1990.

17. Sohn, Bruce, *"Maximizing Profits by Applying Engineering Resources at Constraints"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

18. Sohn, Bruce, *"Dealing with Constraints in a Semiconductor Fab"*, Digest of Technical Papers, Intel Manufacturing Excellence Conference, 1993.

19. Wolf, S, *"Silicon Processing for the VLSI Era: Volume I"*, Lattice Press, 1986.