Increasing Experiment Velocity in a Production Environment

By
Shafali Rastogi

B.S. Electrical Engineering and Computer Science Minor, University of Calgary, 1995

Submitted to Sloan School of Management and the Department of Electrical Engineering and Computer Science in partial fulfillment of the degrees of

Master of Science in Management
And
Master of Science in Electrical Engineering and Computer Science

At the
Massachusetts Institute of Technology
May 1999
[June 1999]

Signature of Author _____
Department of Electrical Engineering and Computer Science
Sloan School of Management

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

Certified by _____
Professor Larry Wein, Thesis Advisor
Sloan School of Management

Accepted by _____
Professor Arthur C. Smith, Chairman
EECS Department Committee on Graduate Students

Accepted by _____
Lawrence S. Abeln, Director of Master's Program
Sloan School of Management

This Page is intentionally left blank.

Increasing Experiment Velocity in a Production Environment

By
Shafali Rastogi

Submitted to Sloan School of Management and the Department of Electrical Engineering and Computer Science in partial fulfillment of the degrees of

Master of Science in Management
And
Master of Science in Electrical Engineering and Computer Science

---

Abstract

Intel Corporation's D2 semiconductor fabrication facility (fab) is the smaller of Intel's two development facilities. The primary purpose of this facility is to bring new processes from early development to the point where they are ready to be transferred to a full scale fab. In conjunction with their charter, D2 also runs an aggressive production schedule. These two goals conflict with each other, in the sense that while production runs in large volumes with low cycle times, experiments run less often but bring more variation to the line and require more time to run. This thesis explores ways to improve the cycle times of technical development (TD) lots while minimizing the impact on production flow.

Two approaches are taken to looking at this challenge. The first is to improve TD setup and running procedures. To this end it is shown that the current experiment setup procedures at D2 are cumbersome, requiring much communication and signing for small changes. Where these gaps exist is highlighted. Focusing next on photolithography (litho), standards for running experiments for different processes and different litho tool sets are created, in order to ease the flow and reduce the confusion created by incoming experiments to the area.

The second approach is to improve the scheduling and prioritization of TD in the fab. The DUV toolset in litho, which runs the five critical layers for all processes, is the focus of this study. TD is characterized into three phases:

1. *Process Development*
   The earliest phase, this is where parameters are just being set.
2. *Process Characterization and Improvement*
   Here, the experimental aim is to ramp up the process to large volume manufacturing
3. *Sustainment Development*
   These lots are already in production, but experiments are being done to tweak the process.

With the three phases of TD and production as inputs, a simulation model was created to determine some rules for optimally running TD on the photolithography toolset.

The simulation study led to following conclusions:

1. Process Development TD is more complex than the other types of TD and production. It should be given more consideration than regular production.
2. Prioritizing TD is a key lever. TD should be given high priority. Deprioritizing TD will not have a large impact on production queue times, but will have a large impact on TD queue times.
3. There is a negative impact to holding lots to reduce set up or processing times.
4. Of the scenarios tried, semi-soft dedication running TD first gave the best results. The semi-soft dedication result is confirmed by a pilot conducted independently on these tools. There is an optimum where tools are most fully utilized while set up times are minimized, which further experimentation should lead to.

Thesis Advisors:
Larry Wein, Sloan School of Management
Stan Gershwin, Department of Mechanical Engineering

Acknowledgements

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

This work was made possible through the help and support of people at Intel's D2 site and LFM. I would like to thank my supervisor, Rob Edmondson, for helping me outline and solve the TD-Production issues presented in this thesis. I would like to thank Melinda Manente, who got me involved in the photolithography area, a central focus of my thesis. I would like to thank Laura Whitaker for taking the time to help me with the simulation and provide the necessary input data and Arena manual. This paper was greatly assisted by the people I worked with in Integration -especially Linda Erlich-, Automation, and the Photolithography group. Finally, I would like to thank the other Intel Santa Clara LFM's (SCILI's): Gary Tarpinian, Ajay Amar, and Manish Bhatia who helped me structure my project throughout the internship and patiently watched my simulation grow.

I would like to thank my two advisors: Larry Wein from the Sloan School of Management and Stan Gershwin form the Department of Mechanical Engineering, who helped me focus and improve the quality of this thesis.

This thesis is dedicated to my mom and dad, for the quarter century of hard work they put into getting me here and to my sister, who helped me solve my research problem and even provided the software and the manual.

This page is intentionally left blank.

Table of Contents

List of Figures

List of Tables

This page is intentionally left blank.

Chapter 1 Introduction

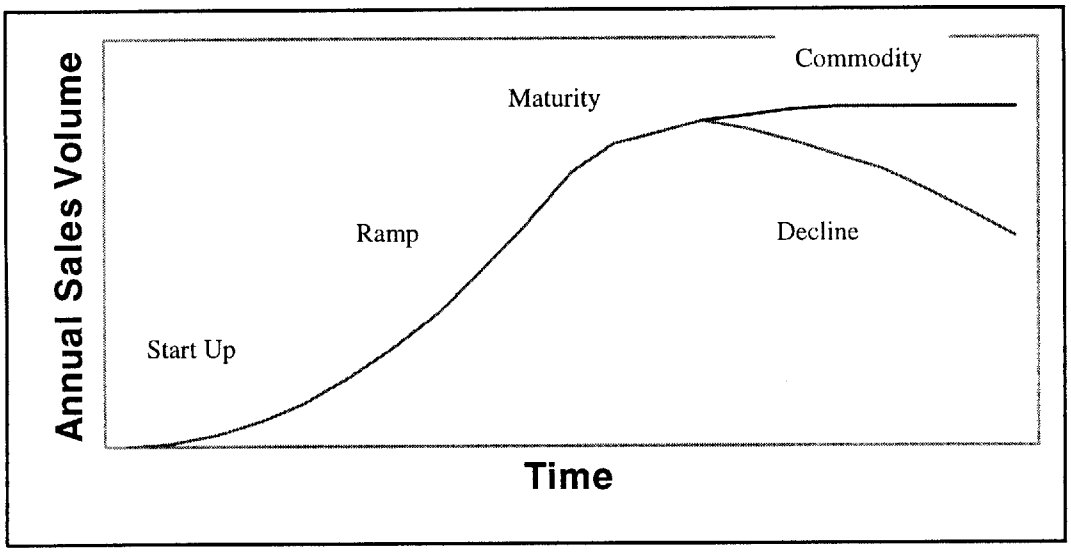
1.1 Overview

This thesis looks at how to reduce experiment cycle times in a joint production-development semiconductor fabrication facility (fab). It looks first at how to make more efficient the setting up of experiments. It then focuses solely on the Photolithography functional area[1], exploring first how to make the worker process more efficient, and then how to reduce cycle times for lots[2] going through the area by scheduling and prioritization changes.

1.2 Problem Statement

The high tech industry can be characterized as one of fast paced product development and exponentially decreasing margins on a given product. Indeed, product life cycles tend to follow the curve shown in Figure 1.

Figure 1: Product Life Cycle Curve



This curve definitely holds true for the semiconductor industry. One of the founders of this industry, Gordon Moore, coined Moore's law which states that approximately every 18 months, the speed of a microprocessor will double and its cost will halve. Furthermore, in almost any high tech development, the highest margins can be achieved in the beginning of the product's life cycle. The problem of short product life cycles are confounded in the semiconductor industry by the expense of R&D and the approximately $2B that must be spent every time a new fab is built. However, it can also take in the range of two years to develop a new semiconductor chip process. There occurs, then, a dilemma between wanting to recoup

[1] Photolithography (shortened to litho for the rest of this paper) is a process in which a mask is applied on a silicon wafer and a pattern is placed on the wafer using electromagnetic radiation such as ultraviolet light or X-rays.

[2] A lot is a set of wafers in a box. In general the lot travels through the fab in the box as a group, though sometimes the lot is temporarily split up. Wafers are round discs of single crystal silicon. Currently they are 8" in diameter with thicknesses in the range of 0.5 mm. Many chips are built on one wafer.

development expenses by ramping up as quickly as possible and maintaining maximum possible production while trying to go forward with development at maximum speed in order to stay at the front of the technology curve.

The purpose of this thesis is to look at the balance between doing development and production on the same line in a fab. It takes the slant of trying to maximize experiment velocity through the fab with minimum impact on production. Velocity, in this paper, refers to the speed at which experiments and production lots can make it through the fab. It is equivalent to the term WIPturns[3] used at Intel. When looking at organizational improvements, increasing experiment velocity means cutting down the time taken to set up and run the experiments. In the simulation model built as part of the research, increasing the velocity means minimizing queue times in front of the tools, which in turn reduces overall cycle times for those tools.

1.3 Research Description

This research was conducted at Intel Corporation's Santa Clara D2 fab. At the time of this project D2 was facing a number of challenges. In 1999 they anticipated an increase in production and development work coming into the facility. They were also in the midst of a massive cost reduction program, which in part meant a focus on reducing head count in the fab.

Through the 6 months time period the research was conducted, D2 produced Pentium II chips for sale, was transferring their first generation of flash memory from pure development to send to another fab for production, and began development work on their second generation of flash memory. This research uses 6 months worth of data from these three processes. It also treats the three processes as representative of the different phases of development and production in a typical fab. Historically, D2 had not expended a great deal of effort analyzing technical development (TD) work. This was changing a little, as the industrial engineering group looked for ways to better analyze the capacity needed to run TD. This research takes a stab at an in depth characterization of TD and its affects on one area, photolithography. Photolithography was focused on because at this time they were having the largest cycle time problems and they tend to be the designed bottleneck of the fab. At the time of this research, photolithography was just bringing on line seven new tools to run their five most critical layers on. The simulation model built to emulate these tools assumes that all seven tools are available.

Organizationally, D2 had long been set in how experiments were being set up and run in the fab. It was recognized, however, that the current methodology was too cumbersome and slow. In fact, one development deadline was missed in this time. There was a real drive to change the organization and procedures of setting up development to increase experimental set up speed, and to make smoother the running of experiments in the fab. In particular, the photolithography group was concerned about being able to cross train their workers on two different tool sets. They wanted standardized procedures for the running of experiments to help them achieve this goal. In part this goal was driven by the need to reduce head count in that area. The organizational research conducted in this paper set the groundwork for reducing experimental set up times and for helping photolithography smooth out the running of experiments in their area.

[3] WIPturns at Intel are a measure of how quickly the work in process moves through the fab.

Conducting this research meant working closely with different groups at D2. In particular, it was necessary to involve the Integration group that set up all the experiments in the fab; the Automation group that ran the automation software used in the fab; the photolithography technicians and engineers; and the industrial engineers who did the capacity, scheduling, and prioritization work for the fab. The customers for this work were the Technical Development (TD) Operations Manager, whose job it was to coordinate experiments in the fab with the rest of the operations people, and the Photolithography Area Manager, who was concerned about the running of experiments in photolithography. The inputs and outputs of the model were delivered to the Industrial Engineering group.

1.4 Thesis Objective

Specific objectives of the research were to:

1. Provide a framework for TD set up procedures as they existed in order to show where gaps were theat were leading to long experiment set up times.
2. Provide a basis on which to standardize photolithography experiment running procedures.
3. Characterize TD at D2.
4. Show through the use of a simulation model the cycle time impacts of various scenarios under which they can run TD leading to some general do's and don'ts for running TD.

1.5 Thesis Structure

This thesis is broken up as follows:
Chapter 2 gives an overview of Intel and the semiconductor fabrication industry. Chapter 3 moves on to focus on Intel's D2 site and the photolithography group at D2. With this background, Chapter 4 goes into details on the work that was done to flowchart how experiments are set up at D2. It highlights the delays that are caused as people chase signatures and paperwork sits in queues. Chapter 5 goes into details on the standardization of photolithography experiments. The rest of the thesis is dedicated to the simulation study conducted for the Deep Ultra Violet (DUV) tools in photolithography. Chapter 6 details the model. Chapter 7 discusses the scheduling policies that were run on the model. Chapter 8 talks about the model validation and Chapter 9 is a discussion of the results. Chapter 10 explains the conclusions reached from the model. Chapter 11 is a summary of what was learned overall through the research.

Chapter 2 Introduction to the Semiconductor Industry and Intel

Semiconductor technology began in 1875 when it was observed that selenium exhibited rectification and photoconductivity[4]. By 1935 rectifiers, photodetectors, and diodes were on the market. In 1947 the first transistor was invented at Bell Telephone Laboratories. By 1966 840 million transistors were sold a year. In the early 1960's, the integrated circuit was developed. As processes and equipment became more sophisticated, scientists were able to reduce the feature sizes on chips. They went from 10 mils in 1957 to 0.024 mils in 1990. Along with decreasing feature sizes, equipment costs rose. By 1990 equipment cost per machine ranged from $500,000 to $3,000,000[5]. Thus, increasing capital expenditures and increasing sophistication have been the main characteristics of the semiconductor industry. Since the late 1980's there has also been ever growing competition in the semiconductor industry. The Japanese entered the memory and device markets in the late 1980's, continuously bringing down prices. Japanese entry into the semiconductor industry led to the formation of Sematech in the early 1990's to help American firms compete with the Japanese. The last couple of years have shown growing competition even within the United States. Companies like AMD have been able to lower chip costs to the point where PC's are now available for less than $1000. Firms have found themselves either competing on cost or on advanced technology, and today they find themselves competing on both fronts. This is certainly true for Intel, which is for the first time in a situation of cost competition coupled with a need to push the technology curve.

Intel designs, develops, makes and markets advanced microcomputer components and related products. Its principal components are silicon-based semiconductors etched with complex patterns of transistors[6]. With a market capitalization of $191B, Intel is arguably the most powerful company in the United States.

Intel was founded in 1968 by Robert Noyce and Gordon Moore. Andy Grove, employee number 4, became their most famous CEO, writing such books as *Only the Paranoid Survive*. His successor, Craig Barrett, is the current CEO. Since developing their first microprocessor in 1971[7], Intel has worked to design, develop, and bring new logic and memory technologies to high volumes as quickly as possible; and to proliferate the Intel name. Their *copy exactly* methodology has given them this ability to rapidly ramp up production on new technologies, and their *Intel Inside* campaign made a chip that no one can see and few understand, a household name. In short, Intel is a business and a technical success.

Intel is composed of two technical development (TD) and a number of production semiconductor fabrication (fab) facilities. Processes, such as logic or memory chips, are made in what Intel calls its *virtual factory*. A virtual factory is a combination of production facilities and a development facility. The development facility does the work necessary to bring a new chip design to full scale production. The process is then handed off to the production facilities. At this point, any change made to the process must be approved by everyone in the virtual factory. This is what is known as the copy exactly method. The method has the advantage of consistency and speeding up ramp up between all the fabs, but the disadvantage of not allowing local optimization and slowing down change in the process. For example, once a process has been established, a fab cannot buy a different type of equipment without approval from everyone in the virtual factory.

---

[4] Rectification refers to the diode like characteristic of Selenium allowing it to either invert the negative or chop off the negative part of a sinusoidal wave. Photoconductivity means Selenium is a better conductor of electricity when exposed to light.

[5] Campbell, Stephan A.. *The Science and Engineering of Microelectronics Fabrication,* New York: Oxford University Press, 199, Chapter 1.

[6] http://quicken.excite.com/investments/snapshot/?symbol=INTC

[7] http://www.intel.com/intel/museum/25anniv/index.htm

A given fab is split up into seven functional areas. These areas are Diffusion/Implant, Etch, Litho, Planar & Yield, Thin Films, and Fab Support Group (FSG). All the areas except yield and FSG are grouped to correspond to one of the steps in the process flow of making a chip. The process steps are outlined in Table 1. Definitions for many of the technical terms are given in Appendix A.

Table 1: Process Flow Steps to Make a Chip

Process Step	*Functional Area*
1. Thermally grow starting oxide	Diffusion/Implant
2. Use Low Pressure Chemical Vapour Deposition (LPCVD) to deposit nitride	Diffusion/Implant
3. Pattern isolation mask (isolates n and p well transistors from each other)	Lihography - coat photoresist, expose to patterned reticle using UV light, and develop resist
4. Nitride Etch - transfers the pattern from the resist to the nitride	Etch
5. Mask the p well area, leaving the n well area exposed	Litho then Etch
6. Implant phosphorus ions into p-well	Diffusion/Implant
7. Grow another layer of field oxide and drive in Ph ions in diffusion furnace	Diffusion/Implant
8. Mask n-well, implant Boron in p well, drive in positives in diffusion furnace	Litho for mask, Etch, then Diffusion/Implant for ion implantation and drive-in of ions.
9. Preclean wafer surface with hydrogen flouride and grow a layer of oxide called gate oxide	Diffusion/Implant
10. Use LPCVD to depositoin a layer of polysilicon. This will be patterned into the transistor gates	Deposit in Diffusion/Implant, mask in litho, transfer patter in Etch
11. Grow a layer of side oxide to seal gate against possible contaminants	Diffusion/Implant
12. Implant a pattern of Boron to suppress runaway electrons	Boron Lightly Doped Drain (LDD) mask in litho then Etch, Implant Bron in Diffusion/Implant
13. Phosphorus LDD Mask and Implant	Litho, Etch, Diffusion/Implant
14. Deposit nitride to create spacers that separate the gate from the source and the drain	Diffusion, Litho, Etch
15. Grow a layer of oxide to protect wafer from damage	Diffusion/Implant
16. Heavily dope the n and p-channel source and drains. A channel is simply the well referred to earlier.	Mask in litho, transfer pattern in etch, implant in diffusion/implant
17. Anneal source and drain - meaning integrate dopant into Si lattice	High temperature diffusion
18. Add first metal interconnects between layers	Deposit in diffusion, Mask in litho, transfer pattern in etch
19. Deposit interlayer dielectric (ILD) to insulate between metal layers	Use CVD in diffusion, pattern openings in litho and etch
20. Add second metal interconnects	Deposit in diffusion, Mask in litho, transfer pattern in etch
21. Add a passivation layer to protect chip from moisture or cracking	Diffusion

22. Add bond pads, which is metal used as conduction paths for the tests	Litho Etch
23. Add polyimide for insulation	Diffusion, Litho, Etch
24. Backgrind wafer to remove built up layers and contaminants	Planar
25. Add gold to back of wafer to provide and ohmic contact	Sputter on in Thin Films
26. Test wafer	Yield

The key point to notice in this complex process flow is that it is cyclic. Layers of chemicals are added on and patterned by running repeatedly through diffusion/implant, litho, and etch.

Of the functional areas, Intel tries to keep photolithography the bottleneck if possible. This is because the equipment in this area is the most expensive and technically complicated. Table 2 breaks down the steps that happen in photolithography:

Table 2: Basic Steps in Photolithography

Step	*Description*
1. Spin coat photoresist	Spin a layer of an organic photosensitive chemical
2. Expose the wafer with resist on it to UV light, shining through a mask	This transfers the pattern from the mask onto the resist
3. Develop the photoresist	The resist that was exposed to UV light has been altered. You can now wash away the altered photoresist.

Intel breaks up its semiconductor work into what are called *processes* and *products*. A process is the series of steps that a given type of semiconductor, whether it be a logic or a memory chip, must go through to come out as say a Pentium II processor. Within a process there can be multiple products (also called dot processes since a process is given a number like 856, and then a product within that process is given a '.number ' like 856.7), which go through the same steps, but have some parameters that are unique to just that product. The series of steps that are performed in a process are called a *route*.

Chapter 3 Intel's D2 Site

Intel's D2 site is the smaller of Intel's two development facilities. The two sites split up the product development for the various processes that Intel is developing. D2's charter is to ramp a process from design to full scale manufacturing; or as one engineer put it, from 'wiggle' to high volume. The experimentation necessary to bring up a process and ramp it to large volume is called technical development (TD). TD is one of Intel's technical competencies. TD is a silicon based competence - it is a determination and tweaking of the physical steps necessary to put multiple layers of masks on a chip[8]. At the time of my internship, D2 was running 0.25 micron technology[9]. They were running one Pentium II logic processor and two generations of flash memory.

There are many groups within Intel responsible for TD. Engineers from one of the functional groups design experiments. The Integration group takes the experiments and decides which experiments will be run, how many wafers an experiment will receive, and then makes sure that the proper paper work and communication has been done to allow the experiment to run smoothly. The automation group makes sure that the Workstream automation system they use in the fab has been appropriately updated. D2 limits the number of experimental wafers allowed in the fab at any one time, which means choosing the experiments is important. The running of experiments is a complicated procedure because there are really two goals in running TD:

1. To collect information from running experiments.
2. To not bring down equipment on the line or delay the production process.

The second goal is mainly due to the fact that D2 runs production and development on the same lines. They can do this, because the one logic and two memory processes they are running were designed to be able to use the same equipment. The fab runs production for a number of reasons, including:

1. To most closely simulate a production environment for development. They want to do this because the goal of the experiments run at D2 is to bring the new process into a production environment.
2. To keep the factory heavily loaded. Again, this is to help simulate a real production facility and to keep people challenged to behave as though it is a production environment.
3. To remain competitive with the other fabs. Fabs at Intel are rated by their cost per wafer, their wafer starts, and their die and line yields[10]. Being a smaller fab, D2 has struggled to keep their cost per wafer down. Increasing wafer starts is one way to do this.

The relationship between TD and production is of paramount importance at D2. Despite a charter focused on TD, TD is not the primary driver at D2. Many people at D2 focus on production as the main importance of the fab. In fact, last year, 75% of the people in the fab did production work. This thesis focuses on TD as the primary job of the fab, but the relationship between production and TD plays a key role in the analysis presented. Figure 2 is a D2 TD System Overview drawn up by the TD Operations Manager responsible for managing the interface between TD and the factory operations. Following is an explanation of some of the issues he outlined as facing TD, and then a narrowing in on the thesis focus.

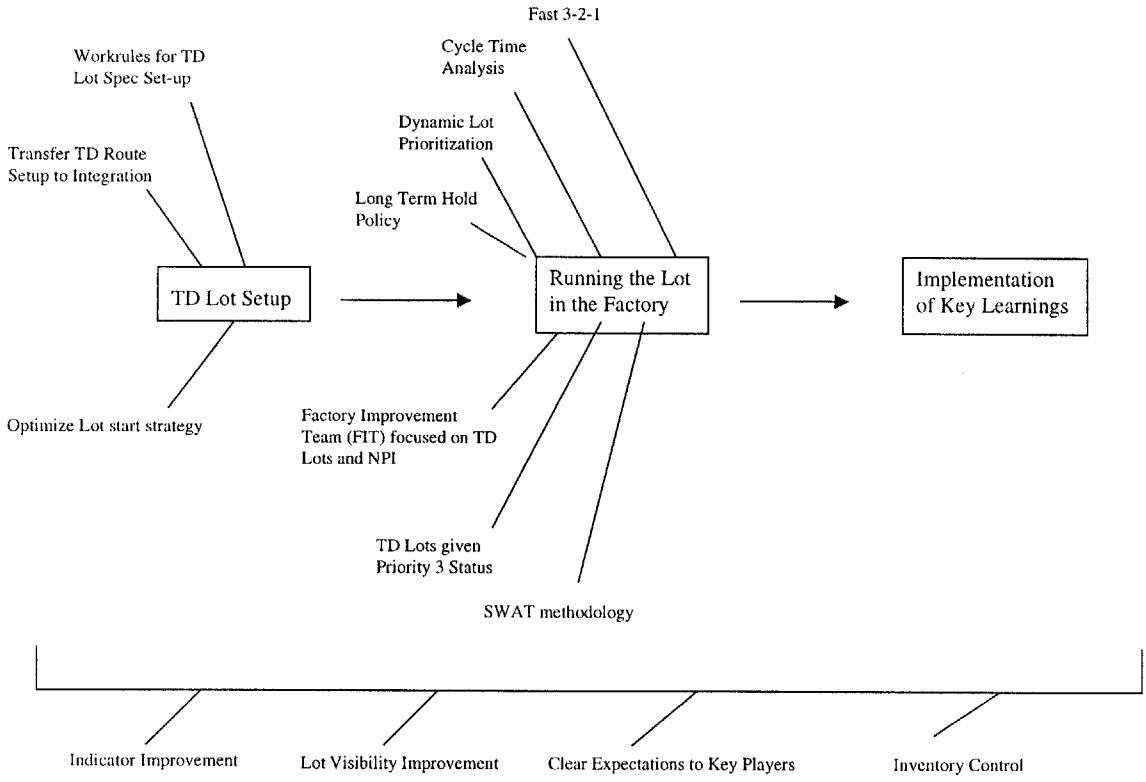
---

[8] Burgelman, Robert A.. *Fading Memories: A Process Theory of Strategic Business Exit in Dynamic Environments.* Stanford University: Administrative Science Quarterly, p. 32, 1994.

[9] 0.25 micron refers to the minimum width of a line that can be put on a chip. The semiconductor industry is continuously working to drive this down.

[10] Line yield is a measure of how many wafers per lot make it through the process. Die yield is a measure of the number of good chips on a wafer.

Figure 2: TD System Overview



3.1 TD Lot Set Up

This is the front end of experimentation. Issues at this point include:

Workrules for TD Lot Spec Set-Up: Each lot has a lot spec attached to it. This lot spec details the route that the lot will be travelling. For TD the lot spec may have steps that are modified, added or deleted. The work rules are necessary to prevent confusion on the floor caused by the changes in the normal or *Plan of Record* (POR) lot spec.

Transfer TD Route Setup to Integration: It is the job of Integration to set up the routes that TD lots follow and to make sure that the fab is able to handle the changes. This is a complex process and D2 still struggles with misprocessing and lots being put on hold because they cannot be processed by the technicians.

Optimize Lot Start Strategy: D2 has a maximum limit on how many TD wafers can be in the fab at one time. Integration and Production Control must figure out which experiments they will allow to enter the fab so as not to exceed this limit.

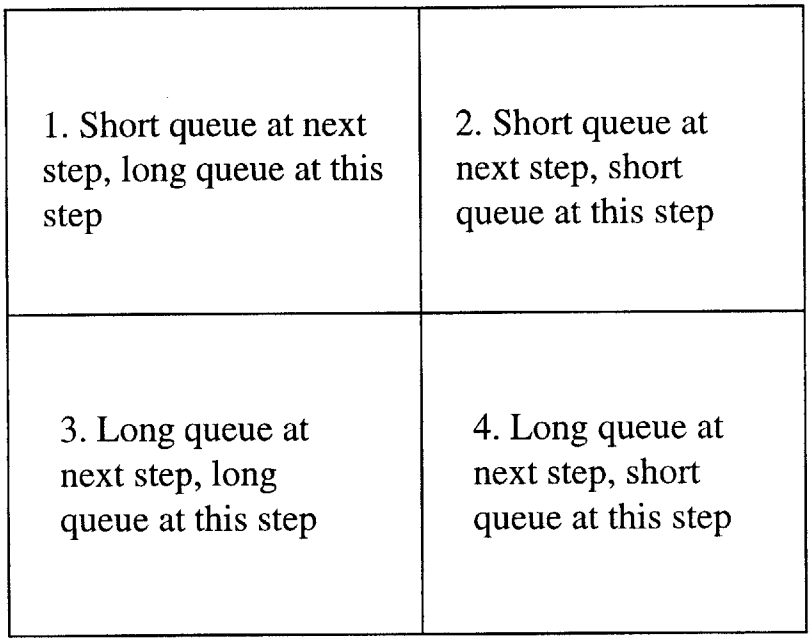
3.2 Running the Lot in the Factory

When TD lots are in the fab, they are tracked as they make there way through the process. Some important considerations in running the lots are:

Fast 3-2-1: Fast 3-2-1 is the generic method used at Intel to decide which type of lot to look at next. Each lot is put into a quadrant as illustrated in Figure 3. The highest priority is given to Quadrant 1,

then 2, 3, and lastly 4. Based on the queue at the next step and the queue at the current step, a lot is prioritized as to when it can run. While this method works quite well, it is also being evaluated to make sure that it is the best method.

Figure 3: FAST 3-2-1 System



Cycle Time Analysis: The Industrial Engineering group at Intel is responsible for doing cycle time analyses. They look at what the cycle time should be based on tool availability and cycle time goals. Last year cycle time goaling went through a major revamp to better meet factory needs. This group continues to look at cycle time goaling algorithms and how to figure out cycle times. This is especially challenging for TD work, which tends to have a lot more uncertainty as to what will be required when it hits the floor.

Dynamic Lot Prioritization: Lots are prioritized according to Table 3.

Table 3: Lot Prioritization

Priority	*Description*
1	Will hold a tool open for this lot. Highest WIPturn goal
2	Second highest WIPturn goal
3	Most TD
4	Monitor wafers
6	Production wafers. Lowest WIPturn goal

This prioritization is periodically revisited especially for TD, where they would like to lower the prioritization. Lot prioritization is linked to the fab's main measured called *WIPturns*. WIPturns are the average number of activities[11] a wafer goes through per day. The higher the

[11] An activity is an irreversible step a wafer goes through. For example, after a wafer has been etched it is difficult or impossible to undo the work. On the other hand, just patterning a mask on photoresist in litho can be undone fairly easily.

WIPturn goal the greater the priority of the wafer. This priority supercedes FAST 3-2-1 priority.

Long Term Hold Policy: This is an interesting problem at Intel. Lots that are put on long term hold are taken off the wafer count for TD wafers. Thus, engineers can increase the number of experiments they have running in the fab by putting some lots on hold. These lots are then outside the system and at some level defeat all the work that is put into controlling TD lots. This is because putting lots on long term hold complicates the measuring of the real cycle times for the wafers and the planning of experiments. For example, when a lot has been put on long term hold, should they stop the clock on that lot as far as cycle time goes? Does that lot still count towards the experimental wafer limit at D2? At the moment it does not count towards the cycle time of the lot and it does not count towards the limit, but whether this is proper is still being considered. Another issue is that engineers can take lots off of long term hold and insert them in the system. This disrupts the scheduling efforts and the smooth experimental set up system that Integration is trying to maintain. In a sense, the long term hold policy is a loop hole in the experimental set up system that, though it may be necessary, disrupts the systematic running of experiments.

TD lots given P3 status: D2 is still experimenting with how to weight priorities. In 1998 a Priority 3 lot could not sit in queue for more than six hours. If this limit was encroaching, the lot had to be moved to the front of the queue. They have been experimenting with increasing that wait to 12 hours.

3.3 Ongoing Issues

Indicator Improvement: Are they using the right metrics to measure success of TD?

Lot Visibility Improvement: TD lots continued to be misprocessed or put on hold. How can this be prevented?

Clear expectations to key players: The entire TD process takes cooperation and communication between many different groups, including engineering, the functional areas, and integration. Improving the process so that it is more efficient and so that mistakes do not happen is an ongoing effort.

Inventory control: How do they deal with long term holds and what is the best use of the wafer limitation?

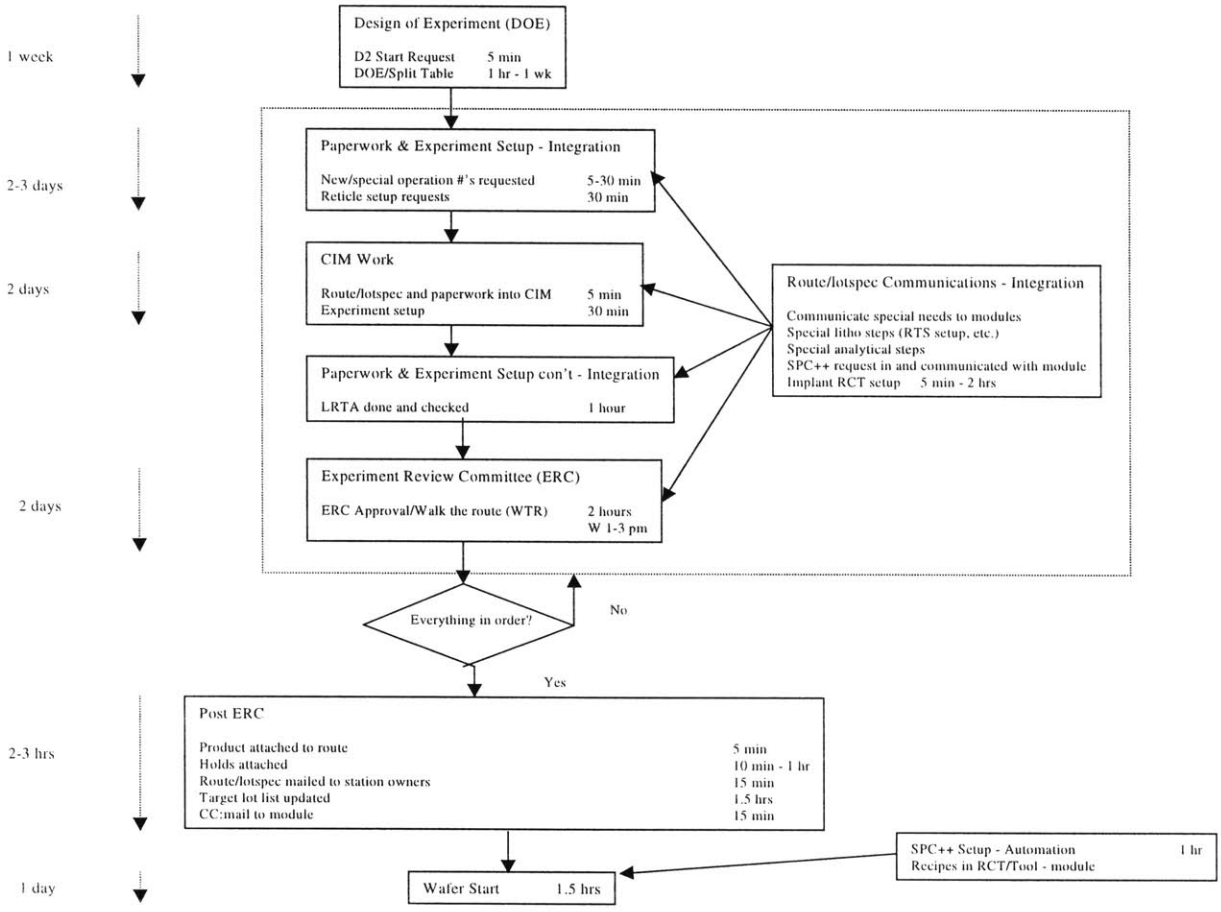
Chapter 4 The Setting up of Experiments

The first part of the thesis project to help D2 reduce experiment cycle times was to look at the setting up of an experiment. This is what is called the front end of an experiment. At D2, one of the primary goals in the setting up of an experiment is that it does not negatively impact production[12]. This could happen, if for example, a measurement device shut down a tool because it mistook a change caused by an experiment for a tool error. Another part of the effort of setting up an experiment is to make sure that the experiment is properly integrated into the fab system and that the experiment runs smoothly while it is in the fab. Now, it takes about 8 weeks to make a chip. The setting up of an experiment only takes two days to a week. Even though this is just a small percentage of the overall cycle time for a wafer to leave the fab, it is a completely person made and beaurocratic addition to the cycle time. That is, while all the paperwork is being completed to get the experiment into the fab, the actual experiment is not being advanced at all. Also, every time an engineer wants to make a change to an experiment, the entire set up procedure is repeated. Thus, for a given experiment, the setup procedures may have to be repeated multiple times.

It is understood that a disciplined set up procedure is necessary at D2 to prevent unnecessary confusion or machine failures in the fab. The question was, given the set up procedure is as it is, what causes the delays? To answer this, the procedure to set up an experiment was flowcharted and comparison made between how long the actual steps took to the 'window' people wanted or needed to do the job. The window could be required because of queues, because people are multitasking, or because they are waiting for someone else to finish their part before they can go. The flowchart is shown in Figure 4.

[12] This goal is part of an overriding issue at D2, which is that of conflicting missions. As a development facility, D2 should emphasize TD, but instead, it is put more in the role of a necessary activity that should not impact the more important goal of producing sellable chips. There is no easy resolution to this conflict, nor is D2 planning to change this goal.

Figure 4: Setting Up a TD Experiment for the D2 Fab



Appendix B explains in detail what the different steps are. Most of the technical and Intel specific jargon is defined in Appendix A. The main driver of this process is Integration. The experiments are sent to them, they start the paperwork, and submit the package to ERC for approval.

The key learning from this study was that paperwork spent a great deal of time waiting for signatures and in queues. Consider these parts of the process from the perspective of Integration:

- Communicate special operations to litho
- Get signatures from litho for special steps
- Talk to module engineers about special analytical steps
- Signatures required for new lot specs can require signatures from entire virtual factory
- Litho must return the RTS table
- LRTA sits in queue at CIM group waiting to be put into Workstream
- If ERC approval is not received, certain steps must be redone

Every time a step goes to another group, the progress of the lot is taken out of the hands of Integration, yet they are considered responsible for getting experiments into the fab. This creates a great deal of stress for this group. They spend a lot of their time chasing and hounding people, that they could spend working on

setting up other experiments. At the same time, the process is very paper and signature intensive. Since there are many experiments being set up or changed every week, people build up queues.

Using this flowchart, one finds that the main box to attack is the large dotted box. It is the CIM queue and the communication done by integration that is adding all the extra time. Things D2 is looking at changing to improve this are:

- Eliminate unnecessary signatures! The paperwork and corresponding signature work that currently exists in the system is the result of years of patchwork solutions to problems that occurred in the fab. They aim to spread the blame for possible failures, when it would be more efficient, and according to people who work in the system, more desirable, to give people full responsibility for their part of the process.
- Move to a more web based system. This will speed up the paperwork process.

To really improve the system, D2 needs to look at each part of the procedure in detail, and ask the question: Why are we doing this? What would happen if we eliminate this step or this signature? They should try a green field approach, where Integration, Automation, and the TD operations people figure out what they really need to ensure that experiments run smoothly. One of the key problems with this system is that it is really complex and cumbersome. If it can be simplified, the next step would be to aim towards getting engineers inputting more of their own experiments. If this can be done, the advantages will be:

- The process will speed up
- It will prevent engineers from trying to go around the system, which will help from a system accounting point of view
- The Integration group will not be under as much stress
- Automation can shift more focus on improving systems than entering data

Chapter 5 Standardization of Photolithography Experiments

In the next stage of my analysis of cycle time reduction in experiments at D2, I focussed on photolithography. As Campbell explains in *The Science and Engineering of Microelectronic Fabrication*,

> *Lithography is the most complicated, expensive, and critical process in mainstream microelectronic fabrication...Lithography accounts for nearly one third of the total fabrication cost, a percentage that is rising.*[13]

As such, Intel is conscientious to try and make photolithography the bottleneck of the fab. The purpose of this project was, again, to increase experiment velocity by improving the TD process. Photolithography was a natural choice to focus on in looking experimentation in the fab because:

- It is the designed fab bottleneck
- There are 20 layers on a given chip that must all go through photolithography. This reentrant flow means that any improvement is very important in the overall process.

This project, done with the photolithography group, was one that had been of concern for months. The project was to standardize the format of experiments that entered the fab. Litho wanted to do this for a few reasons:

1. There was a real drive to make experiments look like production in the fab. The thinking was that if one can make experiments almost invisible to the technicians on the floor, they will run more smoothly.
2. Litho had two different toolsets and three different processes running through it. The way experiments were coming into the area across the different tool sets and the different processes was not consistent. This meant confusion on the floor that led to holds and misprocessing.
3. As part of their cost reduction effort, they wanted to cross train technicians across the two tool sets.

The most challenging part of creating the standards was achieving consensus among all the involved parties. This included Integration, Automation, and litho engineers for both tool sets. Integration was further divided because a different group set up the flash processes and a different group did logic. They were also split by who did the front end and who did the back end of the processes[14].

The final outcome of the standardization work is shown in Figure 5.

[13] Campbell, p. 152.

[14] The front end of a process are all the steps up to Metal 1 shown in Table 1. The back end is the metal interconnects onwards.

Figure 5: Standardized Format for Photolithography Lots

PROCESS	Standard	Non-standard	*Short term Special processing*	*Long term Special Processing*	Trouble Shooting Special Processing	Owner
EXAMPLES	POR	Test wafers, Blind Steps, Production builds	Splits, processing of one or two lots	Pilots, TD baseline	excursions, special measurements	
REQUIREMENTS						
Operation number	4 numbers	4 numbers	2 or 3 numbers	2 or 3 numbers	POR number	Integration/ Automation
Operation Name	func area+F(if flash)+*/-+3(if 803)+layer + desc (max 10 chars)	func area+F(if flash)+*/-+3(if 803)+layer + desc (max 10 chars)	func area+F(if flash)+*/-+3(if 803)+layer + desc (max 10 chars)	func area+F(If flash)+*/-+3(if 803)+layer + desc (max 10 chars)	func area+F(if flash)+*/-+3(if 803)+layer + desc	Integration/ Operations
Operation Description	e.g. Flash Litho S/E/D FL6 803	e.g. Flash Litho S/E/D FL6 803	e.g. Flash Litho S/E/D FL6 804	e.g. Flash Litho S/E/D FL6 805	e.g. Flash Litho S/E/D FL6 804	Integration/ Operations
Lot Spec	yes	yes	yes	no	no	Integration
Lot attrib (SPR) flag	No	No	Yes	yes	Yes	Integration
Specs	Layer spec	Layer spec	Layer Spec	Layer Spec	Layer Spec	Integration
IUSC Script attach	No	No	No	No	No	Automation
SED Move in message	No	No	yes - should stop lot until acknowledged	yes	yes - should stop lot until acknowledged	Integration/Engineer
Metrology Move-in message	No	No	Yes. Should include: 1. Special or additional measurements 2. Need F5/F4 intro 3. Recipe name 4. Dispo criteria 5. # of sites for recipe	Yes. Should include: 1. Special or additional measurements 2. Need F5/F4 intro 3. Recipe name 4. Dispo criteria 5. # of sites for recipe	Yes. Should include: 1. Special or additional measurements 2. Need F5/F4 intro 3. Recipe name 4. Dispo criteria 5. # of sites for	Integration/ Engineer
RCT Setup	yes	yes	yes - include expiration date	yes	yes - include expiration date	Module engineer
RTS Setup	yes	yes	yes	yes	yes	Integration/ litho
RCS: New Stepper Jobfile (micrascan)	yes	yes	No - make edits directly to RCT	Make changes in special files (eg. 2x802eng) on RCS. RCS files must be kept current - include expiration date.	No - make edits directly to RCT	Module engineer

One can see from the table that these rules define the format for production lots, which can be standard or non-standard; experimental lots, which can be short term or long term; and trouble shooting lots. Short term experiments are ones where only one or two lots are used in the experiment. The idea is to modify the system as little as possible for these types of experiments. The long term experiments are lots that are going to be running through on something like a pilot. The idea is to incorporate these lots into the system as much as possible. For each type of lot, there were some basic standards of formatting, like how long operation numbers should be and naming conventions. There were also some changes made that were meant to simplify the system. For one thing, it was decided to eliminate scripts. As is explained in the Script section of Appendix B, scripts are attached to lots that cannot be run automatically by the station controller. They are a headache for automation and do not add value. Since all lots were moving towards being run by the station controller, they were eliminated. It was also decided to formally use move-in messages to tell the technicians what to do with the experiment. A move-in message looks like:

> *Please update NSR parameters. Any questions, please page J. Doe (123-3456).*
> *$EXPMTD=TEST-2$EXPTIM=500$EXPTMS=20$FCSOFT=-0.3$FCSOFS=-0.2*

It tells the technician what the parameters for the experiment are and who to contact if they have problem. The second line, with the parameters, is called a recipe string, and can be cut and pasted to the proper place in the tool recipe and then run. There are two stages in the running of experiments where move-in messages are needed. First is at Spin Expose Develop (SED), where the mask is set and the wafer exposed. The second is in metrology, where CD and Registration measurements are taken.

A final major change was that it was decided that SPR (Special processing) flags must be set. In the automation system on the floor, setting this flag makes running the experiments easier for the technician. Currently, the setting of the flag is at the discretion of the person on who set up the experiment.

Creating these standards did the following:

1. By using standard operation numbers and names, the technician could easily recognize an experiment.
2. By using move-in messages, setting the SPR flag, and eliminating the scripts; all experiments would be run on the floor in the same manner and technicians would know where to go to find the right information.

It took two months to create these standards, but after they were done, the story did not end there. An important issue raised by this exercise was yet another attack on how experiments are set up and who is responsible for what. The next stage in implementing such standards was to get people to agree to change the way they do it currently and to revisit who was responsible for what part. This is a much larger issue being visited by the managers of the Integration, Automation, and Engineering groups.

Chapter 6 Simulation Model Description

The final stage of the research departed from improving the TD processes and moved to scheduling and prioritization of TD lots relative to production. This was done by creating a simulation model using the educational version of Arena. The purpose of this model was to answer the following question:

> *Given that we are going to increase both the amount of production and the amount of TD in 1999, what is the best way to run TD such that we can meet both development and production cycle time goals?*

At this point the focus was narrowed to one tool set called the Deep Ultra Violet (DUV) toolset in photolithography. This toolset runs the five most critical layers in the flash memory and logic processes. The purpose of the simulation model was to be predictive for generic types of TD. A conscientious effort was made not to link the types of TD in the model very closely to what was running in the fab, but to try and use the TD in the fab as the basis for general phases of TD.

Briefly, the model consists of:

- A characterization of the incoming wafers
- A characterization of the toolset
- Prioritization and scheduling scenarios

The model is described more fully in the following sections.

6.1 Inputs

Traditionally, the workload impact of TD at D2 is only approximated. People are not really sure how many resources are needed to handle it or how much time it takes, but they know that it is different and requires more effort. There are currently initiatives underway to try and handle TD resourcing in a more sophisticated way, but there is still some uncertainty. In the simulation, TD is divided into three categories. These categories were meant to be generic, but they are modeled after three different processes running in the factory. The three phases are:

1. Process Development
   This is the first stage in development. It was modeled after the latest generation of flash memory that was being run in the fab. At this point in development, they are just defining the characteristics and even the tools they will need to make the chip. From a litho point of view, they might at this point be trying to figure out what kind of photoresist to use.
2. Process Characterization and Improvement
   This is the heart of what is done at D2. It is modeled after the first generation of flash memory being developed at D2. Here they have already established what needs to be done to make the chip, but now they are trying to get it from wiggle to full scale manufacturing. One example of what they might look at in litho are the CD and Reg measurements to see what the effects of different exposures might be.
3. Sustainment Development
   This is development work that is done on chips that are already in production. It is modeled after the Pentium II chip that was currently in production at D2. The experiments here are mostly tweaks to do things like increase yield.

Aside from the three types of TD two additional inputs were Production, based on Pentium II production data and Engineering wafers. Engineering wafers are test wafers that are put in the machine to qualify it for a process or to make sure that the machine is still qualified[15].

Table 4 shows the input characterization parameters used in the model.

Table 4: Input Characteristics For Three Phases

PROCESS	ARRIVAL DISTRIBUTION	Lots/Batch Distribution	Layer Distribution	Lot Size Distribution (wafers)
Process Development	EXPO(1000)	40%=1, 50%=2, 9%=3, 1%=4	70% = Layers 2,3 30% = Layer 1,4,5	40%=1, 40%=UNIFORM(2,24), 20%=25
Process Improvement & Characterization	EXPO(750)	45%=1, 50%=2, 0%=3, 5%=5	70% = Layers 2,3 30% = Layer 1,4,5	10%=1, 55% = UNIFORM(2,23), 10%=24, 25%=25
Sustainment Development	EXPO(500)	5%=1, 75%=2, 15%=3, 5%=4	20% = Layer 1 20%=Layer 2 20%=Layer 3 20%=Layer 4 20%=Layer 5	10%=1, 30%=UNIFORM(2,23), 10%=24, 50%=25
Production	30	1	20% = Layer 1 20%=Layer 2 20%=Layer 3 20%=Layer 4 20%=Layer 5	10% = 1, 5% = 23, 10%=24, 75% = 25
Engineering	EXPO(480)	1	20% = Layer 1 20%=Layer 2 20%=Layer 3 20%=Layer 4 20%=Layer 5	25

The input parameters shown in Table 4 are detailed below.

6.1.1 Arrival Distribution

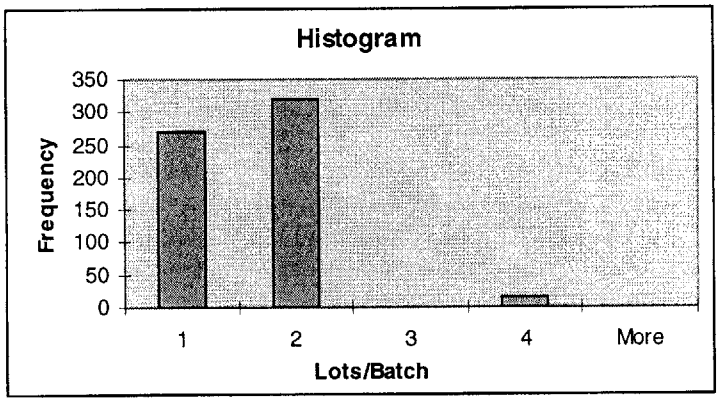
The units for arrival distribution is minutes. What is most important is the relative rate of arrival. One can see for TD that the rate of arrival of the Process Development TD to litho is the lowest, and as one progresses towards Sustainment Development the lots come in twice as often. The exponential distribution is used to account for the variation in the arrival of the lots. Even though he making of a chip is a cyclical process, the model does not allow for reentrant flow. The arrival rate combined with the type of lot accounts for the number of a specific type of lot that shows up at the tools. Appendix C shows an example of the type of spreadsheet data used to find the arrival rates, and explains the calculations.

[15] Layer qualification is a very important step. If a process is not run in photolithography for about a week the tool drifts and needs to be requalified so that it will align the mask properly on the wafer.

6.1.2 Lots/Batch Distribution

This parameter took into account that often many lots showed up at a tool at the same time. This was accounted for by counting the number of lots that arrived at the same time for each type of TD. As shown in Figure 6, a histogram of the number of lots arriving at a time (called a batch) gave a batch size distribution that was entered into the model as percentages. For example, Figure 6 is a histogram for Process Development TD. From it, one can surmise that 40% of the lots come in one at a time, about 50% come in two at a time, and the remaining come in three or four at a time.

Figure 6: Example of Lots Per Batch Distribution Histogram



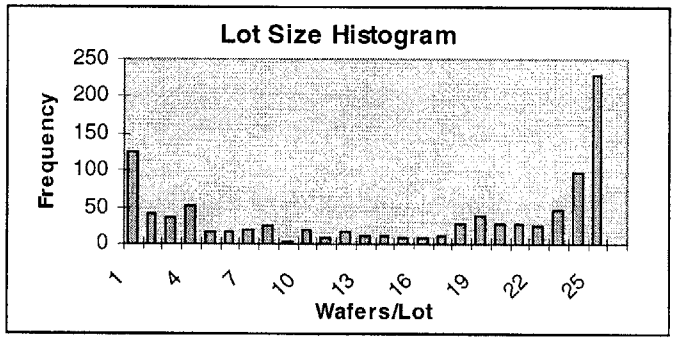
6.1.3 Layer Distribution

Unless told otherwise, it was assumed that for a process all layers come in in equal amounts. The shift supervisors in photolithography said they receive more layers 2 and 3 for the earlier stages of TD, and this was reflected in the layer distribution.

6.1.4 Lot Size Distribution

This refers to how many wafers are in a given lot. As was explained earlier, lots were often split, especially for experiments. Different lot sizes mean different processing times. Lot size distributions were found by taking database data for the number of wafers in a lot and plotting them on a histogram as shown in Figure 7. These values were then converted to percentages of lots of different sizes. Sizes went from one to 25. A lot size of one was given a special characteristic as a *send ahead* wafer. A single wafer from a lot was often sent ahead to make sure the tool would be properly set up for the whole lot when it reached that point. Separate processing time data was available for send aheads, as they tended to take a little longer to get through the system.

Figure 7: Example of Layer Distribution Histogram



6.1.5 Layer Identification

There is one other characterization necessary purely for modeling reasons. Each lot and each layer is given a number. Layers are numbered 1 through 5. Lots are also numbered from 1 to 5. Looking at the table, Process Development is numbered 1, Process Characterization is 2, Sustainment Development is 3, Production is 4, and Engineering is 5. This is important to the model, because as shown later, the lot and layer determines set up times on the tool and processing times.

6.2 Tool Parameters

There are a number of parameters that had to to be set up for both the tools and for the processes running on the tools. These are:

6.2.1 Layer Qualification Tables

For physical reasons, not all layers and lots can be run on all tools. This is the main constraint that causes queues. In the model, each layer is mapped to what tools it can run on. This mapping is called a Layer Qualification Table.

A further complication is tagged onto the layer qualifications. In the simulation model, there are three different tool strategies for which experiments were run. The first one is what has historically happened in the fab. It is called the Theoretical Layer Qualification Table. In this strategy, as long as a tool can be qualified for a type of lot, the lot is allowed to go there. The second strategy is Soft Dedication. This was based on a pilot that was being run at the time. In this scenario, one artificially limits where lots can go to one primary tool and two secondary tools. The third strategy is called Semi-soft Dedication. This reflected what was really happening in the fab, which was that the area was only dedicating two or three of the critical layers and letting all other lots follow the Theoretical Layer Qual Table. Appendix D shows all three layer qual tables. The difference between the primary tools and the secondary tools is the length of queue allowed to grow in front of the tool before being switched to another tool. In the simulation the queue in front of a primary tool had to to be at least twice as long as the queue in front of one of the secondary tools before a lot would go to the secondary tool instead of the primary.

6.2.2 Machine down and repair times

This accounts for preventive maintenance and unscheduled down times. The data and distributions used are the same as those used by the industrial engineering group in their models. Table 5 shows the machine

down and repair times used. The daily gap is a fudge factor used to account for times when the tool is up but starved or waiting for another lot before it can run.

Table 5: Machine Down and Repair Times

REASON	DISTRIBUTION	FAILURE MEAN TIME	FAILURE STD DEV TIME	REPAIR MEAN TIME	REPAIR STD DEV TIME	UNITS
DAILY 1	Uniform	8	0.8	0.2	0.02	hrs
DAILY 2	Uniform	24	2.4	0.8	0.08	hrs
WEEKLY	Uniform	168	16.8	1.8	0.18	hrs
MONTHLY	Uniform	730	73	3.7	0.37	hrs
SEMI-ANNUAL	Uniform	4380	438	15	1.5	hrs
UNSCHED 1	Uniform	168	16.8	2.2	0.22	hrs
UNSCHED 2	Uniform	168	16.8	2.2	0.22	hrs
UNSCHED 3	Uniform	168	16.8	2.2	0.22	hrs
Daily Gap	Exponential	8	0.8	1.6	0.16	hrs

6.2.3 Layer Change and Setup Times

When a new type of lot comes to the machine, there is some time required to clear out the track, change resists, and setup the reticle for the new lot. This time is accounted for in the Layer Change and Setup Time tables below.

Table 6 maps how long it takes to change between layers. The production rows and columns also include Sustainment Development because Sustainment Development is done on processes that are already in production. PII rows and columns are the first two phases of development. This was done because no data was available on Process Development TD at the time, so it was assumed to have similar characteristics to the Process Characterization and Improvement TD.

Table 6: Original Layer Change and Setup Time Matrix

856 TIME MATRIX (minutes) for Changing from Layer to Layer

	856 STR	856 POLY	856 CON	856 MT1	856 VIA1	802 STR	802 FLG	802 SMS	802 MT1	802 CON
856 STR	0	20	10	20	10	10	20	20	0	10
856 POLY	20	0	20	20	20	20	10	10	30	30
856 CON	10	20	0	20	10	10	30	0	20	10
856 MT1	30	30	30	0	20	20	0	0	10	20
856 VIA1	10	20	10	20	0	10	20	0	20	10
802 STR	10	20	10	0	10	0	0	0	0	10
802 FLG	20	10	30	30	30	0	0	10	0	0
802 SMS	0	10	0	20	20	0	10	0	20	0
802 MT1	20	0	0	10	20	20	0	0	0	20
802 CON	10	20	10	20	10	10	20	0	20	0

Because the above table is too complicated to put in the model, it is simplified to Table 7 below. Table 7 is simply the average of all the layer change setup times for a single layer. Thus, Layer 1 to Layer 1 is an average of changing from Layer 1 to Layer 1 for different processes. Layer 1 to Layer 2 is the average of changing from a Layer 1 of one or the same process to a Layer 2.

Table 7: Averaged Layer Change and Setup Time Matrix

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Layer 1	5	23.5	10	11.5	10
Layer 2	22.5	5	26	28	26.5
Layer 3	5	18	5	24.5	18
Layer 4	27	16	15.5	5	26
Layer 5	10	23	10	21	5

The formula used to calculate layer change penalties is:

$$\text{Setup Time} = \frac{\text{ABS(Departing Lot*Departing Layer - Incoming Lot*Incoming Layer)*Setup Factor(Incoming Layer, Departing Layer)}}{3} \quad (1)$$

What Formula (1) does is take into account a change of layer and process and then multiply it by the appropriate Setup Time matrix cell shown in Table 7. The division by 3 is a fudge factor meant to keep the setup time values reasonable. So, for example, if the Departing Lot was type 1, Departing Layer type 1 and the Incoming Lot type 1, Incoming Layer type 2; the Setup time would be ABS(1*1-1*2)*22.5/3 = 7.5 minutes. On the other hand, if it is the same lot and layer coming in, the setup time is 0 minutes.

6.2.4 Processing Times

Once a lot is on the tool, it takes a certain amount of time to process it. This time is given by the formulas shown in Table 8.

Table 8: Lot Processing Times

Process	Send Ahead	Regular Lot
Phase I	150/40 min	NORM(30,10) + (No of Wafers - 1)*3.24
Phase II	100 min	NORM(30,10) + (No of Wafers - 1)*3.24
Phase III	100 min	NORM(15,2) + (No of Wafers - 1)*3.24
Production	40 min	NORM(15,2) + (No of Wafers - 1)*3.24

This table shows two different types of lots. One is a lot that only has one wafer in it, called a *send ahead* wafer. The purpose of this wafer is to make sure that when the rest of the lot shows up, the tool will be able to run the lot smoothly. Since this wafer ties up the tool until the rest of the lot shows up, it takes up a lot of time on the tool. For Phase I (Process Development), there are two different send ahead values, with the lower one being for experiments that are run by engineers. The assumption is that if an engineer is running the experiments, he will not tie up the tool with a send ahead for as long as if the experiment is running unsupervised. The formulas for the regular lots first take into account the *first wafer effect*. This simply says that the first wafer in a lot takes the longest amount of time. After the first wafer, the rest of the wafers take 3.24 minutes. This time is based on discussions with the industrial engineers at Intel, combined with what the model could handle without crashing.

6.3 Model Assumptions

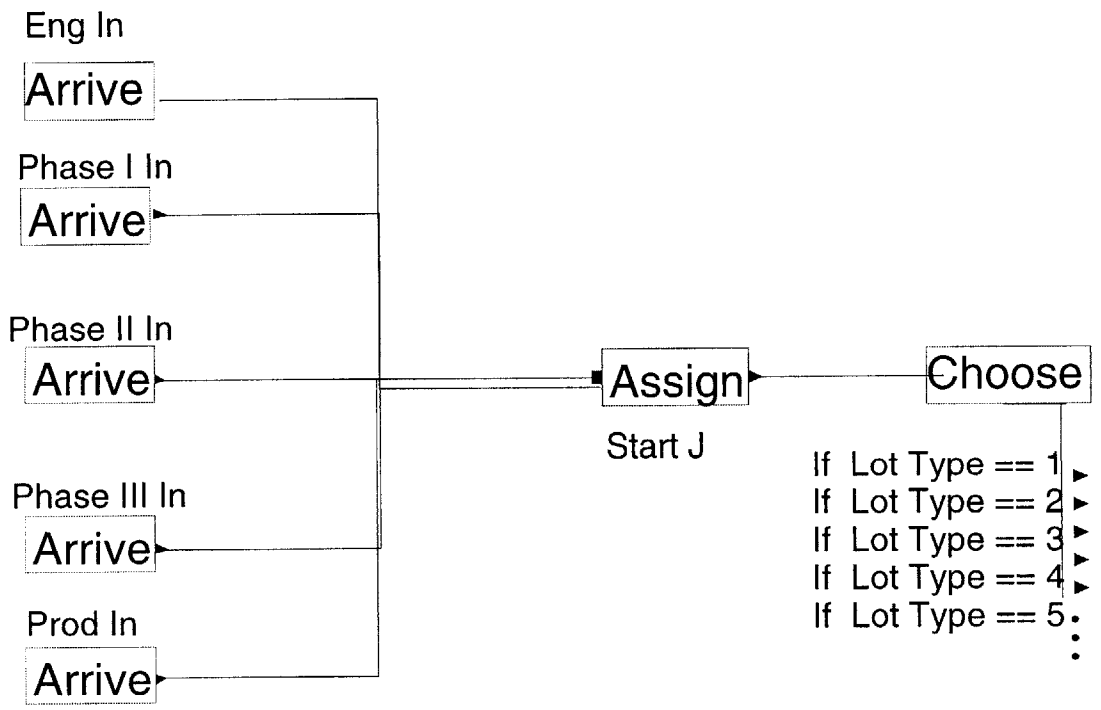
As with any model, this one is based on a number of assumptions and simplifications. The ones in this model are:

- Processes running today reflect processes running in the future
- The arrival, processing, and setup times and distributions do not change once the model is set. They are also simplified to accommodate the constraints of the educational version of the model, and to keep the model simple and understandable.
- There is no links between lots as that enter the tools. This means that even though the same lot enters the toolset five times, the model has no memory of the lot. Also, the tools do not hold themselves open for a lot to catch up to its send ahead.
- This is purely a tool specific study. It is assumed that the technician has done all his preparations before the lot is put on the tool, and if she hasn't, that time is accounted for in the data given on processing times.

6.4 The Model

The following figures breakdown what the model actually looks like. Figure 8 shows the creation of lots:

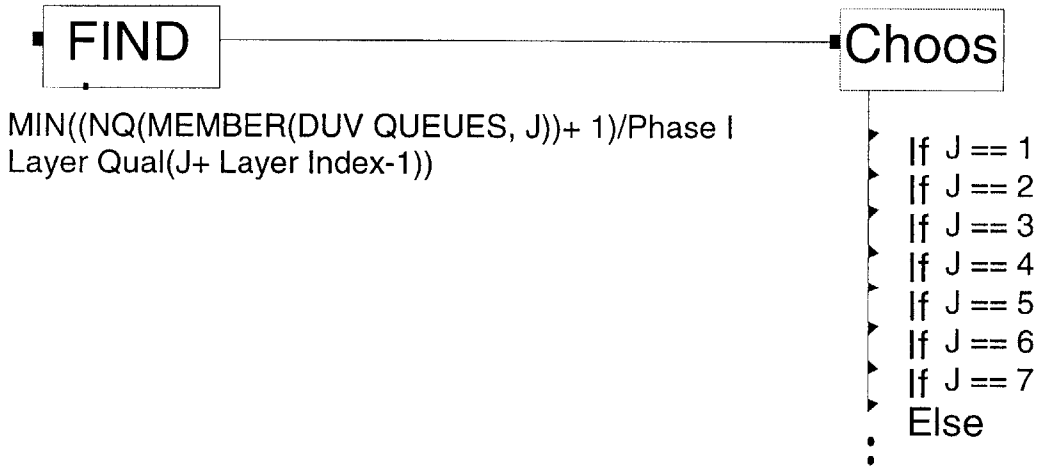
Figure 8: The Creation of Lots



As lots are created, they are given a priority, lot type, layer number, lot size, and arrival time. The priority assignation tells the program if the lot is going to get any special treatment when it gets to the tool. If the priority is 0, then it is simply first in first out (FIFO) at the tool. All other attributes are based on the input parameter discussed in the previous sections.

Figure 9 is the next step in the model. It is one of the key decision making components of the model.

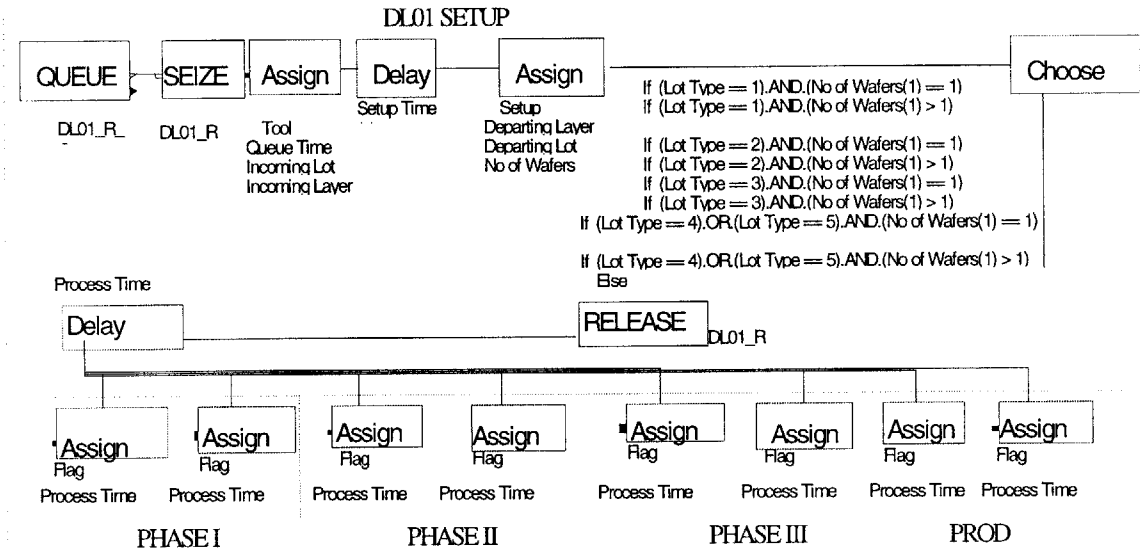
Figure 9: Choosing a Tool



What Figure 9 shows is that once a lot is create, it must find a tool. In the program, the rather complicated formula shown above the <FINDJ> block simply tells the lot that of the tools it is qualified to go to, choose the shortest queue. The Layer Qualification Table for each Phase (stage in the development or production) is stored in the Phase Layer Qual array. The lot indices the appropriate spots in the Phase Layer Qual array and looks to for the minimum queue in front of the tools for what tools that layer is qualified to go to. The <FINDJ> block then assigns the number of the appropriate tool (one through seven) to a variable J. The <Choose> block looks at J and sends the lot to the appropriate tool. One modeling specific constraint is that the lots go to the first shortest qualified queue it finds rather than randomly choosing from the shortest queues.

Once the lot has chosen a tool, it must be processed. The code to do this, is shown in Figure 10.

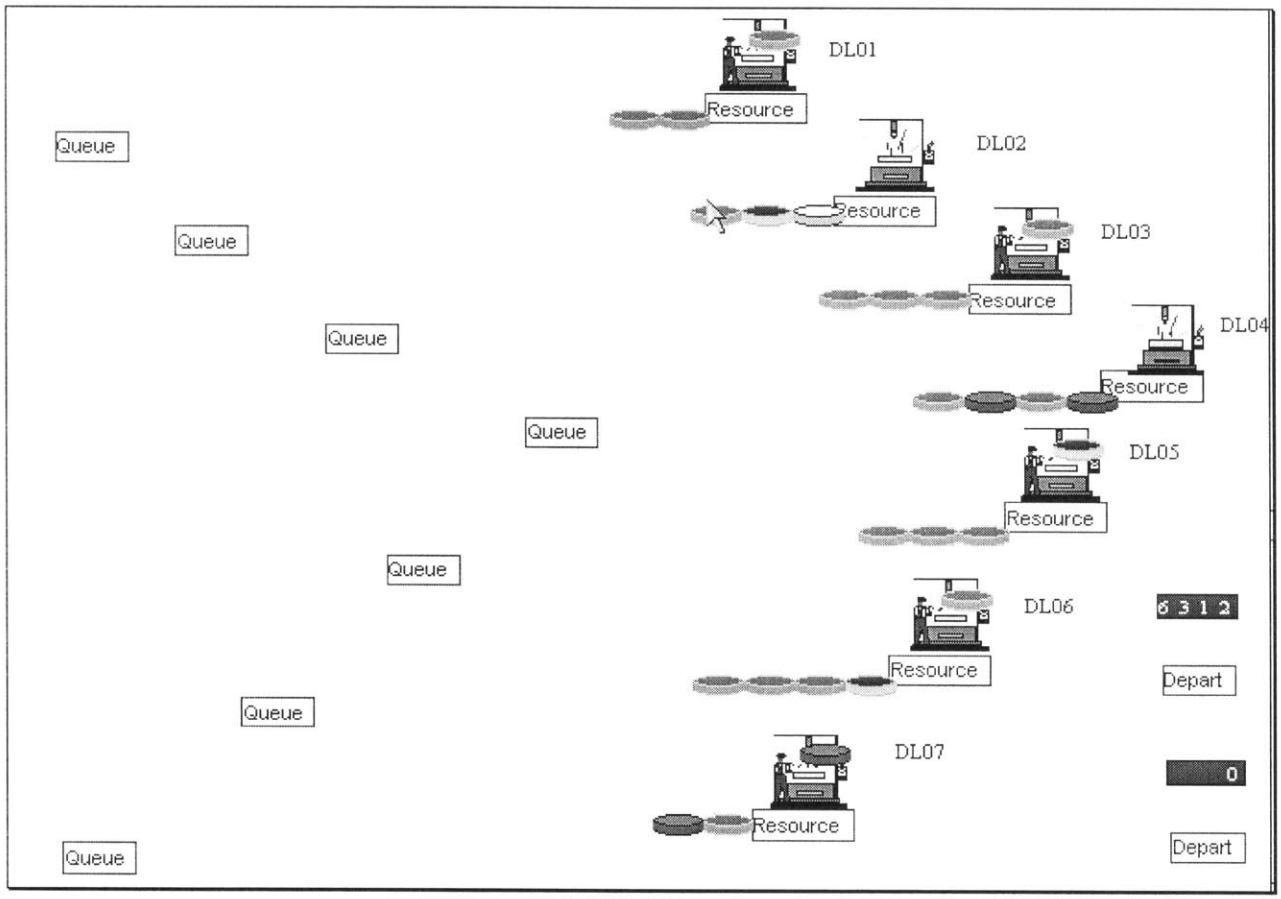
Figure 10: Processing a Lot



Each of the seven tools has such a block. The lot lines up in queue and when the tool is ready for it, it seizes it. There is then a delay based on what the incoming lot is and what the departing lot is. This delay is the setup time the tool needs to change between lots and layers. Setup times are kept in an array in the program. Next the lot must go to the appropriate module to be processed. The processing times are assigned using the appropriate <Assign> block for each phase. Phase I is for Product Development TD, Phase II for Process Characterization, Phase III for Sustainment Development and Prod for Production. For each phase, there is an array that stores the processing times. Depending on whether the flag indicates that the lot is a send ahead wafer (the lot only has one wafer in it), or whether it is a regular lot, the processing time is either set to the send ahead wafer time or to the sum of the first wafer processing time plus the time per each additional wafer. The Engineering wafers go through the production <Assign> blocks. When it is done processing, the lot type, layer, priority, and cycle time data is collected and then the lot exits the system.

A snapshot of the model running is shown in Figure 11. The lots, color coordinated to match their process type, are lined up in front of the tool. If there is an operator at the tool, it means the tool is processing. If there are flashes coming out of the tool, it means the tool is down. When a tool is down, lots simply wait for the tool to come back up. They do not go to another tool. If the tool is just sitting there, it means that it is idle.

Figure 11: Sample Simulation



Chapter 7 Simulation Model - The Experiments

Table 9 outlines the experiments run on the model. These experiments were chosen based on what people in the fab and the industrial engineers thought would be interesting. Each experiment was run for the three dedication strategies. All experiments were baselined to the FIFO scenario. This means that second order effects from combining policies were not looked at.

Table 9: Experiments Run

Experiment	*Description*
Base Case	Lots go to qualified tools with the shortest queues. This FIFO scenario most closely models what happens in the fab.
Prioritization	In this scenario, TD always went to the front of the queue of whatever tool it chose.
Batching Scheme	In this scenario, four production lots were batched together before they would be released to a tool. The purpose of this scenario was to reduce setup times by lumping together lots. The batch size is four because this is the number of lots that can be put on the track at one time.
Split by Shift	Here, the earliest phase, Process Development, was only run during the day when the engineers are available. This reduced the send ahead time from 150 minutes to 40 minutes.

Each experiment run for each of three dedication strategies gave 12 experiments run in total. Experiments were run multiple times, but the data sets generated were so large, that there were no significant differences found in the values between runs. Upwards of 13000 lots came out of one run. The simulations were run for about 6 months. This seemed a reasonable time frame as the data was from a six month time frame and processes seemed to shift from one stage to the next at about that rate. Since the entire data set was used to calculate queue time values, the result were not subject to statistical errors from sampling either.

Ideally it would have been desirable to look at the impacts of tweaking arrival rates, lot sizes, or processing times, but the educational version of Arena could not handle this. It would also have been more accurate to implement FAST 3-2-1, but from a programming stand point, the educational version of Arena could not handle additional code. It can be assumed however, that FAST 3-2-1 is an intelligent prioritization scheme that can be used on top of the prioritization schemes run on the model. One sensitivity run that was performed was the impact of changing the layer distributions. As will be shown later, this did not have much of an impact.

Appendix E shows an example of what the output files looked like for each run. For each experiment, the data points collected are the mean, minimum, 25th percentile, median, 75th percentile, and maximum queue time separated by TD and Production. Using this range one can see if there is statistical significance to the difference between queue times for the various policies. In the simulation, TD was further broken out to the different phases, but Intel management was only concerned with the overall TD and Production queue times.

Chapter 8 - Model Validation

The final step in the simulation model was a model validation. The validation spreadsheet used is shown in Table 10. Data for the actual lot history values came from Intel's lot history database. A sample of the input used is shown in Appendix G.

Table 10: Model Validation Data

LOT HISTORY (ACTUAL) DATA

Process	Time Frame	Average Arrival Rate (hrs)	Average QT (hrs)	Average PT (hrs)	Average CT (hrs)	Average wafers out (wafers/wk)
Phase I TD	7/2/98 to 9/19/98	17.68	11.5	1.2	12.7	74
Phase II TD	12/29/97 to 9/19/98	15.82	10.7	2.5	13.2	378.5
Phase III TD	12/28/97 to 9/19/99	8.9	13	2.3	15	664.8
Production	WW18	0.82	10	2.5	12.5	6067.1
						7184.4

SIMULATION DATA: THEORETICAL LAYER QUAL TABLE FIRST COME FIRST SERVE

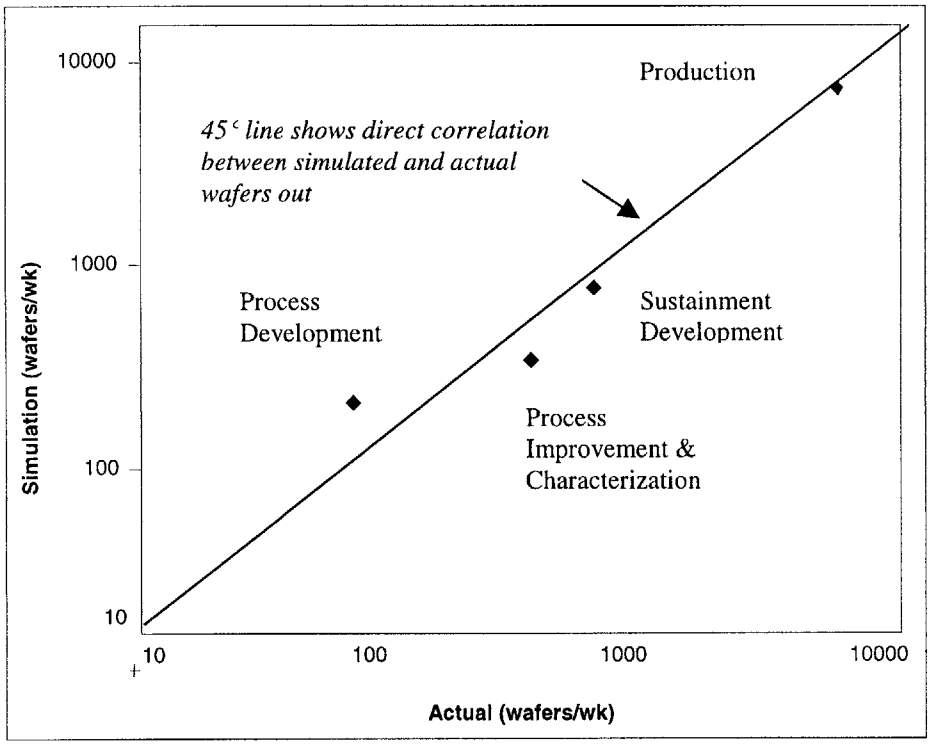
Process	Time Frame	Average Arrival Rate (hrs)	Average QT (hrs)	Average PT (hrs)	Average CT (hrs)	Average wafers out (wafers/wk)
Phase I TD	29.7 WEEKS	15.4	7.6	2.1	9.7	205.9
Phase II TD	29.7 WEEKS	7.4	3.9	1.9	5.8	332
Phase III TD	29.7 WEEKS	3.9	3.9	1.8	5.7	766.3
Production	29.7 WEEKS	0.5	3.2	1.8	4.9	7505
						8809.2

NOTES

	Average # wafers/lot	Simulation:Expected wafers out	Actual: Expected wafers out	
Phase I TD	11	173	151	Formula:
Phase II TD	16	532	249	Expected Wafers Out = (24 hrs/day/arrival rate)
Phase III TD	19	1615	708	*(ave wafers/lot)*(ave lots/batch)*(7 days/wk)
Production	22	7526	4589	

Table 10 compares the actual lot history data to that expected from the simulation. Because the educational version of Arena could not handle the rate at which wafers were really going through the fab, it was expected that the simulation numbers would be a little lower. What was important, however, was that the ratio of the types of lots coming out, whether it be a TD phase of development or production, be consistent. Average wafers out were calculated by multiplying the average arrival rate by the average number of wafers for a given type of lot. Figure 12 illustrates this relationship by showing that there is a direct correlation between simulated versus acutal wafers out of the DUV tools at photolithography. Using this back of the envelope calculation as a comparison, one sees that Phase I (Process Development TD) wafers out were a little higher than the actual indicate it should be. This was considered an acceptable exception because very little data was available for this phase at the time. The other TD phases and production wafer out numbers are reasonable. Also looking at average queue times and process times, while queue times are shorter than actual data, the ratio between different types of lots is about right, and the processing times for the different types of lots are about right.

Figure 12: Proportionality in Wafers Out Between Actual and Simulation



Chapter 9 Simulation Model - Results

Table 11 has the complete queue time data results from the experiments.

Table 11: Simulation Model Results

	TD						PROD					
	Mean	Min	25%	Median	75%	Max	Mean	Min	25%	Median	75%	Max
Theoretical Layer Qual Table												
1. First Come First Serve	5.2	0.0	1.3	3.4	6.7	38.7	3.2	0.0	0.8	2.0	4.5	44.2
2. TD always goes to front of queue	3.1	0.0	0.7	1.6	3.4	34.5	3.8	0.0	0.9	2.2	5.1	59.5
3. Batch 4 same layer production lots before they can go into queue	6.1	0.0	0.0	1.7	5.9	59.1	33.9	4.3	21.3	28.1	38.9	152.4
4. Run Process Development TD only during the day on dedicated tool	25.6	0.0	1.3	3.7	18.6	180.1	7.4	0.0	0.7	1.9	4.7	141.3
5. Process Development and Process Characterization TD Layer Distribution change	4.9	0.0	1.2	2.9	6.3	31.4	3.1	0.0	0.8	1.8	4.2	25.2
SemiSoft Dedication - only dedicate most critical layers												
1. First Come First Serve	4.5	0.0	1.2	3.0	5.9	27.5	2.7	0.0	0.7	1.7	3.6	28.3
2. TD always goes to front of queue	4.1	0.0	0.9	2.0	4.3	36.6	5.5	0.0	1.1	3.0	7.5	81.2
3. Batch 4 same layer production lots before they can go into queue	6.1	0.0	0.0	1.8	6.1	55.6	33.2	4.0	21.3	28.0	39.0	137.5
4. Run Process Development TD only during the day on dedicated tool	19.9	0.0	1.4	4.0	16.5	141.2	5.9	0.0	0.8	1.9	4.8	124.0
Soft Dedication												
1. First Come First Served	5.7	0.0	1.0	3.7	8.5	73.1	12.2	0.0	1.8	7.0	15.7	76.0
2. TD always goes to front of queue	2.2	0.0	0.4	1.3	2.9	27.5	13.8	0.0	1.8	7.8	18.2	93.6
3. Batch 4 same layer production lots before they can go into queue	2.1	0.0	0.0	0.6	3.1	24.5	30.3	4.0	11.8	21.9	38.2	192.5
4. Run Process Development TD only during the day on dedicated tool	9.5	0.0	1.1	4.2	10.5	79.8	9.5	0.0	1.5	5.6	12.3	79.0

This table shows the mean, minimum, 25th percentile, median, 75th percentile and maximum queue times in hours that a lot had to wait from the time it arrived at the DUV tools until it left. More important than the actual numbers are the relative differences between the queue times. The table is divided up by the three different dedication strategies. The four scenarios were run for each strategy. A fifth scenario run only for the Theoretical Layer Qual Table was to see if changing the layer distribution for Process Development and Process Characterization TD to an even split between five layers would help. The results show that it does not have a large effect on the whole, but it does bring in the maximum queue times significantly. This is because a more even distribution of layers allows a more even distribution of tool usage. That is, it prevents certains tools from being more heavily loaded than others. This is an interesting result when

thinking about the order in which to schedule experiments, though it may be hard to control what layers need the most experimentation.

Analyzing the data from Table 11 leads to the following results:

1. *Not all TD lots are the same.*
   As explained earlier, in creating the model, first the data need to be characterized. This characterization clearly shows that there are some distinct differences between the Process Development TD and the two later phases. These differences can be shown in the following figures.

Figure 13: The Number of Wafers In a Lot

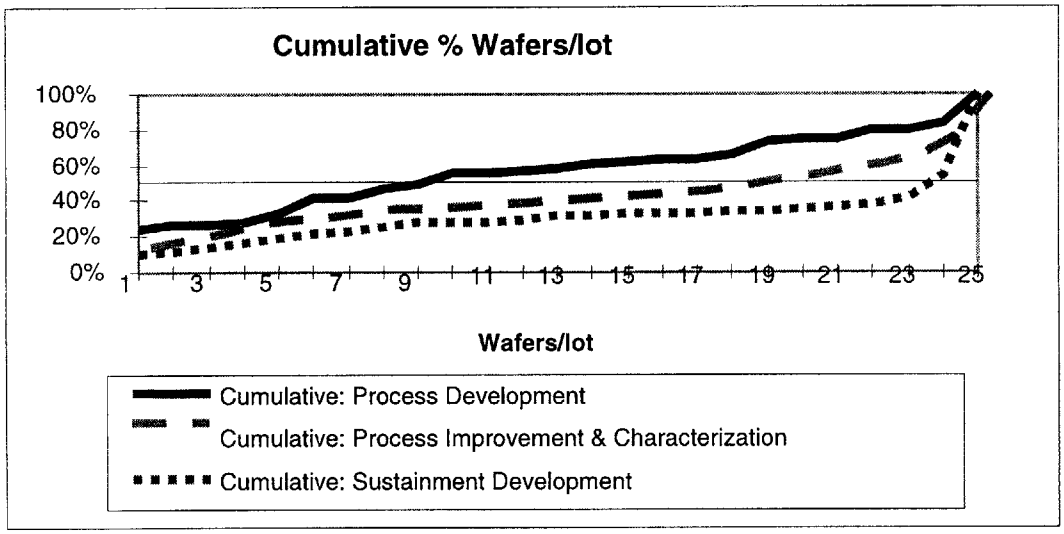


Figure 14: Number of Lots Arriving Simultaneously

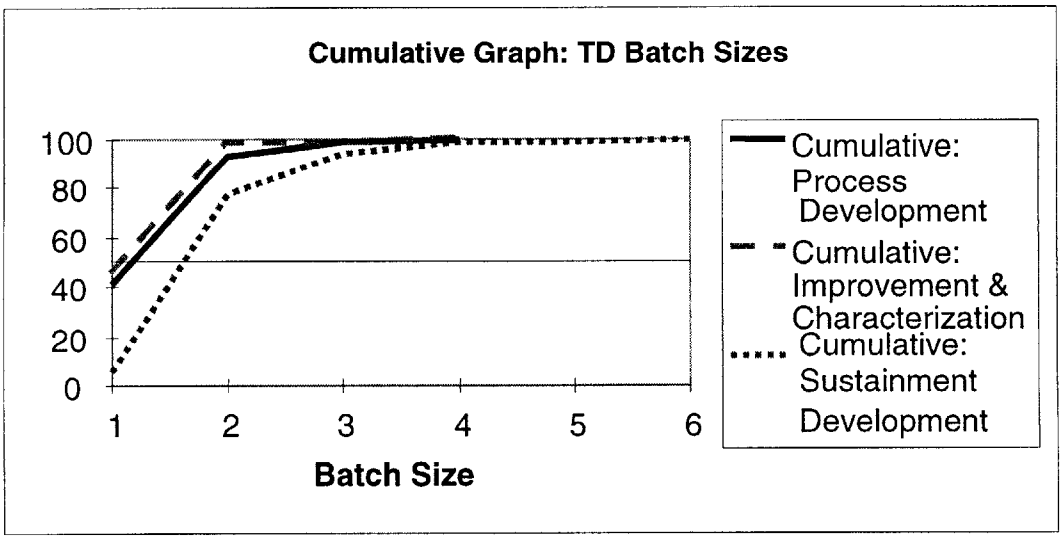


Figure 13 is a cumulative histogram of the number of wafers that are in a lot. Looking at the 50% line, one can see that for Process Development TD, 50% of the lots that come in have nine or fewer wafers.

A full lot has 25 wafers, so what the number of wafers in a lot indicates is approximately how many splits there have been to the lot. The more splits there are, the more complicated the experiment is. This makes sense when one considers that, at a minimum, each split will have the same layer change setup time and a large component of the processing time will be the same due to the first wafer effect. Looking at the 50th percentile for Process Characterization, it is more like 18 wafers and Sustainment Development is around 24 wafers. This says that Process Development experiments are the most complex. Intuitively, it makes sense that the earliest development work would be the most complicated. Figure 14 further supports this by showing that compared to the later two phases, more Process Development lots show up at the same time when they arrive at the tools.

The arrival rate data also shows that Process Development lots come in the least often. What this says is that even though Proces Development runs fewer experiments, these experiments are more complicated and take longer to run. These type of experiments are opposite to production, where there is a constant stream of quickly processed lots. D2 focuses on trying to make experiments look like production. This was the main motivation of the work done in photolithography to help standardize formats for TD lots. However, standardized formatting aside, D2 Process Development does not mirror production.

2. *Changing TD prioritization has a much larger effect on TD queue times than production queue times.* This can best be illustrated by Figure 15.

Figure 15: To a Change in TD priority, TD queue times change more than Production

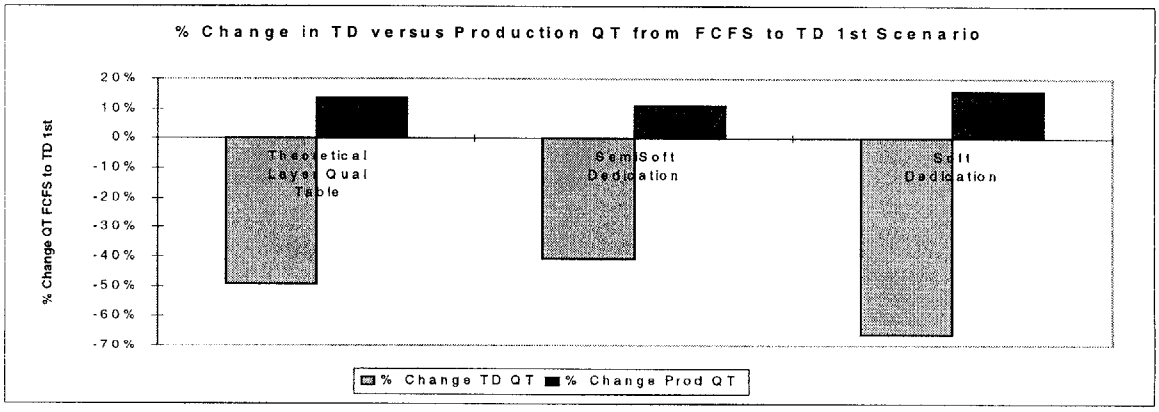


Figure 15 compares the change in queue times for TD and Production under the three dedication strategies when TD priority went from FIFO to putting TD at the head of the queue. What it shows is that under each strategy, the change in queue time for TD was much greater than the change in production queue times. It makes sense that TD queue times would go down and production queue times would go up, but what is interesting is that production queue times do not go up very much relative to the gains in TD queue times. The reason for this result is twofold:

1. TD accounts for only 10% of what goes through the fab. Thus, TD adds only a little bit of queue time to the overall system. Moving TD to the front of the queue, then, does not add much queue time to the production lots.
2. There are some reduced setup times by always putting TD at the front of the queue. This gain is simply from lumping together all incoming TD which leaves longer stretches of production.

This result serves as a warning and a recommendation to D2. The warning is that the queue time impacts of deprioritizing TD will be greater than they might anticipate. The recommendation is that if

you want to decrease TD cycle times, prioritizing TD will give you the greatest gain with the least impact on production.

3. *The lead time increased by batching lots is not compensated for by reduced setup or engineering times.*
   This fact is illustrated in Figures 16 and 17.

Figure 16: Queue Time Impact when Batching Production

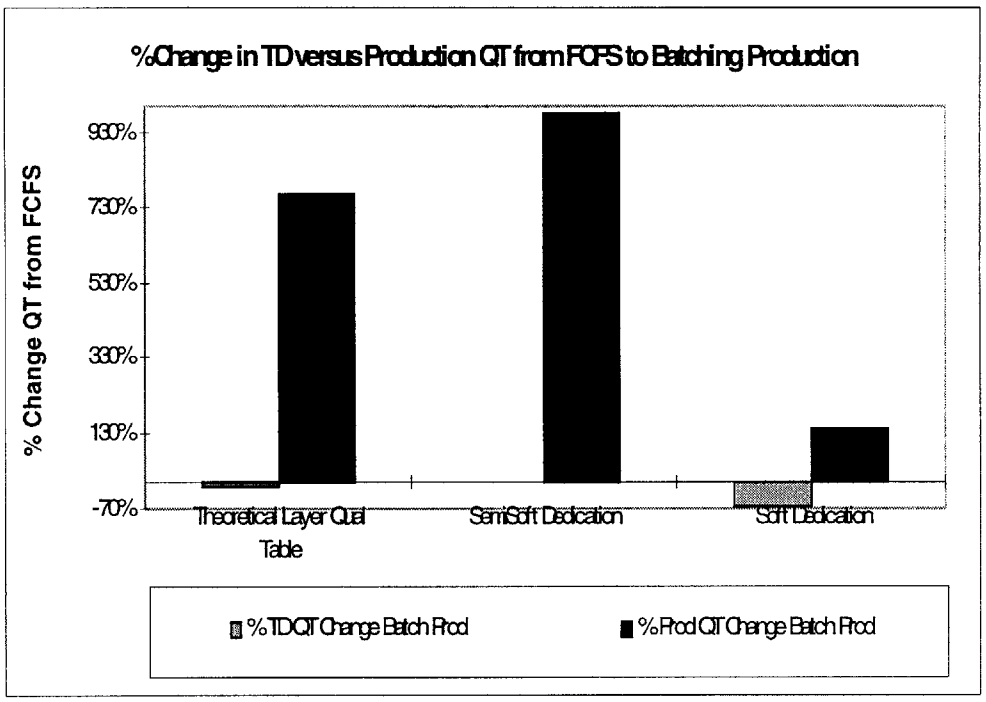
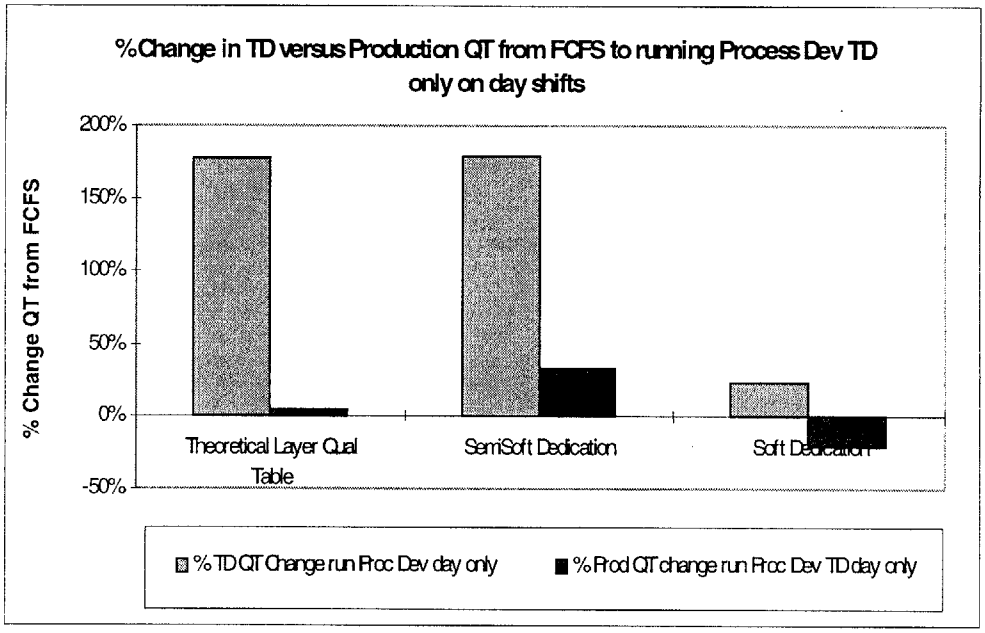


Figure 17: Queue Time Impact of Running Process Development Only During the Day



Figures 16 and 17 are two of the experiment scenarios. Figure 16 shows a batching scheme in which four production lots of the same layer type are batched together and then all sent to the same tool. The purpose of such a scheme is to reduce setup times on the production lots by sending them in groups. Instead what happens is that the lead times for the production lots go up. TD does incur some queue time savings since it has easier access to tools, but it does not compensate for the delay in production lots.

Figure 17 shows a scenario where Product Development TD is run only during the day. This scenario is meant to mimic the idea of only engineers that work during the day running these more difficult lots. In this scenario, at night the tool is allowed to run whatever other lots are qualified to run on it, but in the day it is solely dedicated to Process Development TD. What occurs is a large increase in TD queue times under all scenarios, as the Process Development lots sit around at night waiting to be run in the day. Only under the soft dedication strategy do production queue times benefit. However, looking at Table 11, one can see that the absolute queue times in the soft dedication strategy for these two scenarios are so high (especially for production) relative to the other two dedication strategies, that the soft dedication strategy is not one to pursue.

These two scenarios show, that unlike gains that can be had from holding lots in an area like diffusion, lots are simply not on the tools long enough to benefit from being held back unnecessarily.

3. *There is an optimum strategy that optimizes tool utilization while minimizing setup times. From the scenarios run, Semisoft dedication putting TD at the front of the queue gives the best results.*

Figure 18 shows straight queue times under the TD at the head of the queue scenario.

Figure 18: The Optimal Scenario

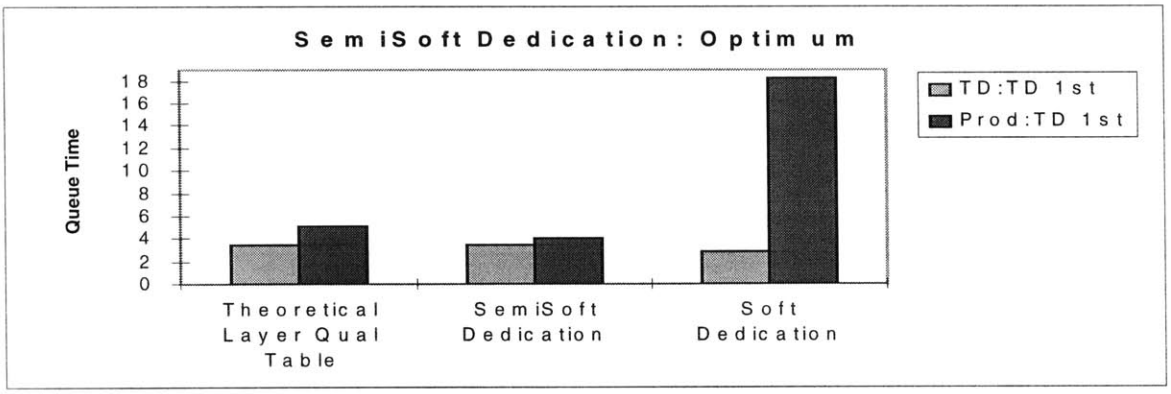


Figure 18 shows that running TD first with semisoft dedication gives the best results from a queue time point of view. This is because in the Theoretical Layer Qual strategy, one does not minimize setup times. Lots have many options on where to go. The pure Soft Dedication strategy is too restrictive so some tools are underutilized while others have long queues. The Semisoft Dedication strategy is a balance between these two extremes. Only the most critical layers are dedicated. Since these layers also have the most experiments coming in (layers 2 and 3 for the first two phases of development), the tools there are longer runs of the same types of lots, minimizing set up times, but the other lots are able to compensate for the dedication by going to the remaining tools. As shown in Figure 15, running TD first is the best scenario for the reasons given before.

Figures 19 and 20 illustrate how the semisoft dedication strategy helps. Tool utilizations are a little lower than for the Theoretical Layer Qual Table, but also a little flatter, indicating better load balancing. Average queue lengths in front of the tools are also slightly lower in semisoft dedication compared to the Theoretical Layer Qual Table. Soft dedication is the worst strategy from a utilization and a queue point of view.

Figure 19: Tool Utilization FIFO Strategies

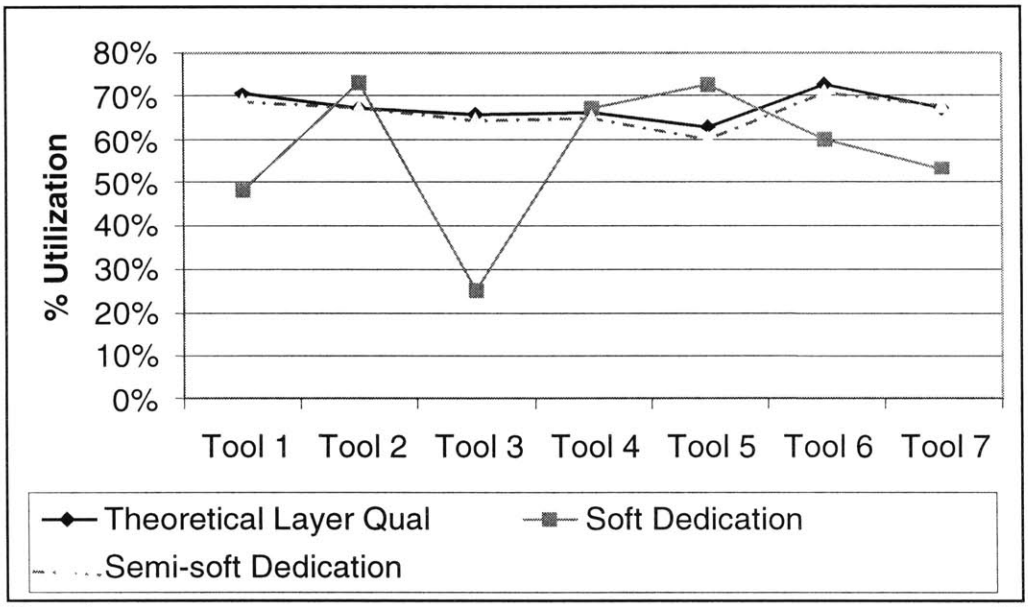
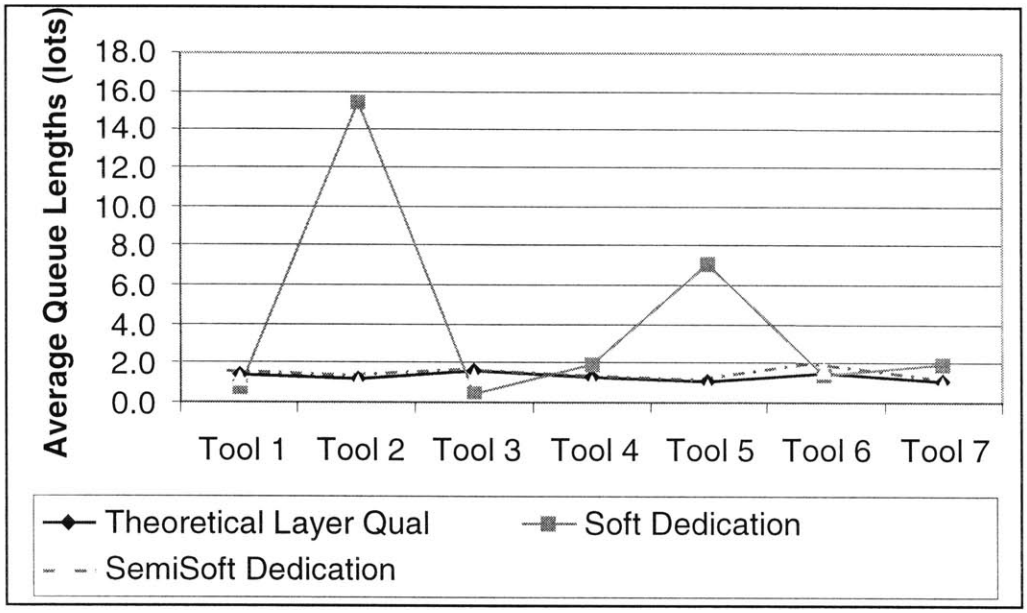


Figure 20: Average Tool Queue Lengths FIFO Strategies



Figures 21 and 22 show the benefits of TD 1st. TD 1st increases tool utilization and decreases average queue lengths.

Figure 21: Tool Utilization - TD 1st versus FIFO for Semisoft Dedication

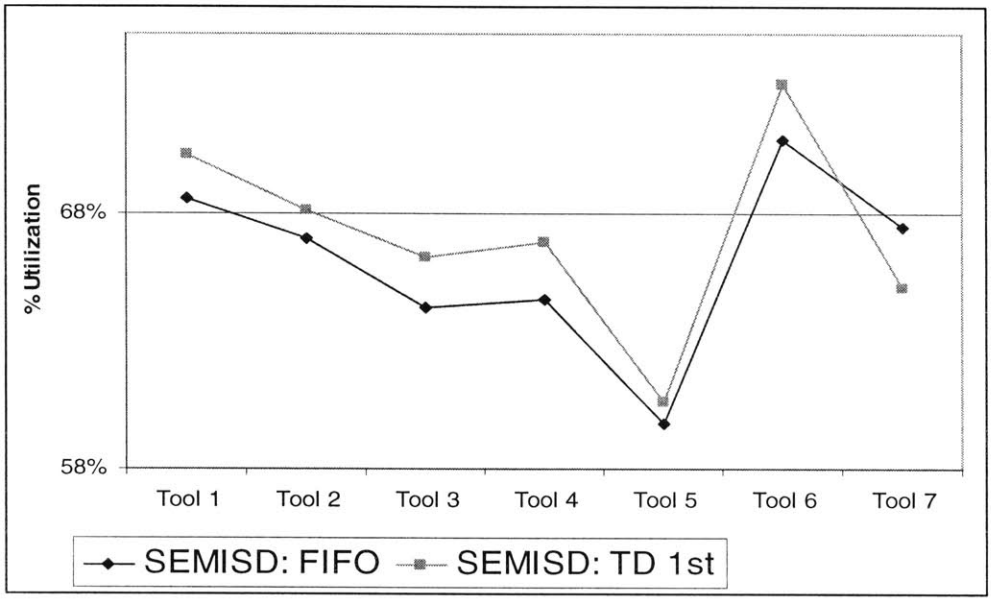
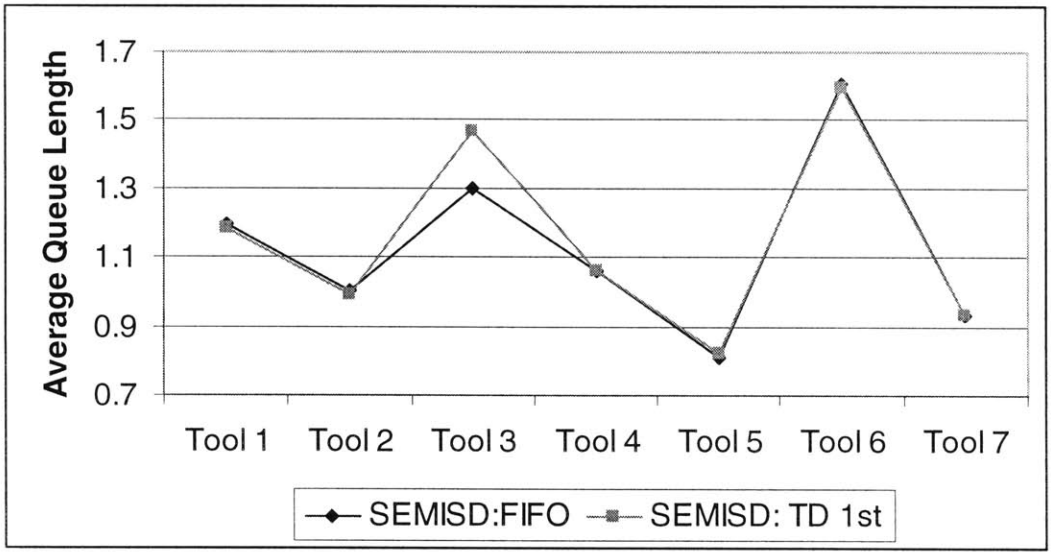


Figure 22: Tool Queues - TD 1st versus FIFO



The industrial engineering at D2 has been piloting the semisoft dedication strategy and have independently come up with the conclusion that the semisoft dedication strategy is more optimal than what they are currently doing.

Chapter 10 - Conclusions

The results above lead to the following conclusions for D2:

1. *Separate out Process Development TD from the later phases of TD and production.*
   Characterizing the data for the lots that go through the DUV tools at litho clearly showed that Process Development lots are a lot more complex than the other types of lots D2 runs. When they hit this area, it can be expected that they will slow down the tool they are on. It is important, then, to focus on this TD to allow it to run as smoothly as possible through the area. One suggestion is to try creating technician-engineering teams for these types of lots. The role of the technician would be to know that such a lot is arriving and to ensure that it gets through the area smoothly. This is already done informally in the fab, where the most experienced technicians tend to run the experimental lots. There is also a concern that creating this special team will require more manpower when they are trying to reduce manpower. However, having lots run more efficiently through the tools should actually help labor requirements by freeing up the people on the floor to do other tasks.

2. *Prioritize TD.*
   The simulation shows that TD prioritization is a key lever in queue time. Giving TD a high priority should ensure that TD gets done as quickly as possible. There is also not much gain in deprioritizing TD to production queue times. Since D2's primary charter is to develop new products, it makes sense to let TD move to the front of the queue as quickly as possible.

3. *Do not hold lots.*
   While reducing set up times is advantageous, strategies that hold lots in order to reduce setup or processing times, such as stringent dedication or batching strategies, negatively impact queue times.

4. *Continue to search for the optimum WIP strategy.*
   The simulation shows that changing WIP prioritization and scheduling strategies does have an impact on queue time. The Industrial Engineering department is already working in the direction of searching for an optimum using semisoft dedication. If this is combined with concepts such as further prioritizing TD, smoothing out the layer distribution that hits the tool sets, and balancing tool utilization, the search can continue for an optimum strategy.

Bibliography

1. Bailey, Troy. "Semiconductor Cluster Tool Availability." S.M. Thesis, M.I.T. Sloan School of Management, 1992.

2. Bohn, Roger E.. *Noise and Learning in Semiconductor Manufacturing.* La Jolla, CA: University of California, San Diego. July 1993.

3. Burgelman, Robert A.. *Fading Memories: A Process Theory of Strategic Business Exit in Dynamic Environments.* Stanford University: Administrative Science Quarterly, p. 32, 1994.

4. Campbell, Stephan A.. *The Science and Engineering of Microelectronics Fabrication.* New York: Oxford University Press, 1996, Chapter 1, p. 152.

5. Kelton, David W., Sadowski, Randall P., and Deborah Sadowski. *Simulation With Arena.* Boston, MA: WCB McGraw-Hill, 1998.

6. Kumar, P.R., Steve Lu, Deepa Ramaswamy. Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants. Illinois: University of Illinois.

7. Liu, K.R. and Johnny Wu. *Troughput Time Reduction through On-Line Simulation and Intelligent Scheduling for Semiconductor Factory Automation.* Intel Corporation, Portland Technology Development.

8. Wein, Lawrence M. Scheduling Semiconductor Wafer Fabrication. IEEE Transactions on Semiconductor Manufacturing Vol 1, No. 3, August 1988.

This page is intentionally left blank.

Appendix A: Definitions

Child Lot	A full lot is sometimes split into child lots. These child lots are processed separately until it is time for them to be recombined into a full lot.
Critical Dimension	This is a very important measurement criteria upon which wafers are passed or failed at a step. It measures feature sizes on the chip.
Gate Oxide	Oxide is actually a layer of silicon dioxide. Gate oxide is the gate dielectric of the transistor. A dielectric separates to conductive materials. You use one when you want to keep two charges separate. This creates a potential difference across the dielectric. Changes in voltage are the basis for how transistors work. One change turns the transistor on, another change turns it off.
Gate, source, drain	A transistor is composed of a gate, source, and drain. Depending on the potential differences between these three points, current flows from source to drain, or does not flow. This is what turns the transistor on and off.
Hold	A lot put on hold means it is temporarily stopped in its route, until given permission to go forward. Normally this is done because something special that may require engineering attention needs to be done to the lot.
Module	A module is one of the functional groups at Intel, like photolithography.
N and p wells	Literally, these look like wells in the silicon that have been doped (charged) heaviliy with negative ions for n wells, and positive ions for p wells
Photoresist	A chemical that is coated on top of the wafer before the wafer is exposed to the mask with the desired pattern. The pattern is initially etched into the resist using UV light. Then, a chemical is used to eat into the desired layer on the wafer. After this the photoresist is washed away.
Recipe	At a specific step, this is the specifications for what should be done. For example, in photolithography the recipe says how long to expose the mask for at what energy level.
Registration	Another critical measurement on a chip. This measures the alignment of masked layers with each other.
Reticle	A mask used in photolithography.
Route	A lot follows a route. A route is the series of steps the lot goes through.
Special Operation	An experimental operation step.
Station Controller	In each area of the fab there is a computer called a station controller. You use this to call up lots. It also automates a lot of the processing of the lot at that station.
Step	One operation done along the route a lot follows. Routes are a series of steps.
Tests	Electical Test (E-test) test for resistance, capacitance, and voltage characteristics. Sort tests for functionality and speed.
Wafer	A round sheet of silicon upon which many chips are built. D2 is currently running 8" wafers, which means the wafers have a diameter of 8 inches.
Workstream	This is the automation system used in the fabs of Intel. All routes and special instructions for experiments are put into this system. This is also the source for much of the information in Intel's in depth databases.

Appendix B: Description of Process Flow Steps for Experiment Setups

DESIGN OF EXPERIMENT

D2 START REQUEST

Responsibility of : Integration
Time: 5 min to create request.
Description:. You must have a product attached to the route, but you can also do this at the end. Process is initiated with a start request. You start with a baseline route and then put holds on where you want to do experiments.

DOE/SPLIT TABLE

Responsibility of: Module Engineer Module engineers own the tools
Time: Table may take days to make. Anywhere from 1 hour to 1 week to complete depending on complications.

Description: Split table is the plan for what you want done to the wafers. Ideally there is a matrix form you fill out, but sometimes Integration Techs get nothing but a scrap of paper with scribbles. Integration techs then turn this plan into routes. LRTA's are done at this time too. An LRTA is a printed route with changes. The LRTA is later put into Workstream.

EXPERIMENT CHANGE REQUEST

ROUTE/LOTSPEC SETUP

Responsibility of: Integration Tech
Time: 2 days are requested to do this, however Integration Techs often receive start requests on Tuesday morning, which gives them a day before ERC review Wednesday at 1:00 to get them done.

Description: turns a split table into a route. There are subcomponents to this process:

1. New/Special Operations: Paper work to request to do a new split
2. Communicate special operations/special needs to module: Make sure that the module engineers are aware of the experiment. Are the analytical steps in place?
3. Special litho steps? You need to figure out the reticles. If there is a special step you must fill out a Litho Processing Special Request Form (a checklist). This list includes telling litho about the experiment, getting their signature to get the reticles and then attaching the reticles. Also need to talk to the tool owners and the layer owners. All this is turned into the ERC. Getting the approvals can take up to a day.
4. Special analytical steps? Talk to Module Engineers. If the results of analysis are out of the ordinary, the lot will go on hold and the tool will shut down, so it is important to talk to the tool owner. This can talk up to 2 hours.
5. After this step, a confirmation mail is sent to all parties.

SPECS FOR NEW OPERATIONS

Responsibility: Process Engineers

Time: 10 minutes to add a line to the spec of a tool. Up to 1 week to create a brand new spec but this is done very rarely.

Spec changes are only done to production operation steps. Experimental steps do not need specs.

Three types of specs:
1. Local: D2 specific. Requires signatures of Engineer + group leader in Document Control.
2. Shared: Process specific. Needs signature loop from each fab in Virtual Factory (Fab 11, 12, 14, D2).
3. Global: Very generic. Not linked to a specific process. Used across the Virtual Factory. Needs signature loop from each fab in Virtual Factory.

Everything done on Workstream. D2 is the parent site (that is, it is responsible for upkeeping this system for the virtual factory). A new spec or a change in a spec requires signatures. Signature getting can take a long time. Need to go through training before can build a spec. There is a spec building checklist you must fill out.

There are no specs for special operations, since once a spec is created it never goes away and these operations tend to be temporary. The module engineer is meant to update the flash memory specific specs. These specs are for documentation purposes only, so no one will check that they are done. If new specs are not attached the Integration Tech will attach the old specs.

RECIPES IN RCT/TOOL

RCT: Recipe Correlation Table. Contains Operation, product, route, recipe
Responsibility: Integration tech does Implant RCT because there are lots of splits that the Module Engineers can't support. Module places recipes in the tool and does the RCT. Litho does the RTS (Reticle Tracking System which attaches reticle to the operation and route).

SCRIPTS UPDATED

Only done on operations that are not controlled by station controller. The script is a manual way of telling the tool what to do with the lot.

ROUTE/LOTSPEC WRITTEN AND SIGNED OFF BY LOT OWNER

Put the Change Integration Management (CIM) package together, have it reviewed and signed, and send to CIM.

SPC++ REQUEST IN

Module owns this, but they communicate it to Integration. These quality control parameters are needed for the analytical steps. The request involves the station owner, Critical Dimension (CD) owner, and Registration (Reg) owner. A new request can take 30 - 40 hours because they must create new charts. If anything, even the naming convention, is wrong, they must start over. 1 or 2 people work on it at a time. Can be done in parallel with Experiment Review Committee (ERC) approval. It is often not done until the lot is at the operation. For a typical experiment, it takes about 1 hour.

They need to coordinate with the Station Owner. They give him the test name, he puts it in the RCT, checks that it is the right model. If there is different die or different probe, will not work.

RETICLE SET UP

Engineers own specific layers of reticles. They use the Reticle Tracking System (RTS).
This is done somewhere between day 1 and day 6 or even after the lot is started - must simply be done before the reticle is needed. But, it must be signed off before ERC on day 5.

Change Integration Management (CIM)

ROUTE/LOTSPEC AND PAPERWORK IN TO CIM

Integration hands paperwork into CIM group.
Time: Takes about ½ hour per change. Given a 2 day turnaround. Usually takes less than a day. Since logic chip ERC review on Tuesday, it is given priority over flash memory at beginning of the week. Otherwise, it is First Come First Serve (FCFS).

WALK THE ROUTE

Responsibility: TD, Integration, Engineering, Automation
Description: Audits the route to make sure there have been no mistakes that might hurt the line.

SPC++ SET-UP

Responsibility of: automation. It is 10 hours of manual labour. Manual input so there can be errors. Usually no advance warning since it often requested when lot is already set up and needs to be attached to a product. The product must already be attached to a route.

EXPERIMENT SET-UP

Time: Completed between days 2 and 4

LRTA DONE, CHECKED, TO ERC

Responsibility of: Integration adds finished routes (LRTA) + CIM package is put in ERC folder
Time: Day 5

EXPERIMENT REVIEW COMMITTEE (ERC)

ERC APPROVAL/WTR DONE

Responsibility of: ERC
Time: Wednesdays, 1 - 3 p.m.
Description: In this weekly meeting, ERC checks that the lots match the routes and that all splits and merges are good. Integration does not find out until a day or two later about the status of their experiments.

POST ERC

PRODUCT ATTACHED TO RT

Responsibility of: CIM attaches product to route. Production control won't start route unless product is attached. If it isn't , they put it on hold.

Time: Day 6

HOLDS ATTACHED

Responsibility of: Integration
Time: Day 6, 10 min to 1 hour
Description: After approvals, put holds in place for lots and child routes.

ROUTE/LOTSPEC MAILED TO STATION OWNERS

Responsibility of: Integration
Time: Between days 4 and 6
Description: Send cc:mail to everyone about needed RCT setup. The RCT Table Setup is an action required for people. It means that a lot is coming in 2 or 3 days.

TARGET LOT LIST UPDATED

Put lot number into TARGET computer system for reporting purposes.

WAFER START

Responsibility: Production Control
Time to process: 1.5 hours

1. Get start request from integration
2. Check inventory level of experiment wafers. Is there enough room to make a start? There usually is.
3. Start the system by entering the product, route, owner. Starts after midnight of that day. Techs decide when, but must start that day
4. Decide which type of Si to use. This is called Pulling the Si. Check if they have it, put the lot #'s on the Si. Pulling Si takes about ½ hour.
5. Send a note to supervisor saying which lots will start.

Appendix C: Input Data for Simulation Model

This is a subset of the data for one of the processes. Arrival rates were calculated as follows:

1. Split up the operations in the database as has been done above. Each operation represents one of the five critical layers that go through the tools.
2. For each operation,
   Arrival Rate of layer (days) = Time(i) - Time(i-1) (1)
3. Arrival Rate (days) = Average of arrival rates for all operations/5 (2)
   Dividing by 5 is an approximation for the fact that there are actually five layers coming in for a single process. The split between layers that was made is important for the other parameters.

Formulas (1) and (2) give an approximation for the time between arrivals. These were then approximated in minutes and played with so the program could handle the rates. The arrival rate values were run by people in the fab as a sanity check.

Table 12: Arrival Rate Data

Owner	Oper Desc	Wafer Quantity	PT/Wafer (hrs/wafer)	Time	Arrival Interval (days)
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	7	0.09142857	35840.83251	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	7	0.09142857	35840.83251	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	7	0.11142857	35841.65789	0.82538
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	7	0.11142857	35841.65789	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	8	0.0925	35852.80255	11.1447
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	8	0.0925	35852.80255	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	8	0.08875	35852.80299	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	8	0.08875	35852.80299	
TD	FLASH SPECIAL STR SPIN/EXP/DEVELOP	14	0.06285714	35888.31116	35.5082
Mean		8	0.95278351		6.46179
Sum		74			
TD	FLASH FLOATING GATE SPIN/EXP/DEV	13		35798.21361	
TD	FLASH FLOATING GATE SPIN/EXP/DEV	23		35813.87554	15.6619
TD	FLASH FLOATING GATE SPIN/EXP/DEV	1		35814.21289	0.33735
TD	FLASH FLOATING GATE SPIN/EXP/DEV	24		35818.95279	4.7399
TD	FLASH FLOATING GATE SPIN/EXP/DEV	25		35819.89303	0.94024
TD	FLASH FLOATING GATE SPIN/EXP/DEV	24		35819.90921	0.01618
TD	FLASH FLOATING GATE SPIN/EXP/DEV	25		35821.73021	1.821
TD	FLASH FLOATING GATE SPIN/EXP/DEV	25		35821.74128	0.01108
TD	FLASH FLOATING GATE SPIN/EXP/DEV	25		35823.91216	2.17088
TD	FLASH FLOATING GATE SPIN/EXP/DEV	24		35830.44921	6.53705
TD	FLASH FLOATING GATE SPIN/EXP/DEV	25		35841.92936	11.4802
Mean		21			4.04435
Sum		234			
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	18		35798.03485	
TD	FLASH SPECIAL SMS S/E/D	4		35798.03563	0.00078
TD	FLASH SPECIAL SMS S/E/D	4		35798.03563	
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	24		35798.04088	0.00525
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	23		35808.31534	10.2745
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	24		35819.25287	10.9375
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	25		35825.06508	5.81221
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	25		35825.17792	0.11284
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	24		35825.66257	0.48465
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	24		35825.71046	0.04789
TD	FLASH SAMOS SPIN/EXPOSE/DEVELOP	25		35828.23147	2.52101
Mean		20	12.3302589		1.52947
Sum		220			
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	19	0.09736842	35796.05808	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	19	0.09736842	35796.05808	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	25	0.0504	35799.58749	3.52941
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	25	0.0504	35799.58749	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	23	0.06652174	35800.50655	0.91906
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	23	0.06652174	35800.50655	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	24	0.10583333	35801.75707	1.25052
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	24	0.10583333	35801.75707	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	20	0.354	35805.04082	3.28375
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	20	0.354	35805.04082	
TD	FLASH CONTACT SPIN/EXPOSE/DEVELOP	22	0.08772727	35805.0941	0.05328
Mean		22	12.3302589		1.52947
Sum		244			

Appendix D: Layer Qualification Tables for Tools

Table 13: Theoretical Layer Qualification Table

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
Prod Layer 1	Q*	Q	Q*	DQ	Q	Q	Q
Prod Layer 2	DQ	DQ	DQ	Q	Q	Q	DQ
Prod Layer 3	Q	DQ	DQ	Q	DQ	Q	DQ
Prod Layer 4	DQ	Q	Q	Q	Q	Q	Q
Prod Layer 5	Q	DQ	Q	Q	Q	Q	DQ

Q=Qualified DQ= Disqualified

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PII Layer 1	Q	Q	Q	DQ	DQ	DQ	Q
PII Layer 2	DQ	Q	DQ	Q	Q	Q	DQ
PII Layer 3	DQ	Q	DQ	Q	Q	Q	DQ
PII Layer 4	Q	DQ	DQ	Q	DQ	Q	DQ
PII Layer 5	DQ	Q	Q	Q	Q	Q	Q

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PI Layer 1	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 2	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 3	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 4	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 5	DQ	DQ	Q	DQ	DQ	DQ	DQ

Table 14: Soft Dedication Qualification Table

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
Prod Layer 1	DQ	Primary	DQ*	DQ	Secondary	DQ	DQ
Prod Layer 2	DQ	DQ	DQ	Secondary	Secondary	Primary	DQ
Prod Layer 3	Primary	DQ	DQ	Secondary	DQ	Secondary	DQ
Prod Layer 4	DQ	DQ	DQ	DQ	Secondary	DQ	Primary
Prod Layer 5	DQ	DQ	DQ	Secondary	Primary	DQ	DQ

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PII Layer 1	Primary	Secondary	DQ	DQ	DQ	DQ	DQ
PII Layer 2	DQ	Primary	DQ	Secondary	DQ	Secondary	DQ
PII Layer 3	DQ	Secondary	DQ	Primary	DQ	Secondary	DQ
PII Layer 4	Primary	DQ	DQ	Secondary	DQ	DQ	DQ
PII Layer 5	DQ	DQ	DQ	DQ	Secondary	DQ	Primary

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PI Layer 1	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 2	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 3	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 4	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 5	DQ	DQ	Q	DQ	DQ	DQ	DQ

Table 15: Semisoft Dedication Qualification Table

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
Prod Layer 1	Q*	Q	Q*	DQ	Q	Q	Q
Prod Layer 2	DQ	DQ	DQ	Secondary	Secondary	Primary	DQ
Prod Layer 3	Q	DQ	DQ	Q	DQ	Q	DQ
Prod Layer 4	DQ	Q	Q	Q	Q	Q	Q
Prod Layer 5	Q	DQ	Q	Q	Q	Q	DQ

Q=Qualified DQ= Disqualified

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PII Layer 1	Q	Q	Q	DQ	DQ	DQ	Q
PII Layer 2	DQ	Primary	DQ	Secondary	DQ	Secondary	DQ
PII Layer 3	DQ	Secondary	DQ	Primary	DQ	Secondary	DQ
PII Layer 4	Q	DQ	DQ	Q	DQ	Q	DQ
PII Layer 5	DQ	Q	Q	Q	Q	Q	Q

Layer/Tool	DL01	DL02	DL03	DL04	DL05	DL06	DL07
PI Layer 1	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 2	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 3	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 4	DQ	DQ	Q	DQ	DQ	DQ	DQ
PI Layer 5	DQ	DQ	Q	DQ	DQ	DQ	DQ

In these tables, a Q means a type of lot is qualified to go to that tool. DQ means the lot is disqualified. A primary tool gets higher priority to a lot than a secondary tool. A PI process is the first phase of development, called Process Development. A PII process is the second phase of development, called Process Improvement and Characterization. Sustainment Development and Production lots have the same layer qualification tables.

Appendix E: Output Data

Table 16: Raw Output Data

LOT	LAYER	WAFERS	ARRIVAL	QUEUE	TOOL	SETUP	PROCESS	CYCLE
3	1	18	0	0	2	5	68.14	73.14
3	4	25	0	0	7	40	91.81	131.81
4	2	24	0	0	6	66.67	87.93	154.59
4	5	24	30	0	5	66.67	86.7	153.37
2	2	1	0	84.13	2	8.33	100	192.46
4	3	25	90	0	1	40	94.3	134.3
5	2	25	0	131.81	7	10	89.34	231.16
4	2	25	120	34.59	6	0	77.04	111.63
3	5	24	147.24	36.13	5	8.33	93.24	137.7
4	1	25	60	132.46	2	0	95.13	227.59
4	3	25	180	44.3	1	0	77.04	121.34
3	4	1	147.24	83.91	7	20	100	203.91
1	1	23	0	0	3	1.67	96.23	97.9
4	4	24	210	74.94	5	26.67	86.27	187.88
4	1	24	150	137.59	2	0	113.02	250.61
4	1	24	300	97.88	5	0	73.83	171.71
4	2	24	420	0	6	0	73.83	73.83
1	3	25	0	384.2	3	6.67	110.14	501.01
4	3	1	480	0	1	0	40	40
4	4	25	330	21.16	7	6.67	92.24	120.07
4	5	25	360	111.71	5	0	77.04	188.75
4	4	1	510	21.74	7	0	40	61.74
4	3	25	540	0	1	0	77.04	77.04
4	1	25	570	0	5	0	77.04	77.04
3	2	9	643.55	0	4	50	39.96	89.96
4	1	25	240	160.61	2	0	111.01	271.62
4	2	23	630	0	6	0	70.62	70.62
4	2	25	660	73.52	4	3.33	90.32	167.17
3	2	1	643.55	106.38	6	3.33	100	209.72
4	1	25	270	468.37	2	0	121.14	589.51
3	3	24	643.55	0	1	5	88.69	93.69
4	1	25	600	47.04	5	0	77.04	124.08
4	4	25	840	25.2	7	0	77.04	102.24
5	2	25	612.11	105.75	3	46.67	89.69	242.1
4	1	25	390	469.51	2	0	108.97	578.49
4	2	25	690	163.27	6	3.33	96.59	263.19
4	2	25	900	0	4	0	77.04	77.04
4	2	25	720	188.47	5	0	77.04	265.51
4	4	25	930	12.24	7	0	77.04	89.28
2	4	7	951.8	37.61	4	0	44.52	82.13
4	1	25	810	175.51	5	0	77.04	252.55
4	3	25	870	106.56	6	33.33	77.04	216.94
4	1	25	450	518.49	2	0	127.09	645.58
1	2	1	632.15	319.06	3	13.33	150	482.4
4	3	25	780	120.39	1	5	92.17	217.55

Table 17: Lot Statistics

	LOT					
Data	1	2	3	4	5	LOT
Average of QUEUE	256.80	253.73	510.26	732.66	619.46	#DIV/0!
Min of QUEUE	0.00	0.00	0.00	0.00	0.00	0.00
Max of QUEUE	1865.64	2100.17	4387.86	4556.47	4329.75	0.00

(Note: Ignore LOT column. This happened when there were fewer lots than columns used in the Excel pivot table)

Table 18: Average Lot Statistics

	TD	Production
Average of QUEUE	5.67	12.21
Min of QUEUE	0.00	0.00
Max of QUEUE	73.13	75.94

(Note: TD is an average of the queue times for LOTS 1,2,3. Production is Lot 4. Lot 5 is Engineering, which is ignored)

Table 19: Sample of Data from Cumulative Frequency of Queue Times for TD and Production

TD			Production		
Queue Time (min)	*Frequency*	*Cumulative %*	*Queue Time (min)*	*Frequency*	*Cumulative %*
0	385	15.45%	0	1065	10.68%
5	15	16.05%	5	65	11.34%
10	20	16.85%	10	72	12.06%
15	20	17.66%	15	55	12.61%
20	19	18.42%	20	87	13.48%
25	19	19.18%	25	60	14.09%
30	13	19.70%	30	52	14.61%
35	18	20.43%	35	56	15.17%
40	18	21.15%	40	72	15.89%
45	13	21.67%	45	68	16.57%
50	22	22.55%	50	103	17.61%
55	19	23.31%	55	62	18.23%
60	21	24.16%	60	72	18.95%
65	29	25.32%	65	84	19.79%
70	22	26.20%	70	56	20.36%
75	20	27.01%	75	59	20.95%
80	17	27.69%	80	82	21.77%
85	25	28.69%	85	59	22.36%
90	24	29.65%	90	59	22.95%
95	19	30.42%	95	60	23.56%
100	29	31.58%	100	56	24.12%
105	29	32.74%	105	51	24.63%
110	25	33.75%	110	46	25.09%
115	19	34.51%	115	53	25.62%
120	20	35.31%	120	51	26.13%
125	24	36.28%	125	51	26.65%
130	18	37.00%	130	49	27.14%
135	22	37.88%	135	47	27.61%
140	18	38.60%	140	49	28.10%
145	21	39.45%	145	53	28.63%
150	19	40.21%	150	53	29.16%
155	18	40.93%	155	51	29.67%
160	24	41.89%	160	48	30.16%
165	14	42.46%	165	51	30.67%
170	25	43.46%	170	48	31.15%
175	21	44.30%	175	42	31.57%
180	19	45.06%	180	48	32.05%
185	15	45.67%	185	57	32.62%
190	11	46.11%	190	41	33.04%
195	17	46.79%	195	43	33.47%
200	12	47.27%	200	45	33.92%
205	13	47.79%	205	43	34.35%

Appendix F: Model Validation Data Input File

Table 20: Sample Input Data File Used in Model Validation

Owner	Lot	Arrived Date	Arrived Time	Wafer Quantity	CT (hours)	PT (HOURS)	QT (hours)
TD	2810832008	8/6/98	20:05	2	44.64	1.12	43.52
TD	2810832008	8/6/98	20:05	2	44.64	1.12	43.52
TD	2810832008	8/9/98	20:25	2	49.11	0.40	48.72
TD	2810832008	8/9/98	20:25	2	49.11	0.40	48.72
TD	28378990	9/17/98	14:08	25	30.08	2.69	27.39
TD	28378990	9/19/98	10:14	25	7.14	7.14	0.00
TD	28268530	7/21/98	3:19	24	39.37	2.17	37.20
TD	28278570	8/13/98	6:59	12	29.50	1.04	28.46
TD	28248940	7/15/98	18:27	25	23.83	2.95	20.89
TD	28248940	7/15/98	18:27	25	23.83	2.95	20.89
TD	28258960	7/15/98	20:30	22	20.40	1.66	18.74
TD	28258960	7/15/98	20:30	22	20.40	1.66	18.74
TD	28278580	8/11/98	20:46	20	85.17	1.32	83.85
TD	28278580	8/11/98	20:46	20	85.17	1.32	83.85
TD	2827858800	8/12/98	23:43	1	9.85	0.43	9.42
TD	2827858800	8/12/98	23:43	1	9.85	0.43	9.42
TD	28278580	8/15/98	18:14	19	7.38	1.37	6.01
TD	28278580	8/15/98	18:14	19	7.38	1.37	6.01
TD	28288590	8/16/98	14:26	25	6.21	1.36	4.85
TD	28288590	8/16/98	14:26	25	6.21	1.36	4.85
TD	28268530	7/5/98	20:57	12	8.74	0.89	7.85
TD	28268530	7/5/98	20:57	12	8.74	0.89	7.85
TD	2826853003	7/5/98	20:57	13	29.38	0.84	28.53
TD	2826853003	7/5/98	20:57	13	29.38	0.84	28.53
TD	28278580	7/10/98	23:01	19	18.83	1.32	17.51
TD	28278580	7/10/98	23:01	19	18.83	1.32	17.51
TD	2827858001	7/10/98	23:23	6	18.47	0.78	17.70
TD	2827858001	7/10/98	23:23	6	18.47	0.78	17.70
TD	2827858001	7/12/98	13:29	6	9.55	0.83	8.72
TD	2827858001	7/12/98	13:29	6	9.55	0.83	8.72
TD	28278580	7/12/98	15:12	19	9.25	1.49	7.76
TD	28278580	7/12/98	15:12	19	9.25	1.49	7.76
TD	28278570	7/18/98	8:20	25	81.46	1.36	80.11
TD	28278570	7/18/98	8:20	25	81.46	1.36	80.11
TD	28298630	7/19/98	9:11	16	76.61	1.29	75.31
TD	28298630	7/19/98	9:11	16	76.61	1.29	75.31