Production Management Policies in a Multiple Product Semiconductor Fabrication Facility

by
Susan E. Perrin

B.S. Mechanical Engineering, Massachusetts Institute of Technology, 1991

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

Master of Science in Management
and
Master of Science in Mechanical Engineering

at the
Massachusetts Institute of Technology
June 1997

©1997 Massachusetts Institute of Technology, All rights reserved

---

Signature of Author ______________________________
Sloan School of Management
Department of Mechanical Engineering

Certified by ______________________________
Dr. Stanley Gershwin, Senior Research Scientist, Thesis Advisor
Department of Mechanical Engineering

Certified by ______________________________
Dr. Donald Rosenfield, Senior Lecturer, Thesis Advisor
Sloan School of Management

Accepted by ______________________________
Jeff A. Barks, Associate Dean, Sloan Master's and Bachelor's Programs

Accepted by ______________________________
Ain A. Sonin, Chairman, Department of Committee on Graduate Students

JUL 21 1997

LIBRARIES
Eng.

Production Management Policies in a Multiple Product Semiconductor Fabrication Facility

by
Susan E. Perrin

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

Master of Science in Management
and
Master of Science in Mechanical Engineering

Abstract
This thesis examines production management policies in a multi-product, multi process-semiconductor fabrication facility. A dynamic simulation model of key functional areas was developed in order to compare and contrast cycle time performance metrics for different work in process selection methodologies. These policies include both global policies and local "functional area" policies, and both lot-specific and non-specific policies. The necessary infrastructure and incentive implications for implementation of top performing policies are examined.

Use of a lot-specific calculated methodology called "X-Theoretical" for choosing lots was the best methodology studied in terms of cycle time and cycle time variation metrics. This global, lot-specific methodology decreased modeled area mean cycle time performance by 11% and 95% Cycle Time by 20% over the baseline "Back-To-Front" methodology.

This model was also utilized to pinpoint appropriate functional areas for performance improvement efforts. Finally, the model was used to predict performance improvement if recommendations were followed.

Organizational metrics and incentives are also examined. Through a series of iterations in model development, the total machine time lost due to technicians not being available was incorporated into the model. This is theorized to be similar in magnitude to the total amount of unscheduled machine down time in most of the modeled areas. This down time was targeted for 50% improvement in bottleneck areas. Excess capacity generated via this improvement could be used to increase fab output by 3.5%.

Thesis Advisors:
Dr. Stanley Gershwin, Mechanical Engineering
Dr. Donald Rosenfield, Sloan School of Management

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

Acknowledgments

Employees throughout Intel were thorough and prompt in their responses to benchmarking information requests. The people at Fab 8 in Jerusalem were unfailingly helpful and patient in supplying information, reviewing output, and generally explaining their operations. They have earned both my gratitude and an immense amount of respect.

There is not enough space here to acknowledge each of them by name. Yonathan Wand's leadership, mentorship and support were all that I could ask for in a manager. Amir Dayan was always willing to make time in a very busy schedule to share his wisdom and perspective. The members of the MFS Group, especially Yossi Eliyha, Shai Yonayov, Avshalom Halevy, Lucie Hamami, Leizy Harris, and Bentzy Shoval were a constant source of humor, information, contacts, and reality checks. Daniel Ohayon provided invaluable historical fab performance data with a quick turn around time and professional results. Pinhas Koren and his entire Industrial Engineering Group (especially Efrat Aran, Aviv Lichtman and Zahi Shaked) provided machine performance data and modeling advice. Inside the clean room, everybody helped! Alon Gal, Shmulik Perez, Eyal Shafir, and Gideon Shteier provided information on how the clean room functioned. Yoel Bleier, Avidan Gal, and Giora Ben-Shoshan displayed patience and humor running shift meetings in English. Eli Sananes provided product demand and wafer start information. Benny and Beverly Adams, Yossi and Nurit Eliyha, Helen Ghould and Orli Meriems provided countless tips on how to live in a country which sometime felt very foreign.

My studies in management have convinced me that a manager does indeed bear some responsibility for the performance of his organization. Therefor, I take off my hat to Gadi Dvir!

Many people at MIT have provided academic support. Don Rosenfield and Stan Gershwin, my academic advisors, provided references and input throughout the internship.

During my internship and throughout the past two years at MIT, every one of my 44 colleagues in Leaders For Manufacturing has provided support in one form or another. I look forward to collaboration with them for many years! Members of Group 7 and the folks from Endicott Avenue have been especially helpful in sharing experiences, planning career activities, presentations, checking equations, and generally helping me to grow as a professional and a leader. I am grateful to Professor Shoji Shiba for teaching us that Fun and Joy are essential elements of professional endeavors.

Finally, I am grateful to my family and friends. Mom, Dad, Cynthia, Barry, Henry and Carrie, you have been my strongest supporters for many years. Thank you. I look forward to many wonderful thoughts and experiences with Christa, Gwen, Mark, Choong, Lillian, John, Ann, Mara and Jenny. Most of all, thanks to my fiancé Jamie for supporting me through two years of a long-distance relationship. We may take a lifetime to prove to ourselves that dual-career relationships can work!

Table of Contents

1. INTRODUCTION **9**

1.1 PROBLEM DEFINITION 9
1.2 WORK IN PROCESS MANAGEMENT METHODOLOGIES 10

2. BACKGROUND INFORMATION **13**

2.1 SEMICONDUCTOR INDUSTRY OVERVIEW 13
2.2 SEMICONDUCTOR FABRICATION PROCESS 14
2.3 FAB HISTORY 15
2.4 FAB PERFORMANCE METRICS EMPHASIS AND COMMUNICATION 16
2.5 OTHER RELEVANT MANUFACTURING POLICIES 19
2.6 PRIORITY LOTS 20
2.7 CYCLE TIME AND VARIATION 21
2.8 FINANCIAL MODELS 22

3. PERFORMANCE DATA AND POLICY BENCHMARKING **25**

3.1 HISTORICAL FACILITY PERFORMANCE DATA 25
3.2 BENCHMARKING WITHIN INTEL AND THE INDUSTRY 26
3.3 ALTERNATIVES TO MODELED METHODOLOGIES 27
3.4 PUBLISHED ACADEMIC INFORMATION 28
3.4.1 WIP Management Methodologies *28*
3.4.2 Loading Change Effect on Mean Cycle Time and Variance Under Several WIP Management Policies .. *29*

4. SIMULATION MODEL DEVELOPMENT **31**

4.1 INITIAL MODEL INPUTS 31
4.2 OUTPUT VERIFICATION AND ITERATION PROCESS FOR INPUTS 32

5. SIMULATION RESULTS **37**

5.1 WIP MANAGEMENT EFFECTS 38
5.2 DOWN TIME EFFECTS 39
5.3 STATISTICAL AND SENSITIVITY EFFECTS 40
5.4 POLICY RECOMMENDATION EFFECTS 40

6. RECOMMENDATIONS **43**

6.1 SYSTEMS DESIGN AND IMPLEMENTATION 43
6.1.1 Design and Implement Metrics Educational Programs *43*
6.1.2 Develop Information Systems To Improve Technician Response Times In Bottleneck Areas *45*
6.1.3 Develop Cycle Time Variation Measurement and Reduction Program *48*
6.2 SEMICONDUCTOR FAB OPERATIONS 49
6.2.1 Retain Back-To-Front Policy For The Fab *49*
6.2.2 Increase Wafer Starts *50*
6.2.3 Run Lithography Layers To 0 .... *51*
6.3 FORM A PERMANENT LINE MANAGEMENT CONTINUOUS IMPROVEMENT TEAM 51

7. UNRESOLVED ISSUES AND FURTHER DEVELOPMENTS **53**

7.1 SIMULATION MODEL IMPROVEMENTS 53
7.2 FACILITY SPECIFIC CONCERNS 53

APPENDIX A: SEMICONDUCTOR FACILITY DATA **55**

APPENDIX B: SIMULATION INPUT **62**

APPENDIX C: APPENDIX OF SIMULATION RESULTS **68**

REFERENCES **69**

Table of Figures

FIGURE 2.1: METRICS EMPHASIS THROUGHOUT THE FABRICATION FACILITY 18
FIGURE 6.1: LACK OF CLEAN ROOM ORGANIZATIONAL ALIGNMENT 44
FIGURE 6.2: CLEAN ROOM ORGANIZATIONAL ALIGNMENT 45
FIGURE 6.3: TECHNICIAN EFFECT ON CAPACITY 46

Appendixes

APPENDIX A: SEMICONDUCTOR FACILITY DATA 55
EXHIBIT 1: WAFER START SCENARIO 55
EXHIBIT 2 FAB + SORT CYCLE TIME DATA 56
EXHIBIT 3: SAMPLE WEEK CYCLE TIME DISTRIBUTION AND NORMALITY ANALYSIS 57
EXHIBIT 4: ACTUAL LOT CYCLE TIME DATA 58
EXHIBIT 5: REGRESSION FIT TO AGGLOMERATE CALCULATED CYCLE TIME 59
EXHIBIT 6: CYCLE TIME HISTOGRAM 60
EXHIBIT 7: CYCLE TIME VERSUS AVERAGE BATCH SIZE 61

APPENDIX B: SIMULATION INPUT 62
EXHIBIT 8: SAMPLE SIMULATION INPUT FILE 62
EXHIBIT 9: SAMPLE STATION DEFINITION FILE 63
EXHIBIT 10: SAMPLE SIMULATION PROCESS FLOW 64
EXHIBIT 11: SAMPLE SIMULATION DOWN TIME TABLE 65
EXHIBIT 12: SAMPLE STATION DOWN TIME ATTACHMENT FILE 66
EXHIBIT 13: SAMPLE SIMULATION RULE EDITOR 67

APPENDIX C: APPENDIX OF SIMULATION RESULTS 68
EXHIBIT 14: TABULAR SUMMARY OF SIMULATION RESULTS 68
EXHIBIT 15: AVERAGE LOT CYCLE TIME DATA FOR WIP MANAGEMENT STRATEGIES 70
EXHIBIT 16: ANOVA AND TUKEY-KRAMER TESTS FOR WIP MANAGEMENT METHODS MEAN CYCLE TIME COMPARISON 71
EXHIBIT 17: 95% LOT CYCLE TIME DATA FOR WIP MANAGEMENT STRATEGIES 72
EXHIBIT 18: ANOVA AND TUKEY-KRAMER TESTS FOR WIP MANAGEMENT METHODS 95% CYCLE TIME COMPARISON 73
EXHIBIT 19: AVERAGE LOT CYCLE TIME DATA FOR SIMULATED DEMAND CHANGES 74
EXHIBIT 20: 95% LOT CYCLE TIME DATA FOR SIMULATED DEMAND CHANGES 75
EXHIBIT 21: AVERAGE LOT CYCLE TIME DATA FOR SIMULATED DEMAND PEAKS 76
EXHIBIT 22: 95% CYCLE TIME DATA FOR SIMULATED DEMAND PEAKS 77
EXHIBIT 23: AVERAGE LOT CYCLE TIME DATA FOR SIMULATED DECREASED DOWN TIMES 78
EXHIBIT 24: 95% LOT CYCLE TIME DATA FOR SIMULATED DECREASED DOWN TIMES 79
EXHIBIT 25: AVERAGE CYCLE TIME FOR STATISTICAL TESTS 80
EXHIBIT 26: ANOVA AND TUKEY-KRAMER TESTS FOR MEAN CYCLE TIME COMPARISON 81
EXHIBIT 27: 95% CYCLE TIME FOR STATISTICAL TESTS 82
EXHIBIT 28: ANOVA AND TUKEY-KRAMER TESTS FOR 95% CYCLE TIME COMPARISON 83
EXHIBIT 29: AVERAGE LOT CYCLE TIME DATA FOR RECOMMENDED ACTIONS 84
EXHIBIT 30: 95% LOT CYCLE TIME DATA FOR RECOMMENDED ACTIONS 85

1. Introduction

1.1 Problem Definition

This research was carried out in order to better understand an Intel semiconductor fabrication facility (*fab*) and to determine the best choice of production methods for this fab. The research was performed at a multi-product, multi-process Intel semiconductor fab in Jerusalem, Israel. Because operations are analyzed and managed by different functional groups, a particularly important corresponding issue is whether the facility performance was being compromised in order to locally optimize within functional areas.

In order to systematically compare different strategies, a model was developed in AutoSched© simulation software. The choice between granularity and accuracy of the model was critical for the project: model input would have to be detailed enough to give believable output, but also allow for data gathering, input and problem solving in the six month project schedule. The choice was made to include eight different functional areas out of approximately twenty. Each process step and corresponding machine performance data is included in the model.

Throughout the project, data was gathered and confirmed by area experts. A support team consisting of the fab manufacturing manager, industrial engineering manager, planning manager, shift managers, several group leaders, and several technicians from the fab was formed. They provided invaluable review and input throughout the project. For machine down times, experts included machine owners (engineers), technicians who operate the machines, and Industrial Engineers. For demand information, expertise in the planning group was utilized. Although the input data is disguised throughout this report, output data's relative performance is maintained. It is disguised by removing axis labels and tabular number specifics.

Semiconductor fabrication processes exhibit a phenomenon know as "Reentry Flow." At any particular station, numerous steps in the process occur. It is not unusual for products to return to some station types more than 10 times. Thus a technician or automated loading system may face a choice of processing one of many different types of products at one of 10 layers. What methodology, if any, should he use in specifying the next product to run?

Different strategies for work in process (*WIP*) management and global output for improved machine performance were examined .

1.2 Work In Process Management Methodologies

WIP moves manually through the fab. Technicians place it on a set of racks near the machine area in which it will next be processed. Next, they move these lots into machines or machine queues, and afterwards off of the machine onto small carts. When a cart fills up, whoever has filled it moves each lot to the following operation, which may be anywhere in the fab.

At any station area, technicians may either choose from lots in queue based on some systematic methodology, or based on none at all. These decisions can be made by a global fab strategy, a local, area-specific strategy, or none at all. Even with a comprehensive global fab strategy, individual areas make some exceptions based on downstream operational time limits, batch sizes and set-ups. During the project, the following strategies were discussed:

- FIFO: First-in-first out. The technician processes the first lot in the local functional area queue.
- Random: When a station becomes available, local lots in queue are assigned a random number. The technician processes the lot with the highest number.
- Shortest Processing Time: The technician prioritizes lots based on its processing time through her area, with shortest lots chosen first. This is a *local* version of the traditionally *global* Shortest Processing Time (*SPT*). Choices between equal priority are determined by earliest arrival time at the functional area.
- Back-To-Front: Technicians prioritize back-end lots (lots processed through the first metal layer) over front-end lots (lots not processed through the first metal layer). Furthermore, each successive operation in the back-end increases a lot's priority. All front-end lots have 0 priority. Lots are processed based on a given lot's priority, with the choice between lots of equal priority determined by earliest arrival to the functional area. This is a variation of a *global* Shortest Processing Time policy, with the change that all lots in the front-end have equal priority.
- Front-To-Back: Technicians prioritize front-end lots over back-end lots. Each successive operation in the back-end decreases a lot's priority. Front-end lots have highest priority.

Lots are processed based on a given lot's priority, with the choice between lots of equal priority determined by earliest arrival time at the functional area.

- X-Theoretical: The technician prioritizes lots based on a calculated algorithm for choosing the lot which is "most behind." The formula for determining which lot is furthest behind is:

(planned processing time to current step)/(actual processing time to current step) *Equation 1.1*

Lots with the lowest value as determined by Equation 1.1 are processed first. Choices between equal priority lots are determined by earliest arrival time at the functional area.

2. Background Information

2.1 Semiconductor Industry Overview

In 1964, Gordon Moore formulated a prediction which later became know as Moore's Law. He predicted that silicon chip density would double every 18 months, and device cost would decrease proportionately. Since that time, this prediction has proven true (Frantz and Tatge, 1997).

Worldwide spending for IT products and services was $610 billion in 1996 (Isaacson, Portia, 1997). The semiconductor industry has grown tremendously in the years since 1971, when a company founded by Moore and Noyce announced the first microprocessor. In 1994, semiconductors alone accounted for roughly $91 billion in the world market (Page, 1994.) By the year 2000, they are expected to account for $300 billion of this market (Robinson, 1997.) In 1995, Intel's revenues were $16.2 billion.

Growth in the microprocessor industry has been largely fueled by device cost decreases and speed increases due to advances in processing technology. These advances in processing technology have created progressively higher demands in equipment sophistication. Projected costs for 300-mm wafer fabs with .35-micron technology are on the order of $2 billion (Frantz and Tatge, 1997.) This price tag is prohibitive for many small-volume producers. Thus economies of scale are causing an industry consolidation trend in wafer manufacturers.

With this magnitude of equipment investment, semiconductor fabs face tremendous pressure to utilize equipment efficiently. However, in many fabs production machinery processes wafers at full speed only 30% of the time, according to Sematech research (Ristelhueber, 1997.) This is partially due to the inherent conflict between the desire to create revenue though high output per machine, and the desire to deliver products quickly to customers. The former involves high in-process inventory, and the latter low in-process inventory. Any operational methods which can either free this expensive machine capacity or decrease in-process inventory without sacrificing output are worthy of study.

Microprocessors exemplify the high-end products of the semiconductor market. Constantly upgraded designs and process technologies allow producers to charge a premium for them. Microcontrollers, on the other hand, are considered a commodity product. They are not produced on the highest technology equipment. A customer can have the reticles (lithography

masks) created at a competing fab for free, and can thus switch suppliers. Competition in the microcontroller market is based on quality, cost and delivery time.

In the quality arena, Intel leverages its excellent reputation, technical support, reliability and a backwards compatibility guarantee on architecture. In cost, this fab's performance is also world class. However, delivery time performance is fair compared with the 6 week industry benchmark. This benchmark is set at co-located wafer fabrication and assembly facilities where fabrication processes are designed to avoid product differentiation until the end of the process.

For Intel processes, most product differentiation occurs within the first few processing steps. With this first lithography step, stocking undifferentiated partially-fabricated wafers to shorten lead time becomes impossible. On the other hand, solving the response time problem by stocking large quantities of finished (but unassembled) dies of hundreds of products is undesirable in terms of inventory holding costs and product quality implications. Within the corporation, business process reengineering efforts are being undertaken in order to reduce order fulfillment time. However, within the fab operations, the time for wafers to move through the fab and sort operations remains by far the largest component of this lead time. Thus, targeting cycle time reduction at the fab level is an important option.

2.2 Semiconductor Fabrication Process

In the semiconductor device fabrication, value-added steps of lithography, diffusion, etch, chemical vapor deposition and implant are repeatedly performed in order to produce a semiconductor device. The initial processing begins with a silicon wafer. Successive operations form the electrical sources, sinks, insulators, and interconnects through a series of layering operations. These operations consist of either material deposition or removal, sometimes applied to select chip areas, and sometimes applied to the entire available surface. In typical semiconductor industry logic fabs, wafers move through the lithography area twenty times.

These steps are combined with in-process and end-of-process inspection steps to complete a wafer in the fab clean room. The name "clean room" is derived from the efforts made to avoid exposing wafers to potential defects due to any sources of contamination. A Class 1 clean room may contain up to 1 particle per cubic foot of atmosphere per minute. The time units in this definition account for constant flow out of the room as the atmosphere is "scrubbed" or cleaned.

Most final fab operations involve plating of the wafers. The plated wafer then leaves the fab area to enter the less clean (and therefor less expensive per cubic foot) sort area. Here, the manufacturing unit transition from wafer to die is made. Electrical tests are performed on each individual die and bad die are marked on the wafer. At the completion of this testing, wafers are shipped to assembly facilities. Assembly processing involves physical cutting of the wafer into individual dies, assembly of the die into packaged transistors, and final electrical tests.

In the semiconductor industry, the term "process" refers to the ordered steps used in creating a functional "die" which becomes a chip after assembly. Typically processes are categorized by the resolution of metal circuit line and gate widths which they can produce. Successive process generations have focused on decreasing this width. The ultimate resolution limit has traditionally been set by lithography, which, as a photography-type operation, is limited physically by the wavelength of light.

2.3 Fab History

This research was performed at a 13 year old Intel fab located in Jerusalem, Israel. In the semiconductor industry, this is considered and "old" fab, and as such this fab has transitioned through five generations of manufacturing processes. Of these five processes, three are currently run. Roughly one third of the equipment must be upgraded to make a transition. This adds to the challenge in running the operation, since equipment transitions are carefully executed in order to minimize effect on output, and since it translates into a significant amount of shared equipment between processes. In June of 1996, most of the newest process output was used for microprocessors, but by the end of the year, the process would be used primarily for microcontrollers.

The constant process upgrading has created several differentiating characteristics for this particular fab. Since process equipment has tended to become bigger, space in the clean room has become extremely valuable. Personnel in the fab have developed an expertise in managing and executing equipment upgrades between processes. Another unusual fab characteristic is due to its Jerusalem location: all processes are shut down weekly for the approximately 24 hours of Shabbat.

The facility produces many end-of-life products and as such is frequently the sole supplier. Customers include dozens of companies ranging from internal Intel customers to automotive, military and personal computer industries.

Product demand exceeds supply, and thus the fab operates at capacity. Due to the expense of capital equipment in the semi-conductor industry, this line is operated as a "balanced line." Translated, this term means that many different machine types have roughly the same amount of capacity, which is almost equal to fab output. This near-capacity loading at numerous stations produces multiple bottlenecks. This tends to lend itself to a roving line management technique. In a given week, the fab starts dozens of products, which must be completed on schedule for customers. This significantly adds to the challenge.

In addition to the weekly shutting down for Shabbat, machine capacity includes processing time, preventative maintenance time, unpredicted down time allowance, an assumed lot size for set-ups and a delta figure for safety margin. Managers of capacity expansion efforts utilize these industrial engineering figures to highlight machines which will not be able to meet capacity.

2.4 Fab Performance Metrics Emphasis and Communication

WIP management and manufacturing strategies are communicated and implemented through a network of personnel. During a once-per-week management update, presentations on specific current efforts and line state are made for the fab manager. During tri-weekly operations meetings, representatives from each functional manufacturing area, planning, facilities, shift managers and engineering meet to present line status and set action items. In reality, most of the day-to-day line prioritization operations are handled by the shift managers.

Following a shift "pass-down," where each person from shift manager to Self Sustaining Technician (*SST*) highlights significant issues and shift performance to his replacement, the new shift manager has a half hour meeting with group leaders from each functional area in the clean room. During this meeting, the shift manager directs discussion on safety issues, priority lots, metrics performance, how to handle WIP, machines down, and any other relevant issues. Decisions are made somewhat by consensus. There is no set procedure for how to handle machines down, so shift managers differ a fair amount in focus. This meeting is followed by a group leader verbally handing down some of the same fab information to his group, further focusing on his area performance. Group leaders do not report to shift managers, but instead to a functional manager, so there is more of a matrix relationship.

The complexity of the semi-conductor fabrication process, the three different processes running in the fab and the information-rich environment combine to give an array of metrics. For the fab manager, the most important metrics are die cost and LIPAS, or line item performance as

scheduled. Die cost represents the facility's fully absorbed cost per die. LIPAS is a binary metric for each order, in which a date is either met or not met. The scheduled date for each product is based on verbal agreements between planning and manufacturing personnel. Forecasts for schedule time include a safety factor plus projected die yield factor (historic by product), line yield factor (historic by process), and average cycle time (based on current cycle time). Occasionally, dates are renegotiated with the customers if an order will not be met. Otherwise, a potentially late item becomes a priority lot.

For the manufacturing manager, the most important metric is LIPAS, followed by line throughput time and line throughput time variability. Throughput time (*TPT*) is an aggregate calculation by process, based on a variation of inventory-turns calculations. For each day, total initial WIP (Beginning on Hand) and total number of wafers processed processed (Outs) and scrapped, are related to agglomerate TPT as follows:

Equation 2.1

$$TPT = \frac{\Sigma(\text{Beginning on Hand})}{\Sigma(\text{Outs} + \text{Rejects})}$$

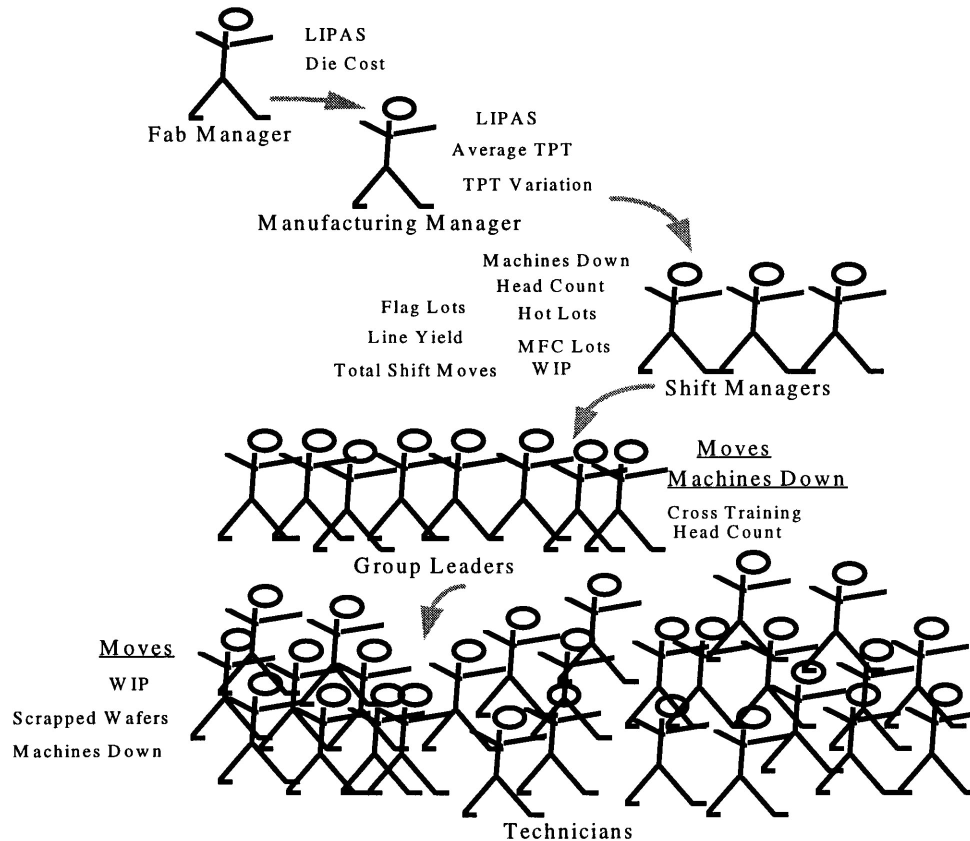
For example, if in a given day Beginning on Hand is 100 wafers, and during the day 9 units are processed (Outs) and 1 wafer is scrapped (Rejects) then agglomerate TPT is calculated to be 10 days. Thus, TPT is an agglomerate calculation rather than a measurement of shipped products, and does not include an allowance for Shabbat. This metric is useful for manufacturing since it gives immediate feedback. However, because it is a calculated figure, it can be misleading due to the fact that it is neither product nor lot specific.

Cycle time variation is currently not measured, primarily because information system aggregate cycle time calculations do no track and record individual lot performance. In order to compensate for this, the planning group commits to the *actual* fab cycle time, plus a safety factor. However, the specific lots shipped are *whichever* lots of the correct product happen to reach the end of the line first. For frequently ordered products, this means that the start dates for actual lots shipped to fill a given order are distributed over multiple weeks. For infrequently ordered products, this means that lots must be prioritized when they begin to fall behind the delivery schedule.

TPT and all other manufacturing metrics are available on-line for day, week-to-date and month-to-date. Line yield, wafers out of the fab, and WIP are the next few metrics on a list of important fab manufacturing metrics.

Focus on metrics varies throughout the fabrication facility hierarchy. A summary of this emphasis is shown in Figure 1.

Figure 2.1: Metrics Emphasis Throughout The Fabrication Facility



The three shift managers operate with some of the same basic metric hierarchy as the manufacturing manager. However, a part of the shift manager's job is the translation of these metrics into practice for the eight group leaders. At this level, the focus tends to be on which wafers to prioritize (by layer), functional area WIP, moves (number of wafers moved into a station during the shift), line yield, and how to get this week's deliverables out of the line. Wafer scraps are discussed on a case by case basis in order to prevent future scraps. In some functional groups of the fab, utilization is posted. There is a friendly rivalry between shifts for highest number of moves.

Each SST in the fab can monitor moves into his group of machines through an in-house developed on-line system. This system gives him a target for the shift, and measures totals against it. This is his primary metric. Numerous other metrics are available on line, including moves out of a station, station throughput time (calculated by the same method as above), etc.

Wafer scraps are strongly fed back. The SST's are not educated on meaning or fab performance in LIPAS and die cost, and die yield is not emphasized

2.5 Other Relevant Manufacturing Policies

Several other manufacturing policies affect WIP management policy. These include local set-up limits, upstream-sensitive scheduling policies, time limit operations, Shabbat, cross-training and down time policies.

In several clean room areas, WIP management rules are modified to decrease time lost to set-ups. These times are considered "significant", as defined by above 5% of production time. Product or layer specific set-ups limitations allow an increase in area utilization. These policies set a lower limit on the number of lots of a given product or layer prior to performing a set-up. However, these policies slow low-volume products as they wait for the minimum number of lots to accumulate in queue. Despite the fact that these products may have begun production together, variation in processing time, rework, and queue times disperses them so that they do not arrive at incoming inventory racks together.

This set-up time reduction is taken one step further in several processing areas, where local scheduling systems look up-stream for incoming products. These pinpoint which product to run based on a minimum number of lots, which includes both lots in queue and upstream lots which will be completed in time to process on the given set-up.

Chemistry of wafer processing necessitates another caveat to stated WIP management methods, due to time limits between some processes. To manage these time limits, a WIP limiting system is used to avoid scrap and facilitate inventory management between these sensitive processing steps. A kanban-type system manages this, with a limited number of tags made available to mark lots between processes. When all tags are in use, wafers can no longer be pushed through production. Thus, upstream machines either run lots from a layer not proceeded by a time limit, or they idle.

Machine down time also affects lot priorities. In most functional areas of the clean room, if enough machines go down to seriously affect production, priorities in other areas change. Upstream machines prioritize operations which will not feed the "down" area. No rigorous procedure governs this; it tends to be an ad hoc decision.

Finally, several organizational issues affect practice of WIP management policy. Because the Fab shuts down for Shabbat, all equipment is eventually turned off during the week-end.

However, some equipment runs for part of the week-end unattended. Other equipment cannot be run due to safety regulations. This affects material choices since batches with no set-ups need to be run. (This also translates to a net reduction in available hours for equipment.)

To run equipment, technicians who are certified must be available for specific operations and maintenance as the need arises. Lot priorities occasionally change when a technician is not available to run a specific layer or operation.

2.6 Priority Lots

Four types of priority lots are run through the line: flag lots, hot lots, MFC lots and special products. The number of each type of lot is limited by policy. Normal production lots, engineering lots and station test wafers are also run.

Flag lot and hot lots are generally used for the same purposes, with weekly highest priorities assigned as flag lots. The uses include:

- new products
- customer's urgent request for product
- need to recover from scrapped wafers
- late wafers on the week due to ship
- late wafers on the second or third week prior to ship
- important engineering lots

Flag Lots run in a special red wafer carrier adorned with a flag. A flag lot requires immediate processing, including breaking necessary set-ups to run. One person is designated to schedule, monitor and manage individual flag lots, including contacting clean room personnel to be sure that set-ups are prepared *prior* to flag lot arrivals in some cases. Flag lots are a planning manager's treasure: not to be wasted.

Hot lots run in a red lot carrier, without the flag. These lots are run first in any set-up, but do not require breaking current set-ups. Frequently, one lot of a product is run as a hot lot with the remaining priority lots of this product processed as hot lots.

MFC lots, or manufacturing first commit, are used by planning and manufacturing to indicate normal priority products which are becoming behind schedule. Potentially late products are identified one to two weeks prior to the ship date by automated systems. These lots are marked

on shift management reports for shift managers to inform group leaders, who in turn inform SST's to prioritize these lots. Occasionally, an MFC lot is marked with a plastic tag to visually identify it. In general, they are handled like hot lots.

Finally, there are special product priorities. These occur only occasionally, on the order of twice per year. However, they can have a large impact on production. An "emergency" product is started in quantities which force the fab to output more wafers than the fab purportedly has capacity to handle. These lots are given priority. These lots are emphasized in verbal instructions by group leaders, and by electronic bulletin boards in the fab saying "choose me first." This policy capitalizes on the ability of technicians to produce more than normal output for a brief and concentrated length of time.

2.7 *Cycle Time and Variation*

During most of this assignment, the primary metrics focus was on Mean Cycle Time and 95% Cycle Time, defined as the time for 19 out of 20 lots to be completed. This was for the following reasons:

1. The original project proposal focused on product mix delivery and decreasing delivery times due to conversion to commodity products.
2. The team chose these metrics.
3. Cycle time is easy to measure.
4. Real data records could be extracted and correlated with a model.
5. These metrics are easily calculated in a simulation environment.

Furthermore, decreasing cycle time and cycle time variability would improve fab performance in the following areas:

- Decreased fab plus sort inventory carrying costs
- Decreased downstream (assembly plus customer) inventory carrying costs
- Reduced scrap due to faster process feed back from electrical testing and assembly
- Decreased time to receive customer money
- Potentially improved customer satisfaction

However, reflecting on these merits leads to the conclusion that cycle time variability is also a critical measure. Firstly, the push to improve cycle time is not as important as having correct products out the door at the right time. These metrics should be intended to improve LIPAS

more than cycle time, and thus cycle time *variability* is quite important. A relatively small percentage of products needed to be delivered quickly, according to planning group estimates. Secondly, large variation in cycle time tends to increase the number of expedited lots, an expense which is difficult to quantify but clearly an operational challenge. In general, a batch of product begins in the line, continues in sync for a few steps, and then becomes separated due to a number of factors. Line yield and lots on hold are also a major problem in terms of delivery variability, since entire scrapped lots in low volume products cause major delivery problems.

2.8 *Financial Models*

All data within this section has been disguised, but for the purpose of discussion order-of-magnitude figures are presented.

Increasing cost pressures with the commodity market make stocking expensive finished (but unassembled) dies undesirable. In supply chain management, recommended inventory levels of finished goods can be related to cycle time and variation for the supply chain by a standard inventory equation. If r is the review period in days, μ is mean demand per day, z is the desired safety in standard deviations, σ is standard deviation in daily demand, L is the lead time in days, and if we assume performance each day is independent, then safety stock inventory for each different product is:

$$\text{Expected inventory level} = r\mu/2 + z\sigma(r + L)^{1/2}$$ *Equation 2.2 (Nahmias, 1993)*

If we assume that assembly is the customer, then lead time includes fab and sort cycle time, as well as time wafers spend waiting to be shipped and actual transport time. Assume that total lead time is 90 days, order review period 10 days, daily demand is 15 wafers of 10 different products, standard deviation in order size is 3 wafers, and z is 1.645 (95% of the time we will not stock out). Solving this, we find that our customers in assembly must store, on average, 56.85 wafers of 10 different products. In others words, downstream wafer inventory is 568.5 wafers.

This equation assumes constant lead time, normally distributed order sizes, and time independence in order size and lead time, all assumptions which are not correct in this case. However, the equation is still very relevant in setting finished inventory levels. Clearly, decreasing cycle time will not only decrease in process inventory, but also the need for finished goods. However, in this example, inventory levels are more sensitive to reduction in order review period and variability in order times.

From a financial standpoint, in-process inventory holding costs can be approximated by the following equation relating average wafer cost (AC), weighted average cost of capital (WAC), and average wafer inventory (WIP) to inventory holding cost (HC) :

$$HC = AC \times WAC \times WIP \qquad \textit{Equation 2.3}$$

If we assume an average wafer cost of $100 (valued midway through production), a cost of capital of 15%, and a constant 10,000 wafers on the line, the annual inventory holding cost for these wafers is $150,000. This cost is directly proportional to number of wafers on the line.

Inventory holding costs for assembly wafer inventory can be calculated by substituting downstream inventory results of Equation 2.2 into Equation 2.3. Here, the annual cost of holding 568.5 completed wafers (at an average value of $200) is $17,055.

For this example, total wafer annual inventory holding cost is approximately $167,000.

This model does not include all of the potential savings due to improving fabrication cycle time. Costs for wafer inventory scrapped due to product design changes can be significant. Savings due to reduced process information feed-back time (from end-of-line testing) can also be significant.

3. Performance Data and Policy Benchmarking

3.1 Historical Facility Performance Data

Throughout the following section, please refer to exhibits in Appendix A which display historical process performance data.

Exhibit 1 shows recent process wafer start and product mix data. As shown in the graph, process wafer starts has remained fairly constant, with an increase in the latest weeks. This jump is allowed because of a recent "ramp" or conversion of older process equipment to increase capacity. The number of different products started in a given week frequently varies by a factor of 2 from the previous week. Thus, wafer starts can be modeled as linear with time, but a sophisticated model must examine the effect of dynamic changes in number of wafer starts on model output.

Exhibit 2 shows historical capacity and process performance on actual and 95% lot cycle time from the time of process initiation to the summer of 1996. The corresponding data is extracted from a historical data base, and contains time for every production lot from the first fab operation to completion of the sort process. Data for 95% Cycle Time is calculated, assuming that sample size is normally distributed, as follows:

*95% Cycle Time = Mean Cycle Time + 1.65*Sample Standard DeviationEquation 3.1*

Note the initial long cycle times, followed by movement down the learning curve to a relatively stable production time. Also note the drop in cycle times after large capacity increases. Week-to-week changes in cycle time appear small due to the large sample size and the number of weeks which lots spend in the facility. Those which do appear are due to either line yield excursions or large scale machine problems.

As seen in the Sample Week Cycle Time Distribution Histogram and Normality Check (Exhibit 3), the normality assumption fits. The Shapiro-Wilk W test results show a p value of less than .05, and thus the data can be assumed normal.

Exhibit 4 displays *actual* lot cycle time extracted from data records on a lot-by-lot basis for a one year period. Exhibit 5 displays regression fits for fab *calculated* TPT over a year long period, based on automated systems data as calculated by Equation 2.1, Section 2.4. This comes from current fab data systems. Agglomerate calculated cycle time is very close to constant. This is

clearly *not* the case with actual cycle time. The line fit to actual cycle time decreases by almost 25% over the one year period! For reasons of data availability, the sample periods overlay for only one quarter of the time. However, even with a precautionary caveat, there is strong evidence to distrust lot cycle times based on this calculated method. However, this is the only data which is used in managing the fab.

Both fab and fab plus sort lot cycle time have relatively high standard deviations, as shown in Exhibit 6. Coefficient of variation in each case is close to 20%. Some points skew this data from being normally distributed, appearing on the right side of tail of the picture. This relatively low volume of lots has probably been placed on hold during the production process due to potential quality problems.

Finally, Exhibit 7 displays this historical data in terms of cycle time versus batch size. 95% Fab Cycle Time is calculated as previously noted. Here, we have given a non-standard definition to the term batch. Batch size is defined as the number of lots of a specific product that is started on a given day. As discussed in Section 2.5, variation coupled with product set-up policies tends to slow low volume products. The correlation is actually stronger than it appears in this analysis, because flag lots and hot lots could not be removed from the data sets. Flag and hot lots are small batches in this data, and they move much faster than average lots since they are expedited. Thus, these tend to raise the average cycle time performance for small lots, while increasing the variation. This correlation is partially due to set-up times

3.2 Benchmarking Within Intel and The Industry

A survey of production management methodologies throughout Intel was performed. Most managers say that they utilize some version of the Theory of Constraints. Many point out that this thought is useful for them in day-to-day operations, but that the method of translating it into specific technician actions is unclear.

Most of the facilities have converged upon some variation of Back-To-Front methodology for specifics in line management. Others run with a focus on maintaining a balanced line. However, most fabs run far fewer products than the Jerusalem facility. Thus, the comparison is not necessarily relevant.

Literature searches provide fairly murky data on *proven* effect of WIP management methodologies in the semiconductor industry. Quite a few papers on theoretical approaches are available, but very few discuss implemented solutions. The question "Why?" remains open. Is it

because the methodology is confidential, the methodology doesn't matter or because nobody knows whether it matters?

Based on the amount of operations management publication in other fields, I suspect that the semi-conductor industry is quite secretive about publishing specific data or operating practices. However, some useful information can be gleaned. Theoretical papers from the industry prove that some fabs run with Kanban policies, and others with calculated Least Slack methodologies (Page, 1994; Wolf, 1986; Di Mascolo, 1996.) It is highly probable that some run with First-In-First-Out.

Due to the lack of publication on semiconductor manufacturing methodologies, one cannot draw many conclusions. However, one thorough benchmarking study performed at University of California, Berkeley, presents semi-conductor industry data from 1992 through 1995 (Leachman, 1996.) In this international, industry-wide study of 15 facilities, WIP management method was not a distinguishing factor for logic fab cycle time. The logic fab with the best cycle time runs with a kanban system, but poor performers also use kanbans (Leachman, 1996.)

From this study, clear evidence exists that some competitors have achieved much better cycle times than Fab 8. However, this cycle time data was not normalized for production capacity. In this same study, the fab performed very well in efficiency measures. However, one of the fabs in the study performed better in both categories, and thus there is proven room for improvement. By today, the competition is probably even better, since this data is between 2 and 5 years old (Leachman, 1996.)

3.3 Alternatives to Modeled Methodologies

Along with the previously mentioned WIP management methodologies, several methods were not examined. The most notable examples are:

- Least Slack
- Kanban Systems
- Equal choice between layers

The Least Slack policy is quite similar to the X-Theoretical policy, with the exception that the X-Theoretical policy is a ratio. The technician prioritizes lots based on a calculated algorithm for choosing the lot which has the "least slack." Slack is defined as

Slack = (due date) - (estimated delay from buffer where lot is situated) *Equation 3.2*

Choices between equal priority lots are determined by earliest arrival time at the functional area. Modeling both was deemed unnecessary, because they are so similar, and the logic for X-Theoretical was coded into the available software, so this was chosen.

In implementation, Kanban information systems utilize a WIP management pull system (inventory-limiting) variation of FIFO. These were not modeled. This is both because FIFO systems were already modeled, and because the author did not consider this a practical and manageable solution. (This facility runs dozens of products with different due dates through a facility operating at capacity with a balanced line.)

Level layer choices, a policy of running equal amounts of each layer during a shift, is generally a sensible policy. This tends to reduce large WIP bubbles moving through the facility. This was not modeled because of the author's time constraints in the project, and would represent a wise future development in the model.

3.4 Published Academic Information

3.4.1 WIP Management Methodologies

Although publications on WIP management methodologies in non-semiconductor industries are common, the results differ largely according to model content and which metric the modelers are attempting to optimize.

Some problems have been solved analytically. For a single station model, processing each product once, Nahmias examines outputs of various WIP management methodologies. In comparing first-come, first-served, shortest processing time, earliest due date, and critical ratio policies, shortest processing time performs best in terms of both mean cycle time and average tardiness (Nahmias, 1993.) He goes on to prove that single station mean flow time is minimized by the rule shortest processing time (Nahmias, 1993.)

However, the optimal solution changes when the single station is replaced by two identical stations in series which operate on a product once. In this case, lots should be run through both machines in the same order. Products should be run in order of decreasing values of the difference between the two machine rates, calculated as follows:

Priority = Processing Time A - Processing Time B *Equation 3.3* *(Nahmias, 1993)*

This solution is very different from the single station shortest processing time approach.

The usefulness of current analytical methods deteriorates with additional machines in a line. Therefore, dynamic simulation approaches are used. Results from these models are highly dependent on the model which has been input. Hence, a dynamic simulation model of the facility was necessary.

3.4.2 Loading Change Effect on Mean Cycle Time and Variance Under Several WIP Management Policies

In comparing Shortest Processing Time policies with FIFO policies, performance changes are highly affected by line utilization. In this simulation, SPT performance in Mean Cycle Time is clearly better than FIFO. When utilization reaches 99%, Shortest Processing Time performance in Mean Cycle Time is approximately 20% of corresponding FIFO performance. However, SPT variance is 16 times that of FIFO! For utilization of less than 70%, SPT variance is lower than that of FIFO (Nahmias, 1993.)

Thus, one needs to be certain that loading in areas to be modeled is well understood.

4. Simulation Model Development

Due to associated disruptions to the fab, the dynamic environment which makes week-to-week data difficult to compare, and the long cycle times in fabs, physically experimenting with WIP management policies would be unwise. Thus, in order to predict the effects of various policies, a model needed to be developed. This problem is far too complex to apply mathematical models. Therefore, in order to *systematically* compare fab WIP management policies, a dynamic simulation model was needed. AutoSched© was chosen as the modeling package based on its capabilities for interfacing and inputting semiconductor facility data.

4.1 Initial Model Inputs

The team chose eight clean room modules to be included in the model. These included both bottleneck stations and areas seen as problems in terms of down time. After this choice was made, data was gathered via team input as discussed in Section 1.1. In between processing steps on modeled stations, a delay time was assigned based on data from the unmodeled steps. This delay was assigned a normal distribution with a coefficient of variation of 20%.

AutoSched© inputs are tabular in format, as shown in the disguised model input exhibits in Appendix B. This appendix systematically steps through model input tables, and modeling method, and is intended to familiarize interested readers with the Autosched format. These include the Wafer Starts Data, Station Definition, Process Flow, Station Down Time, and WIP Prioritization Rule Tables.

Wafer starts data was obtained from the planning group. A sample week's profile was chosen as typical, and wafers were assigned a specific start time in the morning and one in the evening for each day of the week. This initial order was repeated on a weekly basis.

Down time data for unscheduled down time, preventative maintenance and time lost due to technicians were determined. Distributions of exponential for mean time to failure (MTTF), and normal for mean time to repair (MTTR), were assumed based on input from both modeling experts within Intel and the author's academic advisors. The Weibull distribution was chosen as a conservative check on this output. Weibull was also specified for distributions in the machine family with breakdown repair time variation higher than 35%. Initially, Industrial Engineers provided preventative maintenance and down time information in the form of percentage of machine availability. An industrial engineer then used an automated information system to check

historical records on each area, and determine appropriate break-down frequency, repair time and variation in repair time.

Because these distributions are difficult to determine in a manner which does not require extensive amounts of time, assumptions were made. These included the distribution type assumption, the assumption that data over the past year reflected enough of a sampling period, and the assumption that the several machines surveyed in each area were representative of all machines. Because of the uncertainty in these distributions, model runs to determine sensitivity to repair time variability are critical.

Automated systems down time data was then compared to industrial engineering figures, which are agreed upon between the machine owners (equipment engineers) and the industrial engineer who is responsible for each area. In most cases, agreement between figures is within 20%. However, in one functional area, agreement for unscheduled down time differs by a factor of two. A decision was made to utilize the industrial engineering figures, because the automated system sometimes loses data since operator action is required to change machine state to "down." Numbers for the area with the large discrepancy were flagged for investigation by the area industrial engineer.

The total time lost due to preventative maintenance and down time were both large enough so that the cross-product interaction could not be ignored in a dynamic simulation model. Unscheduled down time figures were held constant, and preventative maintenance Mean Time To Repair (MTTR) and variation were scaled up to include the cross-product component of the MTTR.

4.2 Output Verification and Iteration Process for Inputs

The first working model was developed through a series of iterations. First, the basic process flow, order, and station files were defined. The model output was statistically checked for "reasonable" behavior. Next, lot prioritization functions for WIP management were added. Part color changes along the processing steps allowed visual checks of lot prioritization functions. Lot processing times were checked. At this point, model cycle time output was close to theoretical processing time, since machine down time had not yet been added.

Next, preventative maintenance and unscheduled down time were added to each station. The model was run for a 1 year period. Output for amount and distribution of machine down times was compared to input, and found to match.

At this point, simulated mean time each lot spent in modeled stations was roughly 60% of actual time. The initial assumptions were examined for possible corrections. These are shown below:

1. Constant Wafer Starts/6 day period
2. No Shabbat Shutdown (But different clean room areas lose different amounts of time)
3. Technicians are always available to operate machines
4. Down Time is the industrial engineering time unless automated system data is higher
5. Unscheduled, Scheduled down time: MTTR normally distributed, MTBF exponentially distributed
6. Cycle Time for each process step equals the average of past year as calculated from the automated system figures
7. No rework
8. No lots on hold
9. Line Yield 100%
10. Lithography mask set-up ignored (No product set-ups, but product batches are included)
11. No product differentiation occurs during delay times
12. Delay times are normally distributed with constant coefficient of variation (20%)
13. No priority lots
14. Several areas have product and layer differentiation in batching

After examining these assumptions with team members, the following series of corrective actions were taken:

- Each step's processing time was reviewed with Industrial Engineers and fab technicians.
- Each machine area's preventative maintenance and unscheduled down time was reviewed with industrial engineering.
- Appropriate amounts of rework lots in lithography were added.
- A yield factor to compensate for line yield was added. Yield was modeled as equivalent for each successive lithography step. This was confirmed as a reasonable assumption through data review with the engineer responsible for line yield data.
- Sensitivity to machine down time distributions and associated repair time variation was checked. Increasing variability by a factor of two and changing distribution was found to have a negligible effect on output times.
- Since in reality lots spend time in carts waiting to be transported to the next step, a lot transportation step was added to the model after each simulated processing step. The carts were modeled as batch machines, with transport times and methodologies supplied by fab

technician estimates. These estimates were verified using sampling and calculations based on Little's Law. This formula relates average total lots waiting in a queue, L, to the average arrival rate λ, and the average amount of time spent waiting (W) via the relationship:

$$L=\lambda W \qquad \textit{Equation 4.1 (Nahmias, 1993)}$$

After adding these corrections, the model was run again. Although correlation with reality improved, it was still not close to correct. The team agreed that those inputs which were included in the model were correct. Area experts agreed that the model assumption that personnel time had no effect on machine down time clearly needed to be examined and corrected for.

The question of how to gather model input for this correction factor is both critical and difficult. Clearly the best solution in terms of modeling is real data based on technician performance. However, this is an impractical solution since data for the model would need to be gathered over a period of months or years, and on-site observation of dozens of machines requires dozens of observers.

After reviewing the problem with academic advisors and fab experts, the author decided to reflect personnel time losses in machine down times. Based on their experiences, Industrial Engineers agreed on initial machine down times and distributions associated with each machine area. These inputs would be iterated based on model outputs in each area's cycle time and WIP as compared to reality. This iterative approach assumes that *all* unaccounted for inputs should be assigned to machine time lost due to absence of people for loading, unloading or correcting a machine malfunction. Please note that one academic advisor questioned the validity of this approach due to lack of information. However, I maintain that since the other inputs were agreed upon, and we believed that the majority of the "missing" cycle time was due to personnel abscence, this was the only reasonable missing input. Furthermore, in engineering and business one seldom obtains perfect information: we make decisions on the best data that we can obtain. Eventually, all parties agreed that this was a feasible approach.

Modified model assumptions now included the following:

1. Lots are transported from/to next machines in batches of 1-8, requiring 5 minutes to transfer from the previous delay. Batches wait for 30 minutes for completion.
2. Technicians are not always available
3. Line Yield is constant for all layers, and losses occur in lithography only

4. Rework is constant for all layers, and occurs in lithography only

The agreed upon figure was assigned to the model for machine down time after adding the cross-product figure due to interaction with preventative maintenance and unscheduled machine down times. After adding this human factor, the model output much more reasonable numbers for process cycle time. After several iterations, correlation with reality were within 10% in each simulated area. However, in the process, the group discovered that the time lost due to personnel absence is approximately equal to that of unscheduled down times in most areas. This is a great deal of time lost to personnel availability, and calls for a thorough examination of personnel policies!

One cannot prove that technician down time numbers are correct, since they are created through a process of elimination and iteration. Error is also introduced by uncertainty in scheduled and unscheduled machine down times. However, based on the model development process and instinct checks with team members, the team believes that these figures are reasonable.

As a final note, from a technical standpoint the modeling method of losing machine time due to technicians being absent could be improved. Assigning simulated operators to each machine area, requiring them to be present to perform an operation, and assigning them absence distributions is more accurate. However, the amount of time required to include and verify these elements in the model was deemed too great for what would have been essentially the same WIP management and total down time research results.

5. Simulation Results

Cycle time and 95% Cycle Time model outputs are shown in Appendix C. Here, 95% Cycle Time is calculated from weekly standard deviation and cycle time mean as described in Equation 3.1. Note that both Mean and 95% Cycle Times are statistically corrected to include only the modeled areas, since the normally distributed delay times representing the remainder of average processing times between steps is not intended to model every other step. (For instance, this does not distinguish between delay area affect on lot timing for different WIP management scenarios.)

A standard 10 week warm-up period and 42 week run was chosen for each case deemed important enough to run through the 8-hour solution time model. Simulation runs are varied from the base case, which represents current fab performance according to inputs and assumptions detailed in Section 4.2. · The following base case assumptions were changed in the scenarios illustrated:

- Stable weekly order schedule and number of lots at current fab levels
- Back To Front WIP management methodology
- Unscheduled, scheduled and time lost to technician absence as defined and iterated
- Station Mean Time To Repair normally distributed and Mean Time To Failure exponentially distributed
- Time limit area kanbans in effect

Exhibit 14 summarizes results for simulated areas in a tabular format, and the remainder of this appendix summarizes results in graphical form. This summary table compares the model output for each simulation scenario cycle time and 95% cycle time to the base case. Here, a 10% table entry in the row titled "Cycle Time Delta Base Case" represents a 10% increase in Mean Cycle Time. Individual scenarios are discussed in the following sections, but from the initial glance at this table, be sure to notice the significant magnitude of changes we see in model output.

Some Exhibits show results of ANOVA and Tukey-Kramer tests performed using the JMP© Start Statistics package. These analyses were performed when differences in the graphical exhibits were difficult to determine with the naked eye.

5.1 WIP Management Effects

As shown graphically in Exhibits 15 and 17, choice of WIP management policy had a significant impact on both mean cycle time and 95% Cycle Time in the level loading case. Note the week-to-week variation in cycle times in the model output, primarily caused by various machines going down. This pattern is similar to the variation patterns of Exhibit 4. Exhibits 16 and 18 show results of ANOVA and Tukey-Kramer tests to determine which mean differences can be considered significant to the 95% confidence level. (Shortest Processing Time data has not been included in these tests due to the fact that output is visibly different, and one scenario needed to be excluded because of package data size limitations.)

By far the best policy performance is the calculated case of the X-Theoretical lot management. This reduced mean cycle time by 11%, and 95% Cycle Time by 20%. For both metrics, mean differences between the Back-To-Front and FIFO or Random lot choices do not appear significant with 95% confidence. On the other hand, the local choice of Shortest Processing Time is the worst choice, since it adds 23% and 79% respectively. Front-To-Back, as predicted, is a poor choice since mean cycle time and 95% Cycle Time are increased by 10% and 44%, respectively. Because this local SPT is the worst case in this simulation, *at all costs* managers must have a system which does not encourage employees to operate under this policy, which is encouraged by the technician incentive system emphasizing moves.

The next set of runs tested system performance with changes in loading. These loading scenarios were less important than the base case, since these start scenarios are unrealistic at the current time. However, these were important both as a reality check on output, and as an intuitive check on WIP management policies boundaries. Due to this relatively low priority, the model was run only with the base-case and X-Theoretical for most of these tests. Exhibits 19 and 20 show effects of increasing weekly wafer starts by a constant 1.4% and 3.5%, and of decreasing loading by 1.4%. Note that in most cases X-Theoretical performance outperforms Back-To-Front, with the exception of the overloaded capacity case of Mean Cycle Time with +3.5%. At this loading, one can observe that the fab capacity cannot meet demand since cycle time is increasing with time. Back-To-Front manages to decrease Mean Cycle Time in this case, since WIP is removed from the line through prioritizing processing of lots nearing the end of the line.

This result is further evidenced by the peak loading shown in Exhibits 21 and 22. In this scenario, every 10 weeks a peak of lots is introduced to the line when wafer starts are doubled for a one week period. This effectively increases total loading by 10%, clearly overloading the fab.

In this case, the Back-To-Front scenario slightly outperforms X-theoretical for reasons already discussed.

Again, fab "overloading" scenarios are useful for examining the model, and testing one's intuition. However, results from these scenarios are not particularly relevant, since in no case is the fab actually overloaded in terms of wafer starts, without somehow increasing bottleneck machine availability.

From an intuitive standpoint, it is difficult to understand why having a specific prioritization scheme causes such large changes in *mean* cycle time. One way to better understand why a calculated methodology such as X-Theoretical improves performance is to consider the following:

> Some prioritization schemes help to remove WIP from the fab. This, in turn, decreases queue size, since fabs are an example of reentry flow. Thus, lots spend less time in the fab. However, *utilization* should not be affected by this change.

5.2 Down Time Effects

A series of experiments were run to test global performance based on improving machine availability in key areas. These tests were run with Back-To-Front WIP management methodology, and the level loading of the base case. In specific, these tests improved:

- the two main fab constraint areas together
- the two main fab constraint areas separately
- non-constraint areas as a group
- the entire modeled fab

These improvements were modeled by scaling down mean and variation in repair time calendars for either the unscheduled down time or down time due to technicians not being available. Exhibits 23 and 24 show these effects graphically, with down time being abbreviated as DT.

Several important points should be taken away from this analysis. They are:

1. The absolute limit for performance improvement through down time reduction is 39% mean cycle time reduction, and 28% 95% Cycle Time reduction. This is the case where each type of down time is only 1%.
2. Targeting the two bottleneck areas, which represent the most logical areas for improvement efforts, could result in substantially improved fab performance. If the time lost due to personnel absence were reduced to 1% in these areas, the entire modeled area of the fab

would experience a Mean Cycle Time decrease of 20%, and 95% Cycle Time a 12% decrease.

3. A significant machine interaction effect exists. Improving machine availability in areas that are not the primary bottlenecks can result in significant model performance increases. The 19% mean and 14% 95% Cycle Time improvements achieved by decreasing the other six areas' time lost to technician absence are approximately the same as those achieved through the same reduction in the two most sensitive areas. I hypothesize that reduced total set-up time and bigger batch sizes in primary bottleneck areas causes this effect. These are caused by the larger queue sizes in primary bottleneck areas as non-bottleneck queue sizes decrease.

5.3 *Statistical and Sensitivity Effects*

Results from a series of statistical checks are shown graphically in Exhibits 25 and 27. Exhibits 26 and 28 display the results of ANOVA and Tukey-Kramer tests to determine whether these differences are significantly different from the Back-To-Front baseline. Removing all variation from delay times actually increases Mean Cycle Time in both Random and Back-To-Front methodologies, and decreases 95% Cycle Times in the Random case. These are the only cases where Tukey-Kramer tests show mean differences significant to the 95th percentile confidence level.

Note that the removal of model logic to limit number of lots in time limit areas did not effect model cycle time, and thus we can conclude that the numerical limit policy does not appear to worsen fab performance.

A final statistical test was performed to ascertain that the model was indeed repeatable. This is not shown graphically, since results were perfectly correlated. This verifies that the random numbers generated by the model are consistent each time a new run begins.

5.4 *Policy Recommendation Effects*

Output for recommended policies for this fab are shown in Exhibits 29 and 30. These recommendations are to decrease machine down time and use excess capacity generated to increase wafer starts by 2%. The target of reducing machine time lost due to technicians absence by 50% was deemed reasonable. Furthermore, I recommend running the lithography area with a production rule which gives preference to lots with the same lithography layer set-up prior to breaking the set-up down. Model runs under these recommendations yielded the following: Mean Cycle Time is decreased by 26% and 95% Cycle Time increased by 20%. Thus, cycle time mean was much decreased while variation in output time was increased substantially.

Most model tests were run with this assumption. Running Back-To-Front policy without that assumption increased Mean and 95% Cycle Time by 7% and 3%, respectively. Changing this policy had an even more detrimental effect on the Random lot choice methodology.

6. Recommendations

6.1 Systems Design and Implementation

One year ago, all technicians reported to group functional area leaders, who in turn reported to the manufacturing manager. Engineering was a separate organization, also divided into groups by clean room function. At the commencement of this research, an organizational change merged the engineering departments with manufacturing technicians into a functional team organization. The only remaining personnel with a "cross-factory focus" were the newly created shift-manager positions.

This organizational change must be taken into account when implementing any fab-wide programs, since both communication and priorities have largely become functional. However, this change makes the implementation of comprehensive, fab-wide learning programs even more important than previously.

6.1.1 Design and Implement Metrics Educational Programs

Traditional management theory encourages measuring and rewarding people based on factors which they control. This facility follows this advice through constant monitoring of shift moves in each functional area of the clean room. However, in order to achieve an aligned organization, educating technicians on upper management indicators is essential. This understanding is *critical* because technicians have the power to make decisions which negatively affect upper management objectives, while affecting their own indicators positively. Technicians are rewarded via profit sharing based on performance against fab objectives, so the incentive is in place, but the educational component is not.

One simple (and exaggerated) fab example would be the following: a stepper is down and a technician outside of the clean room is summoned into the clean room via her beeper. She rushes inside the clean room to fix the problem, not pausing to check the fastening of her gown. In the ten minute process of fixing the machine to increase moves, either many dies on a wafer or, worse, the stepper, may be contaminated. The machine is fixed 2 minutes faster than if the technician had stopped to check her gown.

Due to the potential for increased moves and some element of a firefighting "hero" culture, the technician probably feels that she did the right thing by responding quickly. However, in the

context of LIPAS and die cost, this is not the right choice. Although there is a policy to dictate how people dress, understanding the effect on the bottom line is critical when people make daily behavioral decisions.

People are generally trying to do what is right, but sometimes what looks right to them is not what is right for the fab metrics as a whole. An illustration representing the problems faced by technicians in an unaligned organization is shown in the following figure:

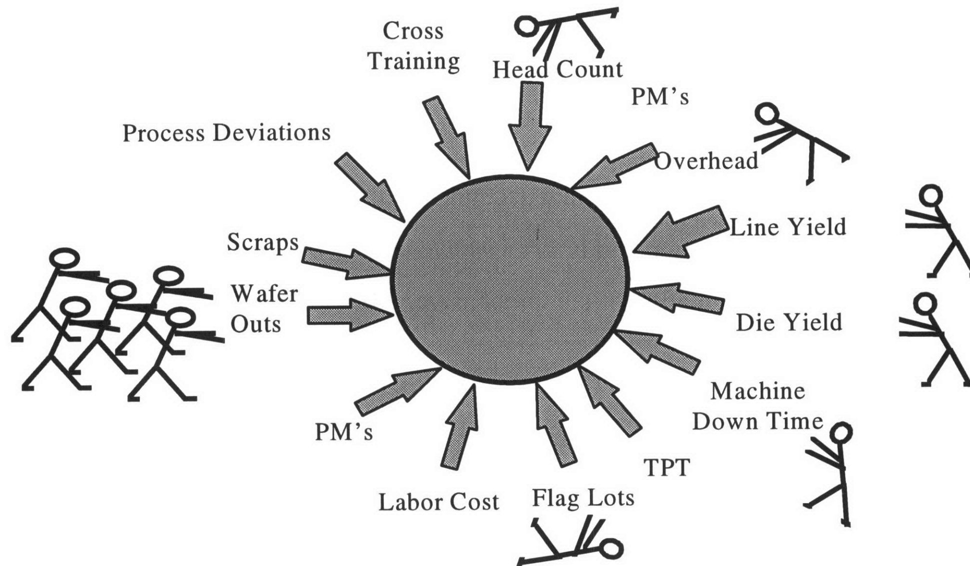


Figure 6.1: Lack of Clean Room Organizational Alignment

This illustration is meant to show a team of people working hard, but occasionally against each other. At times they are frustrated, because management emphasis varies so much from day to day, and they do not understand why. Contrast this with the aligned organization illustrated on the following page:

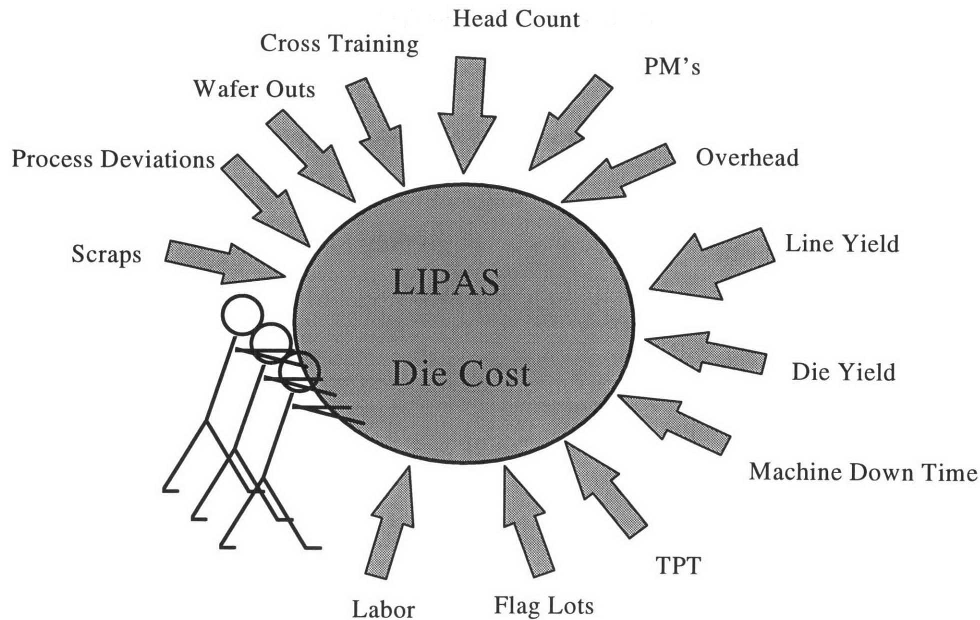


Figure 6.2: Clean Room Organizational Alignment

In this illustration, a team of people works to improve LIPAS and Die Cost. Occasionally they do what appears to be "wrong" in the small picture in order to improve these two metrics. This alignment is relatively simple to achieve through big picture education and performance tracking. Developing and teaching a training class (corporate or site-specific) to educate technicians on Die Cost and LIPAS components would solve this problem. Verbally educating Group Leaders and Technicians on the magnitude of the personnel component of total cycle time, then challenging them to improve would probably yield excellent results.

As a precautionary note, solving this problem requires management to relinquish some information power: technicians can understand and act on their own judgments at times when they traditionally may not have. However, any manager who pretends that this type of judgment is not exercised on a daily basis may be under an illusion.

6.1.2 Develop Information Systems To Improve Technician Response Times In Bottleneck Areas

If one believes the simulation results, personnel issues represent as great an opportunity for increasing capacity or decreasing cycle time as unscheduled machine down time. Therefore, action to improve these issues is critical. Educating employees on low machine utilization numbers can help. "'It's interesting the number of companies that justify why they spend a lot of

money for a tool but can't operate it faster because of their unique process.' Challenging those assumptions can produce an immediate 10 to 15 percent productivity gain" (Ristelhueber, 1996.)

In any simulation, as utilization (ρ) approaches available time for machines, cycle time increases at a rate as (1/(1-ρ) . If people were not involved in running a fab, and the down time due to people responses were not critical, real fab output would look like this. In real life, however, people can increase their response time through many methods. My hypothesis of this relationship is illustrated in the following figure:

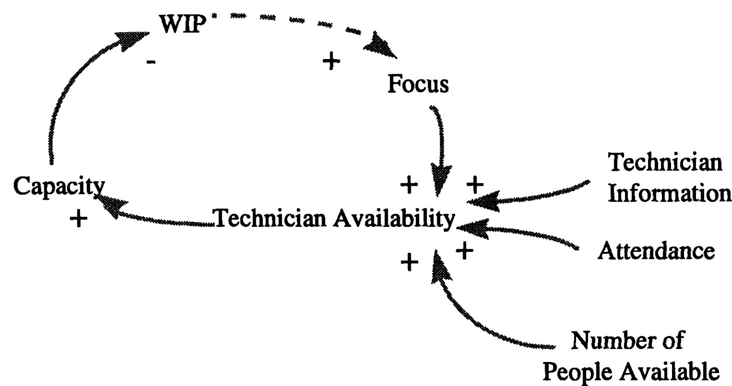


Figure 6.3: Technician Effect On Capacity

This illustration of management challenge, formatted in traditional systems dynamics style, shows a negative cycle. Increasing technician availability has the effect of increasing capacity as shown by the + sign next to capacity. However, increasing capacity in an area decreases WIP after a period of time, as shown by the negative sign next to WIP. Due to basic psychology, this decrease in WIP causes a decrease in focus. (Following systems dynamics sign conventions, a + sign is next to Focus because increasing WIP tends to increase focus.) This in turn decreases technician availability. (Following same sign convention, this is shown as a +.) The dotted line connections between WIP and technician availability are based on psychology. This cycle represents the major challenge to any strategy which depends on energy and focus to obtain results, as opposed to methodology.

The challenge, therefor, is to increase personnel capacity by raising one of the other three inputs, such as attendance, information, or the total technician pool.

To state this differently, in a more traditional manner, the following personnel performance issues served as hindrances to fab performance:

- Variation in technician availability
- Low staffing
- Lack of easily accessed information on machines status

Variation in technician availability can be due to poor attendance planning, illness, meetings or breaks.

Low staffing may be alleviated through hiring more technicians. IBM researchers in France found that that 11% of their fab cycle time was due to technicians being absent, through indirect methods such TAKT time (Boebel and Ruelle, 1996). In order to correct this issue, more technicians were hired.

The information availability may be the easiest issue to attack, and in this case would probably yield the best results. Information at the technician level is gathered either visually (through checking machine status), through audio warnings or through knowledge of when a lot will finish processing. The potential for losing machine time due to malfunctioning is high in each of these methods, since people cannot always see or hear a given machine, and they cannot predict breakdowns. Furthermore, processing time is lost when the responsible technician is dealing with another machine down. Technicians felt that the bulk of the lost machine time due to people occurred after a malfunction which wasn't immediately noticed. Throughout the clean room, relics of former attempts to solve this issue exist in the form of lights which flash and systems to provide information.

In one of the non-bottleneck fab areas, the industrial engineering group has developed a system which accounts for this problem. This system automatically gathers machine status information, based on the fab information system and some programming at the machine level. It allows area personnel to centrally access machine processing information, which in turn allows them to respond quickly to problems. Most importantly, technicians in this area use the system and have succeeded in measurably improving response times. This system should be implemented throughout the fab bottleneck areas, with technicians serving as critical members of the development team. Removal of redundant systems such as lights will help to encourage people to use this system.

It seems ironic that a company with a strategy of CopyExactly© across *international fabs* has not followed this philosophy with *internal* best practice implementation. Then performance can be *measured*, problem areas can be *understood*, and the system can be *improved*.

6.1.3 Develop Cycle Time Variation Measurement and Reduction Program

Since standard deviation in an individual lot cycle time is quite high (20% of the mean), the fab adds a significant time safety margin in order to ensure that LIPAS can be met. In order to expedite individual products which are lagging, fab personnel have created sophisticated systems to efficiently expedite lagging lots.

The fab does not have an information system to measure process cycle time variability or individual step processing time variability. The current system for dealing with LIPAS uses on-line reports to indicate lots which are behind schedule prior to becoming "too far" behind schedule. This information is verbally transmitted to technicians at shift pass-down meetings. Occasionally this is combined with a note attached to lot boxes in order to expedite lots.

The proactive focus on lots with potential to become behind schedule appears to be a good solution to the variability problem. However, it requires a significant amount of management time focus to ensure that these schedules are met. If the number of expedited lots were to increase by a factor of 2, then the management challenge could increase enormously.

A better solution would be to prevent this variability from occurring in the first place. The first challenge here is to find a way to *measure* 95% Cycle Time. A fab-wide measurement of processing time variability by lot would help to locate sections of the process where variability is created, and these could be pinpointed and corrected. Thus, sources of variability would be *understood*, and the decision whether to focus on *improvement* in these areas can be made.

An automated wafer system attached to an information system which prioritized and *physically* sorted lots would be easy for technicians and efficient. However, as previously discussed, the time and monetary cost of implementing this type of system in a fab without an automated wafer handling system is prohibitive.

This analysis points to one solution which is not real time, but does allow relatively easy access to variation information. An automated report which extracted actual lot process time and calculated variability would be helpful in a variation reduction effort. This, of course, depends on whether a team is in place to study data and implement operational changes.

6.2 Semiconductor Fab Operations

6.2.1 Retain Back-To-Front Policy For The Fab

Throughout these analyses, the lot-specific calculated X-Theoretical policy emerged as the definitive best solution for WIP management. Both mean cycle time and 95% Cycle Time were improved by more than 10% over Back-To-Front policy. However, this policy requires lot specific calculations and lot specific choice for running material. In practice, this requires an automated lot information system to perform real-time calculations quickly. This methodology is also much better suited to a fabrication facility with automated material handling systems, where specific lots would be identified and moved to the front of a queue without the necessity of manual lot searches by technicians. Automated systems are considered a cost prohibitive solution for this fab, thus outruling the implementation of X-Theoretical WIP management. However, benefits of this methodology should be carefully considered in those newer fabrication facilities which will eventually support multi-product production. For all of the reasons previously discussed, a calculated method will probably be the best choice for these fabs.

Thus, we are left with a choice between policies which do require lot-specific actions for this fab. In examining the simulation results, one notices that both First-In-First-Out and Random policies frequently perform at the same level or slightly better than Back-To-Front. Based on this result, one might logically ask, "why not implement a First-In-First-Out or Random lot choice policy, or not have any policy at all?" There are several compelling reasons not to do this.

Consider first the fab incentive system. Technicians are measured and rewarded based on number of moves which their area achieves during a shift. Although the addition of some facility-wide metrics to technician performance would improve clean-room performance, individual metrics must always be a part of technician feedback. Therefore, they do and probably will always feel pressure to run lots which will process faster when given a choice. This is the local Shortest Processing Time system. Of all of the policies modeled, this is the worst in terms of both Mean and 95% Cycle Time! Yet, if people are given a random choice of lots to run, they will have an understandable tendency to make the non-random choice of processing by Shortest Processing Time. A similar tendency seems likely to occur if people are asked to choose lots by First-In-First-Out policy. Having a policy of giving preference to random or first arriving lots is probably *not feasible to enforce* in the long term, and the alternative Shortest Processing Time looks very bad.

Secondly, even if these policies are successfully implemented in some clean room areas, it seems unlikely that they can be enforced throughout the clean room. In the early stages of the model building, a mixed-motive strategy was used in simulation runs. This strategy, using different methodologies in different areas, produced disastrous cycle time and 95% Cycle Time results. Although only a few permutations of the total number of potential policy mixtures were performed, the results were suitably negative to avoid spending further valuable simulation time in investigating combinations.

Thus, although several policies displayed better performance than Back-To-Front, by process of elimination, the wisest policy choice seems to be to remain with this choice. If, however, personnel develop ideas on how to implement an X-Theoretical system for this facility in a cost effective manner, fab cycle time and cycle time variation would improve.

6.2.2 Increase Wafer Starts

At the end of the assignment, the team was willing to agree to a obtain a 15% cycle time improvement. This would be achieved through targeting the two primary bottlenecks in this model for 50% improvement in operator response times, which would allow a 20% cycle time improvement. However, it had become clear that people cared more about increasing lot starts than decreasing cycle time.

A rough financial analysis was performed to determine whether this "new" capacity should be used to reduce cycle time or to increase wafer starts. Actual figures are not shown for confidentiality reasons. However, the result shows a 7:1 financial premium on increasing wafer starts, as opposed to decreasing cycle time. In the context of die cost and LIPAS, the decision between increasing wafer starts and decreasing cycle time clearly points to increasing wafer starts. As long as these remain the primary fab metrics, justification of efforts to reduce cycle time at the expense of wafer starts will be next to impossible. Unless, of course, most customers threaten to move business elsewhere in order to gain cycle time. At this point, targets for LIPAS will be changed, and the cycle time efforts will become more important.

Thus, this recommendation is to set a goal of stable TPT and utilize improvements to increase wafer starts by 3.5%. This is dependent, of course, on unmodeled sections of the fab being able to manage this increase. This should be feasible, since the most capacity-challenged areas of the fab were included in the model.

Based on planning group projections for increases in products which require expediting, a dual product approach of expediting and normal production seems feasible. If the fab can absorb a 100% increase in the number of flag and hot lots, then cycle time demands can be met. If shift managers, group leaders, and technicians think that they can absorb a doubling of the number of these lots without ramification for other products, then the system is adequate. I believe that this will be manageable, if implemented in parallel with a focus on improving operations.

6.2.3 Run Lithography Layers To 0

Through simulation, lithography layer set-up times were found to make a significant difference in lithography queue times. Run-To-Zero policies dictate that all products in a given layer are run prior to performing a layer set-up. In the areas modeled, using a Run-To-Zero policy improves (modeled area) cycle time performance by 7%. In reality, the fab lithography area frequently breaks layer set-ups when a large batch arrives from diffusion. This action should be minimized, and lithography layers set-ups run until completed.

Some further reasons for this behavior should be understood. In the lithography area, technicians have a tendency to set-up for large batches instead of running the next batch, even if this means performing a long layer set-up. This is based in a large part on technician availability issues. Even though *machine time* is lost in layer set-ups, *technician* presence during this time is not required. This frees technicians to perform other area manual tasks such as inspection steps. On the other hand, running small batches requires constant technician presence for product set-ups.

6.3 Form a Permanent Line Management Continuous Improvement Team

The key to improving line management is the formation of internal company teams to monitor and improve performance. The use of continuous improvement teams to achieve these results is the best known industry practice for doing so.

Designing an organizational structure and obtaining buy-in to achieve sustainable improvement is probably the most difficult part of any of these efforts. Successfully performing these steps does not guarantee success, but *not* performing them virtually guarantees failure.

One initial key question for this research was "Is facility performance being compromised in order to locally optimize within functional areas?." Although the answer to this has not been proven, I believe that the answer is yes. One way that this can be changed, and negative effects

minimized, is through a line management continuous improvement team. This team could also serve to facilitate internal personnel management best-practices between the functionally-based clean room groups.

7. Unresolved Issues and Further Developments

7.1 Simulation Model Improvements

The goals of the simulation project were met, and from a practical standpoint I believe that fab personnel are better off focusing on implementing the recommendations than improving the model. However, having said this, there are opportunities for improvement. These include areas which would broaden model scope, those which would improve simulation performance, and those which would improve model integrity.

Upgrading the processor speed and available memory in the hardware running the program would immensely help to decrease the 8 hour period necessary for a one-year run.

Modifying the lithography set-up assumptions would help to improve the model. In particular, reticle changes necessary for product set-ups were not included in the model. Addition of this set-up procedure requires writing a brief program and then creating some new set-up tables.

Refining data in all forms of machine down time would improve the numerical accuracy of model recommendations for specific functional areas of the fab. In particular, unscheduled machine down time and time lost due to technicians would be improved.

Including machine operators in the model (as opposed to assigning personnel availability losses to machine down time) would improve the model.

Expanding the model to include certain processing steps would help to make model output even more believable. In particular, diffusion processes were not included in the model. These processes include relatively large batches of products, and thus deviate from the delay step assumption that lot processing times outside of modeled areas are normally distributed.

Finally, several WIP management policies are noticeably absent from this examination. Least Slack, Kanban and strategies which focus on retaining a level distribution of WIP on the line should be included. A more in-depth study of mixed-motive methodologies would also be helpful.

7.2 Facility Specific Concerns

Intel as a company has differentiated itself through bringing high quality designs into high quality, high volume manufacturing very quickly. Its manufacturing processes have been

designed to facilitate this strategy. The question of whether this fab can profitably compete in the cycle-time based ROM market while still producing high volumes of many products with very little spare capacity remains open.

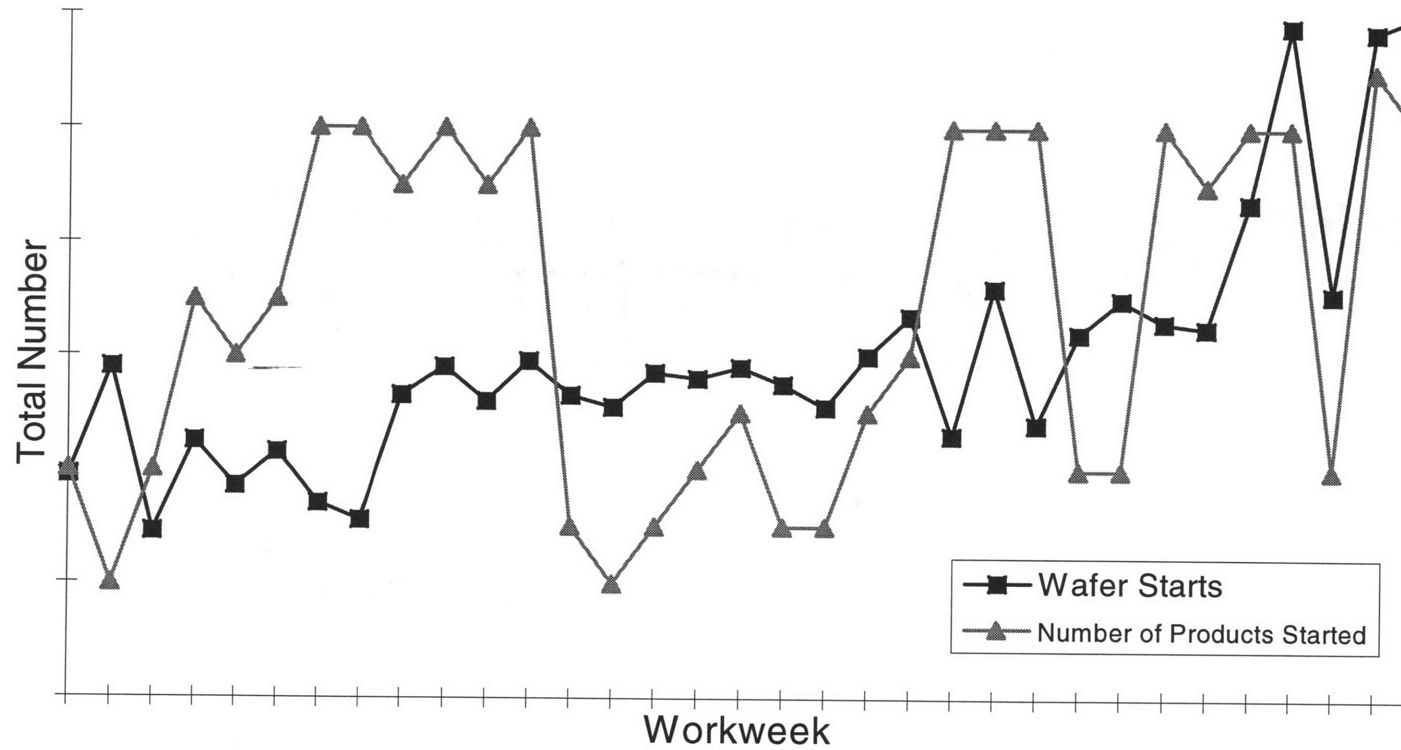
The option to add capacity to functional areas was never seriously discussed. The large expenses of buying semiconductor equipment and the current constraints in clean room space were the reasons for this. However, since these are not new processes, the used equipment market may be quite reasonably priced. As an alternative option, cost for upgrading the sort area to clean room grade operations may not be as expensive as initially assumed.

Finally, as fab personnel widely acknowledge, product demand is not stable in this fast-moving industry. There is no guarantee that there will be very high demand for products produced on current equipment technology levels next quarter, let alone 2 years from today.

The game of demand prediction and capacity expansion is one which management of this facility has chosen to play quite conservatively. So far, this approach has been successful.

Appendix A: Semiconductor Facility Data

Exhibit 1: Wafer Start Scenario



(Note that the Y-axis scales for Wafer Starts and Number of Products Started are different.)

Exhibit 2 Fab + Sort Cycle Time Data

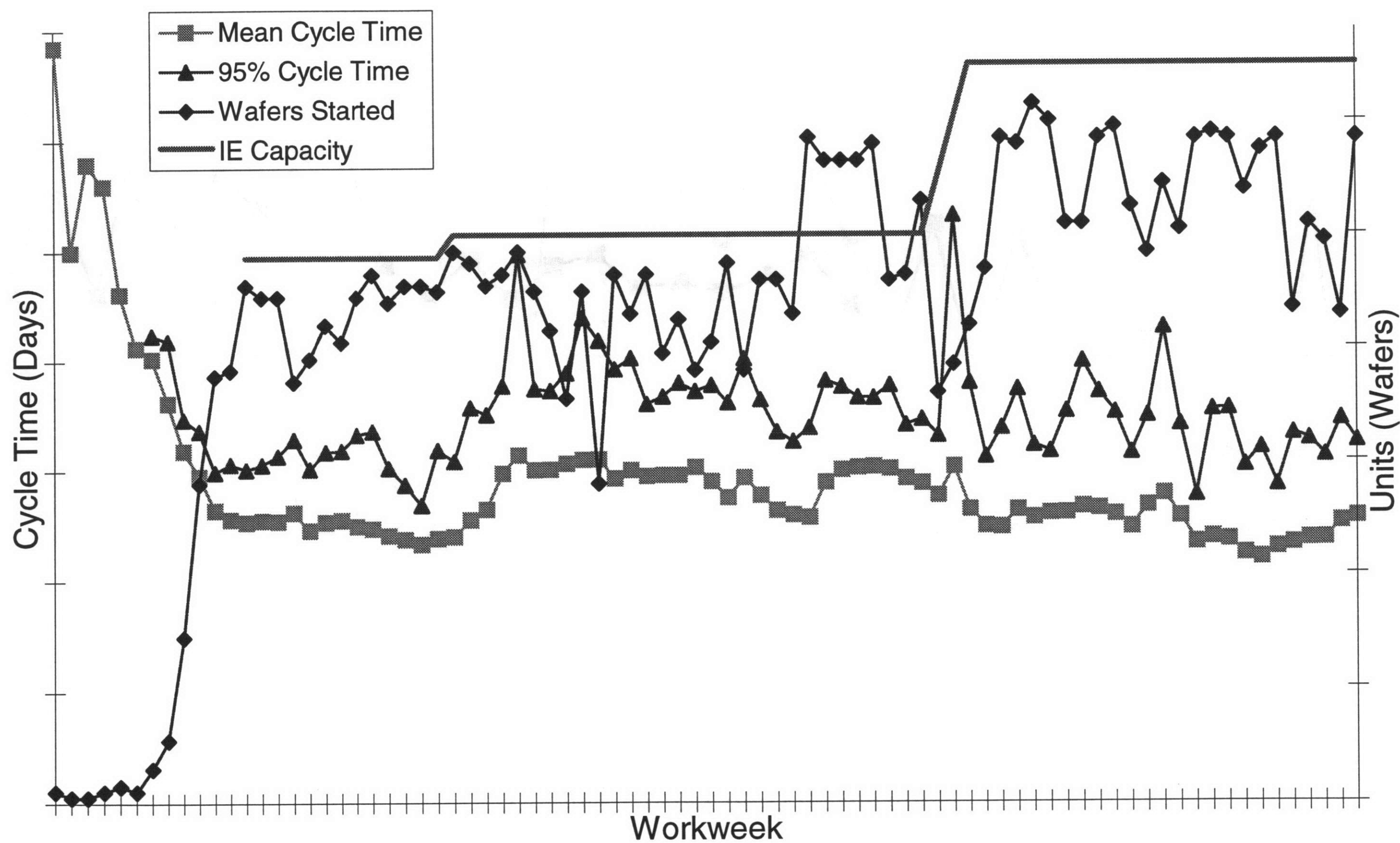
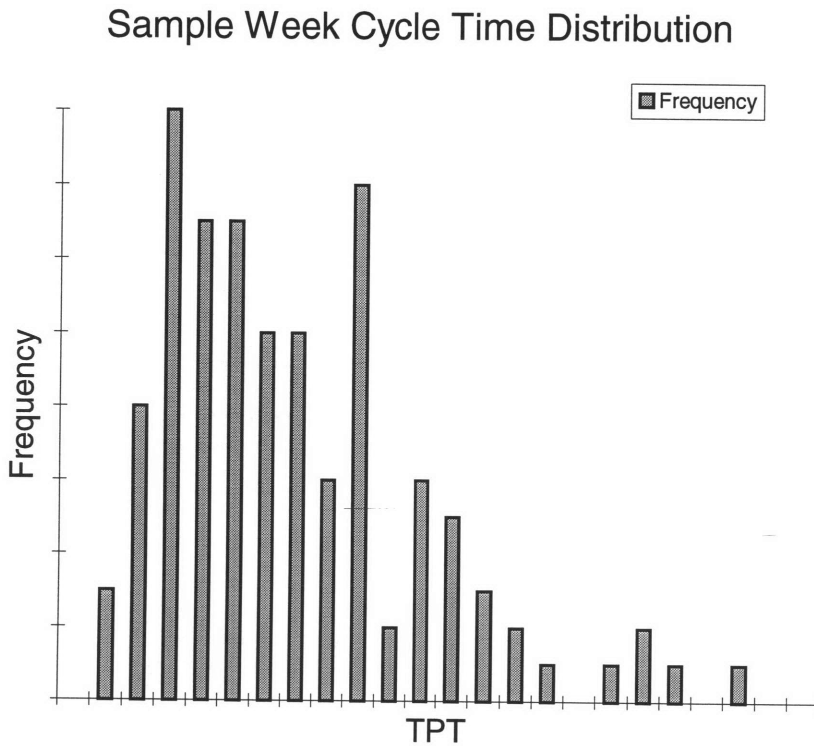
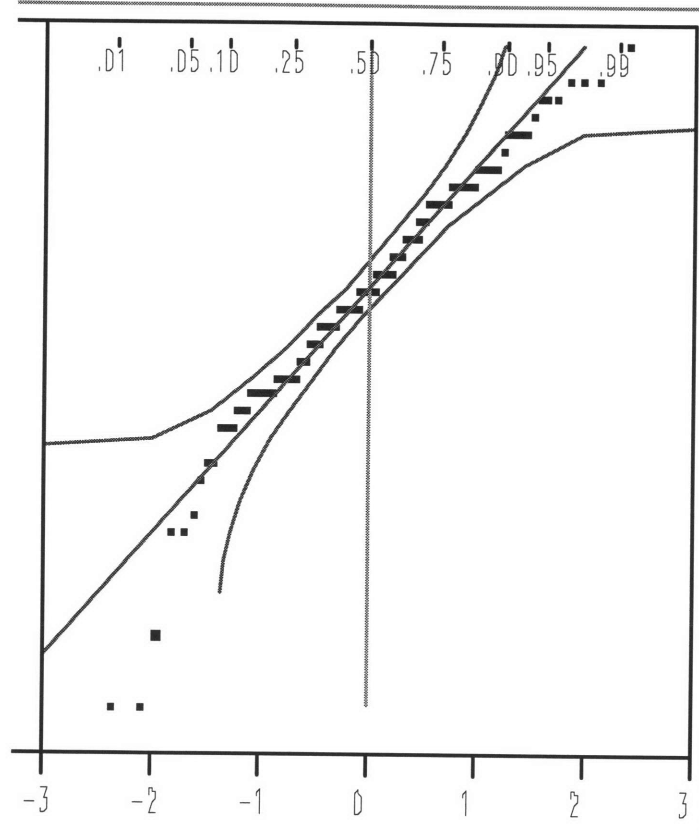


Exhibit 3: Sample Week Cycle Time Distribution And Normality Analysis





Test for Normality

Shapiro-Wilk W Test

W	Prob<W
0.954121	0.0027

Exhibit 4: Actual Lot Cycle Time Data

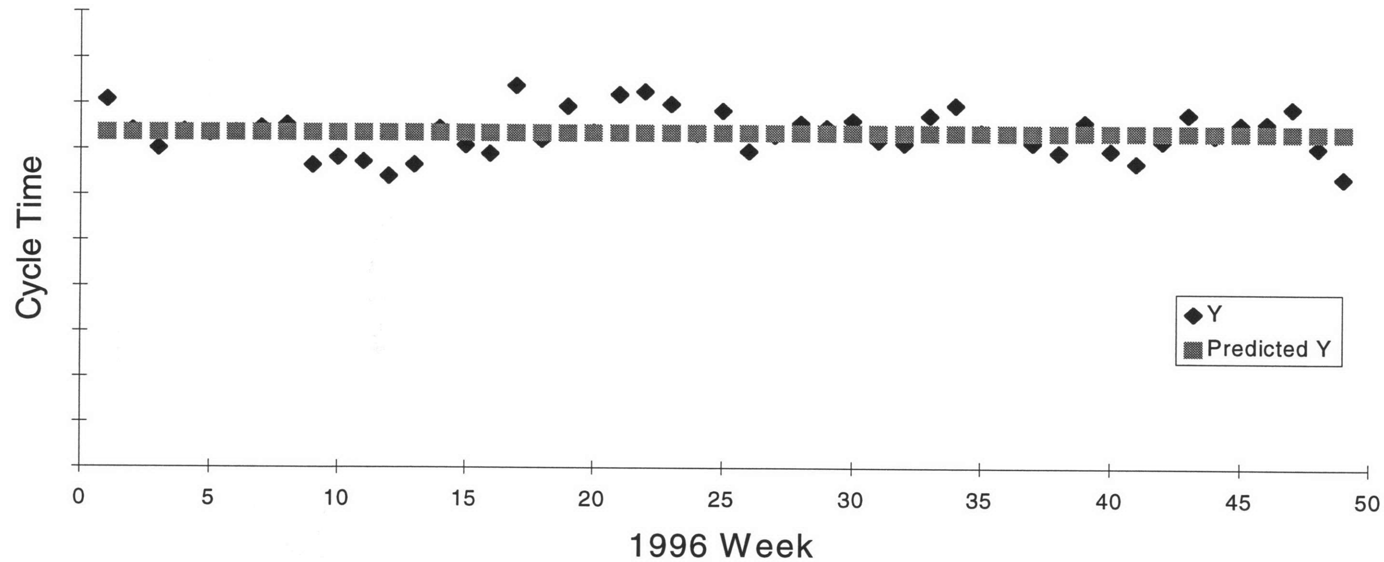
Regression Statistics	
Multiple R	0.8292
R Square	0.6876
Adjusted R Square	0.6813
Standard Error	2.9208
Observations	52

ANOVA

	df	SS	MS	F	Significance F
Regression	1	938.68	938.68	110.03	0.00
Residual	50	426.54	8.53		
Total	51	1365.22			

	Coefficients	Standard Error	t Stat	P-value
Intercept	61.85	0.82	75.25	0.00
X Variable 1	-0.28	0.03	-10.49	0.00

Exhibit 5: Regression Fit to Agglomerate Calculated Cycle Time



Regression Statistics	
Multiple R	0.0116
R Square	0.0001
Adjusted R Square	-0.0211
Standard Error	2.37
Observations	49

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.04	0.04	0.01	0.94
Residual	47	264.05	5.62		
Total	48	264.09			

	Standard Error	*t Stat*	*P-value*
Intercept	0.69	53.48	0.00
X Variable 1	0.02	0.08	0.94

Exhibit 6: Cycle Time Histogram

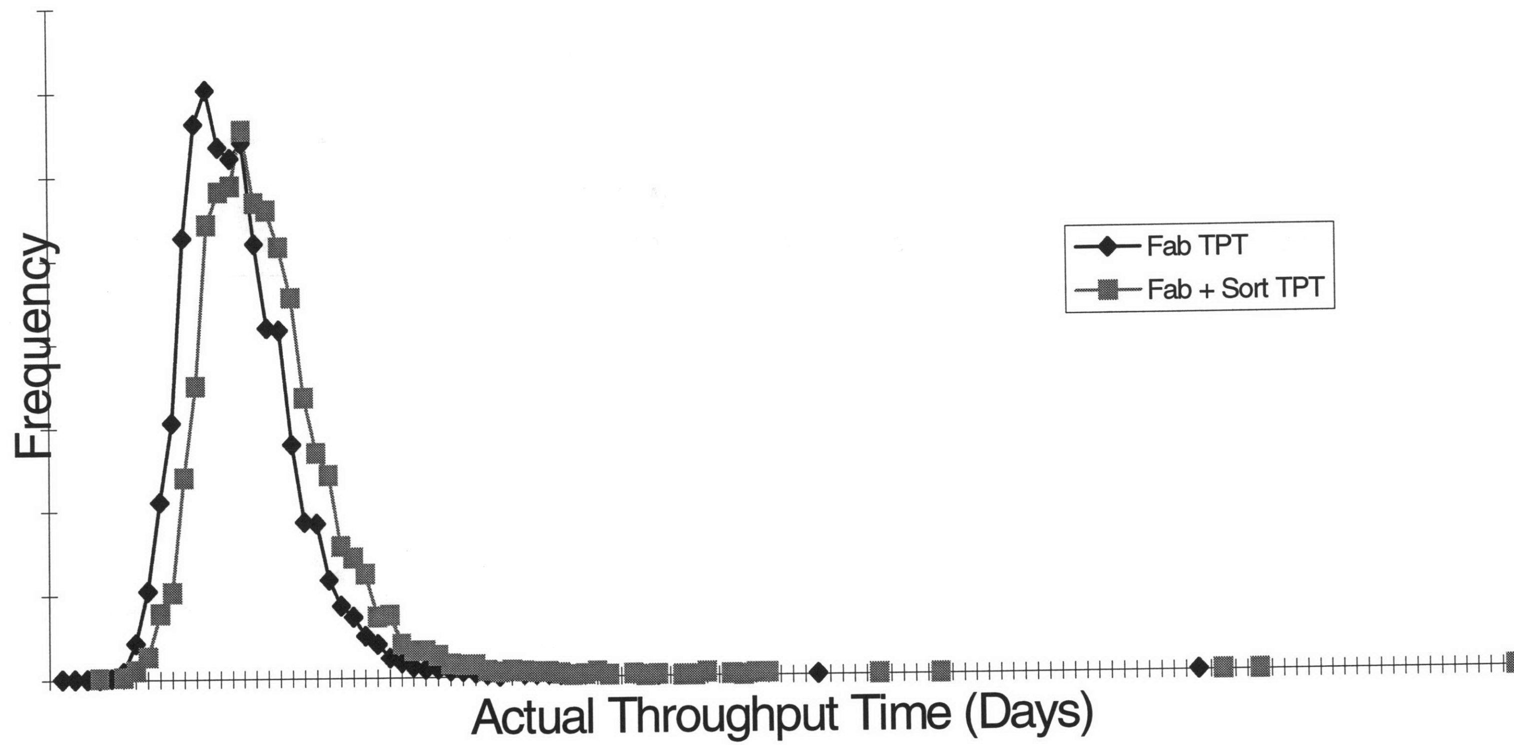
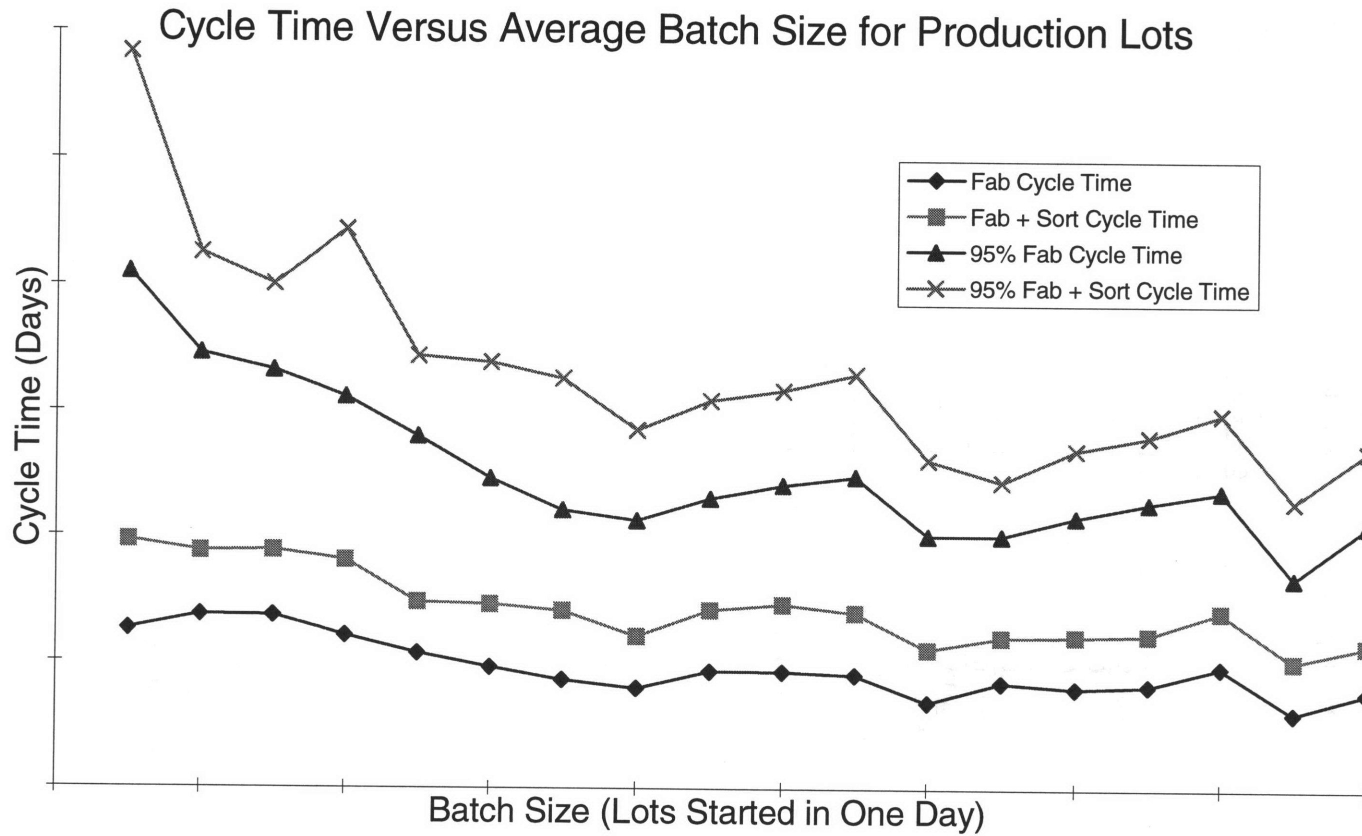


Exhibit 7: Cycle Time versus Average Batch Size



Appendix B: Simulation Input

Exhibit 8: Sample Simulation Input File

Wafers were assigned a specific start time in the morning and one in the evening for each day of the week. This initial order was repeated on a weekly basis. The tabular format for these inputs is shown below:

Order	Lot	Start	Initiatl Value	Repeat Interval	Units	Repeat Number
O1	Lot1	08:01:96:00:00:00	1	6	day	80
O1	Lot2	08:01:96:00:00:00	1	6	day	80
O1	Lot3	08:01:96:00:00:00	1	6	day	80
O1	Lot4	08:01:96:00:00:00	1	6	day	80
O1	Lot5	08:01:96:00:00:00	1	6	day	80
O1	Lot6	08:01:96:00:00:00	1	6	day	80
O1	Lot7	08:01:96:00:00:00	1	6	day	80
O1	Lot8	08:01:96:00:00:00	1	6	day	80
O1	Lot9	08:01:96:00:00:00	1	6	day	80
O1	Lot10	08:01:96:00:01:01	1	6	day	80
O1	Lot11	08:01:96:00:01:01	1	6	day	80
O1	Lot12	08:01:96:00:01:01	1	6	day	80
O1	Lot13	08:01:96:00:01:01	1	6	day	80
O1	Lot14	08:01:96:00:01:01	1	6	day	80
O1	Lot15	08:01:96:00:01:01	1	6	day	80
O1	Lot16	08:01:96:00:01:01	1	6	day	80
O1	Lot17	08:01:96:00:01:01	1	6	day	80
O1	Lot18	08:01:96:00:01:01	1	6	day	80
O1	Lot19	08:01:96:00:01:01	1	6	day	80
O1	Lot20	08:01:96:00:01:01	1	6	day	80
O1	Lot21	08:01:96:00:01:01	1	6	day	80
O1	Lot22	08:01:96:00:01:01	1	6	day	80
O23	LotB1	08:01:96:22:00:00	1	6	day	80
O23	LotB2	08:01:96:22:00:00	1	6	day	80
O23	LotB3	08:01:96:22:00:00	1	6	day	80
O23	LotB4	08:01:96:22:00:00	1	6	day	80
O23	LotB5	08:01:96:22:00:00	1	6	day	80
O23	LotB6	08:01:96:22:00:00	1	6	day	80
O23	LotB7	08:01:96:22:00:00	1	6	day	80
O23	LotB8	08:01:96:22:00:00	1	6	day	80
O23	LotB9	08:02:96:06:00:00	1	6	day	80
O23	LotB10	08:02:96:06:00:00	1	6	day	80
O23	LotB11	08:02:96:06:00:00	1	6	day	80
O23	LotB12	08:02:96:06:00:00	1	6	day	80
O23	LotB13	08:02:96:06:00:00	1	6	day	80
O23	LotB14	08:02:96:06:00:00	1	6	day	80
O23	LotB15	08:02:96:06:00:00	1	6	day	80
O23	LotB16	08:02:96:06:00:00	1	6	day	80
O39	LotC1	08:02:96:14:00:00	1	6	day	80
O39	LotC2	08:02:96:14:00:00	1	6	day	80
O41	LotD1	08:02:96:16:00:00	1	6	day	80
O41	LotD2	08:02:96:16:00:00	1	6	day	80
O43	LotE1	08:02:96:18:00:00	1	6	day	80
O43	LotE2	08:02:96:18:00:00	1	6	day	80
O43	LotE3	08:02:96:18:00:00	1	6	day	80
O43	LotE4	08:02:96:18:00:00	1	6	day	80
O43	LotE5	08:02:96:18:00:00	1	6	day	80
O43	LotE6	08:02:96:18:00:00	1	6	day	80

Exhibit 9: Sample Station Definition File

The station definition table links each station in the model with its functional area, referred to as "station family" in this simulation package. Here the modeler specifies whether to refer to each family as a batch process or a unit processor, and the criteria for forming a batch. For example, the batch criteria crit_spartstep is shown. This is a custom-written criteria which batches lots only if they are the same product at the same step in the process. Batch sizes which have the minimum number of lots but not the maximum defined in this table wait the amount of time specified in the column Wait Limit until being ready to process. The batch criteria crit_anybody allows any products to batch together.

			Batch	Minimum	Maximum		Wait	Limit
Station Family	Station	WIP Rule	Method	Batch Size	Batch Size	Batch Criteria	Limit	Units
Etch	1	rule_FIFO_B	lot	2	4	crit_spartstep	1	hr
	2	rule_FIFO_B	lot	2	4	crit_spartstep	1	hr
	3	rule_FIFO_B	lot	2	4	crit_spartstep	1	hr
Litho	1	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
	2	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
	3	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
	4	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
	5	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
	6	rule_FIFO_B	lot	5	15	crit_spartstep	2	min
Depostions	1	rule_FIFO						
	2	rule_FIFO						
	3	rule_FIFO						
Metal	1	rule_FIFO						
	2	rule_FIFO						
	3	rule_FIFO						
	4	rule_FIFO						
	5	rule_FIFO						
	6	rule_FIFO						
Passivation	1	rule_FIFO						
	2	rule_FIFO						
Adhesion	1	rule_FIFO						
	2	rule_FIFO						
	3	rule_FIFO						
	4	rule_FIFO						
Clean	1	rule_FIFO_B	lot	2	2	crit_anybody		
Web	1	rule_FIFO						
	2	rule_FIFO						
Transport	trolley1	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t2	trolley2	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t3	trolley3	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t4	trolley4	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t5	trolley5	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t6	trolley6	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t7	trolley7	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t8	trolley8	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t9	trolley9	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t11	trolley10	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t12	trolley11	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t13	trolley12	rule_FIFO_B	lot	1	4	crit_anybody	15	min
t14	trolley13	rule_FIFO_B	lot	1	4	crit_anybody	15	min

Exhibit 10: Sample Simulation Process Flow

Exhibit 10 shows a sample process flow definition table. Note that most modeled process steps show processing time as constant, and each delay is normally distributed. Line yield is modeled as a constant number, with each lithography step having an equal probability of causing a scrap on each wafer. Note that some steps have no set-ups, some their own set-ups, and some share set-ups with other layers. Finally, note the value add column. In this column, value is added to a part based on the litho step. This will be referred to in later discussion.

	Value	Station	Processing		Standard	Time	Proc.	Set	Set Up	
Step	Added	Family	Time	Distribution	Deviation	Units	Unit	Up	Time	Yield
1		delay	100	normal	10	min	lot			
2		t1	10	constant		min	lot			
3		Litho	3	constant		min	piece	litho1	20	80
4		delay	100	normal	10	min	lot			
5		t2	10	constant		min	lot			
6		Litho	6	constant		min	piece	litho2	20	80
7		delay	100	normal	10	min	lot			
8		Litho	3	constant		min	piece	litho1	20	80
9		delay	100	normal	10	min	lot			
10		t4	10	constant		min	lot			
11	1	Litho	9	constant		min	piece	litho3	20	80
12		delay	100	normal	10	min	lot			
13		Adhesion	0.1	constant		min	piece			
14		delay	100	normal	10	min	lot			
15		t10	10	constant		min	lot			
16		Deposition	5	constant		min	piece			
17		delay	100	normal	10	min	lot			
18		Web	10	constant		min	piece			
19		delay	100	normal	10	min	lot			
20		t12	10	constant		min	lot			
21	1	Litho	3	constant		min	piece	litho1	20	80
22		delay	100	normal	10	min	lot			
23		Clean	10	constant		min	lot			
24		delay	100	normal	10	min	lot			
25		delay	100	normal	10	min	lot			
26		t14	10	constant		min	lot			
27		Metal	3	constant	7	min	piece			
28		delay	100	normal	10	min	lot			
29		t15	10	constant		min	lot			
30	1	Litho	3	constant		min	piece	litho3	20	80
31		delay	100	normal	10	min	lot			
32		t16	10	constant		min	lot			
33		Etch	60	constant	4	min	lot	etch3	1	
34		delay	100	normal	10	min	lot			
35		t17	10	constant		min	lot			
35		Depostion	3	constant		min	piece			
36		delay	100	normal	10	min	lot			
37		t18	10	constant		min	lot			
38		Web	10	constant		min	piece			
39		delay	100	normal	10	min	lot			
40		t19	10	constant		min	lot			
36		t20	10	constant		min	lot			
37		Etch	50	constant	4	min	lot	etch1	1	
38		t21	10	constant		min	lot			
39		Metal	1	constant	15	min	piece			
40		delay	100	normal	10	min	lot			
41		t22	10	constant		min	lot			
42	1	Litho	3	constant		min	piece	litho1	20	80
43		delay	100	normal	10	min	lot			
44		t23	10	constant		min	lot			

Exhibit 11: Sample Simulation Down Time Table

Station family down times and distributions are assigned as shown below:

Type	Calendar	Method	Distribution	Mean	Standard Dev.	Units
down	Etch	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Litho	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Metal	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Passivation	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Adhesion	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Clean	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Web	MTTF_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
down	Lam	MTTF_by_cal	Exponential	100		hr
		MTTR	Weibull	5	1	hr
pm	EtchPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	LithoPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	MetalPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	PassivationPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	AdhesionPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	CleanPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	WebPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	normal	5	1	hr
pm	LamPM	MTTPM_by_cal	Exponential	100		hr
		MTTR	Weibull	5	1	hr
down	Sabbath	MTTF_by_cal	Constant	143		hr
		MTTR	Constant	1		hr
down	EtchTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	LithoTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	MetalTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	PassivationTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	AdhesionTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	CleanTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	WebTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min
down	LamTechnician	MTTF_by_cal	Exponential	90		min
		MTTR	normal	5	1	min

Exhibit 12: Sample Station Down Time Attachment File

The table used to attach station family down times to stations is shown below:

Resource Type	Resource Name	Calendar Type	Calendar Name
stn	1	down	Etch
stn	2	down	Etch
stn	3	down	Etch
stn	4	down	Etch
stn	5	down	Etch
stn	6	down	Etch
stn	7	down	Etch
stn	8	down	Etch
stn	1	pm	EtchPM
stn	2	pm	EtchPM
stn	3	pm	EtchPM
stn	4	pm	EtchPM
stn	5	pm	EtchPM
stn	6	pm	EtchPM
stn	7	pm	EtchPM
stn	8	pm	EtchPM
stn	1	down	EtchTechnician
stn	2	down	EtchTechnician
stn	3	down	EtchTechnician
stn	4	down	EtchTechnician
stn	5	down	EtchTechnician
stn	6	down	EtchTechnician
stn	7	down	EtchTechnician
stn	8	down	EtchTechnician
stn	1	down	Litho
stn	2	down	Litho
stn	3	down	Litho
stn	4	down	Litho
stn	5	down	Litho
stn	6	down	Litho
stn	1	pm	LithoPM
stn	2	pm	LithoPM
stn	3	pm	LithoPM
stn	4	pm	LithoPM
stn	5	pm	LithoPM
stn	6	pm	LithoPM
stn	1	down	LithoTechnician
stn	2	down	LithoTechnician
stn	3	down	LithoTechnician
stn	4	down	LithoTechnician
stn	5	down	LithoTechnician
stn	6	down	LithoTechnician

Exhibit 13: Sample Simulation Rule Editor

WIP prioritization rules are assigned to station families in the following table. From a programming standpoint, this is the most complex portion of the model, both in terms of understanding and implementation. During a model run, when the simulator clock moves forward, the program checks each station to see if it has become available. If it has, the area rule is invoked to choose the next action to process. Actions pass through a series of filters until a match condition is satisfied, and the true action is performed. For example, for any station which is not down or processing, preventative maintenance is the highest priority order. Thus the first filter invoked in every rule is the filter_PM, which checks the entire station queue to find out if a preventative maintenance work order is waiting. If it is, this station invokes the true action of Do_PM, and the processor moves on to check whether the next station is available.

In the cases where a lot is processed, a function has been programmed to give each lot a priority. This function assigns a rank to each lot in queue based either on some function such as the random number generation rank_Random(this), or based on some lot attribute such as the value of the lot shown as gAvalue(this). The default priority assignment is to give the highest priority to the first lot in the queue. In this program, the part attribute "lot value" is incremented with each lithography step after metal deposition begins as discussed in Section 0. This allows ranked choices between lots of varying position in the process, used in the station rules for Back-To-Front and Front-To-Back WIP management priorities.

Batching rules are more complicated, with filters to check that batching conditions have been met. Additional programming is used in some rules to give priority to lots which run on the same set-up. These rules were used in the lithography area, to examine the effects of assigning priorities partially based on layer set-ups.

Rule	Same	Source	Rank Function	Priority	Filter Name	True Action	False Action
Random		Link			filter_PM	Do_PM	
		FWL & Link	rank_Random(this)	Lowest	filter_CanDo	Do_First	Wait
Random batch		Link			filter_PM	Do_PM	
		FWL & Link	rank_Random(this)	Lowest	filter_CanDo		Wait
					filter_FirstBatch	Do_Batch	Wait
High Value Batch		Link			filter_PM	Do_PM	
		FWL & Link	gAvalue(this)	Highest	filter_CanDo		Wait
					filter_BatchReps		Wait
	sameone				filter_SameSetup		SKIPTO: notsameone
					filter_TransGroup	Do_Batch	
	notsameone	SRCE: sameone			filter_NoBroSetup		SKIPTO: firstone
					filter_TransGroup	Do_Batch	
	firstone	SRCE: sameone			filter_TransGroup	Do_Batch	Wait
Fifo Batch		Link			filter_PM	Do_PM	
		FWL & Link			filter_CanDo		Wait
					filter_BatchReps		Wait
	sameone				filter_SameSetup		SKIPTO: notsameone
					filter_TransGroup	Do_Batch	
	notsameone	SRCE: sameone			filter_NoBroSetup		SKIPTO: firstone
					filter_TransGroup	Do_Batch	
	firstone	SRCE: sameone			filter_TransGroup	Do_Batch	waitdsd
Random Batch		Link			filter_PM	Do_PM	
		FWL & Link	rank_Random(this)	Highest	filter_CanDo		Wait
					filter_BatchReps		Wait
	sameone				filter_SameSetup		SKIPTO: notsameone
					filter_TransGroup	Do_Batch	
	notsameone	SRCE: sameone			filter_NoBroSetup		SKIPTO: firstone
					filter_TransGroup	Do_Batch	
	firstone	SRCE: sameone			filter_TransGroup	Do_Batch	wait

Appendix C: Appendix of Simulation Results

Exhibit 14: Tabular Summary of Simulation Results

All Positive Percentages Show Average Increase Over Base Case

1. WIP MGT	Back To Front	Random	FIFO
Cycle Time Delta Base Case	0%	-6%	-6%
95% Cycle Time Delta Back To Front Base Case	0%	1%	-7%

	Shortest Processing Time	Front To Back	X Theoretical
Cycle Time Delta Base Case	23%	10%	-11%
95% Cycle Time Delta Base Case	79%	44%	-20%

2. Demand Changes	Back To Front -1.4%	X Theoretical -1.4%	Back To Front + 1.4%
Cycle Time Delta Base Case	1%	-6%	12%
95% Cycle Time Delta Base Case	32%	-11%	54%

	X Theoretical +1.4%	Back To Front + 3.5%	X Theoretical +3.5%	FIFO + 3.5%
Cycle Time Delta Base Case	11%	19%	26%	42%
95% Cycle Time Delta Base Case	8%	80%	28%	50%

3: Peak Demand

TPT Results	Back To Front Peak	Random Peak	X Theoretical Peak	FIFO Peak
Cycle Time Delta Base Case	-	7%	2%	6%
95% Cycle Time Delta Base Case	-	17%	10%	7%

Exhibit 14 (Continued)

Tech Down Time = Machine Down Time Due To Technician Absence, 1% Technician Down Time = Reduction In Total Down Time Due To Technician To 1%

	All Ops	All Ops	All Ops	Deposition
4. Down Time Results	**Down Time 1%**	**1% Unsched Down Time**	**1/2 Tech Down Time**	**Down Time 1%**
Cycle Time Delta Base Case	-39%	-25%	-15%	-12%
95% Cycle Time Delta Base Case	-28%	-16%	6%	-9%

	Litho	Litho and Deposition	All Ops Except Litho and
	1% Tech Down Time	**1% Tech Down Time**	**Deposition 1 % Tech Down Time**
Cycle Time Delta Base Case	-13%	-20%	-19%
95% Cycle Time Delta Base Case	-8%	-12%	-14%

5: Statistical Checks			Down Time MTTR	
TPT Results	**No Variance**	**No Variance Random**	**Standard Dev Double**	**No WIP Limits**
Cycle Time Delta Base Case	4%	-5%	-1%	0%
95% Cycle Time Delta Base Case	-2%	-10%	3%	0%

6: Recommendations	Increase Wafer Starts by 2%,		
TPT Results	**Decrease Tech DT by 50%**	**Random No SSU**	**Back To Front No SSU**
Cycle Time Delta Base Case	-26%	10%	7%
95% Cycle Time Delta Base Case	20%	13%	3%

SSU Is the Set Up assumption which dictates that lots with the same lithography layer set-up as the previous lot processed on a machine are given preference over others in the lithography area.

Exhibit 15: Average Lot Cycle Time Data for WIP Management Strategies

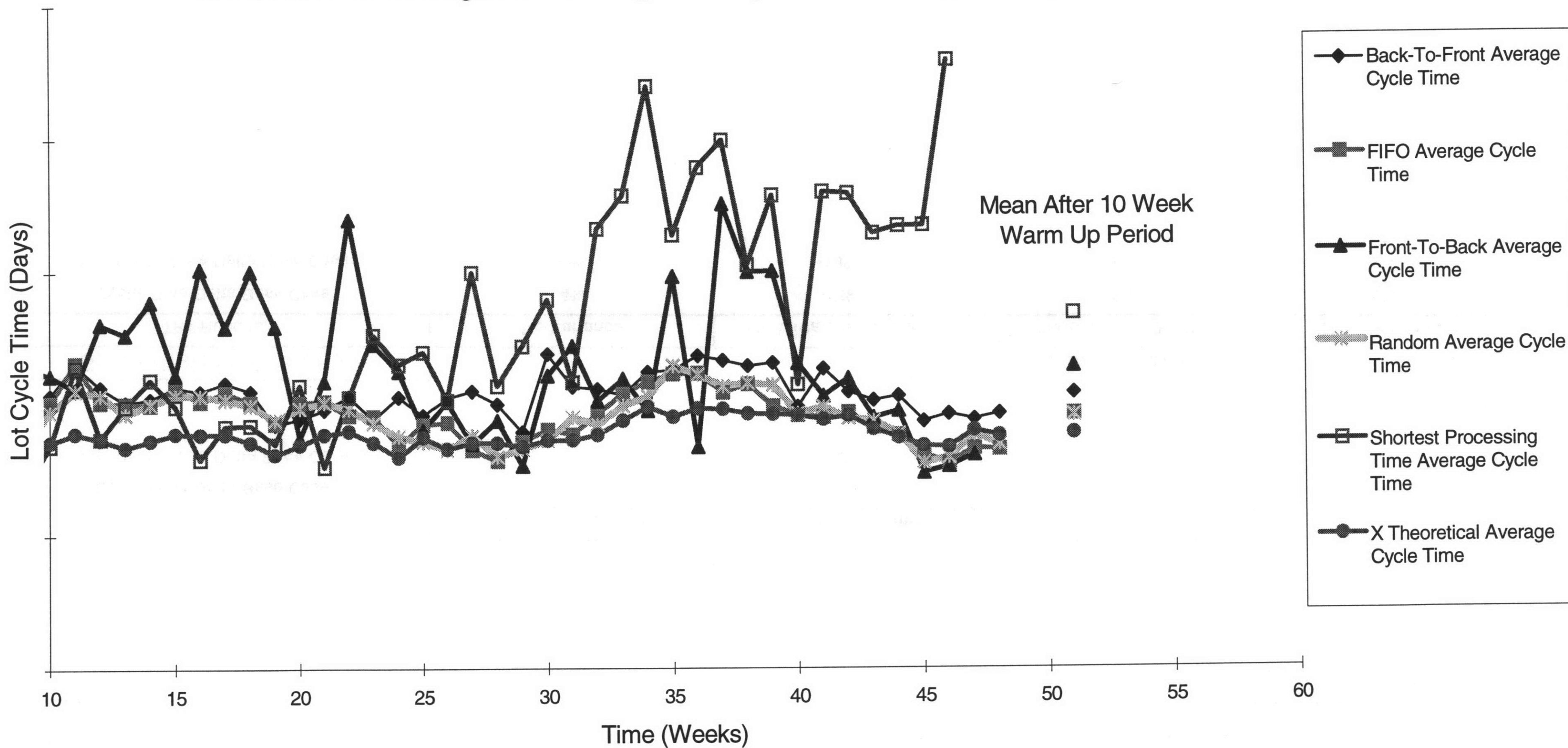
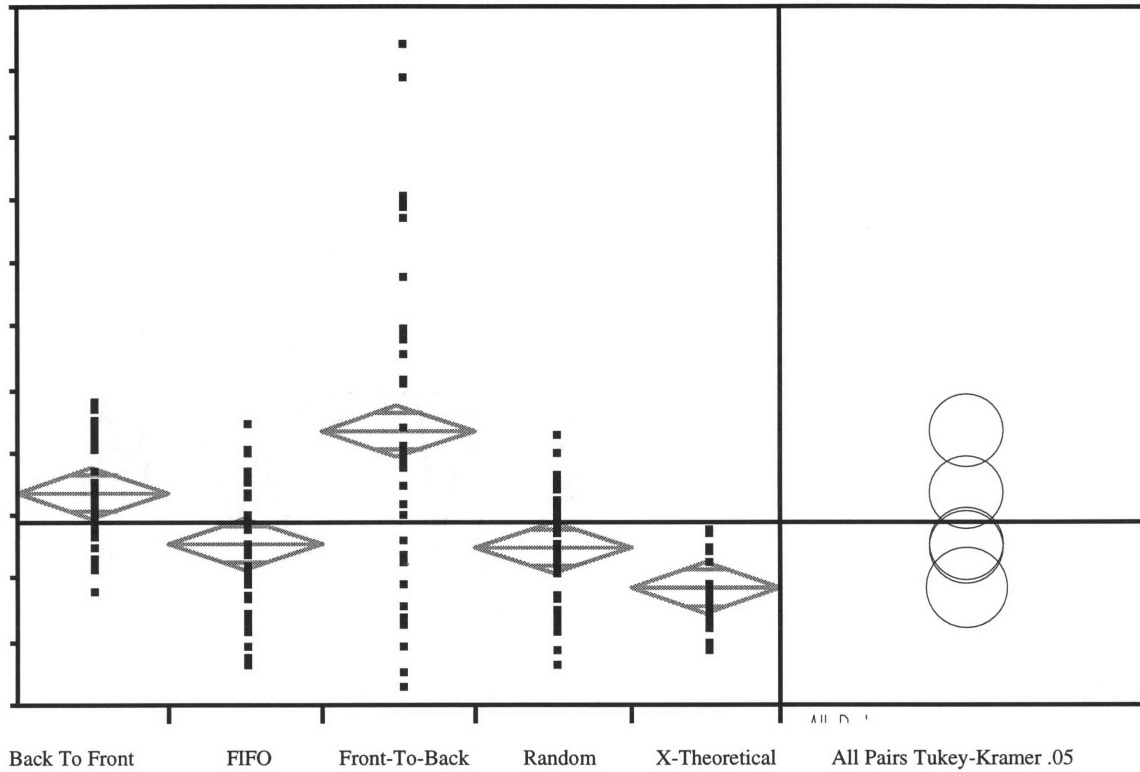


Exhibit 16: ANOVA and Tukey-Kramer Tests for WIP Management Methods Mean Cycle Time Comparison



Comparisons for all pairs using Tukey-Kramer HSD

Test:	Front-To-Back	FIFO	Random	X Theoretical
Back To Front	0.12343	-0.07981	-0.04008	0.60895

Positive values show pairs of means that are significantly different to the 95th percentile confidence.

Exhibit 17: 95% Lot Cycle Time Data for WIP Management Strategies

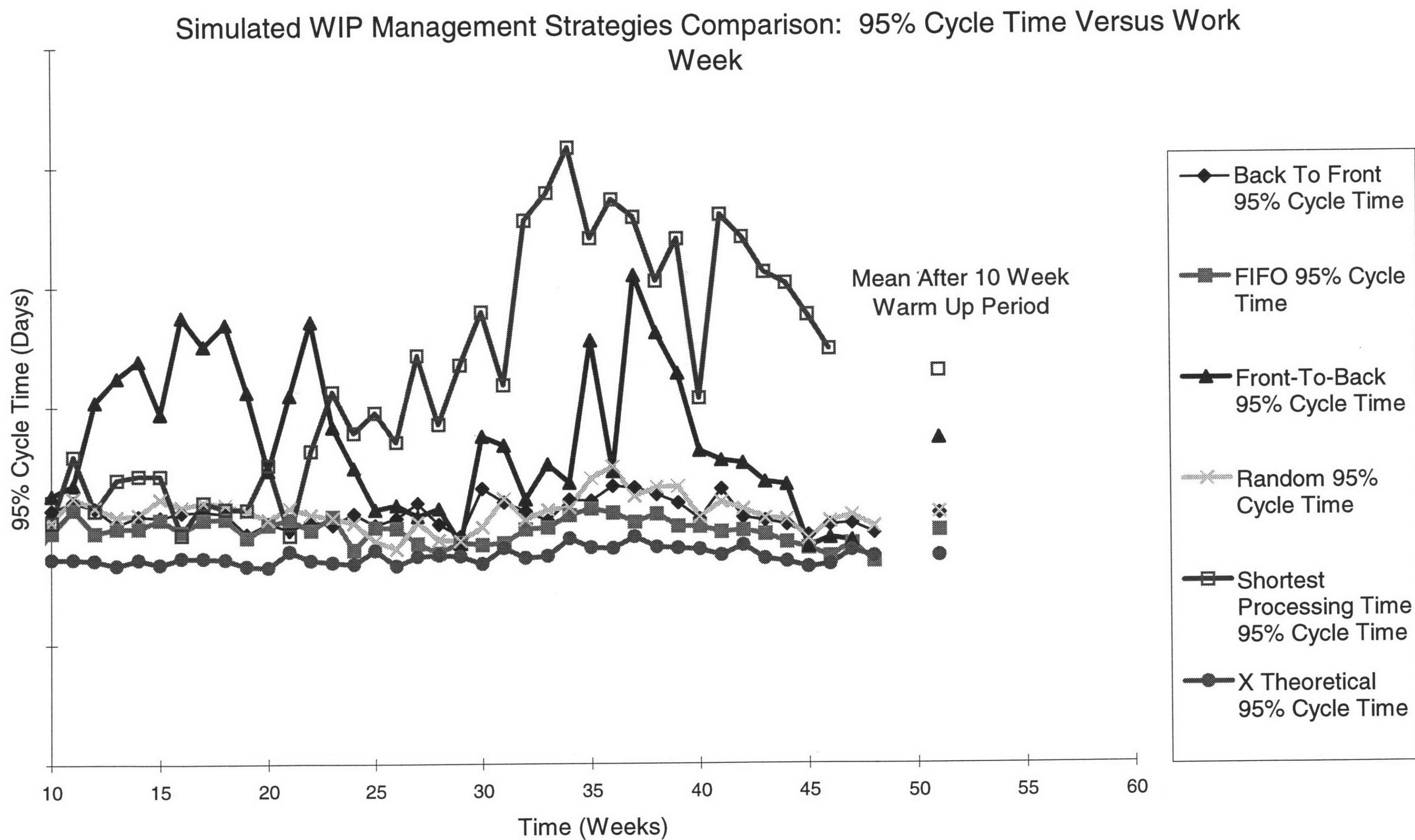
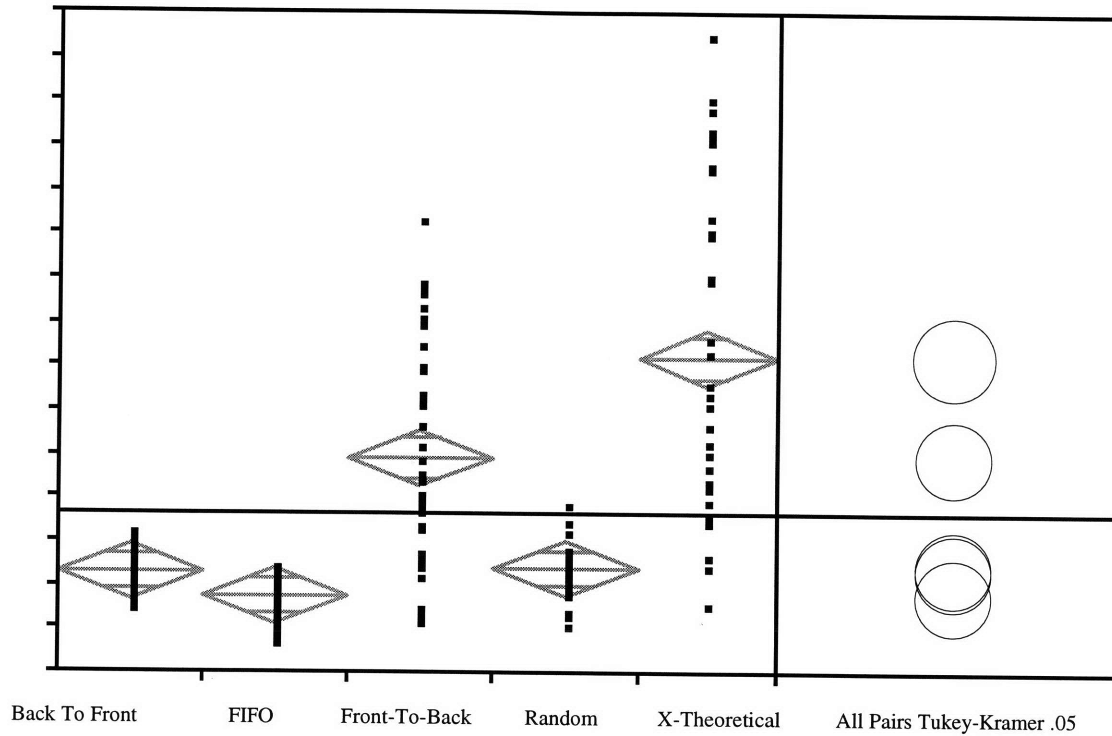


Exhibit 18: ANOVA and Tukey-Kramer Tests for WIP Management Methods 95% Cycle Time Comparison



Comparisons for all pairs using Tukey-Kramer HSD

Test:	Front-To-Back	FIFO	Random	X Theoretical
Back To Front	3.2090	-1.8721	-3.1353	8.9600

Positive values show pairs of means that are significantly different to the 95th percentile confidence.

Exhibit 19: Average Lot Cycle Time Data for Simulated Demand Changes

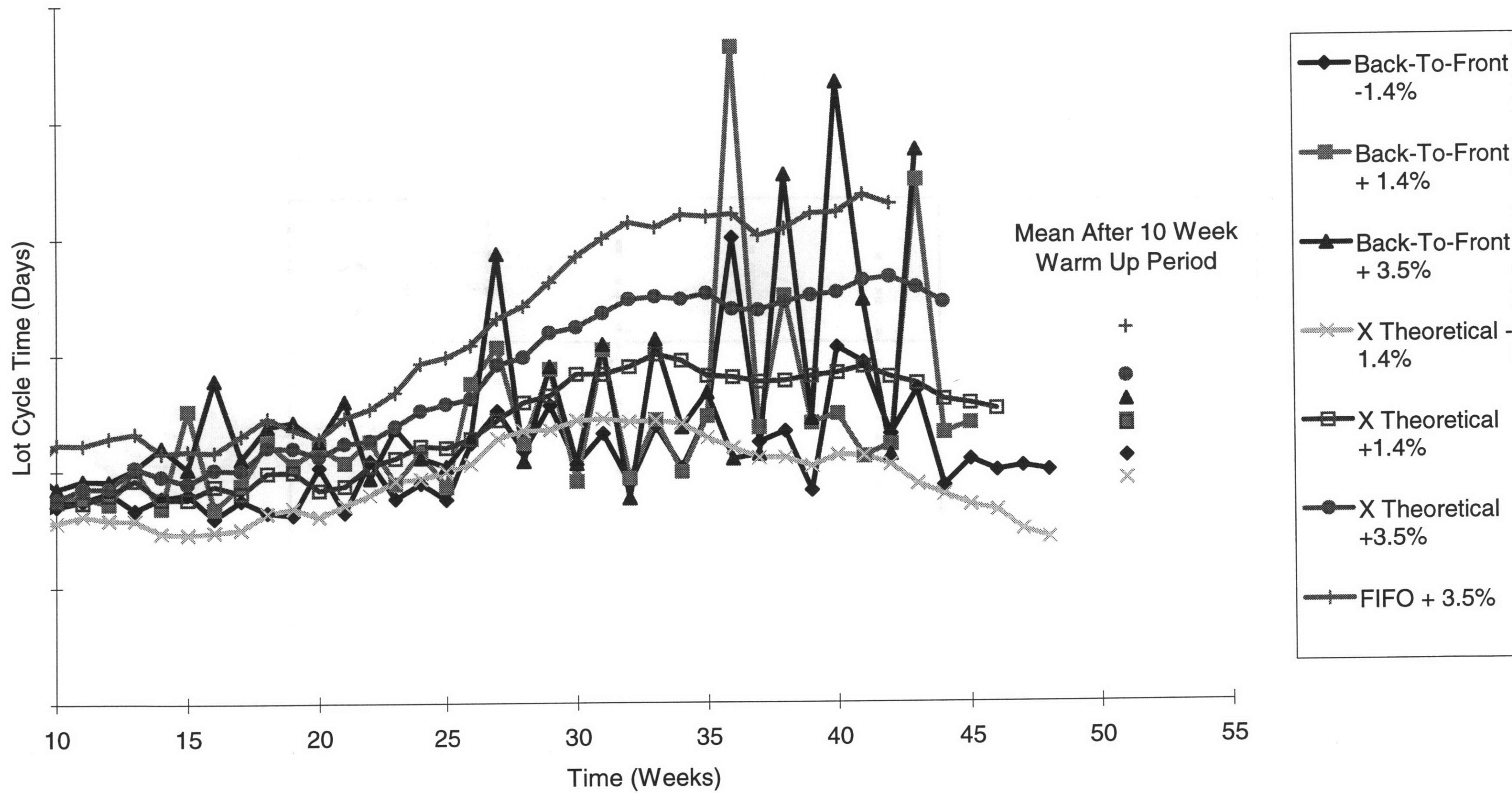


Exhibit 20: 95% Lot Cycle Time Data for Simulated Demand Changes

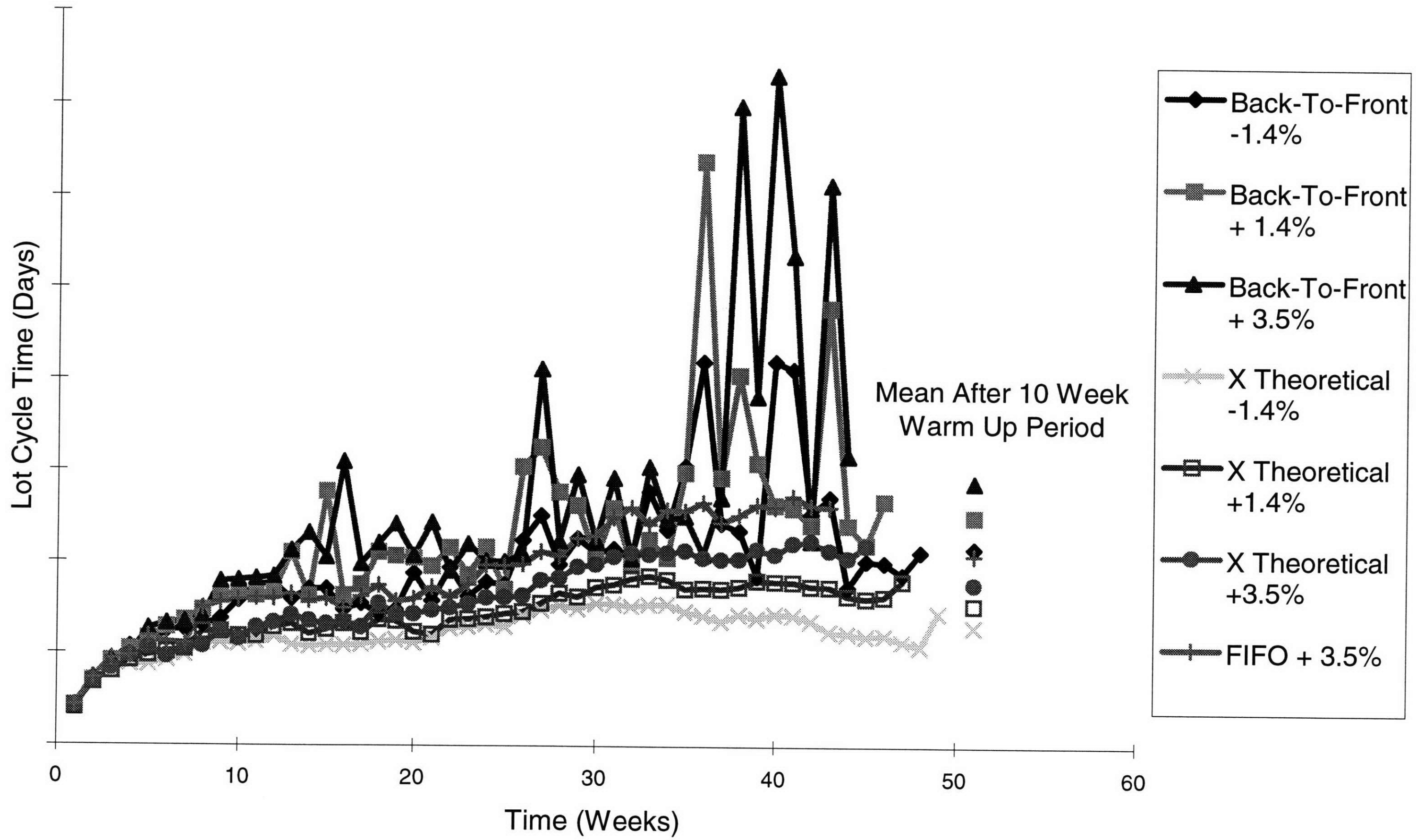


Exhibit 21: Average Lot Cycle Time Data for Simulated Demand Peaks

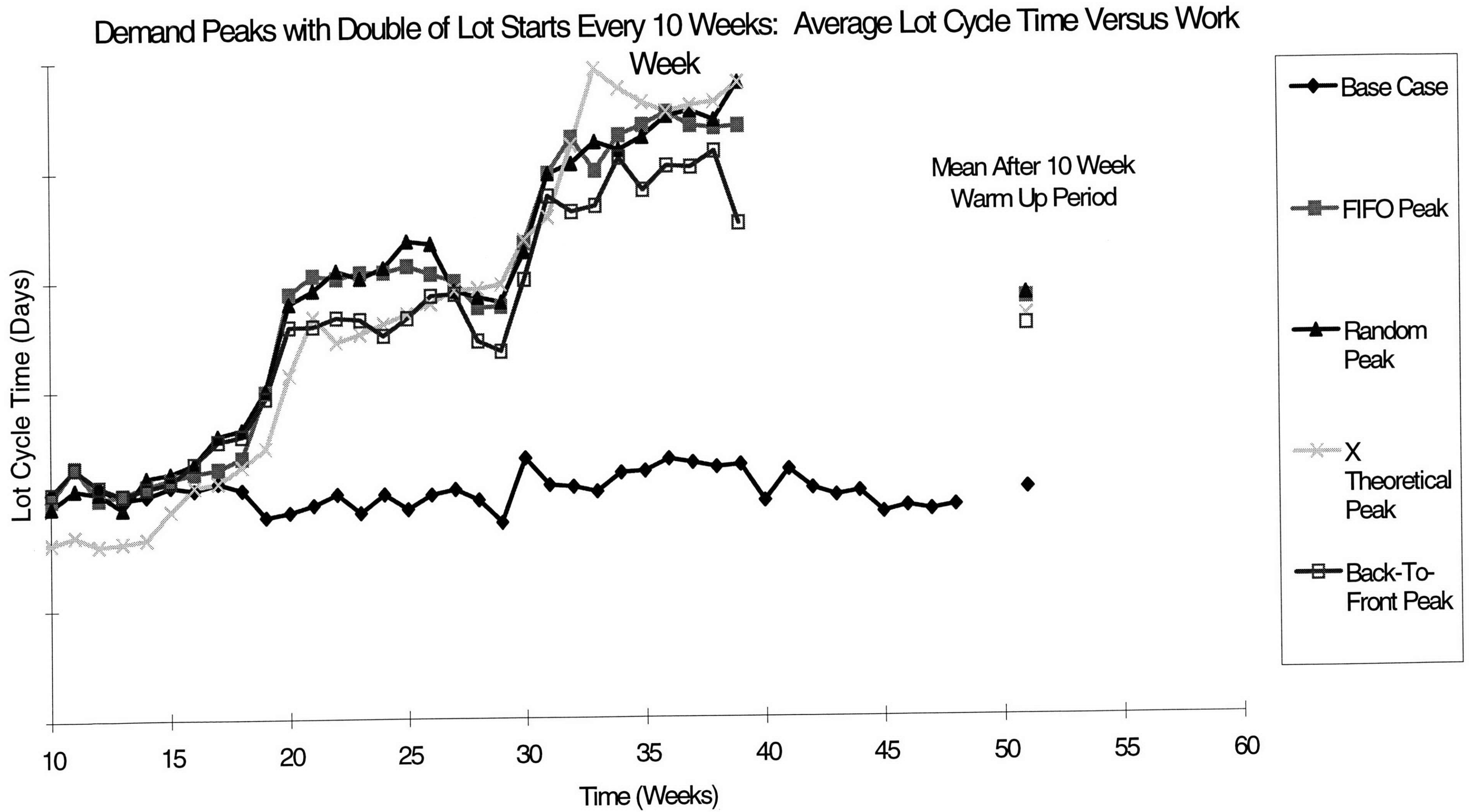


Exhibit 22: 95% Cycle Time Data for Simulated Demand Peaks

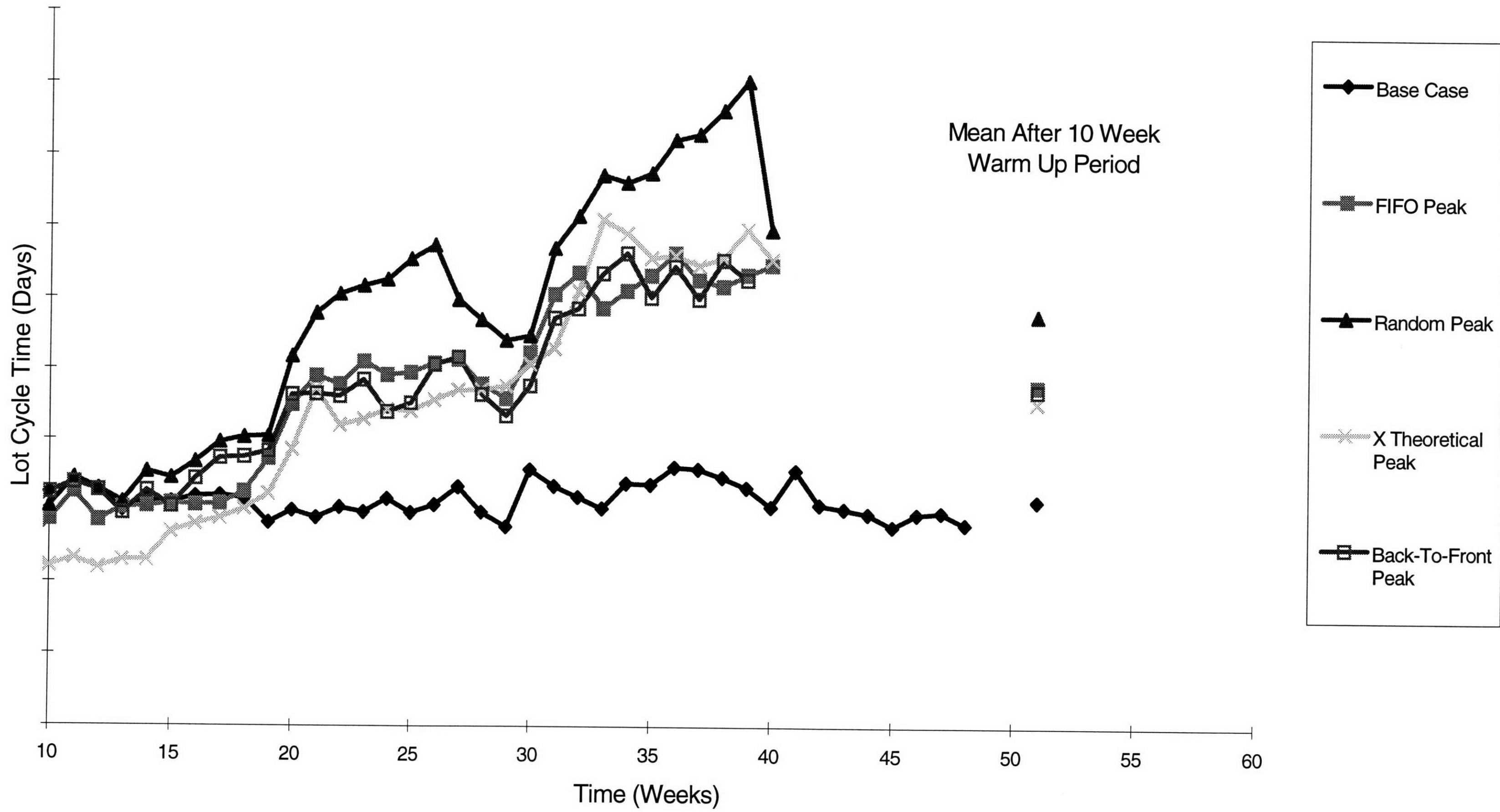


Exhibit 23: Average Lot Cycle Time Data for Simulated Decreased Down Times

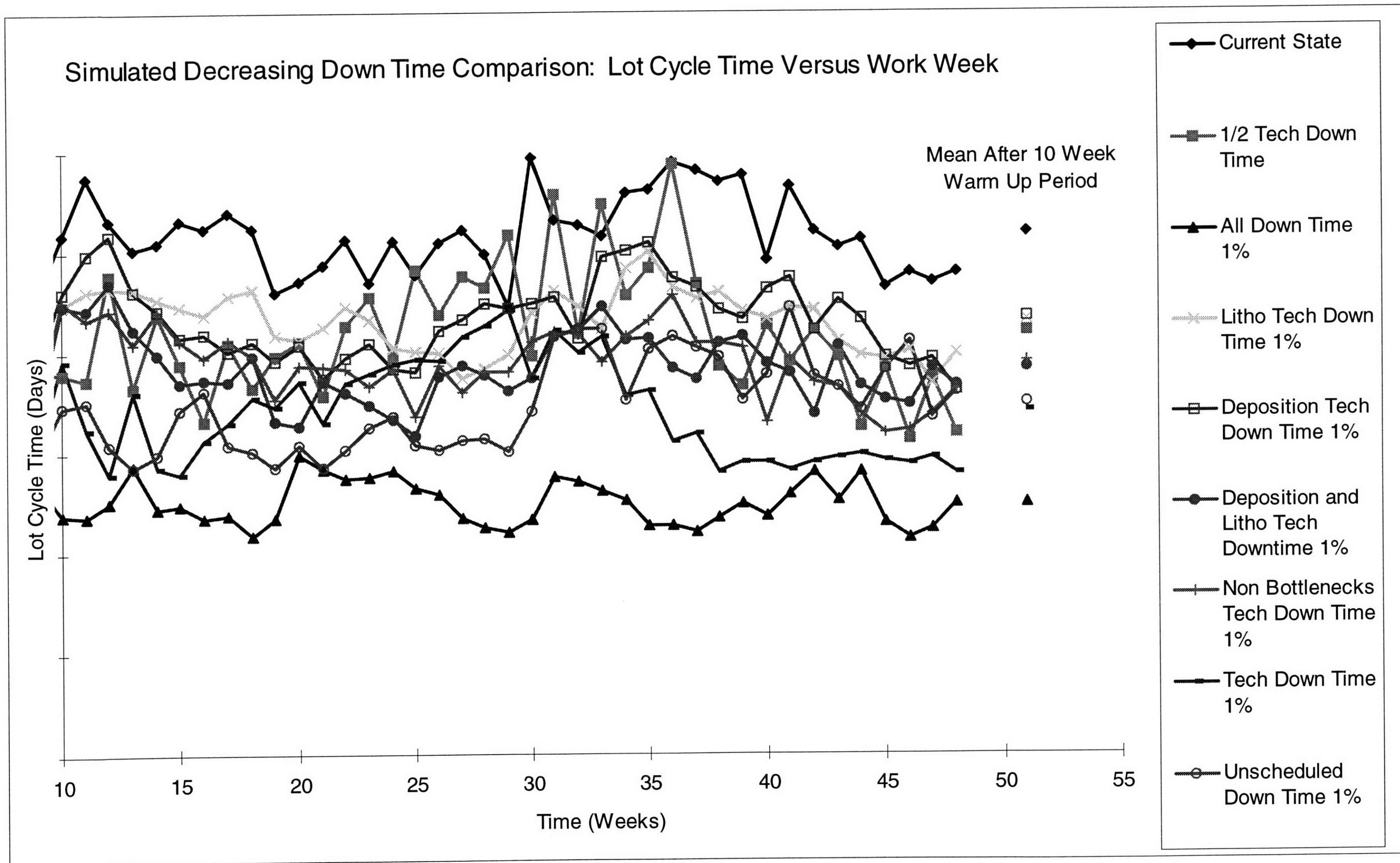


Exhibit 24: 95% Lot Cycle Time Data for Simulated Decreased Down Times

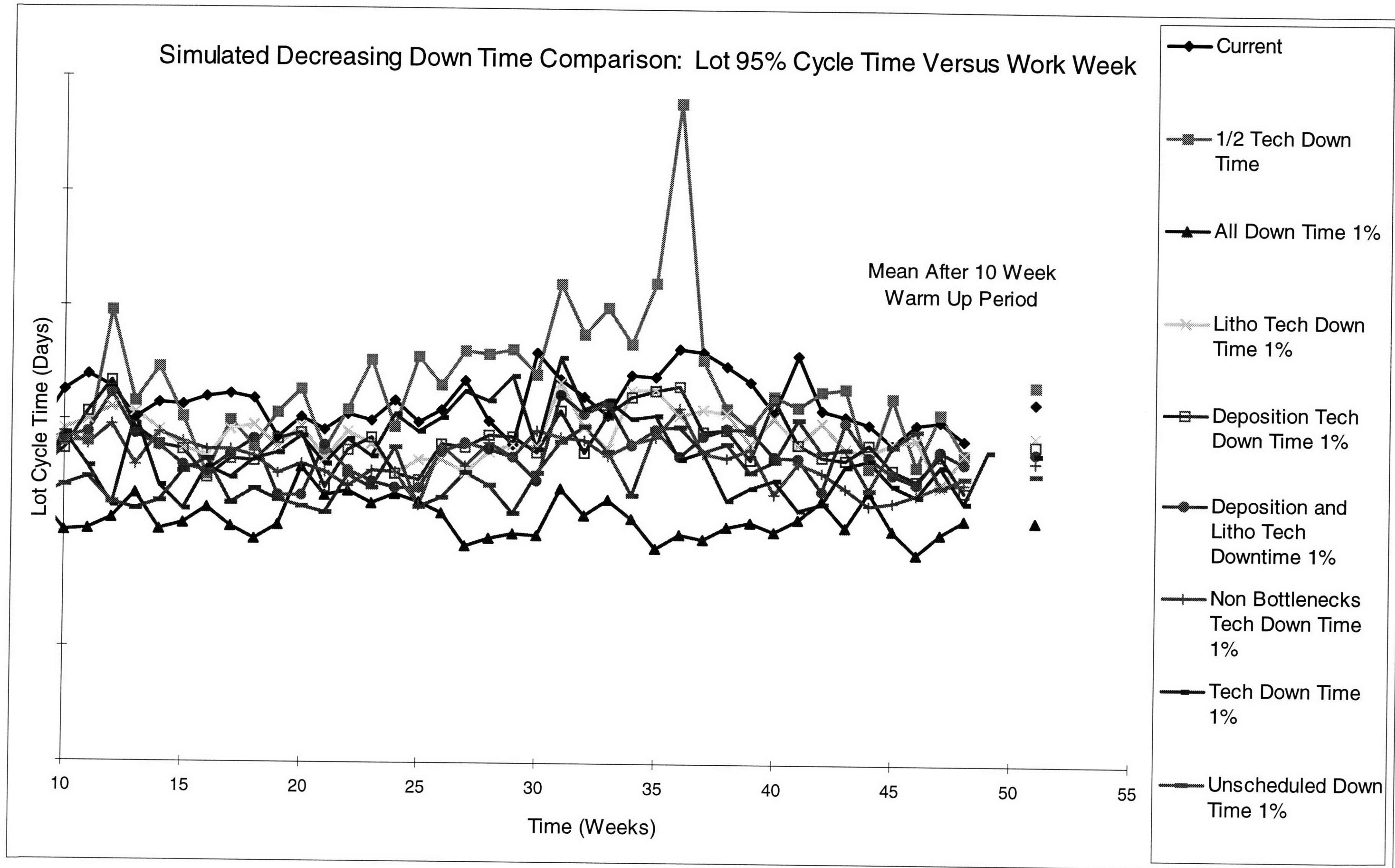


Exhibit 25: Average Cycle Time For Statistical Tests

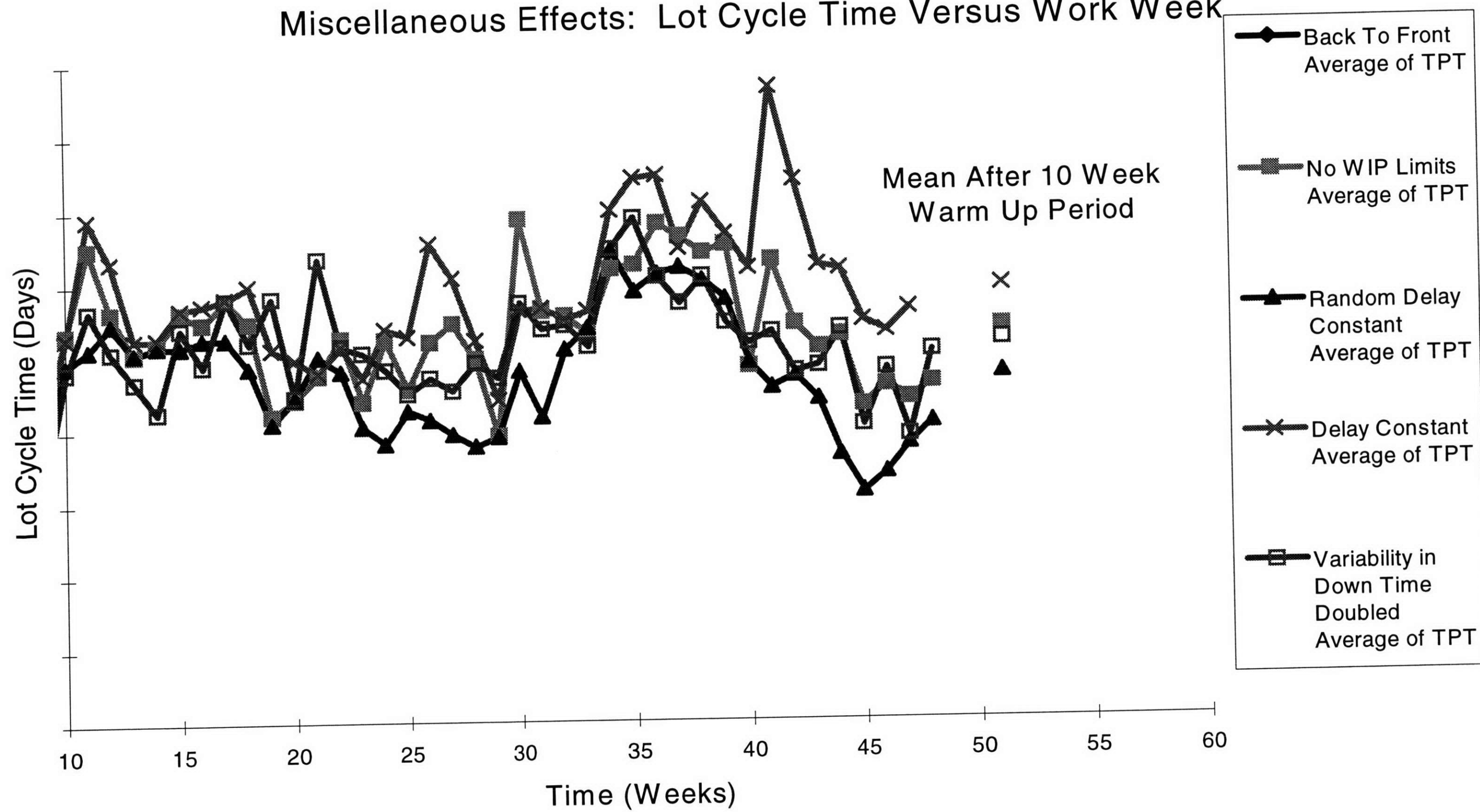
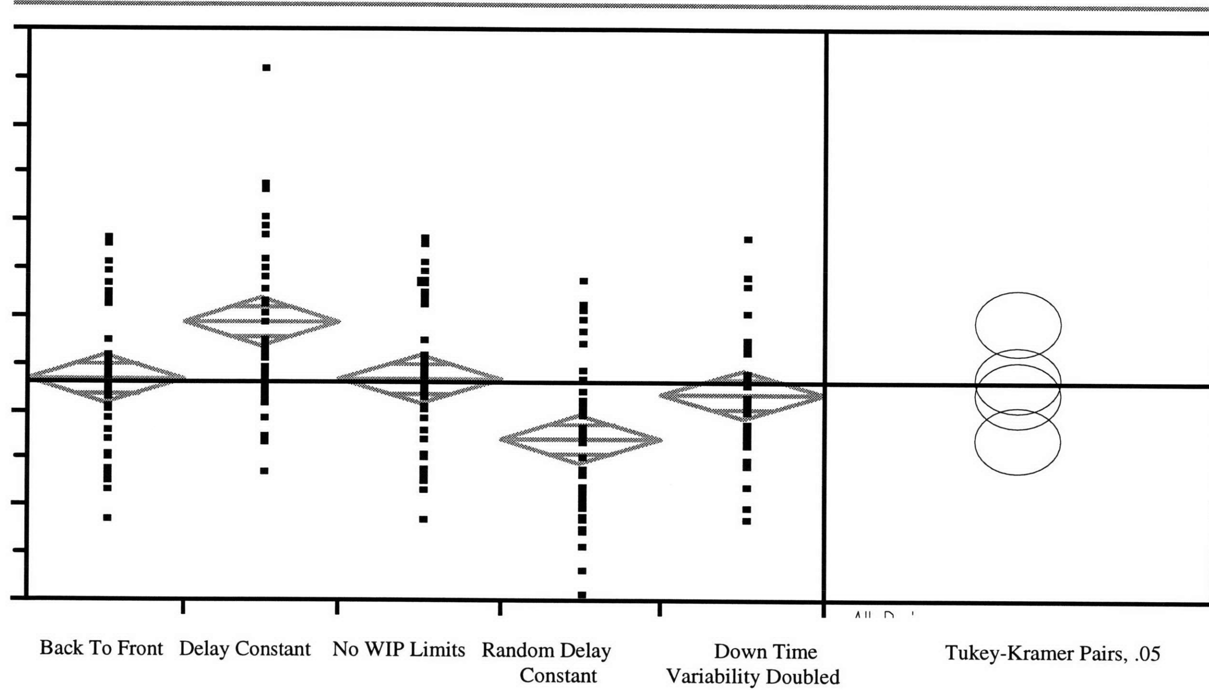


Exhibit 26: ANOVA and Tukey-Kramer Tests for Mean Cycle Time Comparison



Comparisons for all pairs using Tukey-Kramer HSD

Test:	Delay Constant	No WIP Limits	Variability in Down Time Double	Random Delay Constant
Back To Front	0.074042	-0.50221	-0.32958	0.129892

Positive values show pairs of means that are significantly different to the 95th percentile confidence.

Exhibit 27: 95% Cycle Time For Statistical Tests

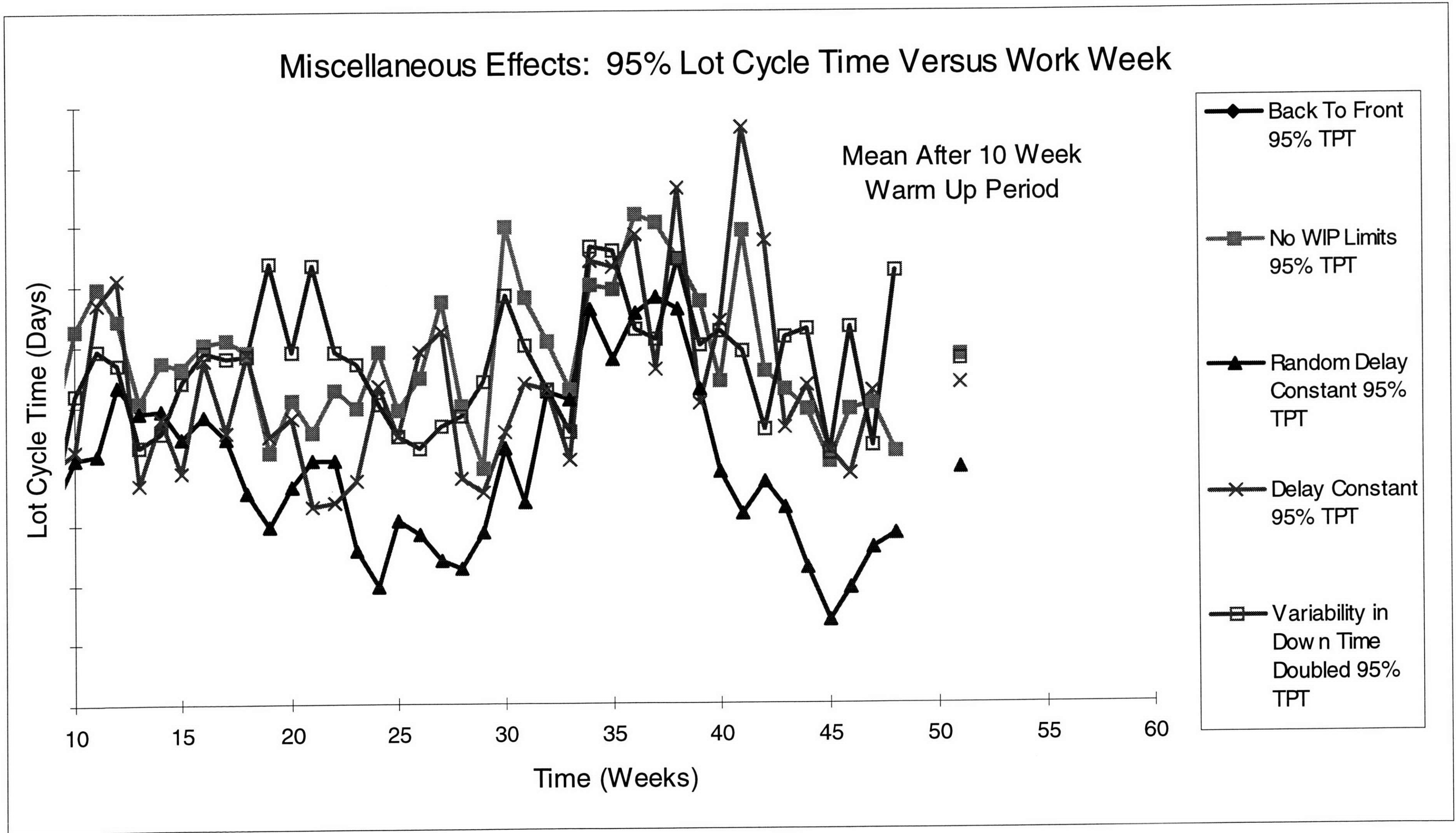
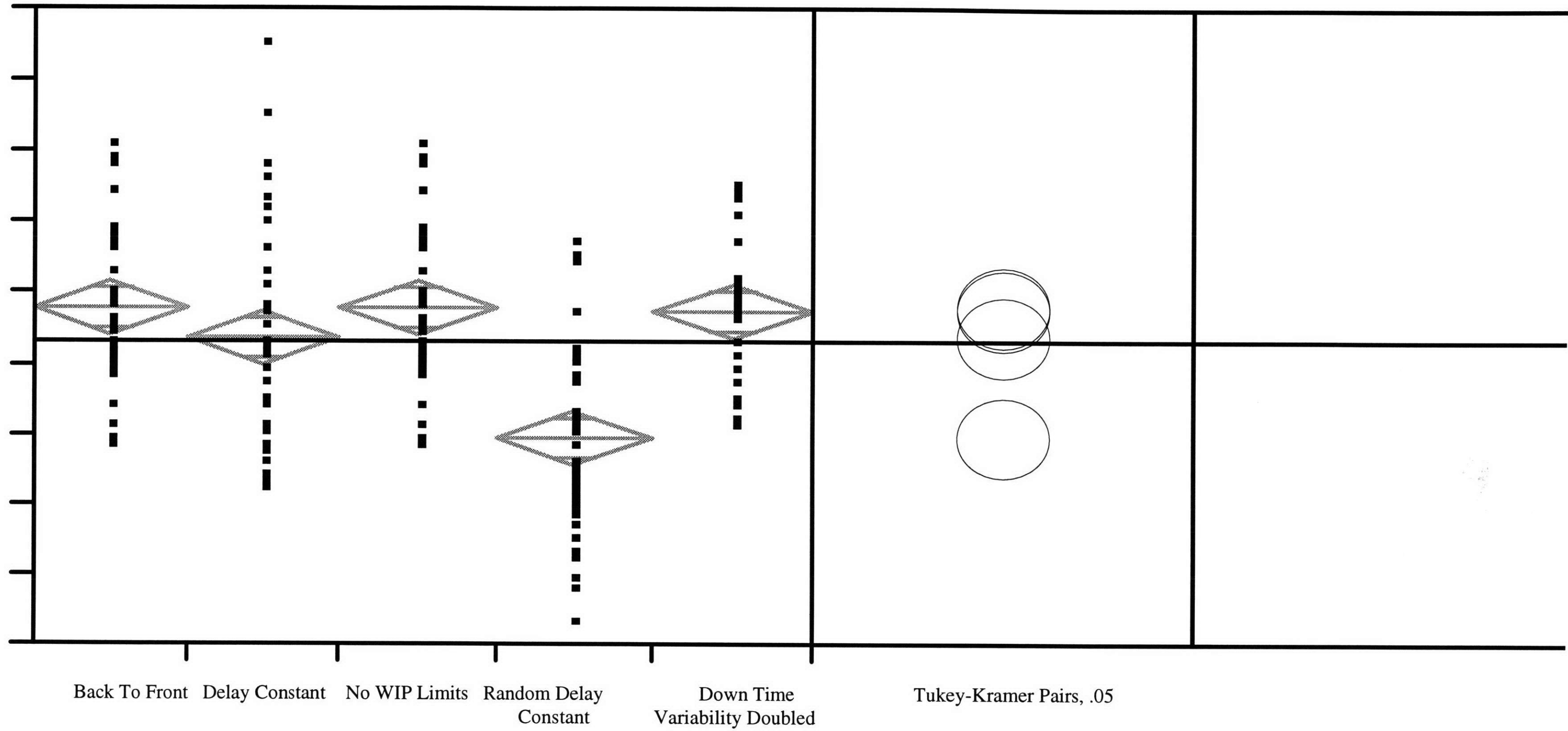


Exhibit 28: ANOVA and Tukey-Kramer Tests for 95% Cycle Time Comparison



Comparisons for all pairs using Tukey-Kramer HSD

Test:	No WIP Limits	Variability in Down Time Double	Delay Constant	Random Delay Constant
Back To Front	-0.80824	-0.75034	-0.37897	1.06018

Positive values show pairs of means that are significantly different to the 95th percentile confidence.

Exhibit 29: Average Lot Cycle Time Data for Recommended Actions

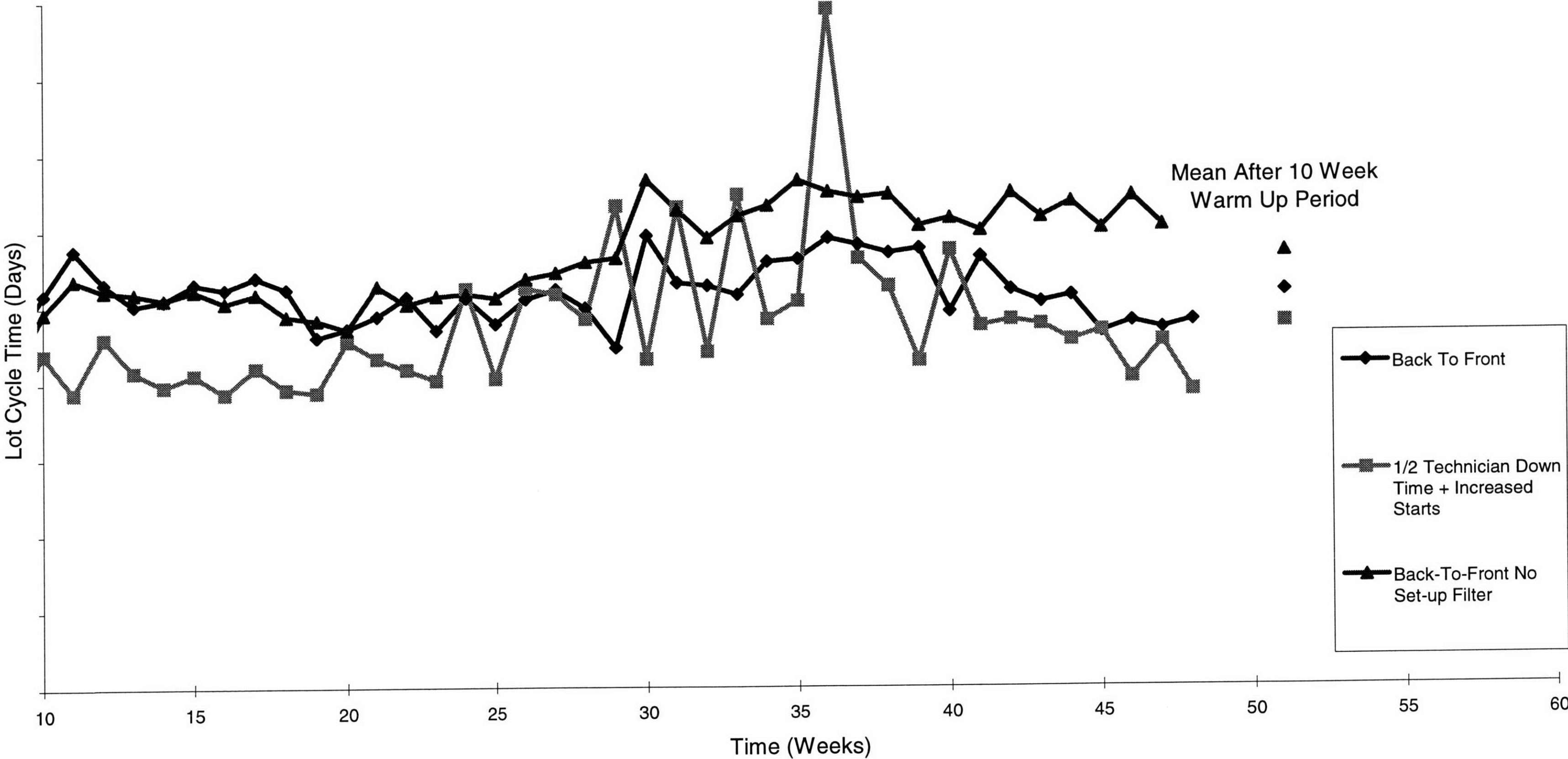
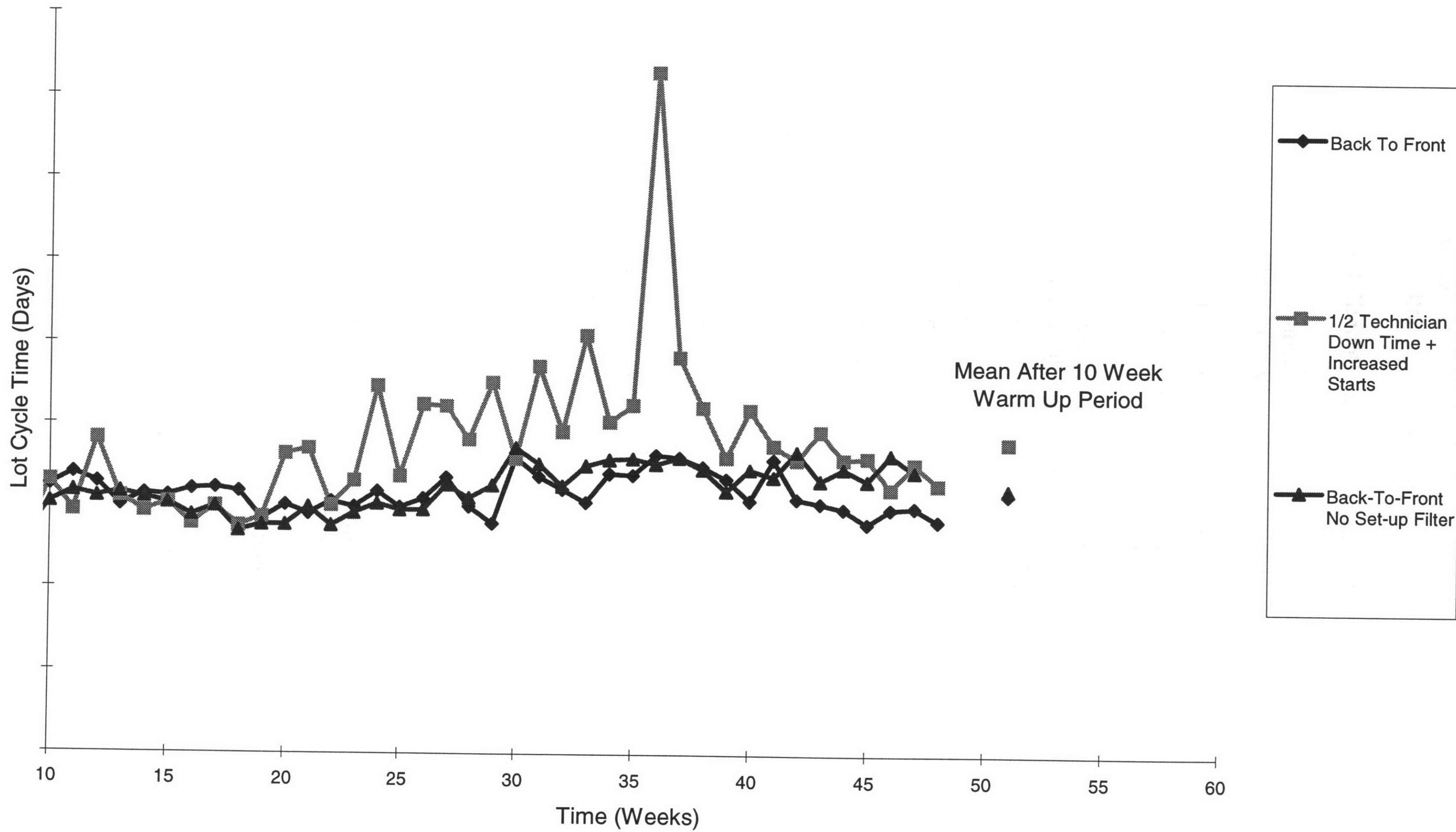


Exhibit 30: 95% Lot Cycle Time Data for Recommended Actions



References

Extensive information from personnel throughout Intel have provided the bulk of the information for this work. Over 40 people have contributed, especially those at Fab 8, Jerusalem.

AutoSched User Manuals I-IV. Autosimulation: Bountiful, Utah.

Boebel, F.G. and Ruelle, E. (1996), "Cycle Time Reduction Program at ACL," 1996 Advanced Semiconductor Manufacturing Conference and Workshop. IEEE: Piscataway, NJ, 1996.

Di Mascolo, Maria, Frein, Yannick and Dallery, Yves (1996), "An Analytical Method For Performance Evaluation of Kanban Controlled Production Systems," Operations Research, Volume 44, No. 1, January-February 1996.

Doering, Robert (1996), "PICO Frontier – In the drive toward smaller, better chips, the fear is not that we'll run out of gas but that we'll be unable to meet its price at the pump," Electronic Engineering Times. CMP Media, Inc.: December 30, 1996, p. 36.

Frantz, Gene, and Tatge, Reid (1997) "System-on-a-chip challenges designers," Electronic Engineering Times, CMP Publications, Inc.: February 10, 1997, p. 78.

Gershwin, Stanley (1994). Manufacturing Systems Engineering. PTR Prentice Hall: Englewood Cliffs, 1994.

Graves, S.C., Meal, H., Stefek, D., and Zeghmi, A (1983), "Scheduling of Re-entrant Flow Shops," Journal of Operations Management, Volume 3, 1983, pp. 197-203.

Hogg, Robert and Ledolter, James. Applied Statistics for Engineers and Physical Scientists. Macmillan Publishing Company: New York, 1992.

Isaacson, Portia (1997), "Will Emerging New Digital Technologies Ever Stop," Computer Reseller News, CMP Publications, Inc.: February 24, 1997, p. 69.

Leachman, Robert, (1996) Editor. The Competitive Semiconductor Manufacturing Survey: Third Report on Results of the Main Phase. University of California, Berkeley, 1996 Report CSM-31.

Nahmias, Steven (1993) Production and Operations Analysis. Boston: Irwin, 1993, pp. 359-393, 429-430.

Page, Paul (1994), "Texas Instruments Using Total Logistics In Worldwide Campaign To Cut Inventory," Traffic World, May 16, 1994, p. 9.

Peuto, Bernard L (1997), "Mainframe History Provides Lessons," Microprocessor Report, VOL. 11, No. 4, March 31, 1997, p. 19.

Ristelhueber, Robert (1996), "Wafer fabs: Getting more bang for the buck; increasing the productivity of semiconductor wafer fabs; Industry Trend or Event," Electronic Business Today, Vol. 22, No. 7, July, 1996, p. 37.

Robinson, Paul (1997), "Prepared Statement of C. Paul Robinson, Director Sandia National Laboratories Before the Senate Committee On Armed Services Subcommittee on Strategic Forces," Federal News Service. March 19, 1997.

Sall, John and Lehman, Ann, (1996) JMP Start Statistics: A Guide To Statistics and Data Analysis Using JMP and JMP In Software, Belmont: Wadsworth Publishing Company, 1996, p.140.

Thompson, Michael (1996) "Simulation-based scheduling: Meeting the semiconductor wafer fabrication challenge," IIE Solutions, Vol. 28, No. 5, May, 1996, p. 30-34.

Wolf, Stanley (1986) Silicon Processing for the VLSI Era Volume 1: Processing Technology. Sunset Beach: Lattice Press, 1986.

Leaders for Manufacturing Program
Massachusetts Institute of Technology
77 Massachusetts Avenue
Room E40-422
Cambridge, MA 02139

Attention: Donald Rosenfield, Director, Fellows Program

Subject: SM Thesis of Susan E. Perrin

The attached proposal, Production Management Policies in a Multiple Product Semiconductor Fabrication Facility, April 27, 1997, describes the research study that Susan E. Perrin intends to carry out as an MIT Leaders for Manufacturing student using, at least in part, the facilities of the Intel Corporation. Susan E. Perrin further intends to prepare a thesis on this research study to be submitted to MIT as one of the requirements for a Master of Science degree. We understand that, if such a report is to be acceptable to MIT, faculty regulations require the following:

1. The research study must be supervised by a member of the MIT staff who can have full access (during the course of study and in confidence) to the pertinent background, methods of investigation and results.

2. Research studies and thesis topics involving subject matter (including date, results, or methods) subject to restriction for reasons of either proprietary interest or national security are unacceptable as the basis for a thesis.

3. The actual thesis document becomes the permanent property of MIT, and will be placed in the MIT Library within one month of the date of submission.

It is our belief that the proposed research can be carried out and an acceptable thesis prepared under these conditions.

We understand that the thesis will not be accepted by MIT unless accompanied by a letter from an authorized official of Intel confirming that the thesis is within the approved scope, and does not contain details objectionable from the Intel standpoint. A copy of the thesis will be submitted by Susan E. Perrin sufficiently in advance of the date of submission to permit review.

We also understand that except as noted above, at MIT, all rights in the results of the research including any inventions made by Susan E. Perrin are subject to the host agreement between MIT and Intel.

Yonathan Wand,
Intel
Manufacturing Manager

Susan E. Perrin
Leaders for Manufacturing Fellow

Leaders for Manufacturing Program
Massachusetts Institute of Technology
Building E40-422
77 Massachusetts Avenue
Cambridge, MA 02139-4307

MAIN TEL 617-253-1055
FAX 617-253-1462

TELEFAX COVER SHEET

Date: ______________

TO: Yonathan Wand / Amir Dayan
Dept: ______________
Phone: ______________
FAX: ______________

FROM: Susan Perrin
MIT Room E40-422
Phone: 617-253-1055
FAX: 617-253-1462

RE: Attached are thesis release documents
I am sending email to you w/ thesis corrected
with Don't Stan's Input
Susan

Number of pages, including cover sheet: 5

Leaders for Manufacturing Program
Massachusetts Institute of Technology
77 Massachusetts Avenue
Room E40-422
Cambridge, MA 02139

Attention: Donald Rosenfield, Director, Fellows Program

Subject: SM Thesis of Susan E. Perrin

The attached proposal, Production Management Policies in a Multiple Product Semiconductor Fabrication Facility, April 27, 1997, describes the research study that Susan E. Perrin intends to carry out as an MIT Leaders for Manufacturing student using, at least in part, the facilities of the Intel Corporation. Susan E. Perrin further intends to prepare a thesis on this research study to be submitted to MIT as one of the requirements for a Master of Science degree. We understand that, if such a report is to be acceptable to MIT, faculty regulations require the following:

1. The research study must be supervised by a member of the MIT staff who can have full access (during the course of study and in confidence) to the pertinent background, methods of investigation and results.

2. Research studies and thesis topics involving subject matter (including date, results, or methods) subject to restriction for reasons of either proprietary interest or national security are unacceptable as the basis for a thesis.

3. The actual thesis document becomes the permanent property of MIT, and will be placed in the MIT Library within one month of the date of submission.

It is our belief that the proposed research can be carried out and an acceptable thesis prepared under these conditions.

We understand that the thesis will not be accepted by MIT unless accompanied by a letter from an authorized official of Intel confirming that the thesis is within the approved scope, and does not contain details objectionable from the Intel standpoint. A copy of the thesis will be submitted by Susan E. Perrin sufficiently in advance of the date of submission to permit review.

We also understand that except as noted above, at MIT, all rights in the results of the research including any inventions made by Susan E. Perrin are subject to the host agreement between MIT and Intel.

Yonathan Wand,
Intel
Manufacturing Manager

Susan E. Perrin
Leaders for Manufacturing Fellow

Leaders for Manufacturing Program
Massachusetts Institute of Technology
Room E40-419
One Amherst Street
Cambridge, MA 02139

Attention: Donald Rosenfield
Director, LFM Fellows Program

Subject: SM Thesis of Susan Perrin

I have received the attached thesis of Susan Perrin on behalf of Intel. The thesis is within the scope of the thesis proposal as previously approved and does not contain any material that is objectionable to Intel. It is also approved for its technical content.

It is understood that the actual thesis document will be the permanent property of MIT and will be placed in the MIT library within one month after the date of submission. Intel agrees that MIT shall have the nonexclusive right to reproduce, publish, and distribute the thesis.

(Authorized Official of Company)

Massachusetts Institute of Technology
Leaders for Manufacturing Program

Proposal for Thesis Research in Partial Fulfillment of the Requirements of the Degrees of Master of Science in Management and Master of Science in Mechanical Engineering

Title: Production Management Policies in a Multiple Product Semiconductor Fabrication Facility

Submitted by: Susan E. Perrin
80 Upland Avenue, Apt. 2
Boston, MA 02140

(Signature of Author)

Date of Submission: April 27, 1997

Expected Date of Completion: May 5, 1997

Company Site of Internship: Intel
Electronics Ltd.
P.O. Box 3173
Jerusalem 91031 ISRAEL

Brief Statement of the Problem:
The relationship between production management policies in a multi-product, multi process-semiconductor fabrication facility and global fab performance metrics is not well understood. The problem will be examined through the development of a dynamic simulation model in order to compare different WIP management methodologies. This will be used to compare and contrast cycle time performance for different work in process selection methodologies. This model will be utilized to pinpoint appropriate functional areas for performance improvement efforts.

Supervision Agreement:
The program outlined in this proposal is adequate for a Master's thesis. The supplies and facilities required are available, and I am working to supervise the research and evaluate the thesis report.

Don Rosenfield, Senior Lecturer

Stan Gershwin, Senior Researcher

Thesis Statement

This thesis will examine production management policies in a multi-product, multi process-semiconductor fabrication facility. A dynamic simulation model of key functional areas will compare and contrast cycle time performance metrics for different work in process selection methodologies. These policies include both global policies and local "functional area" policies, and both lot-specific and non-specific policies. The necessary infrastructure and incentive implications for implementation of top performing policies will also be examined. Along these line, the manufacturing organizational metrics and incentives will be examined.

This model will also be utilized to pinpoint appropriate functional areas for performance improvement efforts. Finally, the model will be used to predict performance improvement if recommendations are followed.

Project Impact

Project impact will be to better understand the effect of production management policies, especially WIP management methodology, on the fab cycle time metrics. Building the model will also help to improve understanding of fab performance.

Outline

INTRODUCTION

PROBLEM DEFINITION
WORK IN PROCESS MANAGEMENT METHODOLOGIES

BACKGROUND INFORMATION

SEMICONDUCTOR INDUSTRY OVERVIEW
SEMICONDUCTOR FABRICATION PROCESS
FAB HISTORY
FAB PERFORMANCE METRICS EMPHASIS AND COMMUNICATION
OTHER RELEVANT MANUFACTURING POLICIES
PRIORITY LOTS
CYCLE TIME AND VARIATION
FINANCIAL MODELS

PERFORMANCE DATA AND POLICY BENCHMARKING

HISTORICAL FACILITY PERFORMANCE DATA
BENCHMARKING WITHIN INTEL AND THE INDUSTRY
ALTERNATIVES TO MODELED METHODOLOGIES
PUBLISHED ACADEMIC INFORMATION
WIP Management Methodologies
Loading Change Affect on Mean Cycle Time and Variance Under Several WIP Management Policies

SIMULATION MODEL DEVELOPMENT

INITIAL MODEL INPUTS
OUTPUT VERIFICATION AND ITERATION PROCESS FOR INPUTS
SIMULATION RESULTS
WIP MANAGEMENT EFFECTS
DOWN TIME EFFECTS
STATISTICAL AND SENSITIVITY EFFECTS
POLICY RECOMMENDATION EFFECTS

RECOMMENDATIONS

SYSTEMS DESIGN AND IMPLEMENTATION
SEMICONDUCTOR FAB OPERATIONS
FORM A PERMANENT LINE MANAGEMENT CONTINUOUS IMPROVEMENT TEAM

UNRESOLVED ISSUES AND FURTHER DEVELOPMENTS

SIMULATION MODEL IMPROVEMENTS
FACILITY SPECIFIC CONCERNS