

Accelerating Time-to-Money for a Semiconductor Wafer Fab Capacity Ramp

by
Christopher K. Keith

S.B. Materials Science and Engineering, M.I.T., Cambridge, MA, 1991

Submitted to the Sloan School of Management and the
Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degrees of

Master of Science in Management
and
Master of Science in Materials Science and Engineering

at the
Massachusetts Institute of Technology
June 1996

(c) Massachusetts Institute of Technology, All rights reserved

Signature of Author _____
Sloan School of Management
May 10, 1996

Certified by _____
Professor Rebecca Henderson, Thesis Advisor
Sloan School of Management

Certified by _____
Stanley Gershwin, Senior Research Scientist, Thesis Advisor
Department of Mechanical Engineering

Certified by _____
Professor Lionel Kimerling, Department Reader
Department of Materials Science and Engineering

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

Accepted by _____
Professor Michael Rubner, Chairman
Department Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 14 1996

ARCHIVES

LIBRARIES

Accelerating Time-to-Money for a Semiconductor Wafer Fab Capacity Ramp

by

Christopher K. Keith

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

Master of Science in Management
and
Master of Science in Materials Science and Engineering

ABSTRACT

Companies in the semiconductor industry plan to build over one hundred semiconductor fabs during the next four to five years. Each of these fabs requires approximately one to two billion dollars of investment. The pressure to quickly get the fab up and running and recover the initial investment is enormous. The pressure is most intense during the factory start-up, also referred to as the capacity ramp.

This thesis focuses on identifying factors that influence the success of a semiconductor fab start-up. Data on semiconductor fab start-ups was gathered from several sources within Intel Corporation for the purpose of better understanding these factors. The data includes both quantitative data, such as wafer starts, yield, and headcounts, and qualitative data that was gathered during interviews with fab personnel who had been involved with previous fab start-ups. In addition, more detailed analysis was undertaken concerning equipment installation schedules for a fab start-up. Four different equipment installation schedules, each based on a different strategy, were compared using a discrete-event simulator. The results of the simulation are also presented and discussed.

The analysis suggests that there are many important factors that determine the success of a semiconductor fab start-up. Examples of important factors include staffing and training, equipment installation, and process capability. The relevance of the results to start-ups at semiconductor fabs in other companies and to factory start-ups in other industries is discussed.

Thesis Advisors

Dr. Rebecca Henderson, Professor, MIT Sloan School of Management

Dr. Stanley Gershwin, Senior Research Scientist, Mechanical Engineering Department

Dr. Lionel C. Kimerling, Professor, Materials Science and Engineering

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. Many people played significant roles in all phases of the thesis. Without the assistance of numerous people at Intel, MIT professors, students, and others, the data 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.

Special thanks go to Don Myers for initiating and supporting this internship at Intel. His willingness to try new things provided me with the opportunity to explore many different areas that were not strictly related to the Operations Group.

I am thankful to Rich Watkins and the entire Manufacturing Engineering Group at D2, including Paul Sura, Rick Schmidley, and Ralph Kiuttu. Rich and the rest of his staff provided valuable guidance, support, and data for focusing my research efforts. Many thanks go to Jacques Vuye for his help and support throughout the internship.

I would also like to thank my thesis advisors, Professor Rebecca Henderson and Professor Stanley Gershwin, for providing me with very valuable support and guidance and for keeping me focused on the bigger picture. Their critical analysis and pertinent comments greatly aided my thinking. In addition, I am very grateful to Professor Lionel Kimerling for his time and effort in reviewing the draft and providing helpful comments.

Lastly, I thank my parents for their support and encouragement throughout my life. They encouraged me to keep learning and get an education.

Bibliographical Note on Author

Christopher K. Keith was graduated from the Massachusetts Institute of Technology, Cambridge, MA in June 1991 with a Master of Science degree in Materials Science and Engineering. Chris joined Intel Corporation in July 1991 and worked in various process engineering roles at Intel Mask Operations (IMO) in Santa Clara, California. Upon his acceptance to MIT's Leaders for Manufacturing program, Intel Corporation decided to sponsor him, and he will be joining Intel's Technology and Manufacturing Group in July 1996. He is currently working towards masters degrees in management and materials science and engineering and expects to graduate in June 1996.

Table of Contents

1.0	Introduction.....	11
2.0	Problem Statement and Research Methodology.....	13
	2.1 Problem Statement and Project Motivation.....	13
	2.2 Research Methodology.....	14
3.0	Literature and Theory Review.....	15
	3.1 Review of Literature on Semiconductor Fab Start-ups.....	15
	3.2 Review of Production Theory.....	17
	3.3 Literature Review of Capacity Expansion.....	18
	3.4 Overview of Semiconductor Device Fabrication.....	22
	3.5 Characteristics of Semiconductor Manufacturing Systems.....	25
4.0	Factors Affecting the Semiconductor Fab Start-up.....	27
5.0	Data and Methods.....	31
	5.1 Historical Data.....	31
	5.2 Data for the Discrete-Event Simulator.....	33
6.0	Intel History and Characteristics of Intel Fabs.....	36
	6.1 Intel's Strategy and Position in the Semiconductor Industry.....	36
	6.2 Understanding Intel's Capacity Planning and WIP Management Strategies...	39
	6.3 Background and Introduction to Intel's Fab A.....	41
7.0	Analysis of Data.....	43
	7.1 Analysis of Data from Fab A.....	43
	7.2 Analysis of Comparison Data from the Ramps at Fab A and Fab C.....	52
	7.3 Analysis of External Benchmarking Data.....	53
	7.4 Limiters to the Start-up at Fab A.....	54
8.0	Analysis of a Capacity Addition Policy that Includes the Effect of Disruptions...	56
9.0	Factory Simulation Model for the 0.25 Micron Technology Ramp at Fab A.....	62
10.0	Trade-off Between Availability and Die Yield at the Capacity Constraint.....	68
11.0	Results of Data Analysis.....	71
12.0	Recommendations for Improving the Upcoming Ramp at Fab A.....	73
13.0	Extension of Results to Start-ups in Other Industries.....	75
14.0	Reflection on Process and Future Direction.....	77
	Appendices.....	79
	Bibliography.....	83

List of Figures

Figure 3.1	Capacity Expansion Strategy Where Capacity Leads Demand.....	19
Figure 6.1	Comparison of Capacity Ramps at Intel.....	38
Figure 6.2	0.25 micron technology generation Virtual Factory Capacity Ramp at Intel.....	42
Figure 7.1	0.6 Micron Process Technology Generation Gantt Chart.....	45
Figure 7.2	Headcount ramp at Intel's Fab A.....	46
Figure 7.3	Capacity Ramps for Intel Fabs for the 0.6 micron process technology Generation.....	47
Figure 7.4	Pareto of activities during ramp at Intel's Fab A.....	48
Figure 7.5	Activity Milestones vs. Wafer Starts per Week (WSPW) Ramp at Fab A.....	49
Figure 7.6	Line Yield during Start-up at Fab A.....	49
Figure 7.7	Cause-and-Effect Diagram for 0.6 Micron Technology Generation Ramp at Intel's Fab A.....	51
Figure 8.1	Sensitivity Analysis and Cost Comparison of Stepper Expansion Scenarios for Medium-Sized Fab.....	59
Figure 8.2	Sensitivity Analysis and Cost Comparison of Stepper Expansion Scenarios for a Large Fab.....	60
Figure 8.3	Sensitivity Analysis and Cost Comparison for an inexpensive process equipment tool at a Medium-Sized Fab.....	61
Figure 9.1	Factory Output Comparison of Equipment IQ Strategies.....	64
Figure 9.2	NPV Comparison of Equipment IQ Strategies.....	65
Figure 10.1	Trade-off Curve for Availability and Die Yield Improvements.....	69

Chapter 1. Introduction

Continued growth in the semiconductor industry, which has been driven by 30% annual growth rates in the personal computer industry during recent years, has led to an ambitious capital spending program by semiconductor manufacturers all over the world. The high growth rate in the semiconductor industry and the need to advance chip performance through the introduction of ever smaller geometries and more complex processes means that semiconductor companies must invest huge sums of money just to maintain their competitive position relative to their competitors. In order to keep pace with projected demand, over 100 new semiconductor fabrication plants, also known as fabs, are planned through the end of the decade. With each of these fabs costing over \$1 billion to build and equip, the construction of all of these new factories represents an enormous commitment of capital and resources on the part of the semiconductor industry.

Typically, a new factory is expected to be up and running quickly, so that it can meet the firm's immediate needs and provide a solid base on which to expand and become increasingly profitable. The ability of a new factory to get up and running quickly in the semiconductor industry is especially critical since the typical productive life cycle of a semiconductor fab is on the order of 4 - 6 years. Contrast that with an automotive engine plant, which is expected to last for 20 years or more. Or chemical plants, whose life cycle is also on the order of 20 or more years. Given a life cycle of about five years, the percentage of time that a semiconductor fab spends in the start-up phase is clearly a larger percentage of its life cycle than a new plant in any other industry. In fact, it is not unusual for a semiconductor fab to spend 10% to 20% of its life cycle in the start-up phase.

Given the huge initial fixed costs, between \$1 billion and \$2 billion, required to build and equip a state-of-the-art semiconductor fab, and the relatively short life cycle of the fab, it is very important for the fab to begin generating revenue as quickly as possible. As a result, one of the most critical periods of time during a fab's life cycle is the start-up, or ramp phase. The ramp phase is a period of time that can last from six to twenty four months during which all the process equipment is being installed and qualified for production and the number of wafers being released into the factory is increasing rapidly. It is during the ramp that the fab begins to generate revenue through the production of functional semiconductor devices.

My thesis focuses on the fab start-up and how to improve it. My goal is develop a framework based on the literature and theory for understanding some of the factors that affect the success of a semiconductor fab ramp and then compare that framework with historical data gathered from actual semiconductor fab start-ups. Most of my data on actual start-ups is gathered from fabs at Intel Corporation. My research is focused on understanding the factors that affect the start-up of a new semiconductor fab and using that information to make recommendations for improving an upcoming ramp at an Intel fab. The primary customers for my thesis are plant managers and operations managers at semiconductor fabs.

The remainder of the thesis is divided into thirteen chapters. Chapter Two presents the problem statement and research methodology. Chapter Three reviews the theory and literature

relevant to the problem statement. Chapter Four summarizes the factors that might affect the performance of a semiconductor fab start-up. Chapter Five discusses the data sources and data collection process. Chapter Six presents a brief summary of Intel's current strategy and discusses some of the characteristics that may be unique to Intel fabs. The data on start-ups collected from Intel fabs is analyzed in Chapter Seven. Chapter Eight goes into more detail concerning the results generated in Chapter Seven that relate to equipment installation. Several capacity expansion strategies for an upcoming fab ramp at Intel are analyzed and compared in Chapter Nine. Chapter Ten presents an analysis of the trade-off between capacity and die yield at the constraint. The results of the analysis are summarized in Chapter Eleven. Specific recommendations for an upcoming start-up at an Intel fab are presented and discussed in Chapter Twelve. Chapter Thirteen discusses the extension of the results to factory start-ups in other industries. Chapter Fourteen considers how the problem-solving process could have been better executed and provides direction for future work and projects. The Appendices and Endnotes can be found after the last chapter.

Chapter 2. Problem Statement and Research Methodology

This chapter presents the problem statement of my thesis and discusses the methodology that was used to research the problem.

2.1 Problem Statement and Project Motivation

My research is concerned with the problem of managing a start-up at an individual semiconductor fab, as opposed to managing capacity expansion at the firm level. This thesis focuses on factors affecting the success of the semiconductor fab start-up. The goal is to identify and understand some of the levers that influence the performance of a semiconductor fab start-up.

As described in the previous chapter, the semiconductor industry has experienced double-digit growth during the last several years. Most semiconductor manufacturers are finding that their current capacity does not meet the increasing demand. The forecast for continued growth, combined with the need for more advanced process technologies, has resulted in an extensive industry-wide capacity expansion. The planned investment in new semiconductor fabs during the next four to five years is expected to be over \$100 billion.

Each of these new fabs must go through a sometimes tortuous process call the ramp. The ramp is the period of time after which the fab is completed and during which equipment is being installed, processes are being developed and tested, and the first product is being manufactured. It is during the ramp that the fab begins to generate revenue. The ramp is critical because up until the ramp phase, the fab is a huge money sink. By the time that the first product wafers are processed, from \$1 billion to \$2 billion has been spent to build, equip, and staff the fab. As the ramp proceeds, the initial investment can be recovered, assuming that the ramp proceeds smoothly and as planned. If the ramp does not proceed smoothly, the return on the rather large initial investment can be significantly reduced.

Time and money, as measured by the speed with which the fab ramps output and hence revenue, are extremely important metrics for a semiconductor fab start-up. Over time, the initial investment can be recouped and the fab can become profitable when cumulative discounted revenue exceeds cumulative discounted cost. The faster the fab can ramp without problems, the faster the initial investment can be recovered, assuming that demand for the product is relatively elastic. If demand is not price-elastic, then it may be difficult to sell additional product from the faster ramp.

What is required for a fab start-up to be successful? How should the fab allocate its limited resources to maximize performance during the ramp? What can be done before and during the start-up to ensure continued success after the start-up phase? These are questions that face the management team of a new semiconductor fab. The purpose of this thesis is to help answer some of these questions by understanding some of the levers or activities that are most important to the success of a fab start-up.

2.2. Research Methodology

This section describes the methodology that was used to investigate and research the problem statement. The research on semiconductor fab start-ups was divided into several steps. The steps were:

1. Review the literature
2. Identify levers that might affect the performance of a semiconductor fab start-up
3. Perform in-depth analysis on a recent semiconductor fab start-up
4. Analyze ramps at other semiconductor fabs
5. Compare equipment installation strategies for the start-up phase
6. Model the trade-off between capacity and yield at the factory constraint

The first step was to review the current literature and theory to develop a framework for understanding semiconductor fab start-ups. The framework was used to identify a list of factors that might affect the performance of a semiconductor fab ramp. Once the factors were identified, historical data from a start-up at an Intel fab - hereafter referred to as Fab A - was analyzed. Additional data was gathered from a more recent ramp at another Intel fab - hereafter referred to as Fab C. Benchmark data from a fab start-up at another semiconductor manufacturer was also analyzed. The factors that were identified in step two were compared and contrasted to the data that was gathered from several actual semiconductor fab start-ups.

The last two steps were attempts to take what had been learned about semiconductor fab start-ups and apply that learning to an upcoming ramp at Fab A. First, several different equipment installation strategies were compared using data from an upcoming ramp at Intel's Fab A. An equipment installation schedule was generated for each of the strategies and then compared using a simple discrete-event simulation model. Last, a simple model was built to investigate the trade-off between wafer capacity and die yield at the fab's constraint.

Chapter 3. Literature and Theory Review

This chapter reviews literature on semiconductor fab start-ups. The purpose of the literature and theory review is to help develop a framework for understanding and analyzing semiconductor fab start-ups. The relevant literature and theory included information on:

1. semiconductor fab start-ups
2. the production function
3. capacity expansion
4. semiconductor manufacturing systems

3.1 Review of Literature on Semiconductor Fab Start-ups

Literature and database searches revealed a limited amount of information and theory concerning semiconductor factory ramps. There were no literature that presented a complete overview of ramping a semiconductor factory; however, there were several sources of information that discussed certain aspects of the start-up. Benfer [1] provided an excellent source of information for some of the critical issues affecting semiconductor factory ramps. The thesis discussed the importance of die yield and yield learning as levers in improving output during a semiconductor fab ramp. The thesis investigated how characteristics of a semiconductor fab start-up, such as noise and variability, affected the ability of the fab to make improvements in die yield. One recommendation that was made was to install capacity in larger chunks during the ramp to minimize the disruptions due to equipment installation and qualification. My thesis builds on this work by applying some level of analytical rigor to the decision regarding the size and timing of capacity installations. Many of the themes discussed by Benfer were used to aid the analysis of capacity expansion strategies in Chapter Seven of this thesis.

Bohn [2] briefly discussed the importance of understanding and managing technological knowledge during ramp-up of new production in high-tech industries. In the paper, Bohn provided a framework to measure how much the firm knows and does not know about its production processes. The framework was a scale of eight stages of knowledge that is useful for measuring knowledge about a process. Stage One represented complete ignorance; while, Stage Eight was complete knowledge. Bohn applied the scale to high-tech manufacturing, such as VLSI semiconductor design and fabrication processes, which requires rapid learning about multiple variables in new products and processes. Since VLSI semiconductor fabrication is very complex and the process is difficult to control, a lot of effort and resources go into raising the knowledge level as quickly as possible. Bohn concluded that “managing in high-tech industries requires both rapid learning and the ability to manufacture with “immature” (low state of knowledge) technologies.” [3]

Because of the limited amount of literature on semiconductor factory ramps, the scope of the literature search was widened to include start-ups in any industry. This literature search led to little additional information. While there was quite a bit of literature that discussed capacity expansion and factory planning and start-ups from a corporate-wide, strategic perspective, there

was a relatively small amount of information and literature that dealt with the issues faced by plant management during factory start-ups in any industry.

Most of the relevant literature on factory start-ups referenced or built on an article by Schmenner[4]. Schmenner discussed the three phases of the a plant's life cycle - the start-up and early years, the mature years, and the failing years. In addition to discussing plant responsibilities during the start-up and early years, the article described some of the issues that affected the success of the plant's start-up phase. The issues identified included:

1. Plant engineering
2. Work force
3. Overhead functions
4. Control systems

Plant engineering includes decisions on factory area, configuration, equipment choice, flow patterns of the products within the plant, materials handling, utilities, environmental regulation compliance, and similar tasks. Schmenner believes that the engineering of a new facility is critical and that it is probably the start-up aspect that manufacturing companies handle most expertly.

Work force issues include the size of the work force, the mix of skill levels, and recruiting and training plans. The article states that in order to smooth the transition to the new plant and speed the start-up, companies may want to begin labor training before the plant opens. Other issues that involve the work force are wage schemes and decisions that affect the hierarchy of the work force, such as the use of self-directed work teams.

Since most new plants do not have in place a full complement of overhead functions in their first years, plant management must decide which overhead functions it will undertake itself and which will be left to headquarters or another plant's staff. Overhead functions include new product engineering, raw materials purchasing, industrial engineering, and production planning.

Control systems are also critical to start-up success. Too often during start-ups, companies neglect the production and inventory control, accounting, and quality control systems. These systems are important because they yield good information about how well the facility is performing. This information can be used to direct continuous improvement activities during all stages of the factory's life cycle.

Other than the three articles discussed in this section, there was very little literature about factory start-ups. The exact reasons for the lack of literature are not known; however, the lack of literature could lead one to conclude that the area of factory start-ups is a fertile area for future research.

3.2 Review of Production Theory

The absence of literature on semiconductor factory ramps, or factory start-ups in general, meant that other areas of theory had to be relied upon to establish a framework for trying to understand the problem statement. In order to understand some of the factors that might affect a semiconductor fab start-up, the literature on production theory was reviewed. The remainder of this section draws from Chapter Six of Pindyck [5].

Production theory attempts to model the relationship that describes how firms transform inputs (such as labor and capital) into outputs (such as cars and integrated circuits). This relationship is typically represented in the form of a production function. The production function is used to show how the firm's output changes when first one and then all the inputs are varied. A production function indicates the output Q that a firm produces for every specified combination of inputs. A simple model might include only two inputs, labor L and capital K . If these are the only two inputs, the production function can be written as

$$Q=F(K,L)$$

This equation relates the quantity of output to the quantities of the two inputs, capital and labor. Each of these broad categories (capital and labor) might include more narrow subdivisions. Labor inputs include skilled workers and unskilled workers. Capital includes buildings, equipment, and inventories. In addition, we could add other broad categories such as materials to the production function to improve its accuracy.

Production functions can take many forms. One of the most common is the Cobb-Douglas production function. The Cobb-Douglas production function is

$$F(K,L) = AK^\alpha L^\beta$$

Another example of a production function is

$$F(K,L) = AK + BL$$

These production functions were generated by fitting equations to historical data.

The production function of a semiconductor factory process technology is very complex. The output of a semiconductor fab is good die. A die is an individual integrated circuit chip. Inputs include silicon wafers, which are the raw material for the dies, the number of factory operators and engineers, the amount of installed equipment, and the square footage of the semiconductor fab.

The purpose of this thesis is not to develop explicit models for production functions in semiconductor fabs; however, the analysis does draw upon some of the basic concepts of the production function. The most important concept is that inputs and outputs are related by some physical relationship which can be described with a mathematical model. In the case of a

semiconductor fab, the most important physical inputs include silicon wafers, work force size, and capital equipment. One complicating factor is that silicon wafers can be treated as either an input or an output. Silicon wafers go into the process at one end of the factory and silicon wafers come out at the other end of the factory. One silicon wafer can hold anywhere from fifty to several hundred die.

The main weakness of the production function is that it is a static model. The production function captures the state of the process technology at a given point of time. However, semiconductor fab ramps are inherently dynamic and much more complex than the production function model allows. As a result, there are some important factors that affect the factory ramp that the production function can not model. Examples of these factors include experience curve effects, organizational learning, and complex yield relationships that affect the number of good die per wafer or the number of wafers that even make it through the process. Another complicating factor is that the output and inputs are changing rapidly over time, making it very difficult to accurately relate the two.

Hayes, et. al. [6] have proposed Total Factor Productivity (TFP) as a useful way to measure factory performance. TFP draws upon some of the principles of production theory. Productivity is a measure of the efficiency with which inputs are translated into outputs. Whereas single factor productivity (sfp) relates the total output to a single input, such as a particular type of raw material, TFP integrates and summarizes the contributions of all the factors of production (according to their contribution to total cost). TFP's main advantage is that it gives managers an integrated perspective on performance by incorporating the trade-offs between various inputs. The basic idea of relating a semiconductor fab's output, as measured by good die, to various inputs, such as labor, using some kind of productivity measure was particularly helpful in analyzing the data gathered from the Intel fabs.

3.3 Literature Review of Capacity Expansion

Capacity expansion is the addition of facilities or equipment to serve some need. Capacity expansion problems arise in a myriad of applications, including communications networks, gas and oil pipelines, public facilities, and manufacturing facilities. The primary capacity expansion decisions typically involve the sizes of facilities to be added and the times at which they should be added. Other major considerations include the type of capacity or the location of the capacity to be added.

There is a rather extensive literature available on capacity expansion and capital investment decisions. This is a basic problem in all industries, especially growing industries, hence much attention has been devoted to it. This section of the thesis reviews some of the more important aspects of the literature and theory concerning capacity expansion and focuses on those parts that are especially relevant to the research. Since the research focuses on individual fabs, capacity expansion literature that discusses the addition of equipment at a single site is the most relevant.

The process of making a capacity expansion decision in the traditional capital budgeting sense is quite straightforward and can be found in any finance textbook. Future cash inflows over the

lifetime of the project are forecasted, discounted and weighed against the discounted cash outflows required for the investment. The net present value is then compared to other investment projects available to the firm, or alternatively, a return on investment (ROI) figure is calculated using the same discounted cash inflows and outflows and compared to the firm's desired rate of return - or hurdle rate. If the ROI is larger than the internal hurdle rate and there are sufficient funds, the project is approved. If the ROI is smaller than the hurdle rate or there are not sufficient funds, the project is not approved.

What other criterion are used to guide capacity expansion decisions. In addition to calculating and comparing discounted cash flows, there are other financial and non-financial objectives. They include:

1. maximizing market share
2. maximizing capacity utilization
3. maximizing profit
4. minimizing costs such that projected demand is met

Figure 3.1 illustrates the strategy of maintaining a capacity cushion during a factory start-up, or capacity ramp. The capacity cushion represents excess capacity that the factory can use to respond to sudden demand surges or to ensure that projected demand for the factory's products is always met. In both scenarios in Figure 3.1 capacity leads demand. Scenario 1 illustrates the case where small increments of capacity are added relatively frequently. Scenario 2 illustrates a strategy where larger increments of capacity are added less frequently than in scenario 1.

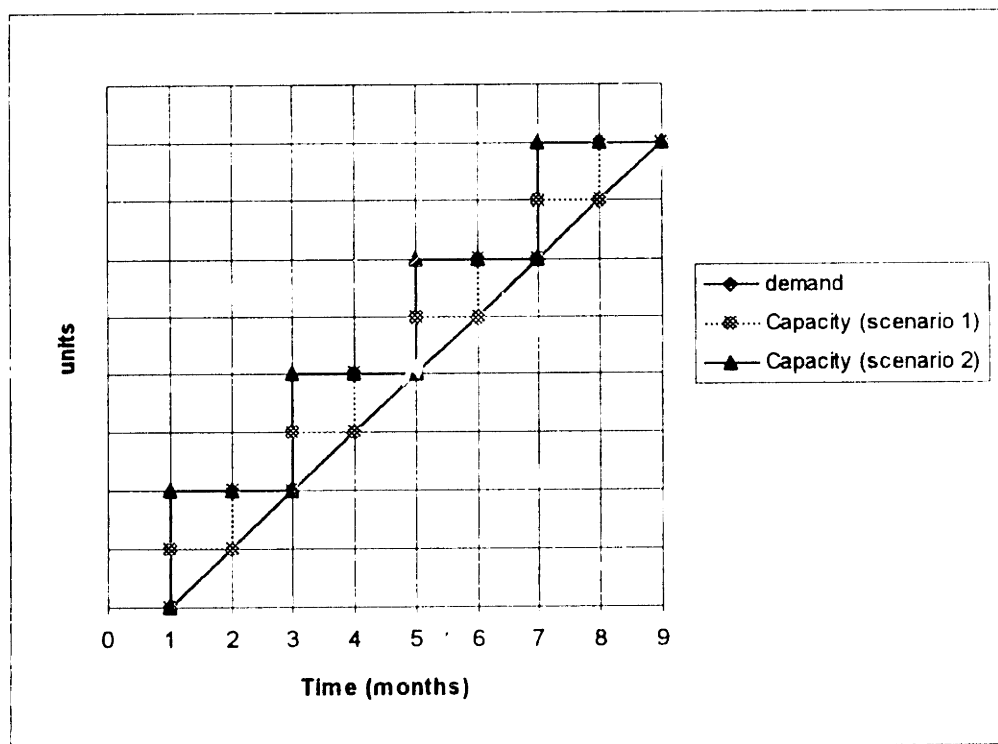


Figure 3.1 Capacity Expansion Strategy Where Capacity Leads Demand

The capacity expansion problem has been modeled analytically for specific situations. Freidenfelds [7] provides a comprehensive discussion and derivation of many of the analytical models. Readers should note that the remainder of this section draws heavily from chapter 3 of Freidenfelds [7]. In general, models for the capacity expansion problem are restricted to situations in which the following apply:

1. The cost of the equipment or facilities added exhibits economies-of-scale (i.e., their cost is less than proportional to size). Stated another way, any addition of capacity is accompanied by some fixed costs.
2. Time is an important factor. That is, there is a continuing (possibly changing) need for facilities, and the facilities or equipment added are durable (i.e., they provide service over more than a short time interval).

One of the simplest analytical models for analyzing the capacity expansion problem is one in which a single capacity type, or piece of equipment, is to be added to serve a known deterministic demand. Since we are adding only a single capacity type, the production sequence only has one stage. The model assumes that the demand for additional units of capacity will grow linearly at rate g over an unbounded horizon, so that starting from time $t=0$, gt additional units will be required at time period t in the future. Typically, additional units of capacity are purchased in bulk, either because they only come that way or because it makes economic sense to purchase them that way. The model also assumes that the cost of additional units consists of a fixed cost A plus a linear cost B per unit of capacity, so that x additional units cost $A + Bx$, where x is a variable for the amount of capacity. $A + Bx$ is the present discounted cost of providing x units of capacity forever. Graphically, this model can be represented on an x - y graph (see Figure 3.1), where the x axis is time and the y axis is demand or capacity in units. The demand line is a straight line with slope g and the capacity line is a staircase function that is always above or equal to the demand line.

At this time the reader should note that the straight line representing the demand curve in Figure 3.1 looks very similar to the growth in wafer starts - wafers started into the production process - experienced by a ramping semiconductor fab. This similarity will be used in later sections to more fully explore the semiconductor fab start-up.

The simple model can be solved to find the values that result in demand being satisfied at all times with minimal discounted cost. Assuming that we can continue to place additional facilities in the indefinite future at the same cost, it should be clear that we always shall wait until existing facilities are full and then place some facility of size x , called the relief size. We wait until the facilities are full because it is better to spend later by the present value criterion. However, in some cases the company may not wish to wait until utilization reaches 100% before additional capacity is added. To include this more complicated dimension into the problem requires an additional concept - congestion cost. While there are more advanced capacity expansion models that do include congestion cost, they will not be used here due to their mathematical complexity. In the simple model we always use a facility of the same size in this formulation because the costs

and the projections of additional demand are identical at every shortage time. We wish to find the x that minimizes

$$C = \sum_{n=0}^{\infty} (A + Bx)e^{-r(nx/g)} \quad (3.1)$$

where the nx/g are the times at which additional equipment will be placed. Equation 3.1 represents the total discounted cost of capacity addition from time $t=0$ to infinity. With a positive discounting rate r , the sum converges to

$$C = \frac{A + Bx}{1 - e^{-r(x/g)}} \quad (3.2)$$

It is simple to show that C is a well-behaved convex function for positive x and so takes on a unique minimum. We can also rewrite equation 3.2 to show how C varies with the relief time interval, t . Substituting $t=x/g$, results in:

$$C(t) = \frac{A + Bgt}{1 - e^{-rt}} \quad (3.3)$$

Unfortunately, it is not possible to write an explicit formula for the x or t that minimizes C . The best that we can do is to get an implicit formula. If we set the derivative of C with respect to t to zero and rearrange terms, we obtain

$$e^{rt} - rt - 1 = \frac{Ar}{Bg} \quad (3.4)$$

The t that satisfies this equation is the optimal relief time. While we cannot explicitly solve for t , we can obtain a useful approximation. If rt is small, e^{rt} can be closely approximated by a second-order Taylor series expansion:

$$e^{rt} \approx 1 + rt + \frac{1}{2}(rt)^2 \quad (3.5)$$

Combining these last two equations, we obtain

$$\frac{1}{2}(rt)^2 \approx \frac{Ar}{Bg} \quad (3.6)$$

or

$$t \approx \sqrt{\frac{2A}{Bgr}} \quad (3.7)$$

$$x = gt \approx \sqrt{\frac{2Ag}{Br}} \quad (3.8)$$

In general this approximation is considered fairly good if rt is small. In later chapters of the thesis, these equations are used to generate optimal equipment installation schedules - for those cases, rt is no greater than .075. The exact value for e^{rt} if rt is .075 is 1.0779; the approximate value from the Taylor expansion is 1.0778. The approximated value is within .0066% of the actual value. If rt is large or greater accuracy is desired, equation 3.4 can be solved explicitly using numerical techniques. These equations are useful as rules-of-thumb guide to scheduling capacity additions for this particular case of linear demand growth, which assumes that manufacturing is carried out in a single stage process.

It is interesting to note the similarity of this capacity expansion model with inventory problems. Readers with some knowledge of inventory theory will recognize that the square root formula above is identical in form with the classical formula for economic order quantity. The similarity is not mere coincidence. The economic order quantity (EOQ) model is the simplest and most fundamental of all inventory models. It describes the important trade-off between fixed order costs and holding costs. Instead of fixed order costs and holding costs, the simple capacity expansion model discussed above optimizes the trade-off between fixed costs and the proportional cost of additional capacity. Both the simple capacity expansion model and the EOQ model use cost minimization as the optimization criterion.

Careful consideration of capacity expansion decisions show that they are very complex in nature. The decision involves not only financial considerations, such as discounted cash flows and rates of return, but also strategic and competitive analysis and the potential impact of the investment on the firm's people, infrastructure, and current capabilities.

3.4 Overview of Semiconductor Device Fabrication

The transistor, which was invented in late 1947, and the integrated circuit, which was demonstrated around 1960, are the basis for today's microelectronic industry. 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. While the first electronics devices were fabricated using germanium, silicon has become the industry standard.

The creation of the integrated circuit and the use of planar technology allowed for an ever increasing density of devices on silicon substrates. The idea of planar technology was to fabricate patterned layers, one on top of the other, made of materials with different electrical properties. The multidecked sandwich of patterned layers are made to form various circuit

elements such as transistors, capacitors, and resistors, and these are connected together by a patterned conducting layer to form an integrated circuit.

The layers are formed by modifying the substrate or depositing a layer of material on the substrate. Examples of substrate modification processes include doping and oxidation. Material can be deposited by evaporation or sputtering. Patterning is usually accomplished by the process of photolithography. The typical processes used to make integrated circuits are [8]:

1. oxidation
2. lithography
3. etching
4. diffusion and ion implantation
5. thin film deposition

Each of these processes are discussed in the remainder of this section.

The fabrication of integrated circuits takes place on silicon substrates possessing very high crystalline perfection. The process for making silicon wafers begins with quartzite. Quartzite, a type of sand, is refined by a complex, multi-stage process which produces electronic grade polysilicon. This polysilicon is used to grow single crystal silicon by Czochralski (CZ) crystal growth. In CZ growth, single crystal ingots are pulled from molten silicon contained in a crucible. The molten silicon contains controlled amounts of impurities which are incorporated into the cylindrical single-crystal ingot which results from the pulling process. The ingot is typically 100-200 mm in diameter and over 1 meter in length. After a single crystal silicon ingot has been grown, a complex sequence of shaping, sawing, and polishing steps must be performed on it to produce silicon wafers suitable for fabricating semiconductor devices. Device fabrication takes place over the entire wafer surface and hundreds of identical chips are created on each wafer at the same time.

The formation of the oxide of silicon (silicon dioxide) on the silicon surface is termed oxidation. A silicon dioxide layer is usually formed on the silicon substrate by the reaction of oxygen with the substrate material at elevated temperatures (900-1200 C). Some of the functions of the oxide film include dopant masking, device isolation, surface passivation, and use as a gate oxide. Oxide thicknesses can vary from tens to thousands of angstroms. The thermal oxidation process is fairly well understood and thicknesses can be controlled fairly well.

Lithography is the process by which the patterns that define the devices on the chip are transferred to the substrate surface. In the lithography process, the pattern in the form of a photomask is projected onto a surface that has been previously coated with a photoresist layer. Positive photoresist materials have two properties. First, when exposed to ultraviolet light, their solubility in one class of solvents is changed, so that after immersion in such a solvent the projected pattern is replicated in the surface. Second, the undissolved regions of the resist are not affected by a second class of etching agents, which are able to etch or modify the underlying material. After the wafer has been coated with photoresist, the photomask and wafer are aligned in a stepper. After alignment, they are subject to UV radiation. The exposed areas of photoresist

are then developed and the wafer is ready for the next process step. A wafer goes through the lithography process from ten to twenty times. The lithography process is critical because it defines the size of the smallest feature on the silicon substrate. Improvement in integrated circuit performance is driven by ever smaller feature sizes. Today, minimum commercially feasible feature sizes range from 0.35 to 0.5 microns.

Etching in semiconductor fabrication is a process by which material is removed from the silicon substrate or from the films on the substrate surface. The film can either be removed via wet etching or dry etching. Etch processes are measured by their selectivity and isotropy. The selectivity of an etched process is the ratio of etch rates of different materials. Any etch process must effectively etch the silicon or film layer with minimal removal of the underlying silicon or resist material. Therefore, a high selectivity is desired. When etching proceeds in all directions at the same rate, it is said to be isotropic. If etching proceeds exclusively in one direction, the etching process is said to be completely anisotropic. Vertical etch profiles are desired, hence anisotropic etching is favored.

The diffusion of controlled impurities or dopants into silicon is the basis of device formation and fabrication in integrated circuit processing. The dopants affect the electrical characteristics of the silicon material. Dopants are selectively introduced into the substrate by the diffusion or ion implantation processes.

In the diffusion process, impurities are introduced into the substrate by chemical sources. The chemical sources include gas, liquid, and solid forms. The dopants are then diffused to the desired depths by subjecting the wafers to elevated temperatures (900-1200 C). The depth is determined by time, temperature, and the diffusion coefficient of the dopant species in the substrate material.

Ion implantation has become the major means for the introduction of impurities into the substrate. Ion implantation is capable of placing the dopant species very precisely, but a diffusion step is still required to drive the impurities to specified depths, to electrically activate the implanted impurities, and to remove defects from the implantation area by annealing. Implantation is achieved by accelerating charged dopants through a high voltage field (10-1000 keV) and then choosing the desired dopant by means of a mass separator. Ion implantation is favored over chemical doping and diffusion methods because of its ability to implant exact quantities of dopants at specified depths below the surface in the designated areas of the substrate.

A large variety of thin films are used in the fabrication of integrated circuits. Examples of thin films used in semiconductor fabrication include polysilicon, silicon nitride, aluminum, and silicides. Thin films can be formed in many different ways. The film growth techniques can be divided into two groups: 1) film growth by interaction of vapor-deposited species with the substrate, and 2) film formation by deposition without causing changes to the substrate material. The first group includes thermal oxidation and nitridation of single crystal silicon and polysilicon and the formation of silicides by direct reaction of a deposited metal and the substrate. The second group includes chemical vapor deposition, in which solid films are formed on a substrate by the chemical reaction of vapor phase chemicals that contain the required constituents, and physical vapor deposition, in which the species of the thin film are physically dislodged from a source, to form a

vapor which is transported across a reduced pressure region to the substrate, where it condenses to form the thin film. Physical vapor deposition can be accomplished by sputtering and evaporation.

Once the individual devices have been fabricated on the wafer surface, interconnections are required to form a complete and functional integrated circuit. Interconnections are made of metals, such as aluminum, that exhibit low electrical resistance and good adhesion to dielectric insulator surfaces. The metal is deposited by physical vapor deposition techniques such as sputtering or evaporation. Today's integrated circuits have 2-5 layers of metallization with an insulating dielectric layer separating each metal layer. The interconnect layers are connected through the insulating layers by vias. Contacts establish the connection between the metal layer and the active devices in the silicon.

After metallization, a passivation layer is deposited on the wafer to protect the integrated circuit from the environment. Each of the chips is then tested using an electrical tester. After testing, the individual integrated circuits are separated from the wafer and sorted. The good chips are then packaged and tested one last time.

3.5 Characteristics of Semiconductor Manufacturing Systems

Semiconductor manufacturing systems represent a unique class of manufacturing systems. They are among the most complex and capital-intensive of all manufacturing processes. Semiconductor manufacturing systems are typically classified as job shops. Both semiconductor manufacturing systems and job shops are characterized by physical groupings of similar types of equipment or machinery, as opposed to transfer lines where there is a well-defined progression of product through the manufacturing system. Semiconductor manufacturing systems are characterized by:

1. high variability
2. reentrant material flows
3. batching of product

There are several sources of variability. One source of variability is low yields. During the start-up of a new semiconductor fab that is using leading-edge process technology, less than 50% of the chips that begin the process are functional when they reach the end of the process. Rework of product at various steps in the semiconductor manufacturing process also makes it difficult to predict system performance. In addition, typical machine uptimes range from 70 to 90%. Low availabilities typically require that extensive in-process inventory be maintained to buffer against machine downtimes, which leads to high throughput times and slow information turns. Since most of the testing takes place at the end of the semiconductor manufacturing process, some problems are not caught until large amounts of product have gone through the process and are affected by the same problem. Most of the difficulties arise from the fact that most semiconductor processes and equipment, especially those used to produce leading-edge chips, are not well understood.

A reentrant system is characterized by the fact that product goes through the same equipment multiple times during processing. For example, in a typical semiconductor manufacturing process the same wafer go through a stepper 15 - 20 times. Steppers are pieces of equipment that transfer patterns onto the surface of the wafer. This means that rules must be put in place to determine which wafers get processed first. Should a wafer at the tenth litho step be given priority over a wafer at the fifteenth step? While much attention and research have been devoted to this question, no universally applicable rules or singular solution have been identified to aid in making the decision.

Many of the semiconductor manufacturing processes require batching of wafers. Wafers typically travel through the manufacturing system in groups of twenty-five wafers, referred to as lots. At some steps, several lots are processed simultaneously. For example, diffusion furnaces require that four to five lots of wafers be processed at the same time to maximize utilization. Batching can be problematic because of the restrictions it places on when lots can be processed. Before a lot is processed at an operation that requires batching, there must be three or four other similar lots at the same step that require similar processing - otherwise, lots wait until there are four or five of the same lots at the process step. This can be especially troublesome when a factory must process many different products. Each of these products requires different process recipes. In a factory that processes several different products, there is less likelihood that at any given batching operation there are enough similar lots to warrant beginning processing. This means that lots must wait in queues even longer and leads to higher throughput times.

A semiconductor factory, or fab, is typically measured along several different performance dimensions. The performance dimensions are delineated to specific performance metrics. The metrics are both financial and non-financial in nature. The first set of non-financial performance metrics are related to measuring capacity and output. The metrics include wafer starts, good die out, wafers out, and operational equipment effectiveness (OEE). Financial metrics include cost per good die and return on assets.

Chapter 4. Factors Affecting the Semiconductor Fab Start-up

Based on the literature and theory review, several factors were identified that could affect the semiconductor factory ramp. Benfer [1] identified die yield as an important factor. Another type of yield which is important is line yield - which is simply the percentage of silicon wafers that make it through the entire process to the test area. Bohn [2] discussed the importance of rapid learning and the ability to manufacture with "immature" technologies, which is also discussed by Benfer [1]. The limited literature on factory start-ups [4] discussed the importance of control systems, work force size and skill levels, plant engineering, and decisions concerning which overhead functions would be contained with the facility and which would be borrowed from outside sources. The literature on production functions [5] identified several key input variables: work force size, capital equipment, building size, and raw materials, such as wafers. And the literature on capacity expansion [7] focused on capital equipment as the main input; however, the general theory of capacity expansion could just as easily be applied to the work force if the worker is treated as the unit of capacity. In addition to these factors, two other requirements for a successful semiconductor fab start-up were identified. A start-up requires a capable production process and a product to manufacture. Formally, nine levers were identified that might affect the performance of a semiconductor fab start-up. These nine levers were:

1. manufacturable production process
2. product
3. die yield
4. line yield
5. plant engineering
6. work force size
7. work force skill level
8. control systems
9. installed equipment base

A semiconductor fab requires a manufacturable process before it can ramp successfully. A process is defined by its process capability (C_{pk}). A manufacturable process is characterized by a high process capability. An immature process is characterized by a low process capability. Process capability is important because it can affect the die yields of the fab. High process capabilities imply processes with low variance relative to the process specifications. When the process is within specification limits and the process specification limits are set at appropriate levels, die yield tends to be high. Processes that are always out-of-control require excessive engineering and managerial attention, lead to increased manufacturing variance, and usually result in low yields. The process development group is responsible for developing a process that pushes the leading edge but that can lead to reasonable die yields. So the process development effort is a complex balancing act. On one hand, the group must operate the process near the leading edge to achieve acceptable chip performance results relative to competitors. On the other hand, if the process variance is too high, manufacturing yields and output suffers and it is impossible to reliably manufacture the chips. Some of the cost drivers that affect the process capability include process development engineering headcount, cost of leading-edge processing equipment, and the cost of running experiments to learn more about the process. An engineer costs approximately

\$100,000 per year. Equipment improvements or buying the latest and best capital equipment can cost tens of millions of dollars extra. The marginal cost of running experiments is usually relatively low, on the order of tens of thousands of dollars. However, in order to run the experiments, a complete fab processing line is typically required. The fixed costs of the equipment for just a semiconductor pilot line can be over \$100 million. In addition, time is important because the process development group is typically under pressure to deliver a manufacturable process at a specific date.

Before a semiconductor fab can ramp, it must have a product to manufacture. While it may seem intuitive, the complexity of VLSI integrated circuits means that significant resources must be devoted to the design effort. In addition, significant coordination between the product design engineers and the fab process engineers and development engineers is required to ensure that the chip is manufacturable. The complexity also leads to uncertainty regarding design schedules. If the design is not ready, the fab has nothing to manufacture. While it may cost millions of dollars to speed up the development of a VLSI chip that is behind schedule, it is even more costly to idle a billion dollar semiconductor fab because the product is not ready.

Die yield is the ratio of good die to total die on a silicon wafer. Since the goal of the fab is to maximize output, one way to do this is maximize the die yield. A fab that is running with a 50% die yield can effectively double its output by increasing its die yield to 100%. Or a fab that is running at 100% die yield could produce the same number of good VLSI chips as fab twice its size and cost that is only running at 50% die yield. In practice, die yields of 100% are very difficult and costly to reach. In fact, if the die yields of a new process are running close to 100%, it is likely that the process is not close enough to the leading edge. Die typically fail due to misprocessing, lack of process control, or contamination. Misprocessing results when the wrong process recipes are used and the wafers can not be reworked. Low process capabilities mean that many important physical and electrical characteristics of the devices are not within tolerance and lead to die failure. The threat of contamination leads to clean rooms that are thousands of times cleaner than hospital operating rooms. State-of-the-art clean rooms can cost well over \$1000 per square foot, not even including the capital equipment costs. Process control is tied to the process capability issue discussed earlier. Money spent during the process development phase can have high leverages later on in the process if the additional dollars lead to high process capabilities and higher die yields. When leading edge process technologies are introduced into a production fab for the first time, die yields can be as low as ten to twenty percent. Additional process experimentation and learning are required to improve the yields during the course of the start-up and steady-state production phases. However, once the process has been introduced into a production fab, experiments are more costly because the cost of running an experimental lot of wafers means that a revenue-generating production lot of wafers can not be run. So not only are there the costs of the raw materials but also the opportunity cost of not using the equipment to run production wafers and generate revenue dollars. This issue of fab capacity and yield is explored further in Chapter 10.

Line yield is the ratio of wafers that make it to the testers at the end of the semiconductor process to the number of wafers that begin the process. Wafers can be lost due to misprocessing or breakage. Misprocessing can be the result of human error or machine error. Wafers can not be

reworked or recovered if a human uses the wrong process recipe. In addition, machine breakdowns that occur during wafer processing can also result in losses. Line yield losses have effects on output similar to die yield losses, as increases in line yield lead directly to increases in output. Line yield losses can be reduced through automation, having the computer download the correct recipe rather than having a human operator input the recipe, or by reducing machine breakdowns. Bohn has argued that die yield, line yield, and process capability are determined by a semiconductor fab's ability to learn. Most of the learning occurs by running planned experiments and carefully analyzing data from production wafers.

Another factor that might affect the success of a semiconductor fab start-up is plant engineering. Plant engineering refers to all the planning that goes on before the fab is even built. Issues such as fab size and layout, utility requirements, environmental regulations, and work force requirements must be planned for before construction of the fab begins. In the early stages of the process this can often be a limiter due to headcount considerations. All this work and planning must be done by someone, yet the factory and most of the factory organization and infrastructure do not even exist at this point. Often the burden falls on planners or industrial engineers at another fab to begin planning while the management team for the new fab begins to assemble and grow its staff. Slowly, the planning workload can be offloaded to the staff of the new fab. Problems can occur here because either the staff are newly hired employees and have no experience in such planning, or the staff have been pulled from other organizations which must now scramble to fill the vacant positions within their own organizations. Since the cost of an employee to the firm is typically only on the order of \$100,000 per year including salary and benefits, and the time constraints for getting the fab up and running quickly are enormous, it would seem to make sense to spend the additional money to hire and train planners ahead of time to prevent this step from becoming a part of the critical path.

The start-up could also be limited by the work force size. If the work force is not big enough, factory throughput can be constrained and inventories can build up. While today's semiconductor fabs are becoming increasingly automated, people are still required to run the process. For example, much of the data collection for statistical process control (SPC) requires that a fab technician manually take measurements on the wafers before and after processing and enter the data into the computer. Equipment can also be idled if at the end of a processing sequence, there are no new wafers ready to be processed. An undersized work force can thus lead to lower equipment utilization. Each member of the work force, whether it be an engineer or technician, costs the fab approximately \$100,000 per year for salary and benefits. Fabs can be staffed by 500 to 1,000 people. The cost of hiring one hundred additional people for one year is approximately \$10 million - about the cost of two state-of-the-art lithography steppers. Given the relative costs of hiring and training additional technicians versus the opportunity cost of idled semiconductor capital equipment, it is advisable to err slightly on the high side when estimating labor force size requirements for the start-up phase.

Semiconductor manufacturing is a very complex process. If the work force is not properly or sufficiently trained, the start-up can suffer due to low yields and poor productivity. Training requires extensive planning and resources because not only must the new employee spend time training, but also an experience employee must devote his or her time to training instead of

moving product through the factory. While it is true that some training involves actually moving wafers through the production sequence, it is typically requires much more time than if the experienced person were simply doing it by themselves. However, during a start-up it is typically the first activity to be given lower priority. As the wafer starts increase and the pressure for output increases, myopic managers can push out or delay training in order to get product out the door. Ideally, training should occur before, during, and after the start-up phase.

Control systems coordinate and manage the flow of information throughout the semiconductor fab. Control systems are used to monitor important performance measures such as output, yield, equipment downtime, and the flow of material through the semiconductor fab. This data is used to monitor and direct improvement activities, allocate resources, and look for trends or problems that need to be addressed. The control system is typically some type of computer network running proprietary or off-the-shelf manufacturing software. Simple software would be limited to tracking the wafers as they move through the manufacturing process. More complicated software would not only track work-in-process (WIP) but also track equipment downtime, maintain preventive maintenance schedules, store and manipulate statistical process control (SPC) data, and maintain yield data. Even more complicated control systems would use bar-code scanners to read the bar-codes on lots of wafers and select the appropriate process recipe. Control systems also include automation systems to automate the movement of wafers between bays in the fab or within the bay, eliminating the need for humans to touch or handle the wafers or wafer storage boxes. Elaborate control systems can easily run into the tens of millions of dollars; however, they can be very helpful if used to automate mundane tasks, store data, deliver the data to the right person at the right time, and in general aid in learning about the process and making yield and process control improvements.

Installed and qualified process equipment is required to run wafers through the semiconductor fab. Semiconductor fabs require from thirty to fifty different types of equipment. To successfully process a wafer, one of each type of equipment is required. Semiconductor processing equipment typically accounts for about 75% of the total fixed cost of a semiconductor fab. For a fab that cost \$1 billion, approximately \$750 million was spent on capital equipment, with some individual pieces of equipment costing over \$5 million. Most of the managers at Intel expressed a desire to put the equipment installation process on the critical path due to the huge expenditures required. Coordination costs are also high because the equipment is typically purchased from as many as twenty to thirty external vendors who build and help to install the equipment. Equipment purchase orders must be generated for all this equipment and orders and specification documents must be given to the equipment vendors far enough ahead of time to allow the equipment suppliers time to build and test the equipment.

There are a number of factors that can affect the performance of a semiconductor fab start-up. Issues ranging from control systems to plant layout must be carefully considered and laid out. And it is not always clear which of these issues are the most critical or should have the highest priority. The importance of these levers will be explored in more detail in later sections of the thesis.

Chapter 5. Data and Methods

This chapter discusses the research data that was collected at Intel Corporation. The purpose of the chapter is to discuss the methods that were used to collect the data and the sources of the data so that my research can be replicated by others.

5.1 Historical Data

To help understand all the activities that must precede a semiconductor fab start-up, a Gantt chart was constructed. The Gantt chart consisted of the high-level activities involved in bringing a new process and product to market at Intel. Through discussions with manufacturing engineering managers, operations managers, product design managers, and process development managers and engineers at Fab A, a Gantt chart was developed that illustrated the high-level activities and associated time durations with bringing a new process technology and new products to market at Intel Corporation. The process technology that was selected was the 0.6 micron process technology. This process technology generation was chosen because the process was sufficiently mature to have a complete history stretching from process development to steady-state production.

Headcount data for Fab A came from the Finance Department at Fab A. The headcount numbers included data from seven different departments at Fab A. The departments were Manufacturing Engineering, Operations, Yield Engineering, and the four engineering functional areas: Etch, Diffusion, Litho, and Thin Films. Data from several different departments was collected in order to distinguish real headcount growth at Fab A from simple transfers of headcount between departments. During the start-up, there were several transfers of personnel between departments at Fab A. To ensure that these internal transfers did not affect the ability to distinguish between new personnel and people who were simply moving around within the organization, all the relevant departments were combined to provide an aggregate number.

To model the training activity a constant six month training time was added to the headcount data. The assumption was that any new personnel would require at least six months of training time. So while the hiring ramp ended after X months, the training ramp did not end until time $X+6$ months. While there was some disagreement about the exact amount of time required for training, most people at Fab A agreed that six months was the absolute minimum required.

The product design activities and durations were estimated by the manager of the product design group and one of the product design engineers. The activities related to process development at Fab A came from the process development manager at Fab A. Information on the other process development activities and equipment supplier programs came from an engineer in the Process Equipment Development Group at Intel.

Data on the equipment installation and qualification activity at Fab A and the wafer starts ramps at other fabs came from the Manufacturing Engineering Group at Fab A. Data on the wafer starts ramp at Fab A came from the Fab A Production Control Group. Data on the Fab A Expansion activity came from the Facilities Construction Technology (FCT) Group at Intel.

In addition to the activities on the Gantt chart, line yield data from the start-up at Fab A was collected. The line yield data for Fab A came from the Manufacturing Engineering Group at Fab A. It was difficult to obtain exact data on die yield for Fab A; however, during interviews with the yield manager and operations manager, I was able to get a sense of the historical trends during the start-up.

Due to difficulty in getting data on the actual number of die shipped per month from the various Intel fabs, wafer starts was used as the measure of output. Therefore, any activities that refer to the output ramp in the remainder of the thesis actually represent wafer starts data.

In addition to gathering quantitative data such as activity durations and headcounts, a substantial amount of qualitative data was collected. The qualitative data was gathered during extensive interviews at Intel. The interview process was based loosely on the formal techniques developed for collecting the voice of the customer as part of Quality Function Deployment (QFD). These techniques are described in Shiba [9], who encourages their use for collecting qualitative data. The process that was used shares several characteristics with these formal methods of collecting qualitative data. The shared features include:

1. 360-degree view, where all viewpoints and perspectives are included
2. QFD's four open-ended questions

The interviews were conducted with approximately 20 Intel employees who were involved with the 0.6 micron technology generation ramp. These employees represented a wide variety of positions in the fab. Shift managers and supervisors, pilot line managers, shift managers, manufacturing and process engineers, plant managers, and engineering managers were interviewed.

To gather data about the 0.6 micron technology generation ramp, several open-ended questions were asked.

1. What metrics did you watch most closely during the ramp?
2. What were the problems with or weaknesses of the ramp?
3. What were the major delays or critical limiters that prevented the factory from increasing output faster?
4. What could be improved in the next ramp?

Responses to these questions are summarized in Appendices A, B, C, and D at the end of the thesis. Only those responses that were mentioned more than once were included.

In addition to collecting data related to Fab A, data from start-ups at other Intel fabs was analyzed. The fabs included two fabs (Fabs B and C) that ramped the 0.6 micron process technology after Fab A. Fab B, which was slightly larger than Fab A, ramped the process less than six months after Fab A ramped. Fab C, which was approximately the same size as Fab A, ramped the process about one year after Fab A. Data about the start-ups at these fabs was

gathered from the Manufacturing Engineering Group at Fab A and during interviews with managers at the fabs.

Data regarding semiconductor fab start-ups at other companies was collected through external benchmarking. Intel's Competitive Analysis Group provided benchmarking data comparing a recent ramp at an Intel fab with a ramp at an Asian DRAM manufacturer. The fabs were of similar size and capacity and ramped the same process technology generation.

Data for the sensitivity analysis results in Chapter 8 were obtained from the Manufacturing Engineering Group in Fab A. Data that was obtained from the Manufacturing Engineering Group included equipment cost data and wafer starts ramp rates.

All of the data from Intel was disguised for confidentiality reasons. Intel requested that as much of the data as possible be excluded from the document. Several of the figures and graphs in Chapter 7 do not have data on the y-axis. This data was intentionally omitted at Intel's request.

5.2 Data for the Discrete-Event Simulator

Part of the research involved building a discrete-event simulation model of a planned 0.25 micron semiconductor fab to analyze and compare equipment installation strategies and schedules. A commercially-available discrete event simulation package called AutoSched/AutoMod was chosen to build the model. AutoSched/AutoMod was developed by a company called AutoSimulations, Inc. The package is fairly user-friendly and uses a spreadsheet format to input data.

AutoSched was chosen for several reasons. The package was recently purchased by the Manufacturing Engineering group at Fab A and they planned to use it indefinitely to aid in future capacity planning. Hence any models that were built could be maintained and used by the group after my internship was done. In addition, the package is very flexible and can model a variety of variables. For example, since the plan was to model the wafer starts ramp, there was interest in understanding the effects on output and other performance metrics of various pieces of equipment being installed and qualified for production at various times during the ramp. AutoSched allows equipment to be used or not used for production for various periods of time during the simulation. In addition, AutoSched can handle the typical functions such as downtime, batching of lots, set-ups, yield loss, rework, and others.

The input data was entered in spreadsheets. The first spreadsheet contained information on each piece of equipment. The spreadsheet is called the stn.txt file and simply lists all the pieces of equipment in the factory. Data on the equipment list and the number of each piece of equipment for the 0.25 micron fab came from the Manufacturing Engineering Group at Fab A.

The part.txt file lists all the different types of products that are built. The model included three products. The three products include a high-volume microprocessor, an advanced low-volume microprocessor, and a test chip. The order.txt file contains all the information concerning how much of each type of microprocessor should be released into the factory and when. This data was

taken from Intel's plans regarding which products should be produced and how much of each should be produced; these plans came from the Production Control Group at Fab A.

The `rte.txt` file contains information about the sequence of process steps that each product must go through. Each process step has an equipment family or specific piece of equipment associated with it. It also contains the processing time for each process step and the number of lots that are processed. The file also shows rework percentages for the process steps. The information for the `rte.txt` file came from a confidential Intel process book, which detailed the 0.25 micron process sequence.

The `cal.txt` and `attach.txt` files contain all of the downtime, preventive maintenance, and repair time data. Since not all of the equipment that is planned for the fab is currently in use, data for downtimes, preventive maintenance times, and repair times was gathered from several sources. Equipment that was common to both the 0.60 micron technology generation and 0.25 micron technology generation was modeled by analyzing the equipment's performance in factories using 0.6 micron technology. More advanced equipment that was shared between the 0.35 micron technology generation and 0.25 micron technology generation was modeled by analyzing the equipment's performance in fabs utilizing the 0.35 micron technology. Information for the equipment that is unique to the 0.25 micron technology generation was estimated. Less than 10% of the equipment sets used for the 0.25 micron technology were unique to that technology. Due to the small percentage of equipment for which we made estimates, the estimates should not severely impact the accuracy of the model for our purposes - this is the assumption that was made.

The model assumed operation dependent failures, which means that a machine may only fail when it is operating. All equipment failures were based on the number of units processed. Exponential distributions were used for all means-units-between-failures (MUBF) distributions. The failure data for every equipment set, of which there are more than fifty, was not analyzed due to time and resource constraints. Failure and repair distributions for two equipment sets (the steppers and a physical vapor deposition tool) were analyzed and found to be approximately exponentially distributed. This led to the assumption that all failure and repair distributions were exponentially distributed. Assuming that all failure and repair distributions were exponentially distributed considerably reduced the amount of time needed to collect the pertinent data. Assuming an exponential distribution meant that the only piece of data that was needed was a mean value, which was readily available for most equipment sets. The reason for using units processed instead of time to model equipment failures was to capture one of the important characteristics of a fab ramp. The characteristic was that semiconductor process equipment tends to fail more often as its utilization increases. This characteristic was verified by several fab personnel who had been participated in many fab start-ups. Repair times were based on times, not units processed.

Preventive maintenance was modeled using a combination of time and units processed. Some preventive maintenance schedules occurred daily, or shiftly, or weekly. These schedules were modeled using uniform time distributions. For example a weekly preventive maintenance schedule was modeled with a uniform distribution of 7 +/- 0.5 days. Other preventive maintenance

schedules were based on the number of wafers, or lots, processed. These schedules were modeled using a constant number of units processed. All preventive maintenance times were modeled using normal distributions.

The majority of the equipment in the model uses the first-in-first-out (FIFO) rule to determine which lot to process next. The equipment that does not use the FIFO rule using a rule called SSU, or same set-up. The SSU rule looks at the lots in queue and chooses the lot with the same set-up as the lot currently being processed. The SSU rule was chosen because it approximates the way Intel runs some of its equipment sets in the fab. For certain pieces of equipment that have been identified as constraints or near-constraints, Intel will run large batches of lots in order to minimize set-up times and maximize throughput.

Some pieces of equipment in the fab run batches of lots simultaneously. Examples include wet clean stations and diffusion furnaces. The discrete-event simulator is able to model these batch-processing stations and the relevant processing times. For some of the tools, set-up times are included in addition to processing times. The tools that had set-up times included were generally the tools with the lowest capacities or capacities very close to the tool set with the minimum capacity.

The model does not explicitly represent the technicians in the factory. The effects of operators were accounted for by adding a delay step after every process step. The delay step time is drawn from an exponential distribution. The product has to go through a delay step in between each process step. Besides operators, the delay step is intended to model transportation delays between process steps and lots going on hold during processing.

Chapter 6. Intel History and Characteristics of Intel Fabs

This chapter acquaints the reader with Intel Corporation and the characteristics of Intel fabs. The objective of my research is to understand semiconductor fab start-ups. However, since most of my actual data on semiconductor fab start-ups comes from fabs at a single company - Intel Corporation - it is important to understand some of the unique characteristics and qualities of Intel fabs. The chapter also acquaints the reader with the strategic context in which the research was conducted.

The first section discusses Intel's strategy in the microprocessor market. The second section discusses Intel's capacity planning process and operational strategies at Intel fabs. The last section discusses Intel's Fab A, where I did most of my research, and discusses the role it plays in Intel's future strategic objectives.

6.1 Intel's Strategy and Position in the Semiconductor Industry

One of the semiconductor companies leading the capacity expansion is Intel Corporation. Intel, the world's largest chipmaker as measured by revenue, is the leading supplier of microprocessors for personal computers. With 1995 revenue of \$16 billion and a five-year annualized revenue growth rate of over 30%, Intel, along with Microsoft, is considered by many industry observers to be the major driver of the personal computer industry. Intel microprocessors and Microsoft operating systems can be found on about 80% of all computers in the world. The dominance of Intel and Microsoft is so prevalent throughout the personal computer industry that a term has been coined to refer to the Intel/Microsoft monopoly - "Wintel" (a combination of Windows and Intel).

Not always known for its semiconductor manufacturing capabilities, Intel's traditional strength has been designing new generations of microprocessors for IBM-compatible computers and bringing them to market before competitors. Since its 8088 microprocessor was chosen by IBM in 1981 to power its new line of personal computers, Intel has introduced five successive generations of microprocessors. The latest generation, the Pentium® Pro, was introduced in November of 1995. Beginning in the late Eighties, Intel recognized the importance of manufacturing as a competitive weapon and placed more focus on developing its semiconductor processing and manufacturing capabilities.

In order to maintain its dominance in the microprocessor industry, Intel has plowed a large percentage of its profits back into capital spending for building and equipping new semiconductor fabs. In 1995 alone, Intel spent nearly \$3.6 billion on fabs and process equipment. And during the five years from 1990-1994, Intel spent over \$8 billion for capital additions to property, plant, and equipment.

Intel's basic business strategy, as defined by Microprocessor Report publisher Michael Slater [10]:

Its fundamental asset, one that is not going to be matched by anybody else, is its manufacturing capacity. Intel continues to invest billions of dollars per year to add capacity and it looks like it may well be in a position to own the majority of the very high-performance, high-volume manufacturing capacity necessary to build these processors. The Intel business model is simple. It builds capacity and then tries to fill the fabs. The variable is the price it has to sell the chips for in order to fill the fabs and to maintain its market share.

Currently, Intel builds and equips about one new semiconductor fab every year. The pace is expected to quicken during the coming years. Given the huge fixed costs associated with building a new fab, it is critical for these fabs not only to be as productive as possible, but also to recoup the large initial investment.

The most recently constructed fabs at Intel, including Fab A where much of my data was collected, are involved in the upcoming 0.25 micron process technology ramp. The need to better understand and continually improve semiconductor fab start-ups at Intel are currently being driven by Intel's long range strategic plan for the 0.25 micron technology generation ramp. Intel's 0.25 micron technology capacity ramp is the most difficult yet. A comparison of the total available fab capacity for Intel's 0.6, 0.35, and 0.25 micron technologies is shown in Figure 6.1. The vertical axis shows wafer starts per week (WSPW), a measure of fab capacity. The horizontal axis measures time. To make comparisons easier, the times for the three technologies are normalized to each of their respective process certification dates. The process certification date is a milestone for each process technology generation. The process certification date defines when the process has been developed enough to be certified as a manufacturable process that is capable of manufacturing product for outside customers. In general, the most important metric for measuring the readiness, or manufacturability, of a new manufacturing process is its process capability (C_{pk}). The 0.25 micron technology capacity ramp, as measured by wafer starts per week, is substantially steeper than the two previous generations.

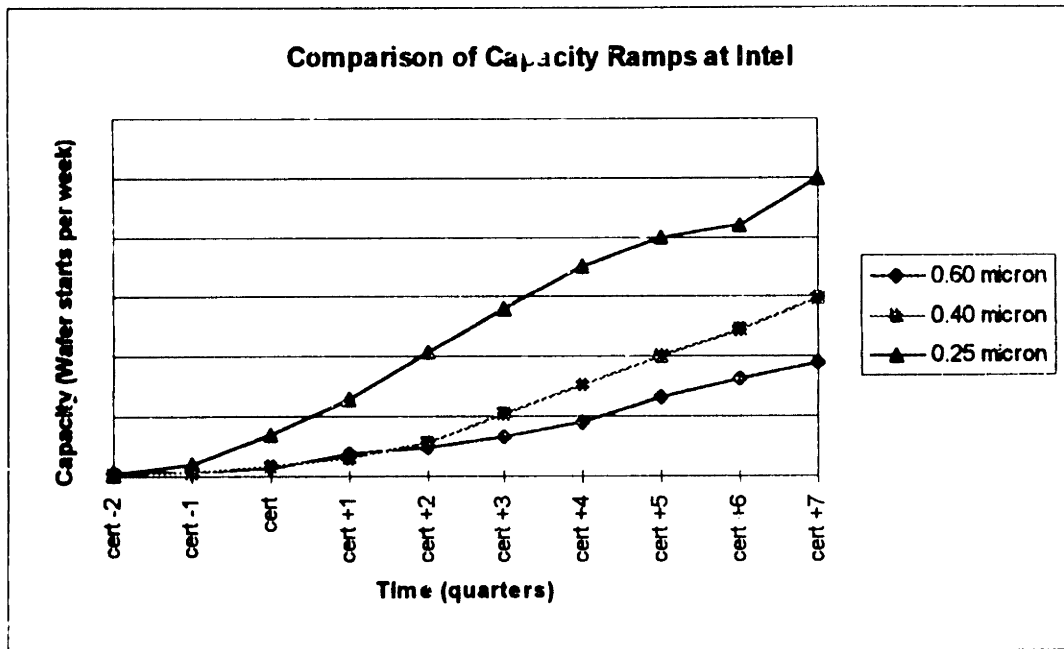


Figure 6.1. Comparison of Capacity Ramps at Intel

The ever-increasing ramp rates are necessary if Intel wants to maintain its dominance in the microprocessor industry. As the dominant player in the microprocessor industry, Intel's strategy is to reach the market before its competitors with each successive generation of X86 microprocessors. Not only must Intel reach the market before its competitors, but also it must reach the market with substantial volume. As a consequence, not only must the product be ready, but also manufacturing capacity must be installed and ready if Intel is to meet its goal of ramping volume quickly. To illustrate the importance of this first-to-market philosophy to Intel, Intel's stated 'Job 1' during 1994 and 1995 was to ramp the Pentium ® processor output as fast as possible.

The capacity ramp is a significant activity in bringing a new process technology and the accompanying microprocessors that are built using that new technology to market in volume. While the capacity ramp is typically one of the last activities to occur in the process of getting a new product to market, it is typically the most complex, costly, and risky activity in the chain of events.

The steeper capacity ramp for the 0.25 micron technology generation requires not only more fabs than previous process generations, but also that each fab ramp faster than the current internal 'best-to-date' ramp rate. In order to ramp at these higher rates, newer and better strategies must be identified and implemented. Ideally, these strategies should not result in significantly higher costs or complexity.

6.2. Understanding Intel's Capacity Planning and WIP Management Strategies

The purpose of this section is to help the reader better understand some of the characteristics that may be unique to Intel's wafer fabs and planning process. These three areas of interest are:

1. capacity planning
2. work-in-process (WIP) management strategy
3. fab metrics

After a decision is made by Intel's upper management to build a new fab, a manufacturing group within the company is chartered to develop the specifics of the plan. Since the design and construction of the fab is a critical path activity, one of the first steps is to estimate the size of the factory needed. This is done by projecting the number of wafer starts per week required and then multiplying that estimate by some factor in order to determine the square footage required. This factor depends on the particular process technology that is being used, the amount of support space required, and the nature and level of automation.

Assuming that the process development group has defined the equipment set required for the process technology generation that the fab will use, the next step in the process is to determine how many of each type of process equipment to purchase. The objective is to ensure that the fab will have enough capacity to meet projected demand and that throughput times will be acceptable. In addition to capacity and throughput time, cost is a major consideration.

The WIP management policy that the fab uses heavily influences this part of the capacity planning process. If a kanban system is implemented as part of a just-in-time (JIT) approach, a balanced line is desired. Capacity is purchased to achieve a balanced line. Most Intel fabs use a constraint management system to manage work-in-process and to schedule wafer releases into the factory. The constraint management system is based on the Theory of Constraints developed by Goldratt [11] [12] [13]. If the fab does use a constraint management system, the manufacturing engineering group identifies a constraint tool set and plans the remainder of the capacity around that constraint equipment set. There are several methods for choosing a constraint tool set - one of the most popular is to simply choose the most expensive equipment set.

At this stage of the planning process, the manufacturing engineering group typically utilizes static capacity models. For example, a typical capacity calculation might look something like equation 6.1. Equation 6.1 is the formula for estimating the weekly capacity of an equipment set.

$$C = \frac{N * CRR * A * 168}{S} \quad (6.1)$$

where,

C = weekly capacity of the equipment set [wafer starts per week]

N = number of pieces of equipment in parallel
CRR = capacity run rate of the equipment [wafers/hour]
A = availability [fraction of 1]
S=number of times that a wafer has to go through this piece of equipment

and there are 168 hours per week. The capacity run rate is an aggregate number that estimates the number of wafers that can be processed during each hour of tool uptime, assuming that the station is never starved for work. The availability is an estimated value. In practice, both the capacity run rate and availability change over time. The variable of interest is N, the number of pieces of equipment.

Having identified the constraint equipment set and its static capacity, the manufacturing group plans for the remainder of the equipment sets to have excess capacity compared to the constraint equipment set. The amount of excess capacity can vary from 10-100%. An additional consideration when deciding how many pieces of equipment to purchase is redundancy. Redundancy means that there are at least two pieces of equipment in each equipment set. So although one piece of equipment may supply more than the constraint capacity, two pieces of equipment are purchased to minimize risk associated with equipment downtime.

After determining how many of each type of equipment needs to be purchased, the manufacturing group lays out an installation schedule to meet the wafer starts ramp. For all equipment sets, installed capacity leads the effective demand of the wafer starts ramp. The installations for the constraint equipment set are chosen to closely match the wafer starts ramp. Some cushion is planned into the schedule to minimize the effect of disruptions or unplanned delays during equipment delivery, installation, or production qualification. The installation schedules for the other equipment sets are planned such that their capacity at any point during the wafer starts ramp is greater than the capacity of the constraint equipment set. Traditionally for the non-constraint equipment sets, the percentage of the capacity installed closely matches the percentage of the wafer starts ramp completed. So, if the wafer starts ramp is 50% complete, approximately 50% of all non-constraint equipment should be installed and qualified for production use.

There are several factors that complicate scheduling installations.

1. equipment qualification typically occurs one to two months after equipment installation
2. installation dates are dependent on equipment suppliers' abilities to meet delivery schedules
3. wafer starts demand at the back end of the production line lags the demand at the front end of the line
4. uncertainty in equipment performance can result in unplanned overcapacity or undercapacity during the capacity ramp

The remainder of this section is devoted to introducing the reader to Intel's policies and techniques for managing their semiconductor fabrication facilities. As stated earlier, nearly all Intel fabs utilize some form of Constraint Management to run their manufacturing operations. Intel's Constraint Management Policies are based Eli Goldratt's Theory of Constraints [11] [12]

[13]. The basic goal of constraint management is to maximize throughput. Throughput is maximized by identifying the bottleneck, or constraint, and making sure that the constraint is never idle. All non-constraints are subordinated to the constraint and inventory is maintained in front of the constraint to minimize the effect of upstream variation on constraint utilization. While most fabs practice constraint management, each fab has a large degree of autonomy in selecting where to place the constraint and how to manage the rest of the factory to support the constraint. However, policies are generally chosen that maximize throughput and linearize wafer shipments from week to week.

Fab management usually emphasizes throughput and utilization of constraint resources. In keeping with the principles of Goldratt's Theory of Constraints, significant resources are devoted to improvement activities at the capacity constraint. The focus is not only on maximizing the utilization of the constraint, but also on increasing the run rate and availability. Resources are also devoted to yield improvement activities throughout the process to increase the number of good die that the fab produces. Yield improvement activities include defect monitoring and preventive maintenance. Process capabilities also contribute to yield performance, so control charts are monitored for out-of-control events

In keeping with the theory that what gets measured is what gets the most attention, a review of the important metrics at an Intel fab should help the reader better understand Intel's fab management policies. Important fab metrics include:

1. the gap at the bottleneck equipment set
2. throughput at the bottleneck
3. number of activities
4. line yield
5. die yield
6. inventory levels

The bottleneck gap is calculated by subtracting the bottleneck's utilization from the bottleneck's availability. An activity is a set of sequential process steps that physically or chemically changes the wafer. A typical microprocessor requires from 50 to 100 activities. Line yield is the percentage of wafers that make it through the factory to the test and sort operation. Die yield is the percentage of die that pass the test operations at the end of the process. And the inventory levels are simply the amount of inventory, measured in wafers or lots of wafers, currently in the fab.

6.3 Background and Introduction to Intel's Fab A

The majority of the data from actual fab start-ups came from a recent ramp at Intel's Fab A. The historical data and observations were drawn from the 0.6 micron process technology ramp at Fab A. Fab A not only developed the 0.6 micron process technology, but also ramped the process to significant volumes before transferring the process technology to other Intel fabs.

The impetus for the research on semiconductor fab start-ups came from the management at Fab A, who were interested in learning more about semiconductor fab start-ups and to apply that learning to an upcoming ramp at Fab A. The plant management at Fab A was interested in understanding how they could improve the upcoming 0.25 micron process technology ramp. Fab A plays a major role in the 0.25 micron technology generation at Intel. Intel's Fab A not only develops Intel's 0.25 micron semiconductor process technology, but also is the first Intel fab to ramp that process technology to significant volumes and transfer the technology to three other Intel fabs. Fab A has several priorities for its 0.25 micron technology ramp and a significant part of the problem is understanding techniques for meeting all these objectives in a cost-effective manner. The objectives include:

1. developing the 0.25 micron technology process
2. transferring the technology to three other Intel fabs within the space of nine months
3. ramping the 0.25 micron technology

Figure 6.2 shows a breakdown of the 0.25 micron technology generation capacity curve in Figure 6.1. Figure 6.2 breaks the total capacity into capacities for each of the fabs involved with the 0.25 micron technology generation. The figure shows that Fab A will initially develop the technology at low capacity levels and then begin its production ramp around the process certification date. About the time that Fab A begins its ramp, Fab B also starts its ramp, with another fab following each of the next two quarters.

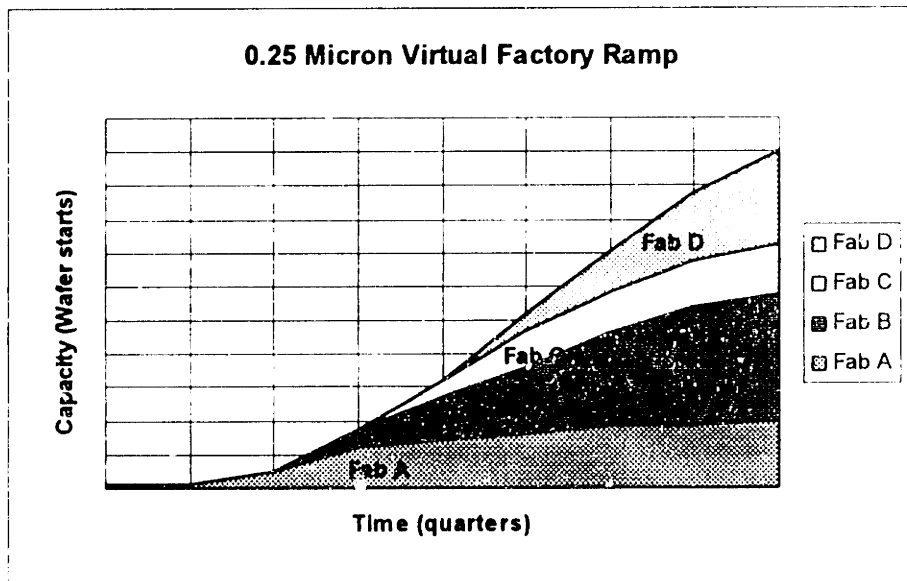


Figure 6.2. 0.25 micron technology generation Virtual Factory Capacity Ramp at Intel.

The majority of the people that I talked to at Fab A were there during the last start-up. In addition, many of them were involved with start-ups at other Intel fabs previous to the ramp at Fab A. So there was quite a pool of experience at Fab A and the recent start-up there was still relatively fresh in their minds.

Chapter 7. Analysis of Data

The chapter describes the data analysis process. The first part of the chapter discusses the analysis of data from Fab A. The second part of the chapter deals with issues discovered when I compared the ramp at Fab A with the ramps at Fab B and Fab C. Both parts discuss how the factors detailed in Chapter 4 affected the start-ups at these fabs.

7.1 Analysis of Data from Fab A

The Gantt chart that was constructed for the 0.6 micron process technology generation at Intel is shown in Figure 7.1. The Gantt chart illustrates the interdependencies and relative positions of the major activities in relation to the semiconductor fab start-ups, which are represented by the last three activities on the Gantt chart. For the purposes of this exercise, the output ramp was defined as the wafer starts ramp for the fab. The reader should note that many of the activities on the Gantt chart include many of the levers that were identified in Chapter 4. Common areas included product design, development of a production process, work force staffing and training, installed equipment, and plant engineering in the form of the fab expansion activity.

The Gantt chart begins with the development of process equipment for 8" silicon wafers. Most of this development work was led by IBM and was carried out in conjunction with equipment suppliers. This activity was important because Intel's 0.6 micron technology generation was the first at the company to use 8" wafers; hence these process equipment development activities were needed before Intel could proceed with its process development on 8" silicon wafers.

The second activity shows the initial process development work that was carried out at another Intel fab using 6" silicon wafers. As the Gantt shows, the responsibility for carrying on this development and transferring it to 8" wafer production was then transferred to Fab A. Fab A then spent approximately two years continuing development and preparing for production. At the same time, two distinct products were being designed in anticipation of manufacturing them using the new 0.6 micron technology. In parallel with the process development activities and product design programs, Fab A was being expanded and new process equipment was being installed and qualified in the fab to prepare for the production output ramp. Not only did Fab A have to develop the process technology, but also it had to ramp output and transfer the technology to two other fabs (Fabs B and C).

The completion of all of the activities on the Gantt chart was required before products could be sold to customers. If one activity was not completed on time, no product could be sold to customers and hence no revenue could be generated. Products could not be sold until the process was certified. Products can not be sold until the product is designed and certified as error-free. No product could be sold if there was no fab or process equipment with which to manufacture product. For a semiconductor company, as with many other manufacturing companies, a massive coordination effort is required to bring new process technology and products to market in volume as quickly as possible. T.J. Rodgers, the CEO of Cypress Semiconductor, summarized it best

when he wrote, "An integrated circuit is the end result of a thousand multidisciplinary tasks; doing 999 of them right guarantees failure, not success." [14]

Note that the Gantt chart does not show many of the activities associated with the output ramps at Fab's B and C. For example, the Gantt chart does not show the fab construction, or staffing and training activities, or equipment installation and qualification phase for either Fab B or Fab C. For the most part, these activities are similar to their respective counterparts at Fab A and putting them on the Gantt would add minimal value to the chart. However, a later section of the thesis does discuss an interesting aspect of the ramp at Fab C, specifically concerning the equipment installation and qualification activities.

The Gantt chart does not clearly identify any one activity or set of activities as the critical path to the start-up at Fab A. A case could be made for any one of several groups of activities. The list of potential critical paths include the process development activities (activities three through seven), the product design activities, and activities specifically related to the start-up at Fab A (activities sixteen through nineteen). Based on discussions with Intel managers, it was not clear if a set of activities was designated as the critical path for the 0.6 micron process technology generation.

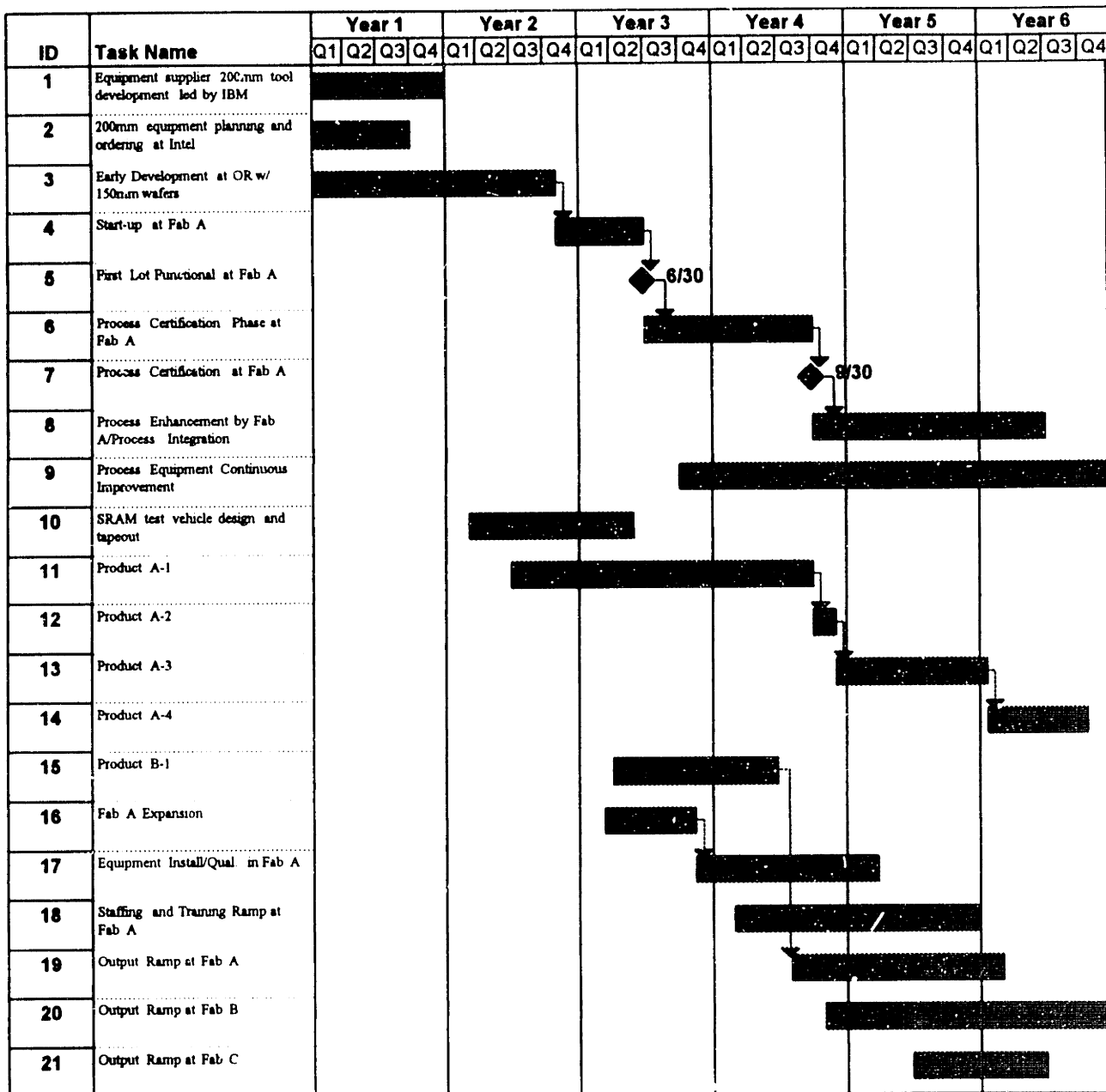


Figure 7.1. 0.6 Micron Process Technology Generation Gantt Chart

An example of some of the data that was analyzed to understand the start-up at Fab A is shown in Figure 7.2. Figure 7.2 shows the headcount ramp at Fab A during the start-up. The headcount ramp in Figure 7.2 illustrates an S-curve which closely mirrors the wafer starts ramp at Fab A - shown in Figure 7.5. Before the start-up, the headcount was fairly stable. Approximately six months before the start-up began, the fab began hiring and training technicians, engineers, and managers. After the start-up, the headcount stabilized and new people were only hired to replace those who left or transferred to other departments or parts of the organization.

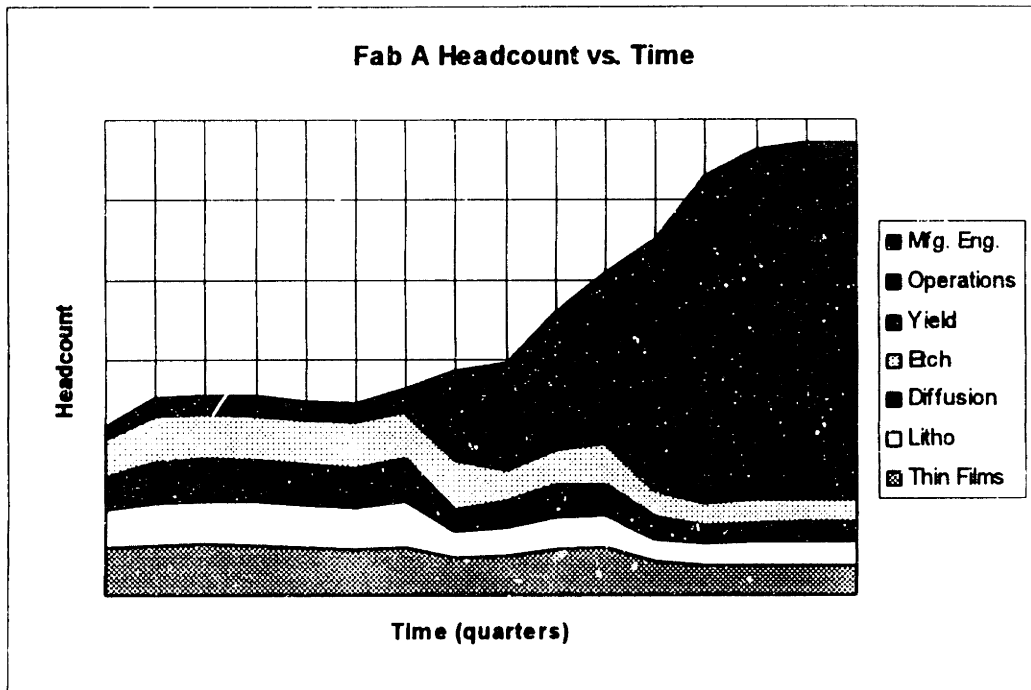


Figure 7.2. Headcount ramp at Intel's Fab A

Figure 7.3 shows the relative capacity ramps, as measured by wafer starts per week capacity, of the three Intel fabs involved with the 0.6 micron process technology generation. The times on the x-axis have been normalized to show the relative capacity ramp rates of the three fabs. Fab A ramped first, followed by Fab B, and then Fab C. Fab B ramped at about the same rate as Fab A and continued to ramp capacity after Fab A stopped because Fab B was about twice the size of Fab A. However, Fab C, which was similar in size to Fab A, ramped much faster than Fab A. Fab C required approximately one year to reach full build-out capacity; while Fab A required almost two years to reach the same capacity level. The reasons for Fab C's faster ramp rate are explored in more detail in Section 7.2.

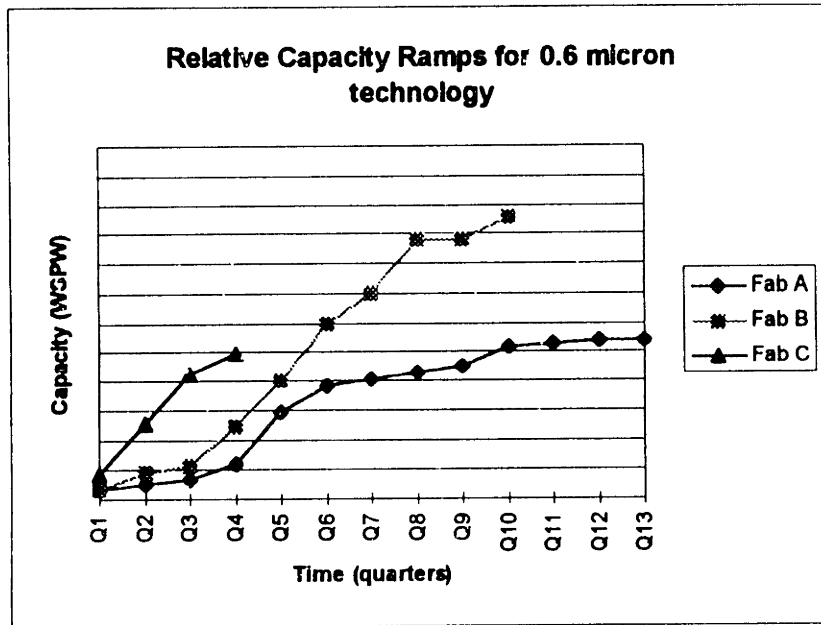


Figure 7.3. Capacity Ramps for Intel Fabs for the 0.6 Micron Process Technology Generation

To help analyze the activity duration data show in the Gantt chart in Figure 7.1, a pareto chart was constructed (Figure 7.4). Four activities from the Fab A start-up were included in the pareto. The four activities that were chosen were the equipment installation and qualification activities, the staffing and training ramp, the wafer starts ramp (referred to in the pareto as the output ramp), and the fab expansion. These four activities were chosen because not only were they major components of the fab ramp based on time and cost, but also they were all within the sphere of influence of the Fab A Manufacturing Group. While there were several other activities that are critical to the fab ramp, such as the product design and process development activities, they were not under the influence or control of the Fab A Manufacturing Group. A key assumption that was made was that none of the other activities were limiters to the wafer starts ramp

The pareto (Figure 7.4) shows that the staffing and training activity required the longest amount of time. The wafer starts ramp and equipment installation and qualification both required slightly less time and the fab expansion required the least amount of time. The pareto analysis showed that the staffing and training ramp was a candidate for the main gating item of the wafer starts ramp at Fab A. The equipment installation and qualification was the second gating item.

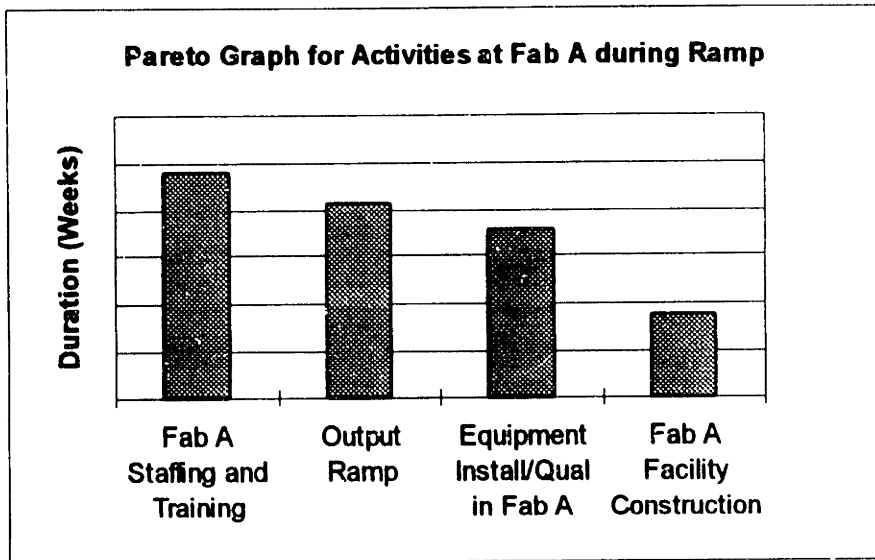


Figure 7.4. Pareto of activities during ramp at Intel's Fab A

Another useful lens for analyzing the data is illustrated in Figure 7.5. Figure 7.5 shows the end dates of three of the activities versus the wafer starts per week (WSPW) over time. The wafer starts per week on the y-axis was normalized by using the fab's long-term rated capacity as the 100% mark and 0 WSPW as the 0% point. The graph shows that there were three different types of wafer starts. Full loop starts were those wafers that go through the entire process. MW/SL represented monitor wafers and short loops. Monitor wafers and short loop wafers traveled through only part of the process. Fab B feeds were wafers that began processing at Fab A, but then part way through the process were boxed and shipped to Fab B to finish processing.

The graph shows that the activity involving staffing and training - "end technician training" - ended much later than all of the other activities. In fact, it did not end until almost 15 months after the process certification - the milestone at which the process is certified as manufacturable and after which microprocessors can be sold. This observation reinforced the conclusion of the pareto analysis which showed that the staffing and training ramp was a significant limiter to the fab ramp.

The graph also shows that the equipment installation and qualification activity ended about six months after the process certification date. The interesting thing to note is that the wafer starts ramp was at about 75% of its final value at this time. This is interesting because the plan for the upcoming 0.25 micron ramp at Fab A requires that the wafer starts level be at 75% of their final value at the process certification date. Based on these numbers, the equipment installation and qualification activity must be complete by the 0.25 micron technology process certification date.

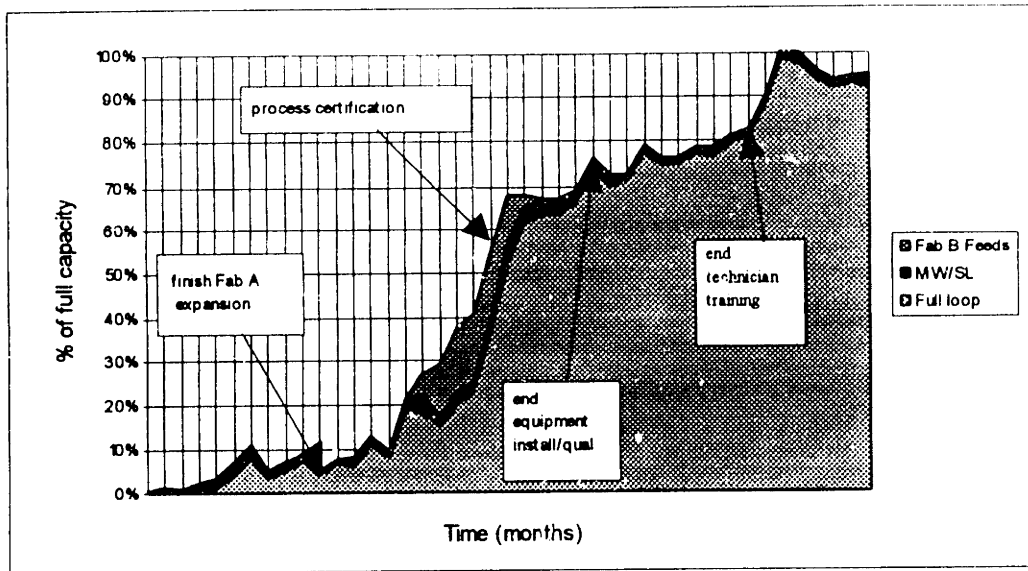


Figure 7.5. Activity Milestones vs. Wafer Starts per Week (WSPW) Ramp at Fab A

The last technique that was considered for looking at the data was statistical analysis. Unfortunately, multiple regression analysis was not possible. The plan was to correlate several input variables, such as: % of training complete or % of equipment qualified with an output parameter, such as % of wafer starts ramp completed or wafers out. However, the S-curve nature of both the inputs and outputs made correlation very difficult. The data could not be normalized and there were no contrasting input data points. For example, when one of the inputs assumed a high value all of the other inputs assumed a high value and there were no instances where some of the inputs assumed a low value while others took on high values.

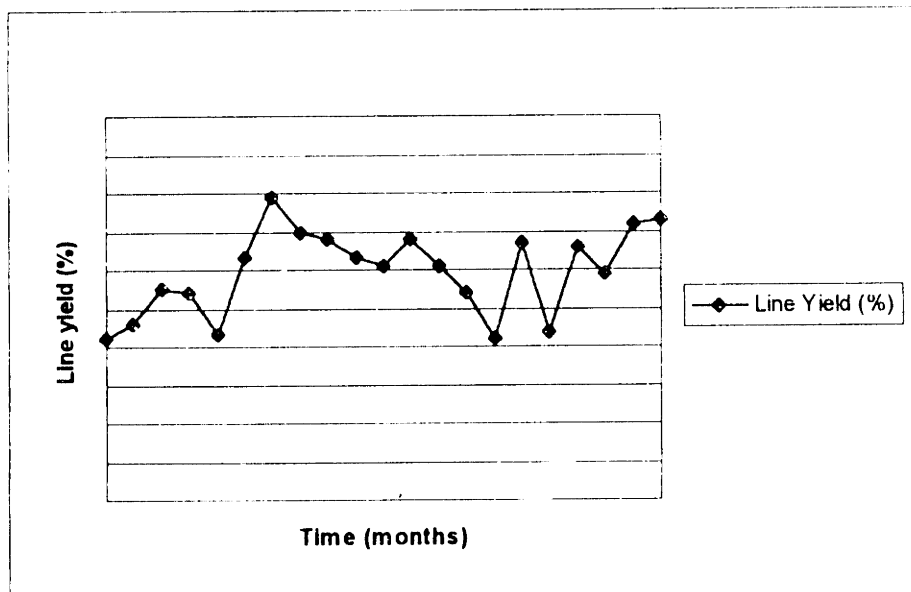


Figure 7.6. Line Yield during Start-up at Fab A.

Figure 7.6 shows the line yield over time at Fab A during the start-up phase. The graph begins about nine months before the capacity ramp (illustrated by the steep part of the wafer starts curve in Figure 7.5) at Fab A and ends approximately twelve months after the start-up began. The graph shows that there was little to no improvement in line yield during the start-up. Not only was there no improvement, but also the line yield fluctuated wildly from month to month. The monthly standard deviation of line yield was over 10%. Based on the lack of improvement in line yield, a case could be made that the line yield was a major limiter to the start-up at Fab A.

The die yield trend at Fab A was characterized by a slow, but steady uptrend beginning before the start-up and continuing for several years. The process development and sustaining engineers at Fab A were focused on die yield improvements and were very successful in making improvements. While the initial die yields were not extraordinary and certainly limited die output at Fab A, improvement was continuous and contributed to the increase in die output from Fab A.

In addition to analyzing quantitative data, such as line yields and headcounts, an extensive amount of data from interviews with fab personnel was collected and analyzed. To help analyze some of the data, a cause-and-effect diagram was constructed using data from the interviews. A cause-and-effect diagram is also known as a “fishbone diagram” or an “Ishikawa diagram.” The purpose of the cause-and-effect diagram was to identify and to clarify the causes of problems. The problem that was investigated using the cause-and-effect diagram was “What were the weaknesses of the 0.6 micron technology generation ramp at Intel’s Fab A?” This question was analyzed in order to better understand some of the factors that limited the start-up at Fab A. The cause-and-effect diagram is shown in Figure 7.7.

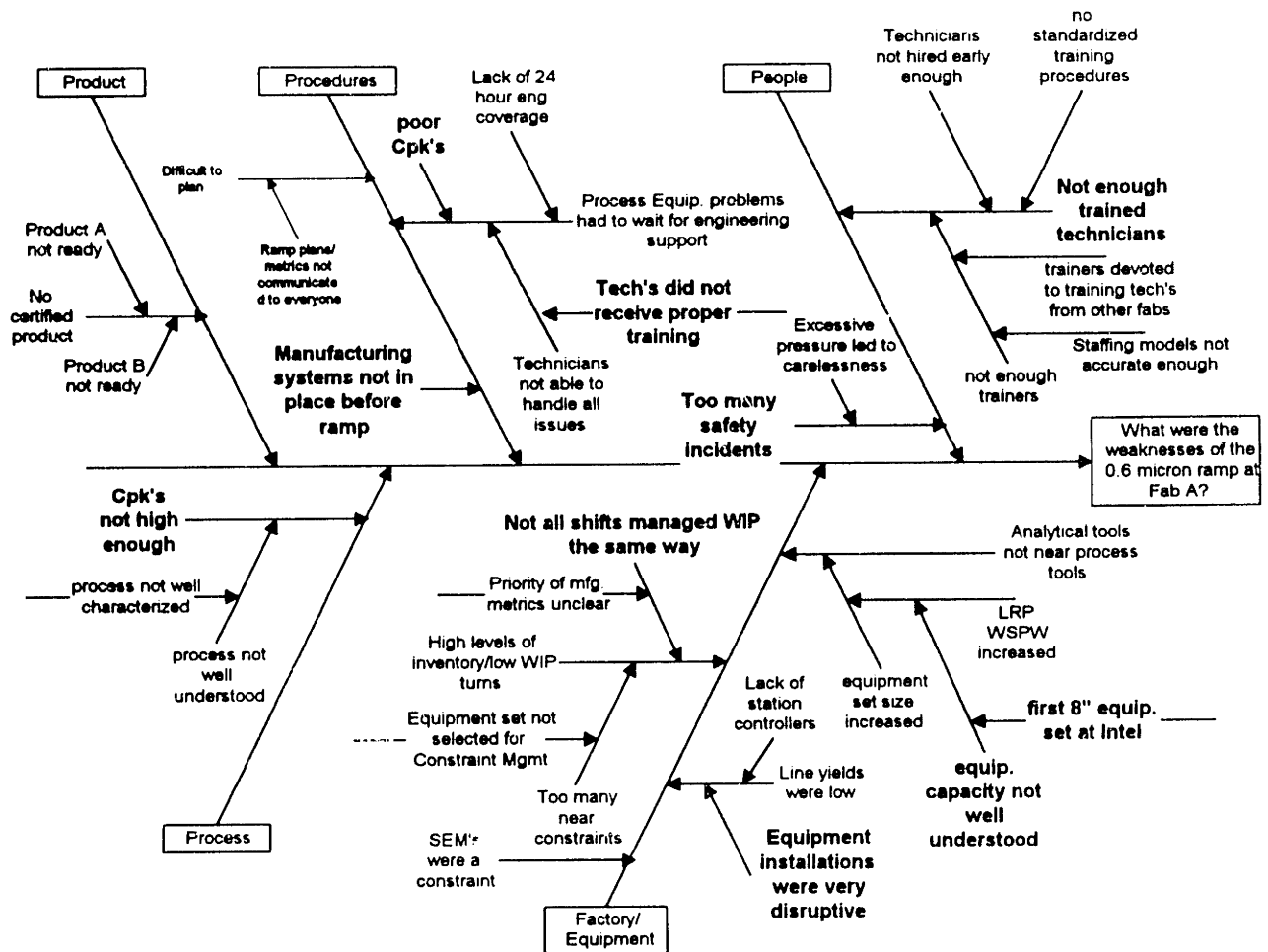


Figure 7.7. Cause-and-Effect Diagram for 0.6 Micron Technology Generation Ramp at Intel's Fab A.

Responses that were mentioned several times are bolded in Figure 7.7. The most frequent responses were related to: 1) the training of technicians, 2) poor process capabilities, 3) lack of manufacturing systems, 4) safety issues, 5) capacity considerations, and 6) disruptions due to equipment installations. The manufacturing systems identified in the cause-and-effect diagram correspond to the control systems identified as a potential level affecting the performance of a fab start-up.

The interview responses to questions involving the start-up were consistent with several of the conclusions from the analysis of the quantitative data. The interviewees, who had not seen any of the quantitative data or analysis, cited work force training, process capability, manufacturing (control) systems, low line yields, equipment problems, and the lack of a product as weaknesses of the start-up at Fab A.

7.2 Analysis of Comparison Data from the Ramps at Fab A and Fab C

Additional data on start-ups at Intel came from analysis of a more recent ramp at another Intel fab - Fab C. Fab C ramped the 0.6 micron process technology approximately one year after Fab A. Fab C was able to ramp output much faster than Fab A (refer to Figure 7.3). Based on wafer starts ramp rate data, Fab C's 0.6 micron technology ramp was identified as the fastest-to-date within Intel. Why was Fab C able to ramp so much faster than Fab A? What special things did Fab C do to prepare for and execute a faster ramp?

In order to better understand Fab C's ramp, several people who had been involved with the ramp at Fab C were interviewed. In addition, people from other fabs were interviewed to get another perspective on the ramp at Fab C. The people who were interviewed at Fab C included the manufacturing manager, industrial engineers, and equipment installation and qualification teams. The people outside of Fab C included other manufacturing managers and industrial engineers.

Not only was Fab C's output ramp faster than the output ramp at Fab A, but Fab C's wafer starts ramp rate was much faster than Fab A's wafer starts ramp rate. Why should this be? Why should two similarly-sized fabs ramping the same technology exhibit such different output and wafer starts ramp rate performance? There were actually several reasons why Fab C was able to ramp much faster. Identifying these reasons should help us gain a better understanding of what factors allow one fab to ramp so much faster than another fab. The reasons are detailed in the next paragraphs.

Based on the research and discussions with Intel personnel, four primary reasons for Fab C's improved performance compared to Fab A were identified. The four reasons were:

1. Relative process technology maturity during Fab C's ramp versus Fab A's ramp
2. Availability of certified product for Fab C's ramp
3. Intelligent use of resources at other Intel fabs by Fab C, especially training resources
4. Techniques and strategies involving installation and qualification of process equipment

Fab C's 0.6 micron technology ramp occurred approximately twelve months after Fab A's ramp. Fab C had the benefit of ramping a much more mature technology than Fab A. Fab A and Fab B had already encountered and solved several process issues with the 0.6 micron technology process before Fab C began its ramp. This meant the Fab C encountered fewer problems and disruptions during its ramp which would have limited its performance. These problems and disruptions were the ones that significantly slowed the ramps at Fab A and Fab B.

The ramps at Fab A and Fab B were also slowed by the lack of certified product. Problems with the design of several microprocessors resulted in no product being available until well after the ramps at Fab A and Fab B had begun. Fab A responded to the lack of certified product by ramping with test product - Static Random Access Memory (SRAM) - instead of logic products during the early part of the ramp phase in order to gain yield learning and proceed with its continuous improvement activities. This strategy was justified by the fact that there are huge

fixed costs and low variable costs associated with running a fab and that learning from the test devices would improve Fab A's knowledge and understanding of the process. Due to the fact that Fab C ramped about a year after the other fabs, Fab C did not encounter this problem.

Fab C's factory management utilized its position as the third fab to ramp the 0.6 micron technology at Intel to get a head start on training its work force. Fab C sent large portions of its work force, including technicians, engineers, and managers, to Fab A and Fab B for training. This strategy enabled Fab C's work force to start down the learning curve before their wafer starts ramp even began. While this training slowed down the ramps at Fab A and Fab B, it enabled Fab C to ramp very quickly and effectively.

The last reason involved the strategies and techniques that Fab C used to install and qualify process equipment. Two areas were of particular interest. The first was the ability of equipment installation and qualification teams to install and qualify equipment at record paces at Fab C. Continuous improvement and focus on non-value-added steps meant that these teams could install and qualify equipment twice as fast at Fab C as they were able to at Fab A. Since the rate of equipment installations and qualifications frequently limit the ability of a fab to ramp wafer starts, any improvements in this activity would enable faster ramps. In addition to being able to install and qualify process equipment quickly, Fab C was the only fab in Intel's history to install and qualify most of the process equipment before the wafer starts ramp began. Approximately 90% of all equipment in Fab C was purchased, installed, and qualified for production by the time the wafer starts ramp began. By comparison, Fab A had installed and qualified less than 50% of its equipment by the time its ramp began. The decision by Fab C's management to install and qualify most of the equipment before the ramp began meant that there were fewer disruptions to production during the ramp due to equipment installations. Fewer disruptions meant that more attention could be devoted to the wafer starts ramp and other continuous improvement activities.

7.3 Analysis of External Benchmarking Data

Additional data collection and analysis focused on external benchmarking. The purpose of the benchmarking was to provide a sanity check for the data that was collected at Intel fabs. By analyzing performance data from a start-up at a fab at another semiconductor company, we hoped to establish a reference point for better understanding the start-ups at Intel fabs. In other words, are the start-ups at Intel fabs representative of start-ups at other semiconductor fabs?

The external benchmarking data came from Intel's Competitive Analysis Group. The data is rather limited, but it does give us some relevant information. The data compares a more recent fab ramp at a large Intel fab with a similarly-sized fab at an Asian DRAM manufacturer. Both fabs used similar process technology generations. The main difference was that the Intel fab produced logic devices and the Asian manufacturer's fab produced memory chips.

For confidentiality reasons, the data can not be included in the thesis. However, the analysis of the data is discussed. Intel's fab required substantially longer than the DRAM fab to install and qualify process equipment and ramp the wafer starts to full capacity. On the other hand, Intel

required much less time to reach mature yields. In addition, Intel required much more time to build their fab than the DRAM manufacturer required to build their fab.

Among other things, the external benchmarking data suggested that there are significant differences between Intel's performance and the DRAM manufacturer's performance. The DRAM manufacturer pursued a strategy of quickly installing equipment and ramping wafer starts and was able to do it very quickly. This contrasts with Intel, which ramped wafer starts more slowly, but was able to devote more resources to yield improvement during that time. The main conclusion of the benchmarking data is that not every semiconductor company uses the same strategy to ramp a semiconductor fab. Since we were not able to follow up on the data, it was not clear why the two ramp strategies were so different. One possible hypothesis is that the DRAM market is a commodity market and that only the most advanced DRAM's command prices substantially above marginal cost. Since advanced DRAM's represent higher margins, DRAM manufacturers may be interested in getting some product to market before the process is mature and while yields are still low.

7.4 Limiters to the Start-up at Fab A

The results of the data analysis showed that there were several limiters to the 0.6 micron process technology generation ramp at Intel's Fab A. The limiters were subjectively ranked according to their effect on the start-up at Fab A.

1. Staffing and training of technicians and supervisors
2. Chip design not ready
3. Equipment installation and qualification
4. Process equipment capability issues that impacted equipment availability, run rates, and yield
5. Miscellaneous items such as line yield and safety issues

Staffing and training was ranked as the major limiter for several reasons. The pareto diagram showed that the staffing and training activities required the longest amount of time to complete. Figure 7.1 showed that other than the wafer starts ramp, the staffing and training activity was the last activity to be completed. In addition, the amount of time required for training during the ramp meant that less time and resources were devoted to other important activities like yield improvement, process capability enhancements, and continuous improvement.

The lack of a chip design was ranked second because it limited Fab A's ability to produce salable product at the beginning of its start-up. Since the design of the product was not completed, Fab A experienced a lag in going down the learning curve. Instead of starting down the learning curve three or four months before the process certification date, Fab A was not able to start down the learning curve until the process certification date. Fabs at Intel that ramped the 0.6 micron process technology after Fab A, such as Fab B or Fab C, were not constrained by product design availability.

The process equipment installation and qualification activity was listed as a limiter for several reasons. First, the fact that the equipment installation and qualification teams could only install

about fifteen tools per month meant that either the wafer starts ramp could only proceed as fast as the qualification rate or that tools had to be installed several months before the wafer starts ramp began. Figures 7.1 and 7.4 showed that the equipment installation and qualification activity required a substantial amount of time to complete and Figure 7.5 showed that it was not complete until about six months after process certification.

The fourth limiter was the low process capability of the 0.6 micron technology. The low level of maturity and understanding of the process contributed to several major process excursions. The lack of technology maturity at the process certification date led to excessive fire fighting by the engineering and operations groups in later months. The excursions took vital resources from long-term continuous improvement activities and led to a slower rate of overall improvement. The excursions also took resources away from programs to improve equipment run rates and availability.

The remainder of the significant limiters to the ramp were grouped together as the fifth limiter under the heading of "Miscellaneous." While none of these limiters were as significant as the first four limiters, they did slow down the ramp in various ways. Low line yields resulted in excessive costs and lost throughput since some wafers were scrapped after being processed through the capacity constraint. Also, several specific issues involving tool ergonomics and safety also impacted Fab A's ability to ramp quickly and effectively. For example, during the initial phases of the ramp, all of the wet stations had to be upgraded due to ergonomics issues. None of these upgrades were planned, and taking the tools down for the upgrades resulted in lost throughput and increased managerial complexity.

The analysis suggests that the headcount staffing and training activity represents the highest leverage opportunity. However, many of the problems associated with the headcount staffing and training ramp at Fab A had already been identified and were being addressed by personnel at Fab A. The next area for improvement was the equipment installation and qualification activity. It was not clear that this area had been a major focus of improvement since the ramp at Fab A so we decided to spend more time analyzing the equipment installation and qualification activity.

Chapter 8. Analysis of a Capacity Addition Policy that Includes the Effect of Disruptions

This chapter discusses one of the major differences between the 0.6 micron process technology ramps at Fab A and Fab C. Fab C installed over 90% of its equipment before the start-up. This chapter analyzes why Fab C, or any factory for that matter, would install equipment before it was needed to meet production schedules. The issue of when to install equipment and how much to install can be treated as a capacity expansion problem. A literature review and some of the theory of the capacity expansion problem were discussed in Chapter 3. Several key points from that section are discussed in this analysis.

There are several ways to understand the capacity expansion problem analytically. One way is to model the problem as a linear program. The objective of the linear program is to minimize the present value (PV) cost of the factory start-up such that the production demands are always met at discrete points in time. This simple model tends to lead to purchasing the equipment using a 'just-in-time' methodology. The 'just-in-time' methodology results because lower PV costs result if the equipment is purchased later rather than sooner, assuming a positive discount rate.

A variant of the linear program approach would be to maximize profit over the life of the factory start-up. This model would use complicated economic supply and demand formula for the products that the factory produces in addition to the costs for the various equipment that the factory uses to produce the products. An important component of the model would include learning effects based on the factory's cumulative production output.

The model that may be relevant to Fab C's decision to install and qualify equipment before production began is a variant of the simple model discussed in Section 3.2. That simple model is a linear program that attempted to minimize the PV cost of the start-up such that the production demand schedule is met at all times during the start-up. The proposed model includes a cost that is associated with the installation of a piece of equipment or pieces of equipment if they are installed during the production ramp.

The negative effects of disruptions can take many different forms. For instance, each time equipment is installed in the clean room, several detrimental effects result:

1. the clean room must be opened up to roll the equipment into place, thus risking contamination
2. clean room bays where the equipment is being installed must be shut down or risk particle contamination
3. valuable resources are diverted from other activities to organize the installations and qualifications
4. continuous introduction of new equipment makes it difficult for the organization to adjust to the new equipment because there is no stable period of time in which to make the adjustments
5. managerial complexity associated with coordinating equipment installation

All of these detract from the factory's ability to focus on other critical issues and aspects of the wafer starts ramp. As the ramp proceeds, there are typically many new issues that surface and

require the attention of the organization. These new problems are a result of the increased production volume, as growth in wafer starts increases the stress on the entire system.

To include the cost of disruptions, the model and equations from Section 3.2 were used. The disruption cost associated with each installation was substituted for the fixed cost, A , in equation 3.1. The optimal solutions to the capacity expansion problem from Section 3.2 are shown below (equations 8.1 and 8.2).

$$t \approx \sqrt{\frac{2A}{Bgr}} \quad (8.1)$$

$$x = gt \approx \sqrt{\frac{2Ag}{Br}} \quad (8.2)$$

The x represents the optimal capacity addition and the t represents the optimal time between capacity additions. This model is a bit abstract as some of the assumptions are not completely realistic. One assumption was that demand grows linearly over an unbounded horizon. A second assumption was that capacity can be bought in very small units. In reality, capacity must be purchased in discrete chunks because it just comes that way.

These last two equations are very critical to the remainder of the thesis so it is probably worthwhile to spend some time exploring them in some detail. The equations are critical because they are used to develop an equipment installation and qualification schedule in a later section. The equations contain four variables: A , g , B , and r . A is the disruption cost associated with each equipment installation or group of installations that occur at the same time. As the disruption cost associated with each installation increases, the lowest cost solution requires that larger and larger chunks of capacity be installed at longer intervals of time. B is the cost per unit of capacity. As B increases, the lowest cost solution requires that smaller and smaller chunks of capacity be installed and qualified more frequently. Increases in the discount rate, r , have the same effect as increases in the B variable; increases in the discount rate lead to smaller, more frequent capacity additions. Finally, increases in the rate of demand growth, g , lead to larger and more frequent capacity additions.

If we attempt to apply these formula to a real life example to determine the real NPV cost, we would use a spreadsheet and run the different possible scenarios, assuming that demand is only linear during the initial part of the factory start-up and that capacity can only be purchased in discrete chunks. And in fact, that is exactly what was done for one equipment set that is critical to the 0.25 micron technology ramp at Intel's Fab A.

Figure 8.1 shows the resulting sensitivity analysis for the calculations. Figure 8.1 shows the discounted cost of different capacity addition scenarios and different values of the fixed cost associated with each equipment installation for a critical lithography tool - the stepper. The graph substitutes W for the fixed cost variable (A). The x axis represents the number of steppers that were installed during each capacity addition. The y axis represents the normalized discounted

cost of the capacity installations, including the hidden cost (W) associated with each installation. Steppers are the most expensive piece of equipment at Intel fabs. They typically cost over \$4 million. In addition, wafers must visit the steppers from 15 to 20 times during processing. Because of their low wafer starts capacity and high costs, steppers generally represent over 20% of the total equipment cost for a new semiconductor fab.

The first thing to note is that some of the curves are not continuous. This makes sense because we are dealing with discrete capacity chunks over a finite period of time. As we would expect from the equations, as the fixed cost (A) decreases, the lowest cost solution is to install a small amount of capacity (one or two steppers) about once a month. As the fixed cost associated with installations increases, the lowest cost solution is to install and qualify equipment in larger batches less frequently. In the extreme as the fixed cost increases to over one million dollars, the lowest cost solution is to install all of the steppers before the wafer starts ramp begins.

The sensitivity analysis shows how incorrect assumptions about the implicit cost due to disruptions caused by equipment installations affect the cost of the wafer starts ramp. For instance, the graph indicates that if the actual cost due to disruptions is \$5 million or more, but we assume that it is something more like \$100,000 and act accordingly, then our costs are more than 50% greater than we anticipated. The curves in the graph show that if we are not sure what the appropriate costs are, the least risky decision is to install capacity in fairly large chunks, eight or more steppers at a time. While we may not achieve the lowest cost solution, we are guaranteed reasonable performance as measured by cost.

It is also worthwhile to compare Intel's current equipment installation and qualification practices against the graph. Intel tends to install stepper capacity in fairly small chunks. This would indicate that Intel believes that the costs due to disruptions are very low - on the order of \$100,000 or less. To put the \$100,000 in perspective, \$100,000 is worth about four product wafers. Many people within Intel that I talked to felt like this was a very conservative estimate, and that the actual cost is probably much higher.

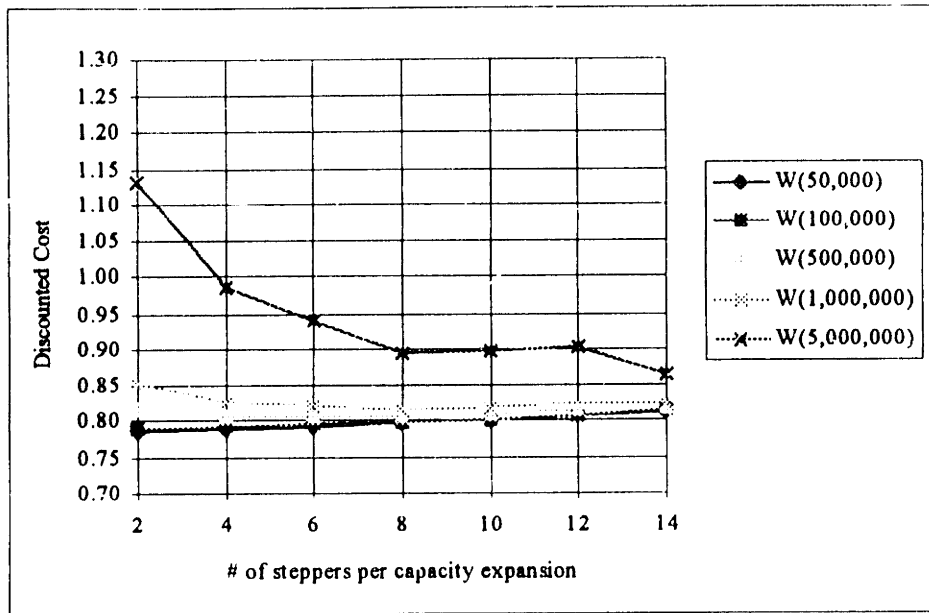


Figure 8.1. Sensitivity Analysis and Cost Comparison of Stepper Expansion Scenarios for Medium-Sized Fab.

Figure 8.2 illustrates the same sensitivity analysis for installing steppers in a larger fab. The larger fab requires 28 steppers at full buildout, compared to the 14 steppers required at the medium-sized fab. The model for the larger fab assumes the same linear growth rate as the model for the medium-sized fab. Since the growth in wafer starts is the same for both fabs, the larger fab's ramp period is much longer - over one year.

The sensitivity analysis for the larger fab has several interesting features. First of all, the discounted cost curve for the \$5,000,000 fixed cost value is very flat for a large percentage of x values. In fact, the actual minimum for the \$5,000,000 fixed cost curve occurs at 14 steppers per capacity expansion, only half of the final buildout value of 28 steppers. The minimum value for the \$5,000,000 fixed cost curve for the medium-sized fab in Figure 8.1 is also at 14 steppers per capacity expansion, but in the case of the medium-sized fab this represents the full buildout value. The cost curves for the large fab are very similar in shape and form to the cost curves for the medium-sized fab up to 14 steppers per capacity expansion.

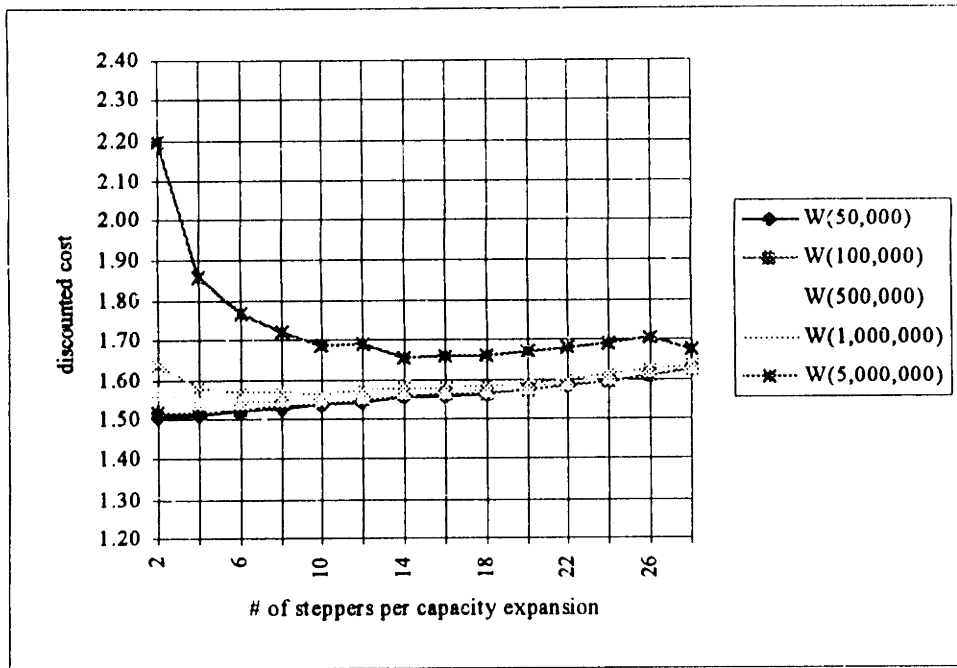


Figure 8.2. Sensitivity Analysis and Cost Comparison of Stepper Expansion Scenarios for a Large Fab.

Figure 8.3 shows sensitivity analysis results for an equipment set other than steppers. This equipment set is substantially cheaper than the steppers, in fact each stepper is more than one order of magnitude more expensive than each of these pieces of equipment. The purpose of this analysis is to demonstrate how the optimal capacity expansion value changes as the cost of the equipment changes. For this particular piece of equipment, ten tools are required at full buildout. As the graph demonstrates, for all values of fixed costs graphed, the optimal policy as measured by cost is to install all ten of the tools before the ramp begins. This result is expected. Intuitively, we expect that as the cost per unit of capacity of the equipment decreases, the fixed cost associated with installations should dominate. And if our intuition is not enough, the equations for capacity expansion should provide support. The approximate solutions for the optimal capacity expansion size and timing show an inverse relationship between the size of the capacity expansion and the cost per unit of capacity for the equipment under consideration. Thus as the cost per unit of capacity decreases and all other variables are held constant, the optimal values for both the size of the capacity expansion and the time interval between capacity expansions should increase.

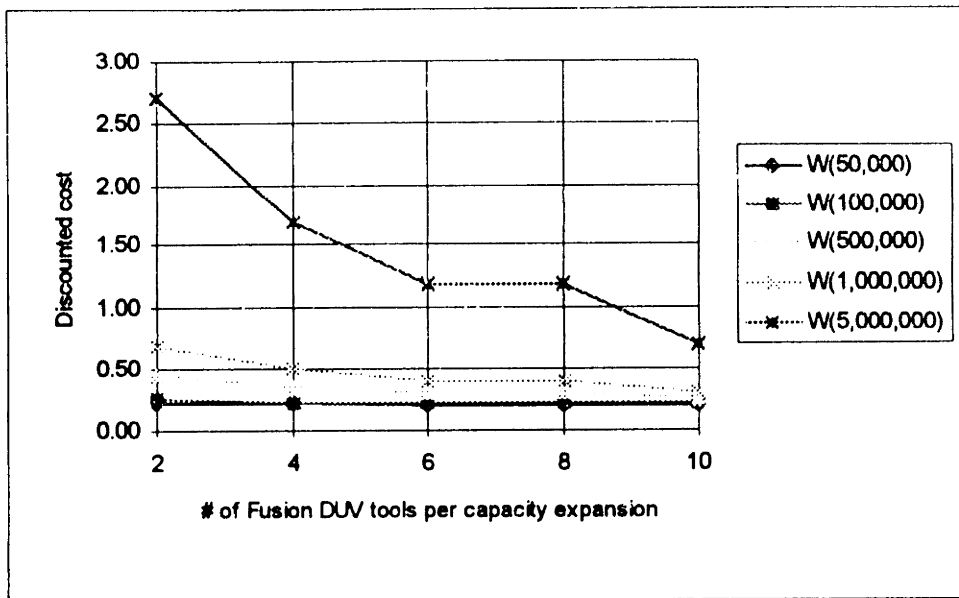


Figure 8.3 Sensitivity Analysis and Cost Comparison for an inexpensive process equipment tool at a Medium-Sized Fab.

The preceding figures showed sensitivity analysis results for two different pieces of equipment. These two pieces of equipment represented the two extremes with respect to cost per unit of capacity. Steppers are extremely expensive and usually many are required because each steppers offers marginal additional capacity. The other equipment set was much less expensive than the steppers and the resulting sensitivity analysis and optimal capacity expansion values reflected the difference in costs. The majority of the equipment in a fab falls somewhere in between these two extremes. The equations for the optimal values of capacity size (x) and the time between capacity additions (t) (Equations 8.1 and 8.2) can be used to get good estimates for x and t.

The sensitivity analysis results illustrated how the cost curves change when different size fabs are considered. While the least risky decisions involve installing large chunks of capacity before the ramp begins, there are practical limitations on the amount of equipment that can be installed simultaneously. These limitations are particularly relevant to larger fabs - for instance, it would be very difficult to install 28 steppers simultaneously. This suggests that the results may be more easily implemented at medium-size and smaller fabs.

Chapter 9. Factory Simulation Model for the 0.25 Micron Technology Ramp at Fab A

This section analyzes the output of the factory simulation model that was built to simulate the 0.25 micron technology ramp at Intel's Fab A. Not only was the simulation model built to answer 'what-if' scenarios and perform sensitivity analysis, but it also can test the effects of various ramp strategies developed during the previous phases of the research.

The simulation model was built to understand the effects of various equipment installation and qualification strategies on factory performance. Four strategies, or scenarios, were selected for testing using the simulation model. Each of the strategies uses a different model to determine when equipment should be added to the factory and how much should be added. The four strategies were:

1. Constant Constraint
2. Batch Capacity Expansion
3. Aggressive
4. Plan of Record (POR)

The Constant Constraint strategy attempts to establish one or two process equipment tool sets as the capacity constraint during the wafer starts ramp. This strategy is based on the published work of Eli Goldratt [11] [12] [13], notably his Theory of Constraints. Among other things, the Theory of Constraints suggests that a specific piece of equipment or a similar group of equipment should be identified as the constraint. After the constraint has been identified, all other equipment and resources should be subordinated to that constraint. Subordination means that product flow and inventory are arranged so that the constraint is never idle due to lack of material.

The first step was to identify the constraint for Fab A's upcoming 0.25 micron technology ramp. Based on capacity run rates, expected availability figures, and plans regarding tool dedication, several pieces of equipment were selected as constraints or near-constraints. The information concerning capacity run rates, availability, and tool dedication represented estimates for the time period of the wafer starts ramp. The constraints and near-constraints were selected by identifying the tool sets with the least amount of capacity at full buildout at the end of the wafer starts ramp. The qualification dates for the tools in these tool sets were then selected so that their capacity was just enough to meet the wafer starts requirements during the ramp. The remainder of the equipment was selected such that each had a sufficient amount of excess capacity during the ramp - at least 10% to 20% extra capacity.

The Batch Capacity Expansion Strategy takes a slightly different approach to the question of when equipment should be installed and how much should be installed. The equation and theory specifics can be found in Chapter 3.2 of this thesis. The general approach is very similar to a linear program, where the goal is to minimize cost subject to having enough capacity at all times during the ramp to meet wafer starts demand. The costs of interest include the discounted cost of the equipment and the cost associated with equipment installations and qualifications. The equipment costs are fairly easy to estimate; the difficult part of this strategy is deciding what cost

to use for the disruptions caused by equipment installations. For the purposes of this exercise, I chose to use \$100,000 as the cost of disruptions due to installations. I believe that \$100,000 is a conservative estimate; the actual cost is probably larger. The strategy is referred to as batch capacity expansion because the lowest cost solution to the problem generally requires that several pieces of similar equipment be installed and qualified at the same time.

The third strategy is the Aggressive Equipment Installation and Qualification approach, also referred to as the early IQ or pre-ramp IQ. In this strategy all the equipment is installed and qualified before the wafer starts ramp begins. In this scenario, no equipment installations take place during the wafer starts ramp. This strategy is very similar to the strategy that Intel's Fab C used during its 0.6 micron technology generation ramp.

The last strategy represents the plan of record (POR). This is the current plan as developed by Fab A's industrial engineering group. This strategy is a baseline against which the other three strategies can be compared. This strategy takes a just-in-time approach to equipment installation and qualification; however, in some cases the equipment installation dates are limited by the equipment suppliers and the equipment qualification dates are not scheduled until after they are actually needed. The other three strategies assume that the equipment installation and qualification dates are not limited by the suppliers' ability to build and deliver the equipment.

Based on these four strategies, an equipment installation and qualification schedule was developed for each strategy. However, several characteristics were shared by all four models. These characteristics were:

1. All models begin with the same start-up equipment set
2. All models run under the same conditions with the start-up equipment set for eight months
3. All models have the full equipment set installed seven months after the wafer starts ramp began
4. All models use the same input file for the wafer starts ramp
5. All models use the same input files for determining downtimes, repair times, and preventive maintenance schedules

All the models were run using essentially the same input files and conditions. All models begin and end with the same equipment set. All models have the same input file for determining when and how many lots are released into the factory. All models run the same product mix. The only true variable for the each of the simulation models is the number of pieces of each type of equipment that are available for running products over time during the ramp.

Figure 9.1 shows some of the output from the simulation models. This figure graphs the output for each month for each equipment installation and qualification strategy. Each strategy represents only one run of the model. Due to problems with the software and time constraints, I was not able to replicate the run for each strategy with a different random number seed. Therefore, each line represents the expected output from the factory by month for each strategy. The number of wafers released into the factory during each month is also graphed as the "input."

The graph has several interesting characteristics. When the equipment installation and qualification dates for the plan of record (POR) strategy were entered into the model, there were some known problems with the plan of record. The problems involved the fact that several pieces of equipment could not be delivered by the supplier until after they were needed to meet wafer starts demand in the later stages of the ramp. The scheduling tool that Fab A's industrial engineering group used had identified these tools as having negative float. In other words, the tool would not be ready until after it was needed. These negative float items were entered into the simulation model as they existed; we were interested to see if the simulation model would reflect these negative float items as losses in throughput. Indeed, the output curve for the Plan of Record (POR) showed a significant throughput shortfall in month six of the ramp compared to the other three strategies. This shortfall gave us, both myself and the rest of the organization, some confidence that the model was reasonably accurate since both the discrete-event simulator and the scheduler that was being used by Fab A showed problems with the plan of record.

Another interesting characteristic of the graph was that all four output curves initially showed close agreement with regard to throughput, but began to diverge significantly as the ramp proceeded. In fact, the output curve for the constraint management strategy showed significant instability in months eight, nine, and ten. In month ten the wafer output for the constraint management strategy overshot the input line.

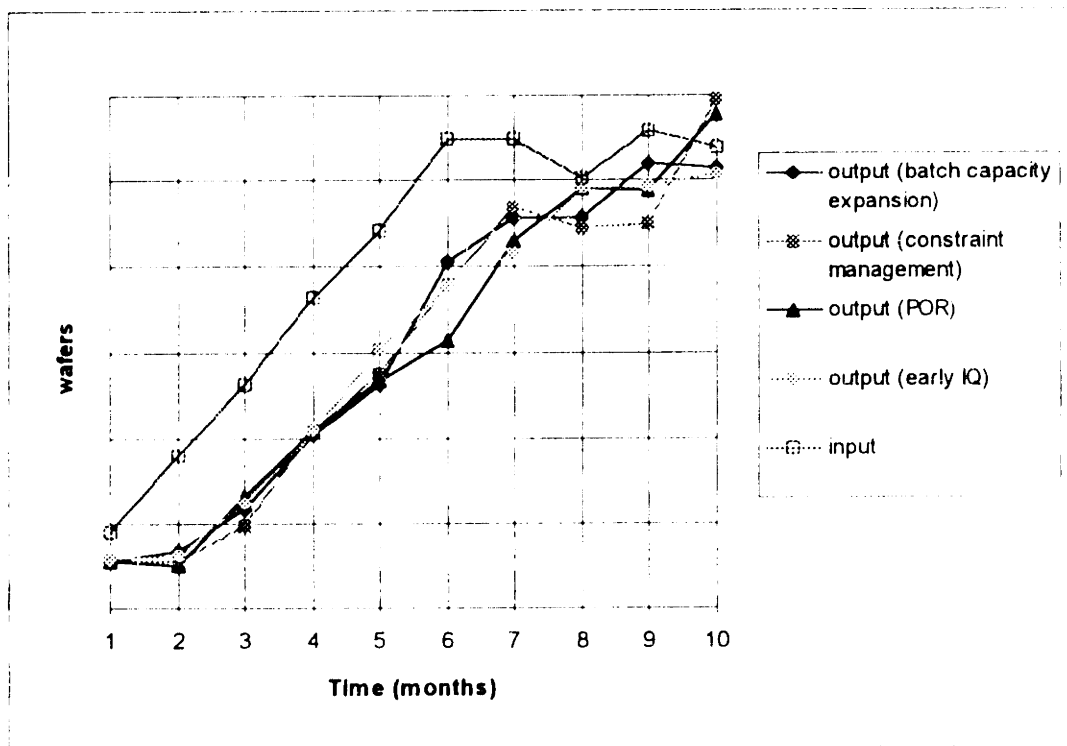


Figure 9.1 Factory Output Comparison of Equipment IQ Strategies.

Figure 9.2 shows a comparison between the four equipment installation and qualification strategies based on net present value (NPV). The plan of record strategy has been defined as the baseline strategy. The three other strategies are compared against the baseline strategy based on their NPV relative to the NPV of the baseline strategy. The assumptions that I made when calculating the net present values for the different strategies are as follows:

1. The discount rate is 15%.
2. Product wafers are worth approximately \$30,000
3. Product produced during the wafer starts ramp can not be sold until the process certification date
4. Equipment cost is incurred in the month in which the equipment is qualified
5. Line yield is 95%.
6. When calculating the schedule for the Batch Capacity Expansion strategy, I chose \$100,000 as the cost of a disruption.
7. Disruptions costs were not included in the NPV calculations for any of the strategies.

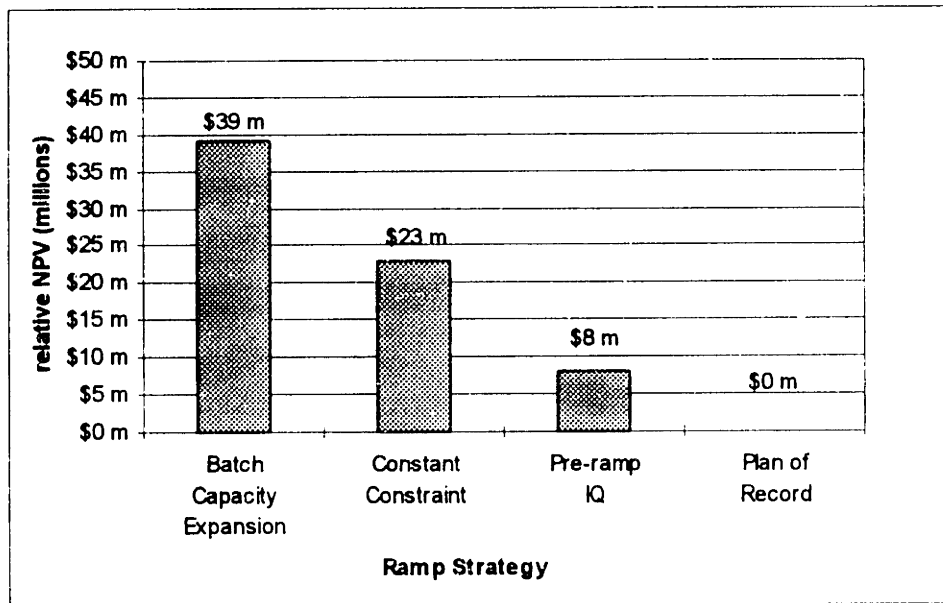


Figure 9.2 NPV Comparison of Equipment IQ Strategies.

Based on net present value (NPV), the bar graph in Figure 9.2 indicated that the best strategy was the Batch Capacity Expansion Strategy - using a \$100,000 disruption cost. The Batch Capacity Expansion Strategy has a NPV approximately \$40 million larger than the current plan of record, which was the baseline strategy. The Constant Constraint Strategy was second with an NPV approximately \$23 million greater than the NPV of the baseline strategy. The NPV associated with installing and qualifying all the equipment before the wafer starts ramp began was almost \$10 million larger than the NPV of the plan of record. It should be noted in defense of the plan of record that there were known shortfalls in the schedule that were being addressed. The

shortfalls led to unrecoverable throughput loss as noted in Figure 9.1. Once the shortfalls are addressed, it is reasonable to assume that the NPV of the plan of record schedule would be greater than the NPV of the Aggressive Equipment IQ strategy.

There is a lot to be learned from the NPV and output results of the four Equipment IQ strategies. Observations include:

1. Throughput is very valuable compared to inventory.
2. Unless the starts ramp is very aggressive, installing and qualifying all the process equipment before the wafer starts ramp is not cost-effective unless disruption costs are included in the NPV calculation.
3. Installing and qualifying large chunks of certain pieces of equipment before the wafer starts ramp can be cost-effective.

The importance of throughput is not a new concept. One of the tenets of Goldratt's Theory of Constraints is that throughput is the most important of the three types of business measurables: throughput, inventory, and operating expenses. In Intel's case, the value of throughput is even more pronounced due to Intel's monopoly in the X86 microprocessor market and the prices that it can charge for microprocessors during the early stages of the product's life cycle. The importance of throughput was emphasized by the output from the simulation. The plan of record showed a significant throughput shortfall and its net present value was the lowest of the four strategies.

Installing and qualifying all the equipment before the wafer starts ramp is very costly. Unless the equipment is utilized quickly because of an aggressive wafer starts ramp, this aggressive strategy does not make economic sense. The one reason why it might make economic sense, even if the wafer starts ramp is not aggressive, is if disruptions due to installations are included when calculating the costs of the equipment installation and qualification schedule. If disruption costs are included, then the cost of the aggressive ramp will not be affected since no installations occur during the wafer starts ramp; however, the costs of the other strategies will increase, making the aggressive equipment installation and qualification schedule more attractive.

The NPV of the Batch Capacity Expansion strategy suggests that it may make economic sense to install large numbers of certain pieces of equipment before the wafer starts ramp. The decision regarding which equipment to install before the wafer starts ramp is based on cost considerations. For each equipment set, the cost in dollars per unit of wafer start capacity is calculated. For example, let's say that a piece of equipment in Equipment Set A costs \$3,000,000 and the addition of each tool adds 500 wafers starts to Equipment Set A's capacity. The effective cost of this equipment is then be calculated by dividing 500 wafers starts into \$3,000,000, which comes out to \$6,000 per wafer start. This cost per unit of wafer start capacity is the B variable used in equations in previous chapters. Equipment sets with large values of B should be installed and qualified using something like a 'just-in-time' approach. All of the tools in equipment sets with a low value of B should be installed and qualified before the wafer starts ramp began.

The equipment installation and qualification schedule generated by the Batch Capacity Expansion strategy was very similar to the schedule generated by the Constant Constraint strategy. The similarity between the two schedules was not foreseen when the strategies were selected for evaluation. In fact, a careful consideration of the two strategies would lead one to predict a distinct difference between the two outcomes. The Constant Constraint Strategy and underlying Theory of Constraints focused on factory throughput and placed minimal emphasis on cost considerations. The Batch Capacity Expansion strategy explicitly focused on minimizing costs, with the requirement that some defined demand schedule must be met.

The purpose of this chapter was to compare and contrast several different equipment installation and qualification strategies. An equipment installation and qualification schedule was generated for each of the four strategies:

1. Constant Constraint
2. Batch Capacity Expansion
3. Aggressive
4. Plan of Record (POR)

A discrete-event simulator was then used to compare the four schedules. Output data from the discrete-event simulator was converted into discounted revenue cash flows. Discounted equipment costs for each of the four strategies was then subtracted from the revenues to calculate a net present value for each schedule. The schedule associated with the Batch Capacity Expansion Strategy had the highest expected net present value (NPV), followed by the Constant Constraint Strategy, the Aggressive Strategy, and the best guess at the Plan of Record (POR). The results of the analysis suggest that there is value to installing and qualifying large numbers of certain pieces of equipment before the wafer starts ramp begins. Those equipment sets that should be installed before the wafer starts ramp begin are those with a low cost per unit of wafer starts capacity.

Chapter 10. Trade-off Between Availability and Die Yield at the Capacity Constraint

This chapter presents the results of additional research involving the 0.25 micron technology wafer starts ramp at Fab A. The research focuses on the trade-off between availability and die yield at critical pieces of equipment. Critical pieces of equipment are those equipment sets that are both fab capacity constraints and die yield limiters.

This analysis was performed for several reasons. First of all, I witnessed some decisions that were made at Fab A early on during my internship that initiated my interest in the topic. One specific incident involved one of the capacity constraints in the 0.60 micron process technology at Fab A. This particular piece of equipment was not only a capacity constraint, but also a die yield limiter. Specific techniques and procedures were identified by the engineering group to improve the die yield contribution of this particular piece of equipment. However, these procedures required additional preventive maintenance time that would take time away from the equipments' ability to run production. The operations department was skeptical of any plans to further reduce the capacity of this equipment set, which was already the factory's capacity constraint. Given the expected reduction in production availability and capacity, what kind of die yield improvement should the operations department expect to make the additional preventive maintenance procedures worthwhile?

Additional evidence came from the contextual inquiry process. During the contextual inquiry, several of the people who were interviewed were very concerned about problems that arose when the factory capacity constraint was also a die yield limiter. These interviewees said that they felt that the capacity constraint should not be the yield limiter in future factories. Their contention was that the pressure to simultaneously increase both equipment availability and die yield at the same equipment set was very difficult to accomplish. They also felt that management's directions to improve both was unreasonable because of the often conflicting nature of the two metrics.

To help the management at Fab A make better decisions in the upcoming 0.25 micron process technology ramp, some simple trade-off curves were constructed. The trade-off is between availability at the constraint and die yield. Die yield and availability at the constraint equipment set are two very important factors that affect the number of good die out. Among other things, output, as measured by the number of good die, is proportional to both the utilization at the constraint and the die yield. The curves in Figure 6.1 are based on simple calculations that attempt to illustrate how different combinations of die yield and availability at the constraint can lead to the same die output. The calculations and the curves were based on estimates regarding equipment sets, process flows, and expected die yields for a planned Intel fab.

The curves in Figure 10.1 are break-even curves for various values of initial die yield (30% - 60%). The curves are intended to help management or engineers answer such questions as: How much of an improvement in the number of good die should I expect for every 1% decrease in availability at the capacity constraint? Or, how much of a reduction in availability is allowable given an increase in the expected number of good die?

The following assumptions were made when calculating the trade-off curves:

1. the current availability of the capacity constraint equipment set is 80%
2. the utilization of the equipment is always 5% less than the availability
3. an additional good die is worth \$350
4. the factory under consideration is a medium-sized fab (3000 - 5000 wafer starts per week)
5. the line yield is 95%
6. there are 200 die per wafer

The graph (Figure 10.1) shows some sensitivity analysis results. The curves represent different values for the current die yield. The current die yield is important because it determines the lost throughput in dollars for every wafer that can not be processed due to the lost capacity that results from lower availabilities. As the current die yield increases, larger and larger die yield improvements are needed to justify lower availabilities because the value of a wafer is now larger. The curves are not linear because there are finite limits on the number of die on one wafer and the equipment availability. As availability approached 0%, we would expect that the additional number of good die required would approach infinity.

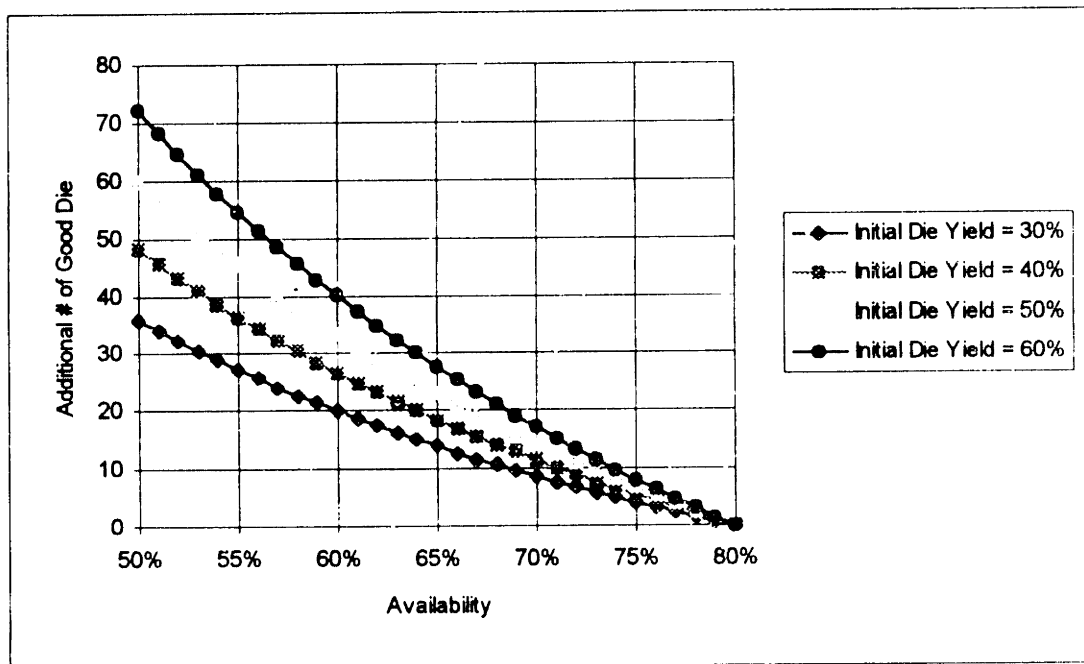


Figure 10.1. Trade-off Curve for Availability and Die Yield Improvements.

The graph shows that the same output, as measured by number of good die, can be achieved through many different combinations of good die, or die yield, and availability at the capacity constraint. While long-term it is desirable to improve both yields and equipment availability, the pressure of short-term production commitments may sometimes force management to make tough

decisions regarding current resource allocation. Figure 10.1 presents curves that help management make some of the tough decisions associated with the factory's capacity constraint.

Chapter 11. Results of Data Analysis

The analysis of start-ups at Intel fabs confirmed the importance of some of the factors identified in Chapter 4 as critical to the success of a semiconductor fab start-up. The analysis of data from the Fab A start-up and a comparison of start-ups at Fab A and Fab C provided insight into the fab start-up process.

Several of the factors from Chapter 4 were identified as important to the start-ups at Fab A or Fab C. The analysis provided evidence to support the importance of several of the factors. The results of the analysis are not conclusive proof; however, they do provide evidence that these factors might be important.

1. work force skill level
2. manufacturable production process
3. product
4. line yield
5. die yield
6. control systems
7. installed equipment base

Work force skill level and training issues were repeatedly mentioned as limiters to the fab ramp during the interviews with fab personnel. When the headcount staffing and training activity was compared to the wafer starts ramp at Intel's Fab A, the headcount staffing and training activity was completed more than one year after the end of the start-up at Fab A. Work force training issues were also mentioned as a limiter to the start-up at Fab C by the operations manager at Fab C. While the trend in semiconductor fabs is toward increased automation, the importance of the human element is still an important factor in determining the success of a fab start-up. Based on the analysis, the existence of a skilled, well-trained work force can improve the chance for success.

A manufacturable production process was mentioned several times during the interviews as a possible limiter to the start-up. It was impossible to verify this claim with quantitative data due to the lack of exact information about process capability. However, there were several process excursions during the 0.6 micron process technology at Fab A, and process excursions are symptoms of a low process capability. Process excursions slow the fab start-up because valuable engineering and management resources must be diverted from continuous improvement activities to firefighting activities. Process excursions did not severely impact the start-up at Fab C. Higher process capabilities lead to fewer process excursions and permit management to devote more time to continuous improvement activities such as yield improvement.

The lack of a microprocessor product design slowed the start-up at Fab A. Since the product design was behind schedule and not available, the fab had to continue to manufacture test chips. While Fab A was able to gain some yield and process learning by producing test chips, it would have been more effective to go down the learning curve by producing the actual product. It did not hinder the start-up at Fab C because the product had been designed and tested by the time of

the start-up at Fab C. The initial start-up of a new semiconductor process technology requires the redesign of an existing chip or the design of a completely new chip. Close coordination of the process and product development activities is critical. Given the relative costs of chip design and fab construction, the design process should never limit the fab start-up. Given the complexity of leading-edge VLSI integrated circuits, it is not unusual for the product design cycle to fall behind schedule and impact the start-up of a new semiconductor fab. In addition to the Intel example at Fab A, the start-up of Advanced Micro Devices' (AMD) Fab 25 in Austin, Texas was limited by delays in designing and debugging the K5.

Line yield and die yield issues are related to process capability because a low process capability usually leads to low yields. Line yield and die yield improvements require real understanding of the process. Line yields and die yields have a linear effect on good die output. Low line yields and low die yields lead to low good die output. Since the majority of the costs of a semiconductor fab are fixed costs, it costs as much to produce good die as it does to produce bad die. Line yield issues were mentioned several times during the interview process as ramp limiters. A graph of the line yield over time during the start-up at Fab A showed great variability and supported the belief at Fab A that line yield limited the ramp at Fab A. Die yield issues were more difficult to analyze; however, management at Fab A would have liked the die yields to be higher and higher die yield would have led to more die output.

Control systems provide a framework for controlling and managing information in the fab. It was difficult to quantify the effect of control systems on the start-up at Fab A or other Intel fabs. The immaturity of control systems was mentioned once during the interviews at Intel as a limiter to the capacity ramp at Fab A. Since control systems have an important role in coordinating and managing the complex activities associated with a semiconductor manufacturing system, it makes sense that a poor control system could limit the fab start-up or the rate of process or yield learning.

All of the equipment at Fab A was not installed until approximately six months after the start-up was completed. The management at Fab A expressed a desire for equipment installation and qualification activity to be the start-up constraint. Their desire was based on the substantial costs associated with the equipment relative to the costs of other activities. Since process equipment is required to process the wafers, the lack of sufficient capacity can severely limit the start-up by leading to output that is less than demand and increased inventories.

Chapter 12. Recommendations for Improving the Upcoming Ramp at Fab A

Based on the analysis of previous ramps at Intel fabs, including the 0.60 micron ramp at Fab A, five I have developed five recommendations for improving the upcoming 0.25 micron process technology generation ramp at Intel's Fab A.

1. Proactively staff and train
2. Install equipment with low cost per wafer starts capacity characteristics before the wafer starts ramp begins
3. Develop and implement manufacturing systems (such as equipment and process specifications, test wafer requirements, and information technology systems) before the wafer starts ramp begins
4. Devote additional resources to process characterization and understanding before the ramp in order to achieve high die yields sooner
5. Ensure that a manufacturable product exists

A common weakness of Intel production ramps was the lack of a trained and experienced workforce. Therefore, the first recommendation was to proactively staff and train. Instead of hiring in a just-in-time manner, training should be done using a just-in-time approach. The number of trained personnel (technicians, supervisors, and engineers) required at each stage in the ramp should be estimated and then plans should be laid out backward in time to ensure that the requirements are met. The plans should allow time for interviewing, hiring, and the appropriate amount of training. The plans should include milestones for when interviewing should begin, when hiring should happen, when all the appropriate training modules should begin and end and who is responsible for the training of the new hire. In addition, training demands from other fabs within Intel should be considered and taken into account when developing the plans. The planning process would require coordination between several departments at Intel's Fab A - operations, training, manufacturing engineering, and engineering - and the operations groups at other Intel fabs who may need to send people to Fab A for training.

Compared to most previous ramps at Intel, more equipment should be installed at Fab A for the 0.25 micron process technology generation before the wafer starts ramp begins. The equipment sets that should be considered for early installation are those with small cost to wafer starts capacity ratios. Using equations and arguments presented earlier, a planner could determine the optimal capacity expansion size and timing based on knowledge of equipment costs, growth in demand, the discount rate, and the fixed cost associated with the disruption of equipment installations. Based on analysis of the simulation output associated with each of the four equipment installation and qualification schedules, one attractive strategy would be to install the majority of the equipment (>80%) before the wafer starts ramp begins and then install the remainder of the equipment to maintain a constant capacity constraint during the ramp. One candidate for the capacity constraint would be the steppers.

Manufacturing systems are used to cope with the inherent complexity of managing a large, interdependent system - such as a factory. Systems provide standards for dealing with day-to-day issues and reduce the burden placed on management to deal with all the decisions that must be

made during the course of a day. My recommendation is that these manufacturing systems should be put in place well before the wafer starts ramp begins and then continuously improved. Developing and installing the manufacturing systems requires coordination between operations, engineering, manufacturing engineering, training, and yield and process integration teams. I did not attempt to include an exhaustive list of manufacturing systems in this paper; however, I will define manufacturing systems as any mechanism used by the factory to control and coordinate actions in a consistent and predictable manner.

Another important prerequisite for a successful fab ramp was a mature and stable process. A mature, stable process is defined as a process that is properly characterized, well-understood, and has process capabilities (Cpk's) significantly greater than one. A large fraction of the fab's resources, especially people, are devoted to firefighting yield excursions during the capacity ramp. These yield excursions distract from the fab's ability to focus on long-term improvements. These yield excursions could be minimized if the process was more mature when the ramp began. The decrease in yield excursions would free up resources which could be devoted to other activities such as continuous improvement of yields, increased equipment availability, and training.

A production ramp requires a product. While the existence of a certified product is not the direct responsibility of the operations and engineering groups at Fab A, they do play a role in working with the product design group to debug and improve the design. Their role is to manufacture product quickly, test the resulting product and devices, and report the results to the design group. Some of the devices are forwarded to the product design group for continued testing and analysis. Based on the results of the testing, the product design is revised and the design/production iteration continues.

Last, the management at Fab A must be careful not to optimize their own production ramp at the expense of the three other fabs that must also ramp the 0.25 micron process technology after Fab A. All of these fabs rely heavily on Fab A as a training site for their own personnel and for setting direction for the technology through the process development that it does. Fab A could justify not training personnel from other fabs because this training slows down the ramp at Fab A; however, not doing this training would slow the ramps at the other fabs and ultimately hurt the company's performance. The primary objective of all the organizations involved with the 0.25 micron technology generation ramp should be to maximize overall performance across the company.

Chapter 13. Extension of Results to Start-ups in Other Industries

Can the factors identified in Chapter 4 be applied to start-ups in other industries? Do the same principles that apply to a start-up at a semiconductor fab also apply to a start-up at a steel mini-mill or auto assembly plant? The purpose of this chapter is to address some of these issues and discuss the applicability of the results of this thesis to start-ups in other industries.

An analysis of start-ups in different industries reveals that some of the factors identified in Chapter 4 could be important to start-ups in other industries. To help understand which factors might be important to start-ups in other industries, the remainder of this chapter is devoted to looking at each of the factors identified in Chapter 4 in the context of auto assembly plant start-up.

Most of the recent greenfield auto assembly plant start-ups in the United States are Japanese transplants. An \$800 million greenfield plant in Georgetown, Kentucky [15] provided a source of information on start-ups of auto assembly plants. Construction of the plant began in early 1986. Volume production began about two years later in July 1988.

Toyota focused on several issues during the start-up. According to the case, "Developing human infrastructure was TMC's foremost priority in transplanting TPS [Toyota Production System] to Georgetown,..." [16] Toyota Motor Manufacturing initiated a hiring and training program while the plant was under construction. Large numbers of trained employees (several hundred) from a similar Japanese plant were sent to Georgetown as trainers. These and other actions showed that Toyota believed that work force size and skill level were critical to the assembly plant start-up.

The Toyota Production System (TPS) was also transplanted to the Georgetown plant. The Toyota Production System is an approach to organizing manufacturing operations. TPS involves aspects of process capability, yield, product design, and control systems. The core of TPS is continuous improvement and the elimination of waste in all activities. At the factory level, TPS is essentially a control system for coordinating material movement through the production system. TPS also provides principles and tools for reducing variability and improving process capabilities. The importance that Toyota attaches to TPS supports the importance of several of the factors discussed in Chapter 4 - product, control systems, and manufacturable production process.

While a certain amount of rework is often required in an auto assembly plant, yields are not major considerations in an auto assembly plant. As a result, die yield and line yield are not transferable to auto assembly plants; however, they may be critical to other industries, such as process industries. The typical strategy for increasing output in a semiconductor fab is to increase die yield; the typical strategy for increasing output in an auto assembly plant is to reduce the line cycle time.

Even though the majority of the equipment in an auto assembly plant is installed before the start-up begins, conclusions involving equipment installation are more difficult to draw because

while semiconductor fabs are set up as job shops, auto assembly plants represent transfer lines. Unless there are multiple, parallel transfer lines planned for the factory, the majority of equipment must be installed before the start-up.

Similar comparisons could be made to start-ups of steel mini-mills. Several of the aforementioned factors are also important to the success of mini-mill start-ups. While auto assembly plants were not concerned with yield issues, steel mini-mill start-ups would be affected by yields. To summarize, several of the factors that are critical to the success of a semiconductor fab start-up are also important in start-ups in other industries, such as autos and steel plants.

Chapter 14. Reflection on Process and Future Direction

This chapter discusses two related topics. First, it presents my thoughts on the strengths and weaknesses of my process for handling the problem. Secondly, I make recommendations for future work. I believe that it is appropriate to put these two topics in the same chapter because the evaluation of the problem-solving process leads naturally to the next steps to be taken.

The main weakness of the problem solving process was that I was not able to test my recommendations. True testing of my recommendations would require implementation in an actual wafer starts ramp at an Intel fab; unfortunately this was not possible during the seven month project. While I believe that some of my recommendations may be implemented in modified form in the upcoming 0.25 micron technology ramp at Fab A, the results of the ramp will not be known until the ramp is complete. The ramp will not be complete for another two years. Ideally, we would like to complete the Plan-Do-Check-Act (PDCA) cycle as quickly as possible; however, the long time horizon associated with such a large capital expansion project does not permit quick feedback.

The second weakness of the problem-solving process involved the use of data. In general, I was not able to use as much hard data as I would have liked. For example, most of my detailed data regarding previous fab start-ups came from one fab within Intel. Ideally, I would have liked to draw the data from a larger sample size of fab start-ups both within the company and from other semiconductor companies. Secondly, I was not able to run replicates of each of the four strategies using the discrete-event simulator. The output data series and net present value calculations associated with each of the equipment installation and qualification strategies represent only one run on the discrete-event simulator. Problems with changing the simulator's random number seed were the reason for the lack of replication. Lastly, there was no rigorous statistical treatment of the factors affecting the factory ramp-up. As I stated in a previous section, multiple regression analysis was not possible due to the nature of the available data.

A case could be made for a third weakness, that the problem was not well defined. However, I would disagree as to whether this was a weakness or not. In fact, I would argue that the vagueness of the initial problem statement made it possible for me to try new things and to go beyond conventional thinking. By providing me with a general framework and a wide latitude for discretion, the management at Fab A encouraged me to explore many different areas and topics. One of the results of this process was that I was able to discover, almost by accident, the simple capacity expansion model that I used to understand why fab management might choose to install all of the equipment before the wafer starts ramp began. Openness to new ideas also made it easier to convince people that a discrete-event simulation was the best tool for comparing the equipment installation strategies and schedules.

The remainder of this chapter is devoted to my recommendations for future work relating to this problem statement and some of my results. First of all, I believe that the simple simulation model that I built can be used to continue to test equipment installation and qualification schedules. As I demonstrated earlier, when schedules which had known problems were run on the model, lost throughput was evident. The simulation model can not only identify poor

schedules, but also quantify how good or bad a schedule is by estimating the lost throughput or additional cost. Of course, continued use of the model means that it must be maintained and updated as new information becomes available. In addition to maintaining the model, the manufacturing group should also devote resources to improving the model and how well it reflects reality.

Secondly, I would recommend that more resources be devoted to understanding the staffing and training requirements for a semiconductor factory start-up. The manufacturing engineering group at Fab A plans to use a simulation model to understand staffing requirements for the upcoming 0.25 micron process technology ramp. Hopefully, the demands of building an accurate simulation model will force the fab to carefully analyze and think about staffing demands and requirements. There are several other things that can be done which do not require the same technical complexity as building a good simulation model. First of all, the staffing and training problem was not treated in a holistic manner. The majority of the staffing model work was done by the manufacturing engineering group and an outside consultant. Other key contributors had minimal input into the design of the staffing plan. Representatives from the Operations, Training, and Engineering groups should be included in the staffing planning process for the upcoming ramp at Fab A. The presence of and input from internal customers, such as the Operations department, would result in up-front buy-in and more commitment, resulting in a better plan.

I would recommend that a more data-intensive approach be used to analyze fab start-ups at Intel. While I believe that my use of qualitative analysis techniques, such as open-ended interviews, added significant value to the analysis, I did not take a rigorous statistical approach to the problem. A more rigorous statistical approach would begin by carefully analyzing all the input and output variables and gathering data from many different fab start-ups within Intel to quantify the relationship between input and output variables. The data would be used to build a large regression model to understand which input variables are significant and seem to have the most effect on the important output variables. Critical to this approach is having access to a wide variety of data, such as cost, headcounts, and schedules, from many different fabs in the company.

Appendix A - Responses to the question: "What metrics did you watch most closely during the 0.6 micron technology generation ramp at Fab A?"

Leading Indicators

- o Head count/Open Requisitions filled/Technicians trained
- o Tools installed and qualified
- o In-line defects
- o Constraint output
- o Process Capability (Cpk's)
- o Number of moves/activities (number of process steps performed)
- o Utilization and Availability of process tools

Lagging Indicators

- o Wafer losses during processing
- o Wafers shipped to sort and test
- o Throughput time
- o Inventory
- o Number of lots on hold
- o Die yield

Appendix B - Responses to the question: “What were the problems with or weaknesses of the 0.6 micron technology generation ramp at Fab A?”

- o Not enough equipment expertise on all four production shifts
- o Not enough time allowed for proper training of technicians
- o Not all shifts managed work-in-process (WIP) the same way
- o Changes in plans not communicated to everyone in a timely manner
- o Lack of 24 hour support by engineering
- o Each department had its own goals
- o Processes not well understood or characterized

Appendix C - Responses to the question: “What were the major delays or critical limiters that prevented the factory from increasing output faster?”

- o Lack of equipment stability and process maturity
- o Lack of extensive knowledge reservoir on all four shifts
- o Manufacturing systems not in place before ramp
- o Excessive training requirements from other fabs
- o Reactive nature of head count ramp
- o Lack of certified product
- o Process excursions

Appendix D - Responses to the question: “What could be improved in the next ramp?”

- o Proactively staff and train
- o Use factory-wide goals to promote global optimization
- o Have the process technology as mature as possible before the wafer starts ramp
- o Communicate performance expectations clearly and in a timely manner
- o Ramp with WIP strategy in mind
- o Standardize WIP strategy across all shifts
- o Minimize distractions during ramp by having manufacturing systems in place, installing process equipment as early as possible, and having training complete as soon as reasonable

Endnotes

1. Benfer, Rich. "*Learning during Ramping: Policy Choices for Semiconductor Manufacturing Firms*", S.M. Thesis, MIT Sloan School of Management, 1993.
2. Bohn, "Measuring and Managing Technological Knowledge." *Sloan Management Review* (Fall 1994). pp. 61-73.
3. *Ibid.*, p. 64.
4. Schmenner, Roger W. "Every factory has a life cycle." *Harvard Business Review* (March-April 1983): pp. 121-129.
5. Pindyck, Robert S. and Daniel L. Rubinfeld. *Microeconomics*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1995.
6. Hayes, Robert H., Steven C. Wheelwright, Kim B. Clark. *Dynamic Manufacturing*. New York: The Free Press, 1988.
7. Freidenfelds, John. *Capacity Expansion: Analysis of Simple Models with Applications*. New York: Elsevier North Holland, 1981.
8. Brodie, Ivor and Julius J. Muray. *The Physics of Microfabrication*. New York: Plenum Press, 1982.
9. Shiba, Shoji, Alan Graham, and David Walden. *A New American TQM*. Portland, OR: Productivity Press, 1993.
10. Slater, Michael. "Challenging the Champ." *Upside* (August 1995): pp. 64 - 75.
11. Goldratt, Eliyahu M., and Jeff Cox. *The Goal*. Croton-on-Hudson, NY: North River Press, Inc., 1984.
12. Goldratt, Eliyahu M., and Bob Fox. *The Race*. Croton-on-Hudson, NY: North River Press, Inc., 1987.
13. Goldratt, Eliyahu M. *The Theory of Constraints*. Croton-on-Hudson, NY: North River Press, Inc., 1990.
14. Rodgers, T.J. "No Excuses Management." *Harvard Business Review* (July-August 1990): p. 84.
15. Mishina, Kazuhiro and Kazunori Takeda. "Toyota Motor Manufacturing, U.S.A. Inc." Harvard Business School Case #1-693-019.

16. Ibid., p. 3.