IMPROVING PRODUCTIVITY OF SEMICONDUCTOR MANUFACTURING CLUSTER TOOLS

by

Randall T. Nickel

B.S., Electrical Engineering, GMI Engineering & Management Institute (1988)

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

MASTER OF SCIENCE IN MANAGEMENT

and

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Signature of Author _____
MIT Sloan School of Management
Department of Electrical Engineering and Computer Science
May 6, 1994

Certified by _____
Charles H. Fine
Associate Professor of Management Science
Thesis Advisor

Certified by _____
Lionel C. Kimerling
Professor of Material Science
Thesis Advisor

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

IMPROVING PRODUCTIVITY OF SEMICONDUCTOR MANUFACTURING CLUSTER TOOLS

by

Randall T. Nickel

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

Thesis Advisors:

Charles H. Fine
Associate Professor of Management Science

Lionel C. Kimerling
Professor of Material Science

Abstract

The AMAT Endura 5500 sputtering machine is used as a case study to improve productivity for semiconductor manufacturing cluster tools. The repetitive and time consuming PVD chamber target and shield exchange procedure is streamlined from nine to 6.5 hours, increasing effective machine availability by 2%. Primary savings are attained by challenging standard process parameters for chamber baking and target conditioning. Modeling and simulation were employed to increase both equipment availability and wafer throughput rates. Static modeling, using linear and integer programming implemented through the Microsoft Excel Solver routine, suggested that up to 5% gains in uptime were possible by implementing specific redundant process chambers. Dynamic computer simulation, using AT&T's WITNESS package, showed that a mix of optimally configured machines could increase net wafer throughput by up to 8%. The total possible increase in productivity achieved by the combination of increased uptime and increased throughput exceeds 12%.

Acknowledgment

The author wishes to acknowledge the Leaders for Manufacturing Program for its support of this work.

Preface

This thesis concludes seven months of on-site research at the Portland Technology Development group of Intel Corporation with the support of MIT's Leaders for Manufacturing program. Many individuals have helped me tremendously over the course of this project, and although I can't name everyone, I would be remiss in not mentioning the following people:

David Marsing, Fab 9 Manager, for his foresight and groundbreaking work in starting Intel's involvement in MIT's LFM program; Gerald Marcyk, PTD Thin Films Area Manager, for his efforts in developing and supporting this project; Charles Fine and Lionel Kimerling, my MIT advisors, for their support and aid in expanding the boundaries of the project; Chi-Hwa Tsang, Metals Group Leader, for his flexibility and patience; Chi-Hing Choi and Brandon Robinson for their engineering insight, practicality, support, and friendship; Bruce Sohn, for his invaluable insight into constraint theory; Dan Goranson, for his aid in all areas relating to Intel/LFM administration; and finally, Judy Timberlake, for her tireless help in the PTD group.

Table of Contents

Table of Figures

Chapter 1: Introduction

Purpose

The purpose of this thesis is to improve productivity of cluster configured semiconductor manufacturing equipment. A cluster configured tool, the AMAT Endura 5500 sputtering machine, is used as a case study for this purpose.

Cluster Tools

A cluster tool links multiple process chambers to perform consecutive tasks on silicon wafers. Benefits that cluster tools derive include greater control over the wafer environment and wafer handling mechanisms, which leads to improvements in process integration and contamination control. Manufacturing processes may be developed in a coupled fashion, with the assurance of consistent wafer handling and a controlled wafer environment. These advantages also allow the resultant process technologies to be successfully transferred to production sites. Controlled ambient processes enjoy gains in contamination control, avoiding intermediate monitoring and cleaning steps by cascading manufacturing steps within the same machine. An example of a vacuum based process that exploits the advantages of cluster tools is sputtering, the most commonly used technique for physical vapor deposition (PVD).

Sputtering

Sputtering is a technique that applies a thin film of material onto a desired product surface. In the semiconductor industry, sputtering is commonly employed in metallization, applying thin films of various metals which provide a conducting path or serve as a barrier layer

between various active devices on the silicon wafer. Sputtering dislodges the desired material from a source, or target, onto the work surface, or substrate. A high energy radio frequency (RF) energy source in a reduced pressure ambient environment creates a plasma of ionized argon gas. A high DC voltage potential at the target causes the gas ions to accelerate toward the target surface physically dislodging target atoms upon surface impact. The reduced pressure ambient environment increases the distance that sputtered atoms can travel, thereby reducing subsequent collisions away from the target surface. The target atoms then travel within the vacuum and deposit on chamber surfaces, including the substrate. Buildup of unwanted target material upon chamber walls is minimized by a removable shroud, or shield, which flanks the substrate and covers the nearby chamber walls. Typically, the substrate pedestal is grounded and isolated from the shield. Because target material is physically removed, the target is a wasting resource that must be replenished. In contrast, the shield must be replaced because excessive surface deposition leads to particle flaking and hence process contamination. To perform metallization for the fabrication of the 486 and Pentium™ microprocessors, one of the tools that the Intel Corporation uses is the AMAT Endura 5500, a sputtering machine manufactured by Applied Materials, Inc.

The Endura 5500 Sputterer

Overview

The Endura 5500 is a cluster configured sputtering machine that supports up to eight process chambers connected to a large machined aluminum base, or mainframe. The 5500 uses a single wafer process, isolating each wafer through segregated fabrication steps. Process chambers can be evacuated to ultra high vacuum (UHV) levels, allowing for improved process control. Sophisticated machine automation allows for precise control of

sputtering energies, gas flows, and wafer routing. An overview of the 5500's sputtering process follows.

As shown in Figure 1.1, wafers are introduced to the machine through the load locks, and are then transported between process chambers via two wafer handling robots. The robot end effectors, or blades, are used to lift the wafer from a pneumatically controlled wafer pedestal, move to another process chamber, and lower the wafer onto that chamber's wafer pedestal. The wafer pedestal allows clearance for mechanical transfers while maintaining a repeatable height for wafer processing. Processing requirements determine the wafer path, or sequence, that the wafer follows. Each robot services the wafers within the robot's mainframe chamber. The buffer robot moves the wafer between process chambers that adjoin the buffer chamber, while the transfer robot is responsible for servicing process chambers in the transfer chamber group.

Process chambers

Process chambers which attach to the buffer chamber include load locks, degas/orient chambers, and etch chambers. The two load locks serve to introduce wafers from the ambient pressure of the fabrication plant to the vacuum pressure levels required for the sputtering process. The degas/orient chamber, located at position E or F, heats the wafer to drive off residual gases, while assuring consistent and symmetric orientation of the wafer within the process chambers by referencing a notch on the wafer circumference. The etch chamber enhances surface quality by using a RF generated plasma etch to remove native oxides that have grown upon the wafer surface. Etching is performed either in Chamber A or in auxiliary chambers located in positions C or D. The PVD chambers, located in Chambers 1, 2, 3, or 4, perform the actual sputtering process discussed earlier. Chamber B is dedicated to the wafer cool down process. A gate valve

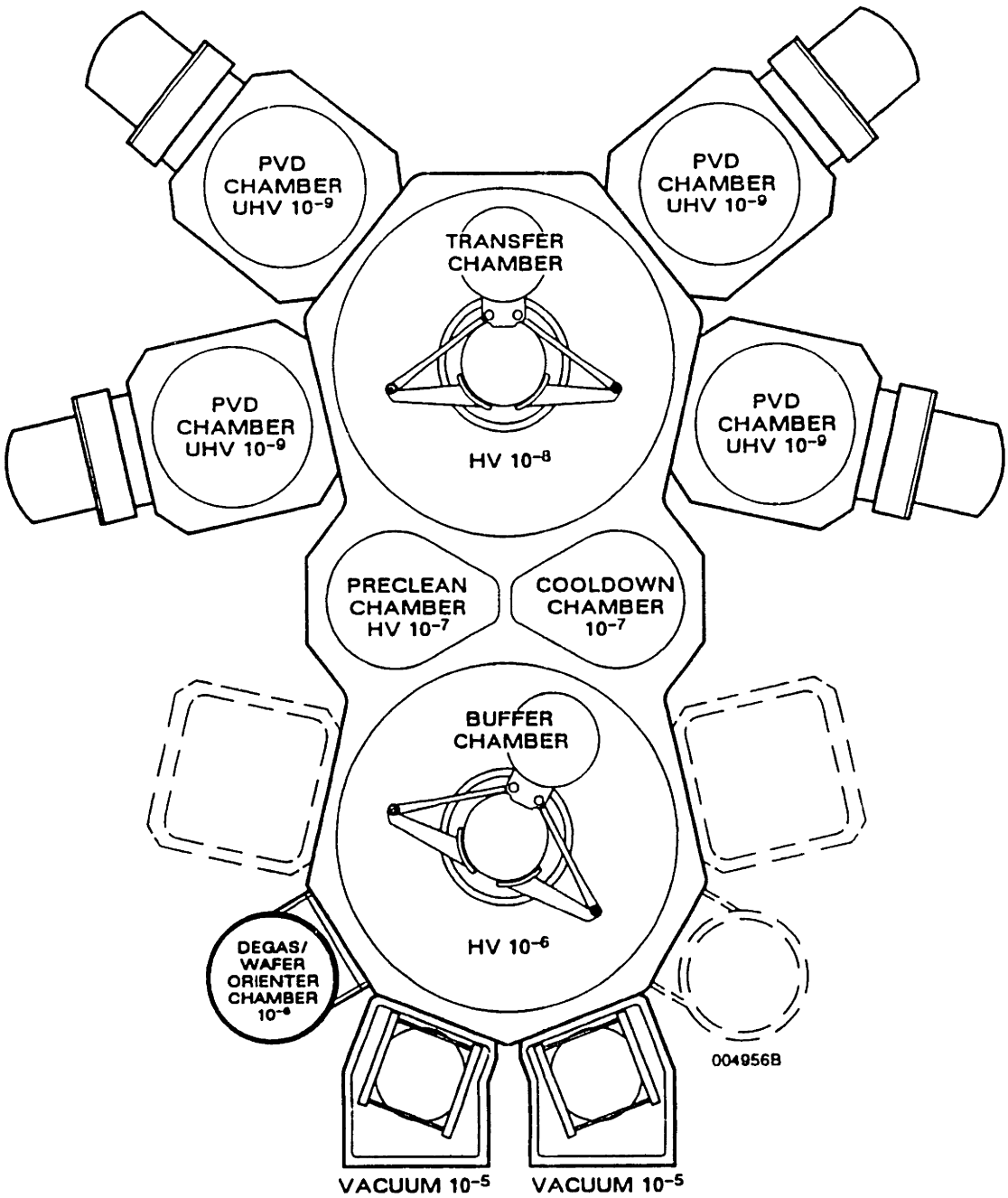

Figure 1.1 – Overview of the AMAT Endura 5500

isolates each of the process chambers from the mainframe. The appropriate power and gas feeds are connected to each process chamber, which are precisely controlled by the process chambers operational instructions, or recipe. Recipes dictate for each step the process time, power level, gas flow, and other critical parameters affecting process quality.

Robots

Robots are used to transport wafers between process chambers, based upon the desired process path. This wafer path, or sequence, is programmed by the user through the operator control terminal. Once the wafer has been delivered to the destination chamber, the recipe associated with that sequence is initiated. Hence, unique recipes may be invoked by multiple sequences, saving programming time for standard operations. It is also noteworthy that the sequence is capable of addressing single chambers or multiple chambers within a group. The latter option allows for parallel processing, as the robot will move a wafer to the first chamber available among those specified. The buffer and transfer robots share access to two chambers, A and B, which serve as mechanical portals between the two mainframe chambers.

Operator control terminal

The operator control terminal consists of a computer monitor and light pen. This triumph of man-machine interface allows for the display and adjustment of hundreds of process parameters, as well as providing extensive machine diagnostics. Operators use this interface for all operations, including changing machine mode, loading and unloading wafers, programming sequences and recipes , monitoring wafer histories, and troubleshooting. Although a review of all these operations is unnecessary, it is instructive to review the typical production sequence of operation.

Sequence of operation

The sequence of operation for a simple application is shown in Figure 1.2. The wafers are introduced into either load lock in lots, or batches, of up to 25 wafers in a single wafer holder, or boat. The load lock pumps down to pressures of 10^{-5} torr, and determines the number and positions of each wafer within the boat. The buffer robot transfers the first wafer from the boat to the wafer pedestal in the degas/orient chamber, then exits. The gate valve closes, and the degas/orient chamber is evacuated further to pressure levels of 10^{-6} torr. As the wafer is centered to a consistent position using the wafer notch, a halogen lamp heats the wafer, driving off residual gases remaining from previous process steps. Once completed, the buffer robot transfers the wafer to Chamber A, where a plasma etch removes native oxide. (Meanwhile, the buffer robot returns to move the second wafer from the load lock to the degas/orient chamber.) After the wafer is etched, the transfer robot transports the wafer from Chamber A to Chamber 1. A thin film is sputtered onto the wafer surface within this PVD chamber, and then the transfer robot moves the wafer to Chamber B for cooling. Finally, the buffer robot again takes the wafer and returns it to the load locks.

Goals

The goal of this thesis is to investigate strategies that will increase the productivity of the AMAT Endura 5500. Increasing machine uptime and increasing throughput rates both serve to increase machine capacity. Machine uptime may be increased by decreasing the length and frequency of the target and shield change process, while throughput may be increased by shortening production cycle times.

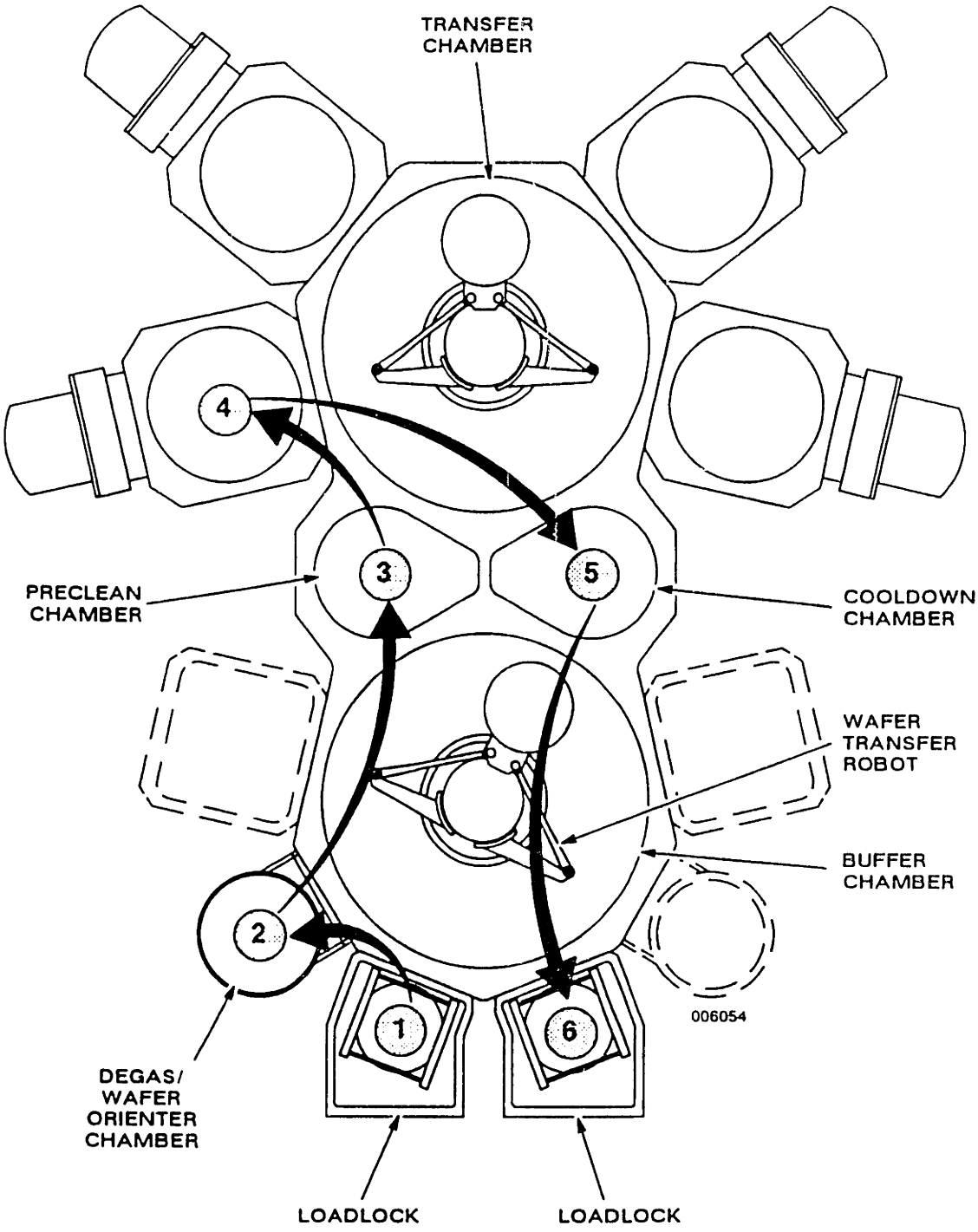

Figure 1.2 – Example of Endura 5500 Wafer Processing

Chapter 2: Conclusions and Recommendations

The productivity of the AMAT Endura 5500 sputtering machine may be improved by applying both engineering and management techniques to increase machine uptime and decrease machine cycle time. Conclusions, listed by category, and specific recommendations follow.

Conclusions

General

- Increasing machine capacity by improving throughput and availability can reduce the number of machines required. The net costs associated with improving productivity must be weighed against measurable gains in productivity. For Intel's next generation process, selecting an equal mix of AMAT Endura 5500 machine configurations optimized for both conductor and adhesion layers increase weighted wafer throughput time by 8%. (Chapter 5, p. 74) The optimized metal configuration requires 5% less maintenance hours than the standard configuration. (Chapter 4, p. 55) Reducing planned target/shield maintenance time can increase tool availability by 2%. (Chapter 3, p. 38) The net productivity improvement can increase machine capacity by up to 12.5%, thereby reducing the number of machines required in a large facility. This 12.5% gain allows a strategy which utilizes 8 optimized machines to replace 9 suboptimal machines. These savings must in turn be weighed against any increased costs of outfitting these "optimized" machines. The most productive AMAT Endura 5500 configurations cost approximately 10-15% more than their standard counterparts. Thus, either combination is nearly equally cost effective in this case.

Nevertheless, this exercise should be performed to determine the best strategy for any given cluster tool.

Target/shield change

- The time required for the exchange of the target and shield in the Endura 5500's PVD chambers may be reduced by 25%, corresponding to an improvement in machine uptime of 2%. (Chapter 3, p. 38) This improvement is generated by updating generic default process parameters to optimized values.

Cluster tool configuration

- Cluster tool chamber configuration impacts machine uptime. Using the Endura to fulfill Intel's requirements, redundant conductor chambers have the most positive effect. The metal configuration requires 5% less maintenance hours than the standard configuration. (Chapter 4, p. 55)

- Cluster tool chamber configuration impacts machine cycle time. An optimized configuration with redundant conductor PVD chambers can reduce metal layer cycle times significantly, compared to a base configuration. (Chapter 5, pp. 64-65)

- Dedicating optimized machines to specific layers improves weighted throughput by 8% compared to the previously planned mix of configurations. The optimized combination of machines contains equal numbers of machines optimized for conductor and adhesion layers. (Chapter 5, p. 74)

- The proliferation of multiple machine configurations dedicated to specific process layers can conflict with Intel's "Copy Exactly" process technology transfer strategy, and may be unfeasible in plants requiring few machines. (Chapter 5, pp. 65-66)

- Processing requirements may restrict the flexibility of cluster tool configuration by requiring redundant chambers for multiple, non-consecutive steps of a similar operation. This constraint may be alleviated by the implementation of intra-machine buffers. (Chapter 5, p. 71)

Production and maintenance scheduling

- A cluster tool with no internal buffers may increase throughput by interleaving wafers requiring different processing sequences, instead of cascading consecutive, identical wafers. However, the Endura 5500 does not yet possess the ability to mix wafers from different batches. (Chapter 5, pp. 67-68)

- If full scale target and shield machine maintenance is used instead of piecemeal chamber-by-chamber techniques, suboptimal schedules may eliminate anticipated uptime gains. (Chapter 4, p. 55)

Tools and techniques

- Default machine parameters established by the original equipment manufacturer may be conservative measures appropriate to a wide variety of customer requirements. Productivity gains may be made through experiments which optimize such variables. (Chapter 3, pp. 33-34)

- Linear and integer programming is a useful tool for finding optimized solutions to problems with interactive constraints. Access to this tool is almost universal due to implementation on standard computer spreadsheet packages. (Chapter 4, pp. 42-43)

- Computer simulation is a useful tool for modeling the impact of cluster tool configuration changes upon machine throughput, avoiding costly and time consuming machine outfitting and experiments. Use of computer simulation may be limited due to initial purchase costs. (Chapter 5, pp. 57-58)

Recommendations

- More film thickness uniformity data must be collected using existing recipes to establish accurate benchmarks for the amount of target conditioning required. Then, new control limits must be established for updated conditioning recipes. (Chapter 3, p. 37)

- Both operational uptime and throughput for Intel's next generation processing would be improved by the use of redundant conductor chambers on the AMAT Endura 5500 sputterer. (Chapter 4, p. 49, and Chapter 5, p. 67)

- Two enhancements could improve the productivity of the Endura: parallel wafer processing (Chapter 5, pp. 67-68) and the implementation of Chamber 5 as a buffer available for wafer sequencing. (Chapter 5, p. 71).

Chapter 3: Improving the Target/Shield Change Procedure

Introduction

The target or shield change is performed because the target material is depleted when deposited onto the surface of the wafer. Eventually all target material would be consumed, but uniformity of the wafer film thickness degrades well before that point. Thus, the target lifetime is quantified by determining the amount of material that may be removed while maintaining film quality. The shields protect chamber walls and wafer pedestals from accumulating unwanted deposition that could lead to flaking and hence the introduction of particles onto the wafer surface. Because different wafer layers require different amounts of target material, the number of wafers processed through a chamber is not always an accurate indicator of remaining target life. Instead, cumulative kilowatt-hours (KWH) of work required during each deposition process is tracked for each chamber within the Endura control system. Thus, the lifetimes of targets and shields are determined by chamber KWH usage. (It is noteworthy that film uniformity or particle variation beyond established control limits will also trigger a target/shield change). This information appears on a chamber maintenance screen available to operations personnel. This screen must be monitored occasionally to determine when a chamber is scheduled for maintenance. The number of KWH that trigger chamber maintenance are tabulated in Figure 3.1. The average number of wafers, assuming that each layer is processed on the machine, is also shown. For Intel's next generation fabrication process, up to 475 complete wafers will be processed weekly through each Endura 5500 sputterer. The target/shield change procedure is the most time consuming maintenance procedure required for the 5500. As such, it is the largest contributor to scheduled production downtime. Figure 3.1 shows that average weekly service time is 12.5 hours, or 7.5% of weekly machine availability. By analyzing and streamlining the procedure, equipment

uptime may be improved. The total process time for a normal PVD process chamber target and shield change is nine hours. This time will serve as a benchmark for improvement.

Target Material	Target Lifetime (KWH)	Shield Lifetime (KWH)	Average # of Wafers before Service	# of Services per week (based on 475 wafers)	Average Service per Week (hrs.)
conductor	600	200	441	1.08	9.69
anti-reflective coating	750	750	2253	.21	1.90
barrier	250	250	4587	.10	.93
Totals				1.39	12.5

Figure 3.1 – Lifetimes for AMAT Endura 5500 Targets and Shields

Review of the Existing Target/Shield Change Procedure

The target/shield exchange process is composed of dozens of steps, each step requiring between 30 seconds and four hours. This sequence is displayed in Figure 3.3 and detailed in Appendix A. Segregating the process into an outline will help to simplify the explanation. An outline of the steps required is shown in Figure 3.2.

1. Pre-work
2. Exchange shield
3. AFS shield treatment
4. Exchange target
5. Pump/purge cycle
6. Bake out
7. Post bake
8. Target burn-in
9. Measurements

Figure 3.2 – Outline of Target/Shield Change Procedure

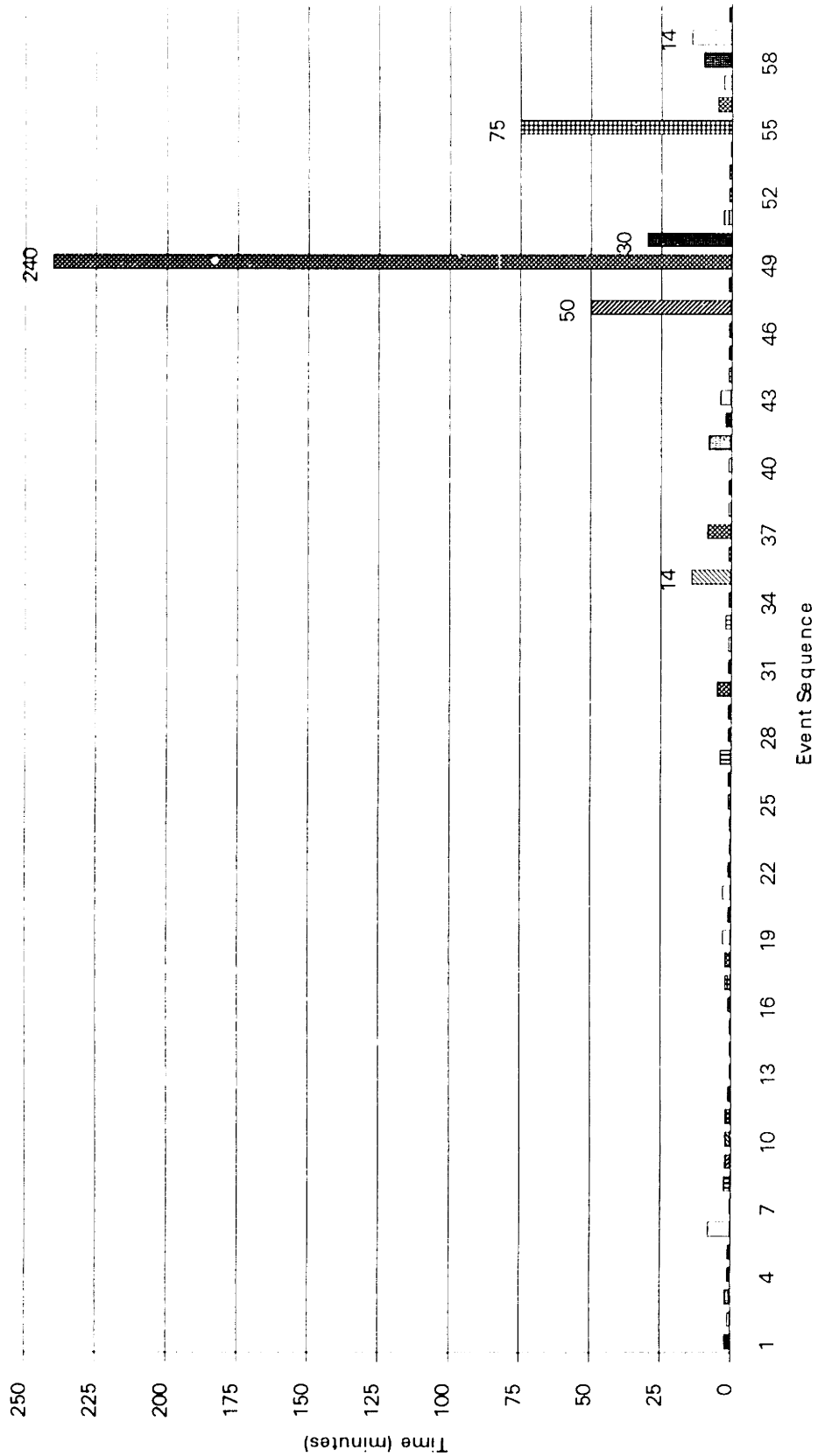

Figure 3.3 – Time Sequence of Target/Shield Change Procedure

Pre-work

After the procedure has been triggered by chamber usage or quality data, the computer aided manufacturing (CAM) control system is used to log the machine into maintenance status. Once parts such as aluminum foil, high temperature gloves, and chamber gaskets are gathered, the breaker switch for the RF/DC power supplies for that chamber is turned off, and the Endura control system system is switched to remote status. The operator now has control of the remote light pen activated control screen located in the service chase. The chamber mode is switched to off line, and chamber venting is initiated. The actual target lifetime and chamber base pressure are logged onto a cleanroom paper copy of a preventative maintenance (PM) sheet designated for the target and shield change procedure. Critical parameters are noted on this sheet throughout the procedure. The PM sheet is stored by the equipment process engineer, and is a source for data to analyze for trends if equipment problems arise. Each of the pre-work procedures takes less than three minutes.

Exchange shield

The process chamber, normally maintained at an ultra high vaccuum pressure of below 10^{-8} torr, vents to atmospheric pressure in eight minutes. (If the target is not required to be exchanged, it is covered with aluminum foil to prevent damage and avoid contamination of the target surface.) The shield, comprised of several pieces fastened together in one assembly, is removed from the chamber. A solution of deionized water and isopropyl alcohol (IPA) is used with clean room approved towels to thoroughly wipe all surfaces of the chamber. The new shield is then retrieved from an oven which maintains the preassembled shield kit at 100°C in a nitrogen gas ambient pressure environment. This serves to drive off liquids that would tend to condense on the shield's

surface and inhibit the chamber's ability to evacuate to ultra high vaccuum conditions. The shield kit is removed from the oven using insulated gloves, and moved into the service bay on a service cart. The new assembly is then inserted into the chamber. The exposed surface of the shield is cleaned with the IPA solution, as are the chamber o-ring gasket and ceramic insulating ring. The latter two are installed within the circumference of the chamber opening. The Endura control screen is then used to move the wafer pedestal into the deposition process position. This allows a visual inspection which assures that the shield is aligned symmetrically within the chamber, and that the shield doesn't provide a ground path to the wafer pedesal.

Start AFS shield treatment

This process is performed on all chambers except those containing conductor targets. AFS, or anti flake shield treatment, is performed to removed native oxide (surface material oxidized in the ambient environment) from the shield. The AFS procedure takes approximately 30 minutes to perform. If required, the target exchange procedure can begin at this point.

Exchange target

The target is changed during each shield change with the exception of the conductor chamber, which normally needs three shield changes for each new target.

Cooling water is drained from the old target. The old target is then unbolted from the chamber lid and the new target is installed. The o-ring surrounding the target is wiped down, lubricated and reinserted, and the target surface is cleaned with IPA. Cooling water is then introduced into the new target. The target surface is covered with aluminum

foil to await the completion of any AFS treatment in process.

Finish AFS shield treatment

When the AFS routine completes the chamber is vented. The chamber is pumped down again, and a leak checking operation is performed, introducing helium near the seams of the chamber lid through a small diameter hose and monitoring helium concentration via the residual gas analyzer (RGA).

Pump/purge process

A pump down/venting cycle called the "pump/purge" is used to drive off residual contaminates to help evacuate the chamber to high vaccuum levels. In this case, the chamber is pumped to a 10^{-6} range vaccuum, and subsequently vented to near atmospheric pressure. This sequence takes about 50 minutes.

Chamber bake out

Following the pump/purge cycle, a halogen lamp bakes the chamber at over 200°C temperatures to drive off liquids that have condensed onto the chamber surface and that were not removed during the pump/purge cycle. Again, the bake out procedure is an automated routine initiated throught the remote terminal. It is noteworthy that at this point the control system may be returned to automatic, and that the system can exit repair status through the CAM system. Any layers on the wafer not requiring the use of the serviced chamber may now be processed through the machine in "handicapped" mode.

The bake out procedure lasts four hours. The cumulative target life is now reset to zero

KWH through the Endura terminal.

Post bake

A 30 minute cool down period is incurred to allow the chamber to return to normal process temperatures.

Target burn-in

A full 25 wafer lot of non product, or "dummy" wafers, are loaded into the machine. The purpose of these wafers is to "burn-in" or condition the target surface until repeatable and uniform films conforming to quality specifications are deposited. This process lasts 75 minutes.

Measurements

One wafer is tested for the addition of particles from the serviced chamber, while every fifth of the 25 wafers are measured for intra-wafer film thickness uniformity. Both particle and uniformity measurements must be within established control limits to consider the chamber "qualified" to run production material. If these test fail, the target burn-in may be repeated. If this fails, a repeat of the target exchange may ensue.

Improving the Target/Shield Change Procedure

Reviewing Figure 3.3 and Appendix A, this maintenance procedure has 60 different steps. In order to gain the most returns on engineering effort, a simple pareto chart is compiled and displayed in Figure 3.4.

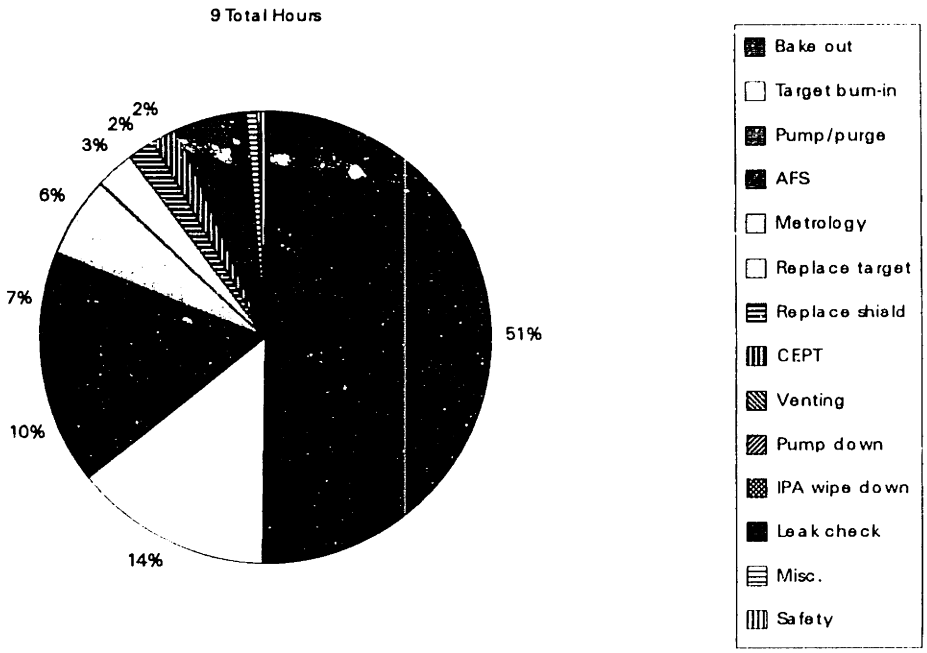

Figure 3.4 – Original Target/Shield Change Pareto

Bake out and target burn-in clearly dominate the list, accounting for 65% of the total target/shield change procedure. It is noteworthy that the target burn-in process occurs within the time region where the machine is considered down for all processes. As mentioned earlier, some processes may continue while bake out occurs in one of the process chambers. Use of the machine under these conditions is known as "handicapped" mode operation. Figure 3.5 shows the effects of chamber maintenance upon machine uptime in handicapped mode.

When this type of chamber is unavailable due to maintenance:	The number of layers that the machine can still process are:
conductor	5 of 9
anti-reflective coating	1 of 9
barrier	none

Figure 3.5 – Handicapped Mode Production

Another subtle reason why effort should be concentrated upon these items is that although these unique procedures are performed only one time within the maintenance sequence, they still dominate the pareto chart.

Reducing bake out time

The purpose of chamber bake out is to evaporate and remove any molecules that may have condensed inside the chamber during the exchange of the target or shield, allowing the chamber to return to UHV pressure levels. It is instructive to note what substances may condense upon the chamber walls – a list is compiled in Figure 3.6.

Liquid	Boiling Temperature (°C)
water	100.0
isopropyl alcohol	84.2
oxygen	-183.0
argon	-185.7
nitrogen	-195.8

Figure 3.6 – Normal Boiling Points of Common Gases

Due to their higher boiling temperatures, water and isopropyl alcohol seem reasonable as liquids that may condense onto the chamber surface, and both are introduced directly during IPA wipe down. To measure their concentration within the chamber, their partial pressures may be monitored by using the RGA during bake out. Figure 3.7 shows the partial pressures of the main ion constituents of water (OH^-) and isopropyl alcohol ($CH3^+$).

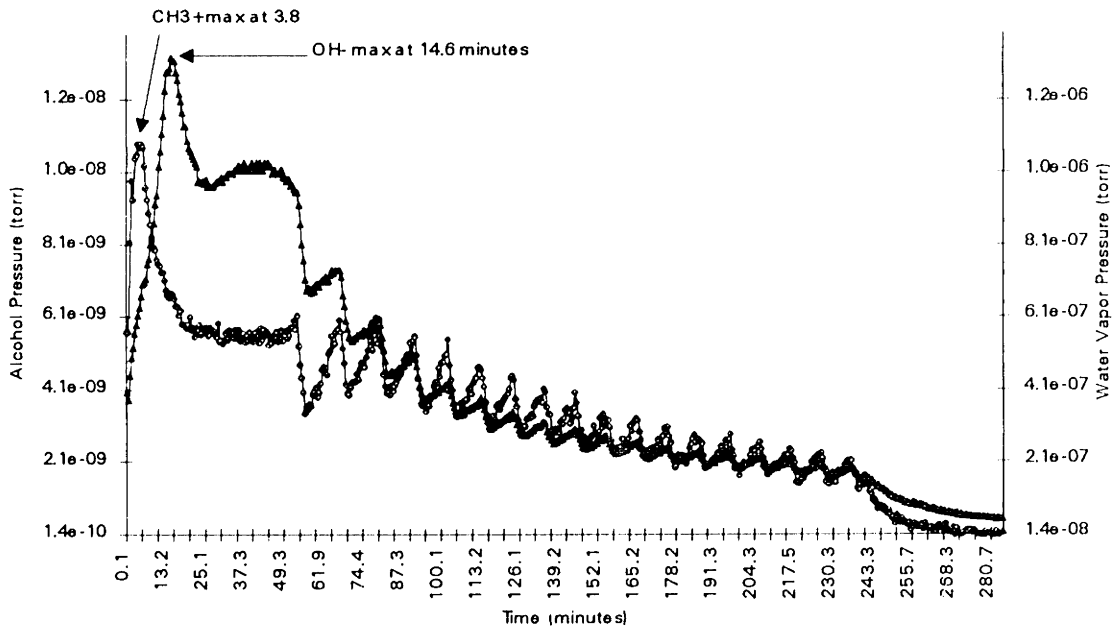

Figure 3.7 – Partial Pressures of Water and Alcohol during Bake Out

Note that alcohol concentration is two full orders of magnitude less than that of water, and due to its boiling point, is driven off much more quickly. Water vapor, then, may be considered the primary driver. Figures 3.8 shows a typical full profile of water vapor pressure during chamber bake outs.

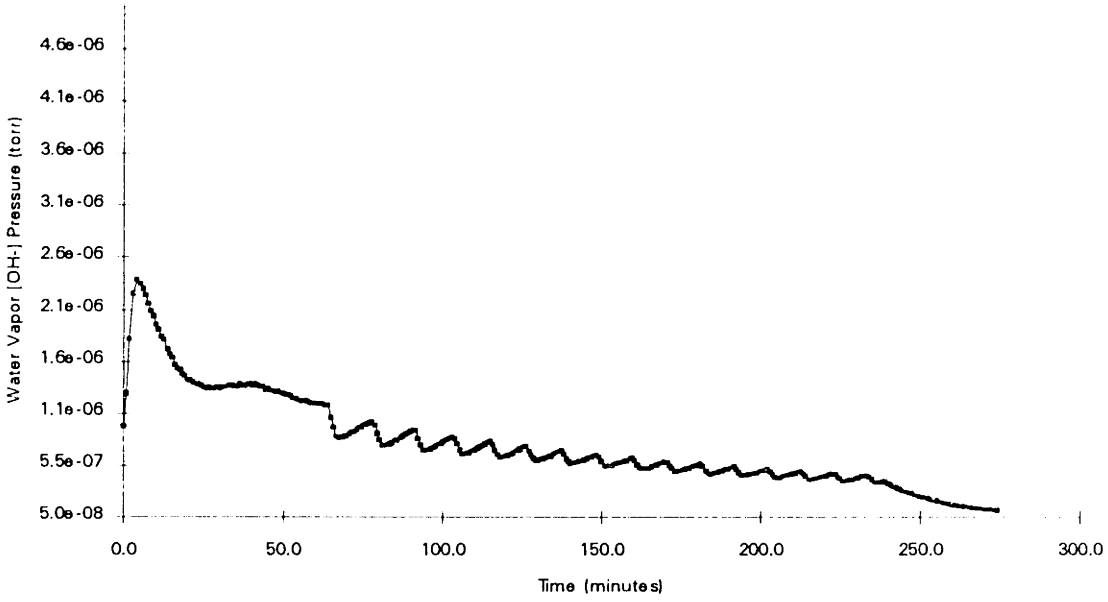

Figure 3.8 – Partial Pressure of Water during Chamber Bake Out

This profile shows that the majority of the benefit of bake out occurs within the first two hours. After this trend was repeated several times, the bake out time was reduced from the excessive default of four hours to two hours. The resultant profile is shown in Figure 3.9. Chamber base pressures continue to return to ultra high vaccuum ranges using the two hour bake out time.

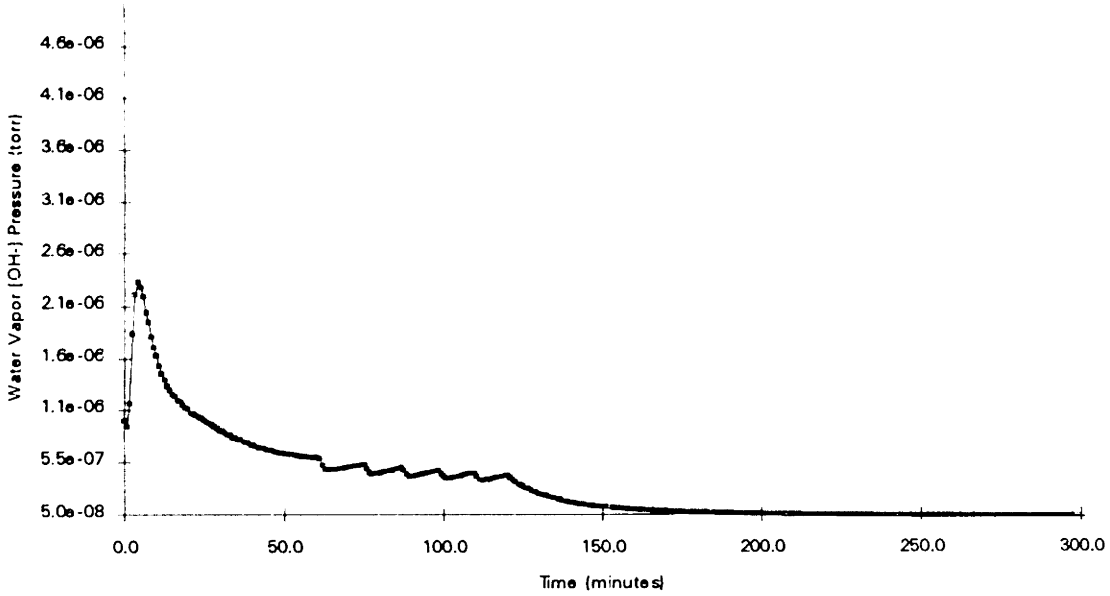

Figure 3.9 – Partial Pressure of Water during 2 Hour Chamber Bake Out

Reducing target burn-in time

As discussed, the target conditioning sequence takes approximately 75 minutes, and during this time no production wafers can run. The time required for the 25 wafer target burn-in lots can be segregated into two categories: process time and transport time. Process time is the amount of time the wafer spends in deposition chambers, while transport time is the time spent traveling from chamber to chamber via the wafer handling robots. Thus two strategies exist for reducing the total time required: 1) reduce process time by reducing the total amount of material removed from the target, and 2) reduce the transport time by reducing the number of conditioning wafers used during burn-in. To that end, film uniformity trends for different target materials are shown in Figures 3.10 and 3.11.

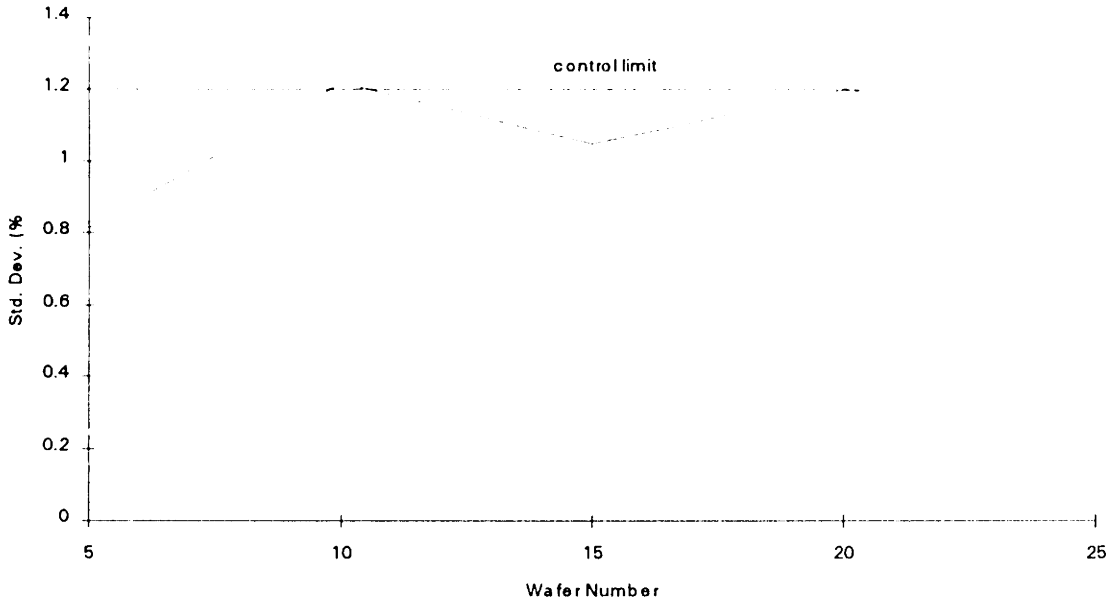

Figure 3.10 – 11 Point Uniformity Measurements for Barrier Layer

The uniformity data for the barrier target is unclear, but appears to require the entire 25 wafer lot to stabilize within control limits. Thus, most of the gains will be found in reducing wafer transport time by depositing more material on fewer wafers. Constraining the total thickness to levels known to avoid particle problems, the same amount of material may be deposited on 15 wafers instead of 25. Although deposition time increases, the corresponding gain in robot transport time of over one minute per wafer is greater. The anticipated savings for burn-in of new barrier target material is 18 minutes. The conductor film shows a better uniformity trend. Approximately 20% less material must be removed from the target than the present burn-in procedures to yield acceptable film uniformity. Another advantage is that film thickness may be safely tripled. The net result is that only seven conditioning wafers are required. Although individual process

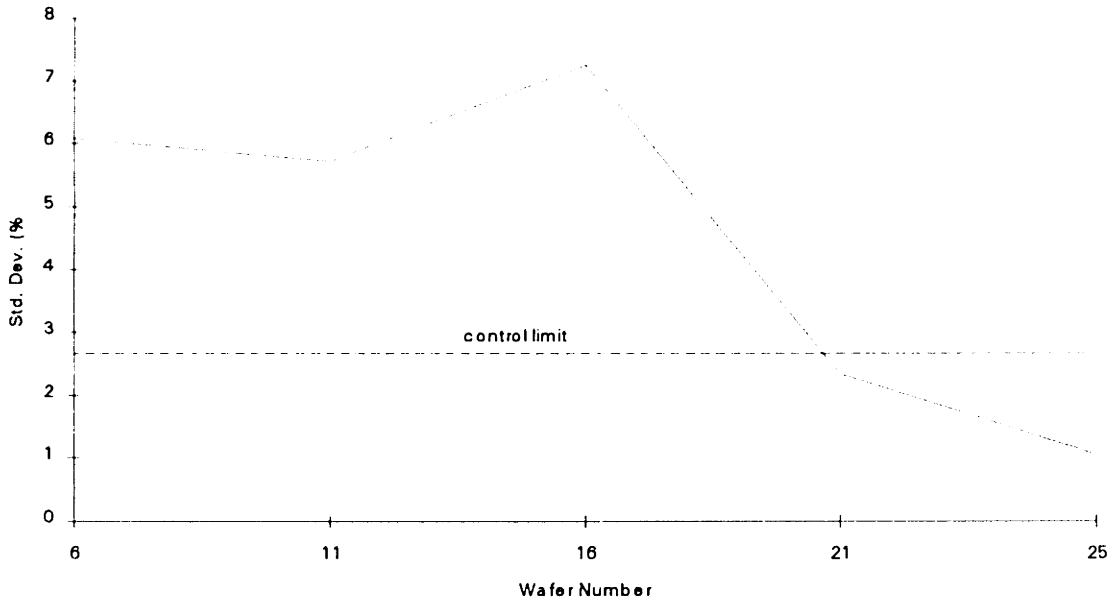

Figure 3.11 – 11 Point Uniformity Measurements for Conductor Layer

time for those wafers is considerably higher, the reduction of both net total deposition time and transport time yields savings of 30 minutes. Furthermore, the conductor chamber requires much more frequent maintenance. (This will be explored further in Chapter 4.) Thus, the weighted average savings for a generic target change is 25 minutes. A summary of these changes and associated time savings is displayed in Figure 3.12. Unfortunately, implementation of this new approach is problematic. The existing conditioning procedure uses standard production recipes, with well defined control limits for film thickness uniformity. The use of thicker films requires the calculation of new control limits based upon successive runs of this film thickness. Additionally, data points beyond those of Figures 3.10 and 3.11 are not yet available. Subsequent target changes must be monitored similarly to validate this data.

Chamber Type	% of Standard Deposition Required	Savings in Deposition Time	Number of Wafers Required	Savings in Transport Time	Total Time Savings
barrier/ ARC	100%	0 min.	10	18 min.	18 min.
conductor	80%	9 min.	7	21 min.	30 min.

Figure 3.12 – Summary of Target Conditioning Changes

Conclusions

The chamber bake out time supplied by Applied Materials proved excessive for Intel's needs, and was successfully halved from four to two hours. Although the target burn-in process shows potential for significant reductions of full machine maintenance downtime, more data is required to confirm initial data, as well as to establish new control limits for the streamlined conditioning process. Assuming that data from future conditioning procedures confirm the procedural change indicated above, the revised procedure now lasts six hours, 35 minutes, a reduction of over 25%. The pareto chart for the new procedure is shown in Figure 3.13. Overall machine availability is increased, as the 7.5% planned maintenance time shown in Figure 3.1 is reduced to 5.5%.

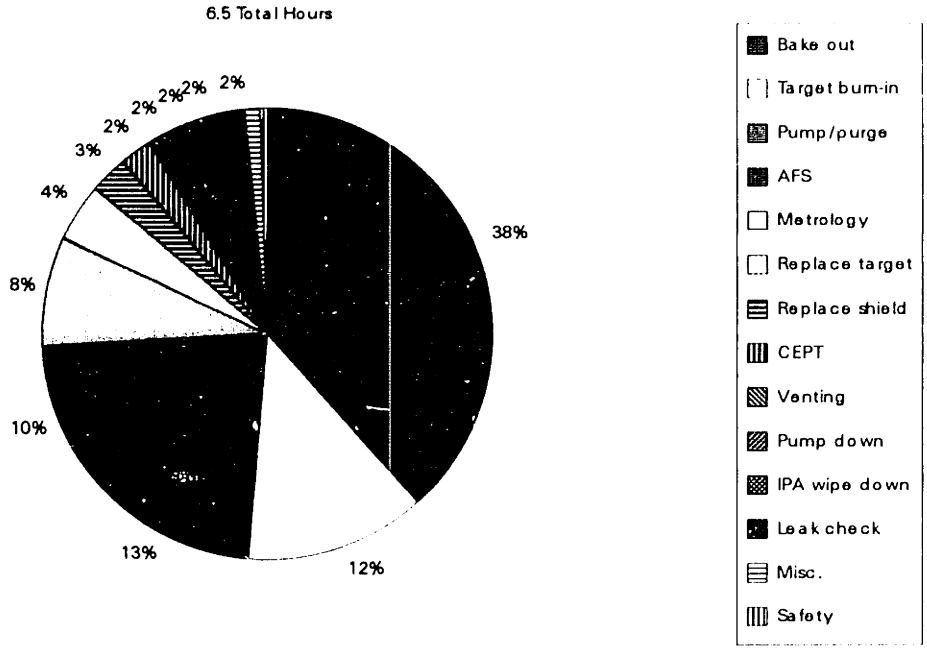

Figure 3.13 – Target/Shield Change Pareto – Revised Procedure

Chapter 4: Increasing Uptime using Static Modeling

Introduction

As seen in Chapter 3, certain procedures and process parameters can be used to decrease the length of target/shield change for the AMAT Endura 5500 sputterer. While shortening this time is important to increasing equipment availability, another productivity enhancing strategy is reducing the absolute number of times such maintenance procedures are required. This section investigates machine configuration and maintenance and production scheduling as methods that may reduce the total maintenance time required.

Contributors to Target/Shield Maintenance

Intel's next generation process technology requires at least nine distinct layers of metals. These layers consume different amounts of machine resources, i.e. target materials. Thus, required maintenance for the PVD chambers depends upon the amount of material required by each layer of the wafer. This amount is in turn influenced by the production schedule for that machine; that is, the type of layers run on a machine determines which target materials are consumed first. In addition, different target materials have different lifetimes. For example, a barrier metal target may provide material to the wafer with acceptable control for 250 kilowatt-hours (KWH), while a metal target may last 600 KWH. Of course, if redundant PVD chambers exist, the effective target life for that type of material increases. Continuing with the previous example, a second barrier target would give the machine an effective lifetime of 500 KWH. Chamber shields are usually replaced along with the target. However, anticipation of particle contamination problems may lead to more frequent scheduling of shield changes.

As mentioned in Chapter 3, the Endura 5500 can run material in "handicapped mode", with a PVD chamber disabled, thus limited the amount of downtime that inhibits production. However, if more than one chamber is pending target maintenance, it may be prudent to synchronize the maintenance of multiple chambers. Although this practice disables the entire machine, the resultant total downtime may be less than that of the piecemeal downtime encountered when servicing PVD chambers individually and operating the machine in handicapped mode. Common sense indicates that the ratio of handicapped mode downtime to full machine downtime is a critical parameter in this evaluation.

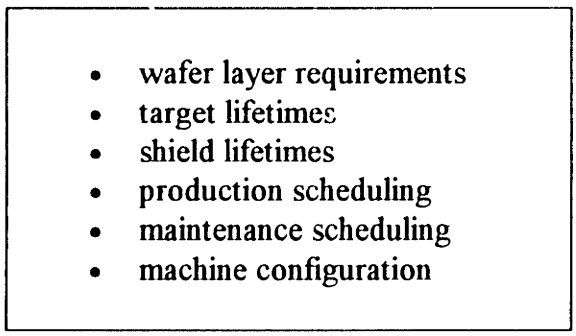

Figure 4.1 – Factors Contributing to Target/Shield Replacement

Optimization

A summary of the factors contributing to target or shield replacement are listed in Figure 4.1. As shown, many factors affect the frequency of this procedure, and some of these factors are interactive. Determining the correct decisions for the controllable variables of machine configuration, production scheduling, and maintenance scheduling is therefore difficult. One possible method is to incorporate this data on a simple computer spreadsheet, and try to minimize the number of KWH incurred by each PVD chamber by adjusting the type of wafers run on a machine. However, to balance that machine, what

are the effects upon other machines that have to run the remaining wafers? In addition, the possibility of redundant chambers must also be considered, as well as the possibility of servicing chambers sooner than required under a "full", rather than "handicapped" maintenance strategy. It is apparent that a spreadsheet solution would require multiple iterations of holding some variables steady, such as production loading, while manually altering other variables within an allowed range, such as number of chambers, in order to reduce the overall maintenance downtime required. Many permutations of these variables make this process extraordinarily difficult and time consuming. An alternative method is the use of optimization.

Optimization is an established iterative technique for solving problems such as these. This method attempts to *optimize* (typically minimize or maximize) the value of the target variable. This value in turn depends upon the magnitude of other variables, subject to reasonable limitions. The limitations on these variables are termed *constraints.* In short, optimization changes the values of the constrained variables in an effort to minimize or maximize the target variable. Optimization avoids checking every possible combination by analyzing the amount of increase or decrease registered in the target variable by a change in a constrained variable.

Solver

The mathematical principles used by optimization – usually categorized as linear programming (LP) or integer programming – are best implemented on a computer, and indeed, higher level languages like LINDO have been developed to leverage computer processing speed. Furthermore, computer processing also yields the additional benefits of sensitivity reports, allowing the user to determine the range of variables allowable for a given solution. However, LINDO shares the nuances (and nuisances) of formatting and

compiling issues, not unlike C, FORTRAN, or BASIC. A much simpler presentation of this tool exists as the Solver routine contained in Microsoft Excel and other spreadsheet programs. Solver avoids the hassles of setting variable names and conforming to formatting standards by leveraging the familiar cell structure, formula-building, and point-and-click features of spreadsheets.

Single Machine Model

The single machine model is shown in Figure 4.2. This model will be explained thoroughly, but only the differences in other, similar models will be discussed later. Each row and column of interest is designated as one of the following: the target variable (goal), a constraint, a machine parameter, a calculated value, or a variable.

kwh remaining – (constraint)
This first row tracks the effective amount of KWH life available for the target in that chamber based upon the type and number of wafers processed, and the number of each type of chamber available. This value is set to the *kwh chamber life* after a target change. This row contains the most significant constraints; each cell must remain above zero.

kwh chamber life – (parameter)
This row determines the total KWH available on a new target, based upon the type of target material in that chamber.

Chamber Location:	1	2	3	4								
kwh remaining	250	41	252	41								
kwh chamber life	250	200	750	200								
full maintenance hours	10.0	10.0	10.0	10.0	10							
# full maintenance	2	2	2	2	2							
total full hours					20							
h'cap/full hours	40%	40%	40%	40%								
h'cap hours	4.0	4.0	4.0	4.0								
# h'cap maintenance	0	4	0	4								
total h'cap hours	0.0	16.0	0.0	16.0	32							

Key:

target

constraint

variable

Total Maint. Hours

52

Chamber Type					type totals	normal loading %	actual loading %	kwh/ wafer	nominal # wafers	effective # wafers	kwh used	kwh life
Barrier	1	0	0	0	1	6%	6%	0.055	6000	6000	327	250
Conductor	0	1	0	1	2	52%	52%	0.453	6000	6000	2719	200
ARC	0	0	1	0	1	38%	38%	0.333	6000	6000	1998	750
location totals	1	1	1	1		100%	100%	0.870	6000	6000		
					4							

Figure 4.2 – Single Machine Solver Model with 40% Handicapped/Full Ratio

full maintenance hours – (parameter)

This row denotes the total time the machine cannot produce due to full scale target maintenance of all chambers simultaneously, based on the standard time for the procedure.

 # full maintenance – (variable)

This is the number of full scale maintenance procedures that are required. Each one contributes the number of full maintenance hours to the overall downtime incurred on the machine.

total full hours – (calculated value)

The total downtime incurred by full scale maintenance, in hours; i.e. (*full maintenance hours*) x (*# full maintenance*).

h'cap/full hours – (parameter)

The ratio of handicapped mode downtime to full maintenance mode downtime. Thus, 40% indicates that if full chamber maintenance inhibits production by 10 hours, then chamber-by-chamber maintenance takes four hours.

h'cap hours – (calculated value)

The number of hours required by the handicapped mode; i.e. (*full maintenance hours*) x (*h'cap/full hours*).

 # h'cap maintenance – (variable)

The number of individual chamber maintenance procedures required.

total h'cap hours – (calculated value)

The number of production halting hours caused by handicapped maintenance; i.e.

(*h'cap hours*) x (# *h'cap maintenance*).

Barrier, Conductor, ARC

These rows contain all relative information for process chambers of the respective types.

Chamber Location vs. *Chamber Type* – (variable)

Binary information in this area indicates the type of chamber used in different machine locations. For example, if Chamber 1 has a conductor target, then the cell in the *Conductor* row and *Chamber Location 1* column has a one; otherwise, a zero.

location totals – (constraint)

These constraints allow only one chamber at each physical location.

type totals – (constraint)

These constraints limit the total number of PVD chambers to four.

normal loading % – (parameter)

The fractions in these rows contain the relative KWH contribution that each chamber type must make to the completion of all nine layers of the wafer.

actual loading % – (variable)

The relative amount of material required to produce the wafer layer production mix entered into the model. Although academic in the single machine model (the single machine must run every type of layer in equal amounts), the *actual loading %* is a convenient method for adjusting production loading in a multiple machine plant. Of course, the weighted average in these plants must still equal *normal loading %*.

kwh/wafer – (parameter)

The amount of material required for the composite wafer, based upon chamber KWH.

nominal # wafers – (variable)

The nominal, or expected, number of wafers run through each chamber.

effective # wafers – (calculated value)

The effective number of wafers introduced in each chamber. This will deviate from *nominal # wafers* only when *actual loading %* differs from *normal loading %*.

kwh used – (calculated value)

The amount of target material of each type required for the effective number of wafers introduced into the model.

kwh life – (parameter)

The base lifetime of a single target of the indicated type.

Total Maintenance Hours – (target)

The total maintenance hours required for the machine, based upon the sum of *total full hours* and *total h'cap hours*. The goal is to minimize this target variable.

Determining Maintenance Strategy using the Single Machine Model

The simplest variation of the single machine model is to determine the maintenance hours required by a set machine configuration. In this scenario, the only variables that affect the goal of total maintenance hours are the number of full and handicapped maintenance routines required to service the machine. These variables are constrained to be positive

numbers, but may also be limited to integer variables to add a sense of realism. In other words, two or three procedures are viable, but 2.3 are not. This type of optimization is called integer programming. Many separate iterations must be run for each integer combination, instead of working over a range of valid real numbers. This approach, however, can result in long run times for complex models. The results of this model, using a baseline configuration of three PVD chambers – one each of conductor, ARC, and barrier type targets – has already been displayed in Figure 4.2.

A reasonable next step is to note changes in total maintenance hours while changing the ratio of handicapped to full maintenance hours. Figure 4.2 shows that the best combination stands at two full maintenance procedures, supplemented by four additional procedures on each of the two conductor chambers, with the latter "handicapped" procedure taking 40% of the full time. Figure 4.3 shows the effect of lowering that ratio to 20%: handicapped procedures are preferred to wholescale ones. In constrast, Figure 4.4 shows that full machine maintenance is preferred at a 55% ratio. Some simple calculations can help to validate the model. Figure 4.3 shows that 30 total hours are required using all handicapped procedures, while using a mix of four handicapped and two full (the numbers from Figure 4.2) would yield a higher total of 36 hours. A similar comparison yields 64 hours for the machine in Figure 4.4, again higher than the optimized value.

Determining Machine Configuration using the Single Machine Model

A simple application of this type of model can indicate the preferable maintenance mode of operation for a cluster tool, wholescale or piecemeal. It may also help to determine the machine configuration that minimizes maintenance downtime, regardless of the maintenance scheduling strategy. Consider again Figure 4.2, but this time allow the

configuration columns to vary, constrained to have a binary choice for each possibility.[1] For example, Chamber 1 must be used, and may have a conductor *or* an ARC *or* a barrier target. Thus, the total under column *ch1* must be constrained to exactly one. Because the machine needs at least one of each type of chamber, each number in the totals column must be greater than or equal to one. The result of running this model is shown in Figure 4.5. Figure 4.5 shows that two conductor chambers should be used to minimize maintenance hours. This exercise can be repeated with different handicapped/full time ratios and different maintenance strategies as in the above example. In each case, two conductor chambers are inserted to obtain a minimum total maintenance time.

[1] Unfortunately, binary integers cannot be directly designated as such within the Solver constraints. Instead, a combination of three constraints must be used to restrict the value of a cell variable to binary: 1) integer, 2) >=0, and 3) <=1. This omission is the most notable inconvenience to using Solver. The other is a limitation of the number of constraints in the model to 200, which is still quite a large model.

Chamber Location:	1	2	3	4								
kwh remaining	173	41	252	41								
kwh chamber life	250	200	750	200								
full maintenance hours	10.0	10.0	10.0	10.0	10							
# full maintenance	0	0	0	0	0							
total full hours					0							
h'cap/full hours	20%	20%	20%	20%		**Total**						
h'cap hours	2.0	2.0	2.0	2.0		**Maint.**						
# h'cap maintenance	1	6	2	6		**Hours**						
total h'cap hours	2.0	12.0	4.0	12.0	30	30						
Chamber Type					type totals	normal loading %	actual loading %	kwh/ wafer	nominal # wafers	effective # wafers	kwh used	kwh life
Barrier	1	0	0	0	1	6%	6%	0.055	6000	6000	327	250
Conductor	0	1	0	1	2	52%	52%	0.453	6000	6000	2719	200
ARC	0	0	1	0	1	38%	38%	0.333	6000	6000	1998	750
location totals	1	1	1	1		100%	100%	0.870	6000	6000		
					4							

Key:

target

constraint

variable

Figure 4.3 – Single Machine Solver Model with 20% Handicapped/Full Ratio

Chamber Location:	1	2	3	4									
kwh remaining	250	41	750	41									
kwh chamber life	250	200	750	200									
full maintenance hours	10.0	10.0	10.0	10.0	10								
# full maintenance	6	6	6	6	6								
total full hours					60								
h'cap/full hours	55%	55%	55%	55%		**Total**							
h'cap hours	5.5	5.5	5.5	5.5		**Maint.**							
# h'cap maintenance	0	0	0	0		**Hours**							
total h'cap hours	0.0	0.0	0.0	0.0	0	60							
Chamber Type					type totals	normal loading %	actual loading %	kwh/ wafer	nominal # wafers	effective # wafers	kwh used	kwh life	
Barrier	1	0	0	0	1	6%	6%	0.055	6000	6000	327	250	
Conductor	0	1	0	1	2	52%	52%	0.453	6000	6000	2719	200	
ARC	0	0	1	0	1	38%	38%	0.333	6000	6000	1998	750	
location totals	1	1	1	1		100%	100%	0.870	6000	6000			
					4								

Key:

target

constraint

variable

Figure 4.4 – Single Machine Solver Model with 55% Handicapped/Full Ratio

Chamber Location:	1	2	3	4								
kwh remaining	250	41	252	41								
kwh chamber life	250	200	750	200								
full maintenance hours	10.0	10.0	10.0	10.0	10							
# full maintenance	2	2	2	2	2							
total full hours					20							
h'cap/full hours	40%	40%	40%	40%		**Total**						
h'cap hours	4.0	4.0	4.0	4.0		**Maint.**						
# h'cap maintenance	0	4	0	4		**Hours**						
total h'cap hours	0.0	16.0	0.0	16.0	32	52						

Key:

target

constraint

variable

Chamber Type					type totals	normal loading %	actual loading %	kwh/ wafer	nominal # wafers	effective # wafers	kwh used	kwh life
Barrier	1	0	0	0	1	6%	6%	0.055	6000	6000	327	250
Conductor	0	1	0	1	2	52%	52%	0.453	6000	6000	2719	200
ARC	0	0	1	0	1	38%	38%	0.333	6000	6000	1998	750
location totals	1	1	1	1		100%	100%	0.870	6000	6000		
					4							

Figure 4.5 – Single Machine Solver Model used for Machine Configuration

Effects of Production Loading on Target/Shield Maintenance

As discussed earlier, in a plant with only one machine, the choice of which machine to use for what type of layers – production loading – is nonexistent. All layers on each wafer must be processed through a single machine. However, in multiple machine plants, different machines may be dedicated to different types of layers. A single tool in a five machine plant is shown in Figure 4.6. Note that two columns have been added as constraints: *upper load* and *lower load.* These columns contain upper and lower limits for the loading of a machine. As an example, there is no combination of layers which require a ratio of higher than 84% conductor layers. Alternately, it is not possible to have less than 3% of the total KWH consumed by the barrier layer. Thus, different production loading schemes may be modeled by allowing the ratio of KWH expended by each type of PVD chamber to vary within realistic ranges. A final constraint is required, however. For the total of the five machines, the *average loading %* must equal *normal loading %*; i.e. the machines must together process an equal amount of each layer to reflect real production schedules. For this experiment, different combinations of machine configurations are set and not varied. Again, the number of maintenance procedures are allowed to vary. Then as the control case, the optimization is run fixing the actual production layer loading ratio, *actual loading %*, to the normal load, *normal loading %*. Then, the trials are repeated, this time allowing the production loads to vary to optimized values. Finally, suboptimal production loads are fixed for each machine. These are values that, while conforming to the overall load balancing constraint, are close to the load limits for individual machines. The results of these trials are displayed in Figure 4.7.

Machine # 1 of 5

Chamber Location:	**1**	**2**	**3**	**4**											
kwh remaining	250	41	252	41											
kwh chamber life	250	200	750	200											
full maintenance hours	10.0	10.0	10.0	10.0	10										
# full maintenance	2	2	2	2	2										
total full hours					20										
h'cap/full hours	40%	40%	40%	40%		**Machine**	**Total**								
h'cap hours	4.0	4.0	4.0	4.0		**#1**	**Maint.**								
# h'cap maintenance	0	4	0	4		**Hours**	**Hours**								
total h'cap hours	0.0	16.0	0.0	16.0	32	52	260								
Chamber Type					type totals	normal loading %	average loading %	upper load	actual load	lower load	kwh/ wafer	nominal # wafers	effective # wafers	kwh used	kwh life
Barrier	1	0	0	0	1	6%	6%	100%	6%	3%	0.055	6000	6000	327	250
Conductor	0	1	0	1	2	52%	52%	84%	52%	0%	0.453	6000	6000	2719	200
ARC	0	0	1	0	1	38%	38%	83%	38%	0%	0.333	6000	6000	1998	750
location totals	1	1	1	1		100%	100%				0.870	6000	6000		
					4										

Key:

target

constraint

variable

Figure 4.6 – Five Machine Solver Model reflecting Production Loading

Machine Configuration	Normal Load Maintenance Time (hours)	Optimal Load Maintenance Time (hours)	Suboptimal Load Maintenance Time (hours)
5 Base	288	288	298
1 Adhesion, 4 Base	288	288	291
5 Standard	262	262	281
3 Standard, 2 Metal	256	256	273
2 Standard, 3 Metal	254	254	269
5 Metal	249	249	279

Figure 4.7 – Effects of Production Loading and Machine Configuration

This data indicates that for this set of process recipes, normal loading (same number of each type of layer processed on each machine) results in the same number of maintenance hours as an optimized load. However, suboptimal loading can result in maintenance hours increasing by up to 12%. Again, the configuration with two conductor chambers fares best, with the metal configuration requiring 5% less maintenance hours (249) than the standard configuration (262).

Conclusions

The static optimization models indicate that production loading, maintenance scheduling, and machine configuration all contribute to total target/shield maintenance time.

The advantage to balancing production loading in this case is to avoid losses, since the normal load happens to be an optimized load. However, some obstacles combine to confound this process. A man-machine interface must direct operators to the best idle machine to load for a particular layer in order to maintain balance. However, this requires a real time exchange of data (remaining lifetime of each target, requirements of each layer,

status of other machines) that simply does not exist on the present computer-aided manufacturing control system. Additionally, surges in work-in-process inventory may result in the wrong layers being fed to unbalanced machines. Delaying processing to balance machine loading would increase inventory and cycle time.

The model also indicates that servicing multiple chambers can be advantageous over using a piecemeal policy. The decision is driven primarily by the ratio of single chamber maintenance time to total machine maintenance time. Unfortunately, the model is weakened by two exclusions: maintenance reliability and manpower availability. The model assumes that every chamber maintenance procedure will be effective. However, especially for high vaccuum cluster tools, this is not always the case. Problems in wafer film quality or particle generation may force a repetition of chamber maintenance, increasing overall time and deterring attempts at "early" chamber maintenance. Also, the ratio of single chamber maintenance time to total machine maintenance time is aggravated by the availability of manpower. The ratio increases greatly if the number of chambers is greater than the number of personnel available.

Nevertheless, the benefit of redundant conductor chambers seems clear. Each model, regardless of production loading or maintenance policy, showed decreased total maintenance time when two conductor PVD chambers were used. Only corresponding production throughput problems should deter the implementation of this strategy. This issue is addressed in Chapter 5.

Chapter 5: Increasing Throughput Rates using Dynamic Modeling

Introduction

The static linear programming models developed in Chapter 4 indicate that increasing the number of conductor PVD chambers reduces total planned target/shield maintenance, regardless of maintenance scheduling strategies. The production balancing constraints considered in the model of Figure 4.6 are limited, however, as that model merely restricts single machine output to realistic ranges and requires overall production to "balance" (or produce equal amounts of every metal layer of the wafer) within a multiple machine plant. A significant drawback to this model is the negligence of difference in throughput times for unique layers on different machine configurations. While Chapter 4 indicated that overall clustered tool availability could be improved through thoughtful choices of machine configuration, this section will investigate the effects of configuration on machine throughput.

The major motivation for using simulation for operations modeling is to gain knowledge about the impact of proposed changes in a production system while avoiding time consuming and expensive experiments or impractical facility rearrangements. A well designed model can reflect proposals which alter critical entities within a manufacturing plant, such as machine cycle time, buffer capacity, failure and repair rates, machine setups, labor availability, and so forth.

Simulation Tools

Many different software packages are available, but they may be classified into two categories: simulation languages and simulators. Simulation languages such as SIMAN,

AUTOMOD, and GPSS are higher level languages with special commands devoted to simulation needs. The major strength of these packages is flexibility; however, these packages require significant programming insight and development time for even simple models. Contrasting simulation languages are simulators, including WITNESS and XCELL. These software packages have configurable entities such as machines, buffers, and parts whose attributes may be tailored to the needs of the model. The flexibility and menu driven development features of simulators simplify and quicken the modeling process. Sophisticated packages developed with simulation languages often provide a limited function set of the underlying language, while others may opt for a pseudo code (similar to BASIC) which also allows for more complex modeling logic. Other factors which influence the selection of a simulation tool are credibility and availability. Even a well-designed model is useless if the results hold no credence with decision makers in an organization. In addition, simulation tools, due to their complexity and limited sales volume, are quite expensive, reaching prices of tens of thousands of dollars. Thus, these last items are highly linked -- management will not allocate such funds unless they acknowledge simulation as a useful tool. Thus, the model builder is usually restricted to software available in house. In this case, the simulator WITNESS was recently purchased by the Intel PTD site, so WITNESS is used for all the dynamic model building that follows.

Modeling Strategy

The objective of this simulation effort is to measure the effects on throughput of different chamber configurations of the AMAT Endura 5500 sputtering machine. Intel uses PVD chambers to deposit three different types of metal thin films upon nine different layers of the silicon wafer. As shown in Figure 1.2, locations 1 through 4 are available on the machine mainframe for PVD chambers. In addition, other processes must be performed

including degas/orient, etching, and wafer cooling. The degas/orient process may be performed at locations C, D, E, or F while etching may be performed in line in Chamber A or externally in positions C or D. Cool down occurs in line in Chamber B. The desired outcome is to determine the most effective use for "spare" chamber positions with respect to throughput rates, and to recommend the best configurations. These recommendations must be tempered by any conflicts with optimal maintenance configurations and conflicts generated by differences in process technologies across various Intel sites.

Ideally, the modeling process allows fast development of a model that completely reproduces the measured outputs of an actual production system. However, a model that replicates production systems to such a degree may be unnecessarily complicated and yield no better information than a "rough" model. To that end, the simulation strategy used here is to quickly develop a working model that determines the relative magnitude of anticipated improvement. If these gains are enticing, more time may be invested to build a model with the precision to accurately quantify these improvements.

Developing the Rough Model

A single machine model of the AMAT 5500 sputterer is a microscopic view – the tool must be modeled as a miniature production plant, with each element modeled as an individual machine. This contrasts with the scope of a plant wide model where a sputterer may be considered a single "black box" of attributes such as cycle time, breakdowns, and labor requirements. In a tool level simulation of the 5500, the process chambers may be modeled as simple, single stage machines with a set cycle time. Note that because the chamber cycle times are fixed within wafer recipes, the processes are not stochastic and are therefore repeatable. This reduces time otherwise required for experimental simulation runs. In contrast, the most complex, and hence most difficult to model, are the wafer handling

robots. These two robots, located in the buffer chamber (flanked by chambers C, D, E, and F) and the transfer chamber (flanked by chambers 1, 2, 3, and 4) also have predictable and repeatable wafer transfer time requirements. However, the robots make complex decisions regarding movement of wafers based upon wafer sequences and chamber availability. As such, the most effective simplifying assumption eliminates the robots from the model. In this case, the wafers "jump" between process chambers without the aid or time delay of the robot. Overall cycle times are further reduced by neglecting the wait time imposed by a robot working on one wafer while another is finished within a process chamber. In other words, the rough model would allow a wafer to move from Chamber 1 to 2, while another moves simultaneously from 3 to 4. This assumption allows for a simplified model building process. This model introduced wafers to the machine in lot sizes (batches) of 25 into the load locks, processed them using existing production recipe times, and returned the wafers to the load locks. The total time required to learn WITNESS software features and complete the rough model was three days.

Using the Theory of Constraints to Build a Cluster Tool

As mentioned, certain chambers are required for different layers of the wafer. A degas/orient and a cool down chamber are required for each of the nine wafer sequences. A barrier metal chamber is also used in each sequence. An etching or "pre-clean" chamber (either internal or external) is required for adhesion layers. An anti-reflective coating is used in all but one of sequences. One conductor chamber is necessary for any of the conducting or "metal" layers. Still, spare chamber locations are available, even when the machine is minimally configured to process each of the layers. Starting with this base configuration, constraint theory may be used to decide what type of process chambers should be placed in the spare slots on the machine mainframe. The Metal 2 and Metal 3

layers (which share a common sequence) make a good case study in configuring the tool. The steps and time required to process 50 Metal 2/3 wafers, are shown in Figure 5.1.

Layer	Total Cycle Time (minutes)
Metal 2/3	68.9

Process	Number of Chambers	Effective Cycle Time (seconds)
degas/orient	1	66
barrier	1	18
conductor	1	50
ARC	1	16
cool	1	30

Figure 5.1 – Wafer Sequence and Cycle Times for Metals 2 and 3

Constraint theory maintains that overall throughput rates may be improved only by decreasing the cycle time of the constraint, or the weakest link in the production chain. Any efforts to improve the performance of other machines in the sequence will be ineffective.

The simulation study begins with a baseline tool comprised of one each of the following process chambers: degas/orient, in line etch, barrier metal, anti-reflective coating (ARC), conductor, and cool down. This configuration can be illustrated using the chart displayed in Figure 5.2 in conjunction with the illustration in Figure 1.2. Since this model neglects robot transfer times and other mechanical delays, total cycle times are artificially low. In addition, cycle times for each step in the process are decreased by exactly one-half when a second identical type of process chamber is added and decreased by exactly two-thirds when a third redundant chamber is added. Figure 5.1 shows that the degas/orient process requires 66 seconds, the most processing time. Thus, the degas/orient chamber is the

Chamber Position	Function
1	barrier
2	ARC
3	conductor
4	none (spare)
A	etch
B	cool down
C	none (spare)
D	none (spare)
E	degas/orient
F	none (spare)

Figure 5.2 – Baseline Configuration

weakest link or the "bottleneck" for this sequence. In order to improve the total cycle time of 68.9 minutes for 50 wafers of this sequence, a second degas/orient chamber must be added to the tool. When a redundant degas/orient chamber is added, the total cycle time of the tool decreases, as shown in Figure 5.3. The effective cycle time of the degas/orient process is now cut by half, to 33 seconds. This results in a new total of 55.8 minutes. To illustrate and confirm the effects of constraint theory, consider adding a second conductor chamber, instead of a second degas/orient, to the baseline tool. Figure 5.4 shows the result. The effective cycle time for the conductor deposition process has been halved. However, the total cycle time of 68.9 minutes is identical to the original cycle time of the machine that had only one conductor chamber. Clearly, adding a redundant conductor chamber to the baseline machine is not productive in terms of the goal of reducing total cycle time. Returning to the configuration with two degas/orient chambers illustrated in Figure 5.4, the conduction deposition process now becomes the bottleneck. Its cycle time of 50 seconds is now higher than the effective degas/orient process, now at only 33 seconds due to the addition of a second degas/orient chamber.

Chamber Position	Function
1	barrier
2	ARC
3	conductor
4	none (spare)
A	etch
B	cool down
C	none (spare)
D	none (spare)
E	degas/orient
F	none (spare)

Figure 5.2 – Baseline Configuration

weakest link or the "bottleneck" for this sequence. In order to improve the total cycle time of 68.9 minutes for 50 wafers of this sequence, a second degas/orient chamber must be added to the tool. When a redundant degas/orient chamber is added, the total cycle time of the tool decreases, as shown in Figure 5.3. The effective cycle time of the degas/orient process is now cut by half, to 33 seconds. This results in a new total of 55.8 minutes. To illustrate and confirm the effects of constraint theory, consider adding a second conductor chamber, instead of a second degas/orient, to the baseline tool. Figure 5.4 shows the result. The effective cycle time for the conductor deposition process has been halved. However, the total cycle time of 68.9 minutes is identical to the original cycle time of the machine that had only one conductor chamber. Clearly, adding a redundant conductor chamber to the baseline machine is not productive in terms of the goal of reducing total cycle time. Returning to the configuration with two degas/orient chambers illustrated in Figure 5.4, the conduction deposition process now becomes the bottleneck. Its cycle time of 50 seconds is now higher than the effective degas/orient process, now at only 33 seconds due to the addition of a second degas/orient chamber.

Layer	Total Cycle Time (minutes)
Metal 2/3	55.8

Process	Number of Chambers	Effective Cycle Time (seconds)
degas/orient	2	33
barrier	1	18
conductor	1	50
ARC	1	16
cool	1	30

Figure 5.3 – Metal 2/3 Cycle Times with 2 Degas/Orient Chambers

Layer	Total Cycle Time (minutes)
Metal 2/3	68.9

Process	Number of Chambers	Effective Cycle Time (seconds)
degas/orient	1	66
barrier	1	18
conductor	2	25
ARC	1	16
cool	1	30

Figure 5.4 – Metal 2/3 Cycle Times with 1 Degas/Orient and 2 Conductor Chambers

To further reduce total cycle time, a redundant conduction chamber must now be added. This addition reduces the effective conduction cycle time to 25 seconds, further reducing total cycle time, as shown in Figure 5.5. Thus, the addition of a redundant conduction chamber reduced cycle time, *but only in the presence of a redundant degas/orient chamber*. Without a second degas/orient chamber, an additional conduction chamber is worthless in terms of total cycle time. Only actions which increase the effectiveness of the bottleneck process increase throughput; any other "improvements" are not productive.

Layer	Total Cycle Time (minutes)
Metal 2/3	41.9

Process	Number of Chambers	Effective Cycle Time (seconds)
degas/orient	2	33
barrier	1	18
conductor	2	25
ARC	1	16
cool	1	30

Figure 5.5 – Metal 2/3 Cycle Times with 2 Degas/Orient and 2 Conductor Chambers

Similarly, adding a redundant barrier or ARC chamber would be non-productive at this point. Continuing with the same methodology, the only way to further reduce total cycle time is by adding a third degas/orient chamber, attaining the optimized configuration for Metals 2 and 3. Figure 5.6 shows the results. A similar exercise may be completed for each of the remaining seven layers. The results are summarized in Figures 5.7 and 5.8.

Layer	Total Cycle Time (minutes)
Metal 2/3	39.5

Process	Number of Chambers	Effective Cycle Time (seconds)
degas/orient	3	22
barrier	1	18
conductor	2	25
ARC	1	16
cool	1	15

Figure 5.6 – Completing the Optimized Configuration for Metals 2 and 3

Note that chambers are added as needed to alleviate constraints. Results of sub optimal configurations are also included to illustrate the effects of constraint theory. Further note that external etch chambers have been assumed for the configurations in Figure 5.8.

Configuration (#degas/orient, #conductor)	1,1	1,2	2,1	2,2	3,1	3,2
Metal 1 Cycle Time (minutes)	68.5	68.5	41.5	41.5	39.1	39.1

Process	Effective Cycle Time (seconds)					
degas/orient	66	66	33	33	22	22
barrier	18	18	18	18	18	18
conductor	27	13.5	27	13.5	27	13.5
ARC	16	16	16	16	16	16
cool	30	30	30	30	30	30

Configuration (#degas/orient, #conductor)	1,1	1,2	2,1	2,2	3,1	3,2
Metal 4 Cycle Time (minutes)	96.2	89.7	96.2	55.9	96.2	55.9

Process	Effective Cycle Time (seconds)					
degas/orient	90	90	45	45	30	30
barrier	18	18	18	18	18	18
conductor	98	49	98	49	98	49
ARC	16	16	16	16	16	16
cool	30	30	30	30	30	30

Figure 5.7 – Metals 1 and 4 Cycle Times for Various Chamber Configurations

Proliferation of Specialized Machines

Note that similar processes have been segregated to separate machines optimized for those processes. This was done to restrict proliferation of many unique machine configurations. Multiple configurations, each dedicated to specific wafer layers, are somewhat

Configuration (#degas/orient, #etch, #barrier, #ARC)	1,1,1,1	1,2,1,1	2,2,1,1	2,2,1,2	2,2,2,2
Adhesion 1-4 Cycle Time (minutes)	78.6	69.6	48.4	47.0	47.0

Process	Effective Cycle Time (seconds)				
degas/orient	66	66	33	33	33
etch	77	38.5	38.5	38.5	38.5
barrier	10	10	10	10	5
ARC	40	40	40	20	20
cool	30	30	30	30	30

Configuration (#degas/orient, #etch, #barrier, #ARC)	1,1,1,1	1,2,1,1	2,2,1,1	2,2,1,2	2,2,2,2
Barrier Cycle Time (minutes)	37.8	37.8	37.8	37.8	37.8

Process	Effective Cycle Time (seconds)				
degas/orient	30	30	15	15	15
barrier	17	17	17	17	8
cool	30	30	30	30	30

Figure 5.8 – Adhesion/Barrier Cycle Times for Various Chamber Configurations

disadvantageous. Small facilities may not have the volume to support multiple configurations. For example, it is impossible to segregate layers into five different types of machines if the overall plant capacity requires only four machines. Additionally, Intel's "Copy Exactly" technology transfer strategy is negatively impacted by tool proliferation. To seamlessly transfer process technology to high volume fabrication sites, process development groups would require extra facilities and more development time to validate wafer fabrication processes on each machine configuration type.

For this nine layer process, consider the benefits of the two machines in question. Each configuration has no spare chamber positions on the machine mainframe. The optimal configuration for metal layers requires three degas/orient and two conductor chambers, while the optimal machine for adhesion/barrier layers has two degas/orient, two etch, and two ARC chambers. Notice, however, the effect of diminishing returns as the effective chamber times become "balanced". For example, a 39% cycle time reduction is achieved by adding redundant degas/orient and conduction chambers for the Metal 2/3 layers, while the third degas/orient and second cool down account for only 3% of further reduction. Similarly, a 38% reduction is achieved for adhesion layers with two degas/orient and two external etch chambers, while only a 2% further reduction is attained by adding another ARC chamber. A common configuration, which services every layer yet provides most of the cycle time benefit of the dedicated layer strategy can be attained by combining the best of the dedicated machines. This configuration has two degas/orient, two etch, two conduction, and a single ARC chamber, and yields a weighted total cycle time which is 63% of the baseline machine.

Alternating Different Wafer Sequences

One useful feature within WITNESS is the ability to introduce parts into the model via an ASCII text file. In this manner, any "run" of production parts may be introduced. This function is especially valuable in using actual production schedules to validate simulations of large factories. In this case, it allows for easy introduction of a finite lot of 25 identical wafers. It is also useful for testing the following supposition: what happens when wafers requiring different recipes are introduced from each load lock in an alternating fashion? For example, a barrier layer requires this sequence in the baseline configuration: E, A, 1, B, with Chamber B as the bottleneck. One metal layer requires the following chamber sequence : E, A, 1, 3, 2, B, with Chamber 3 as the bottleneck process. Typically, both lots

are loaded into the machine and then fed sequentially, or cascaded, one after the other. Wafer 1 from the second load lock immediately follows wafer 25 from the first load lock, as follows: 1, 2, 3, ... 24, 25, 1, 2, 3, ... 24, 25. However, it is possible to use the text file to introduce the wafers in an alternating sequence such as: 1, 1, 2, 2, 3, 3, ... 24, 24, 25, 25. When compared to the former schedule, the latter almost always results in lower net cycle times, as shown in Figure 5.9. It is noteworthy that in order to avoid delays caused by load lock pumping and venting times, a more realistic continuous sequence would be similar to: 1, 2, 3, 4, 5, 1, 6, 2, 7, ... 20, 25, 21, 1, 22, 2, etc. Otherwise, these pumping and venting times would not be transparent and would be included in the total cycle time. The effect of this loading feature is to diminish the effect of bottlenecks by overlapping tasks. For example, the first wafer may be processing in Chamber B, its bottleneck. Instead of the next wafer sitting idle, however, it may be completing processing in its bottleneck, Chamber 3. Although this functionality does not exist in the current revision of the Endura 5500 software, the positive effects of optimized scheduling upon clustered tools is evident by this example.

Processing Two Layers within One Chamber

The rough model indicates that two conductor layers are required for the optimal machine configuration. However, this configuration conflicts with the production requirements of Intel's production sites and the "Copy Exactly" philosophy. The sequences discussed earlier in this chapter are required by Intel's future fabrication process. The present generation process conflicts in an important way: some of the present metal layers require two separate anti-reflective coating layers within the single stack of metal layers. In other words, these metal layers have one conductor and one barrier layer, but two ARC layers. This metal stack is impossible to process using only one ARC chamber. If only one

Configuration: Base									
% change in cycle time when scheduling alternating wafer layers:									
B/ A	B/ M1	B/ M2	B/ M4	A/ M1	A/ M2,3	A/ M4	M1/ M2,3	M1/ M4	M2,3 M4
13.8	0.7	1.0	21.4	0.2	0.2	8.8	0.3	3.2	3.5

Configuration: Base + external etch									
% change in cycle time when scheduling alternating wafer layers:									
B/ A	B/ M1	B/ M2	B/ M4	A/ M1	A/ M2,3	A/ M4	M1/ M2,3	M1/ M4	M2,3 M4
2.4	0.7	1.0	21.4	3.0	3.3	6.3	0.3	3.2	3.5

Configuration: Base + 1 degas/orient, 2 external etch, 1 conductor									
% change in cycle time when scheduling alternating wafer layers:									
B/ A	B/ M1	B/ M2	B/ M4	A/ M1	A/ M2,3	A/ M4	M1/ M2,3	M1/ M4	M2,3 M4
7.4	1.2	3.5	13.9	1.3	1.2	7.9	0.2	1.8	2.2

Configuration: Base + 1 degas/orient, 2 external etch, 1 barrier									
% change in cycle time when scheduling alternating wafer layers:									
B/ A	B/ M1	B/ M2	B/ M4	A/ M1	A/ M2,3	A/ M4	M1/ M2,3	M1/ M4	M2,3 M4
9.0	3.2	17.8	16.6	1.4	13.8	20.1	4.6	2.9	0.3

Configuration: Base + 1 degas/orient, 2 external etch, 1 ARC									
% change in cycle time when scheduling alternating wafer layers:									
B/ A	B/ M1	B/ M2	B/ M4	A/ M1	A/ M2,3	A/ M4	M1/ M2,3	M1/ M4	M2,3 M4
7.5	11.3	6.9	16.6	4.0	16.7	21.0	4.6	2.9	0.3

Key:

Base configuration = 1 degas/orient, 1 etch, 1 barrier, 1 ARC, 1 conductor

B = barrier A = adhesion

M1 = metal 1 M2= metal 2 M3 = metal 3 M4 = metal 4

##.# = % cycle time decrease ##.# = % cycle time increase

Figure 5.9 – Alternating Schedules for Wafers using Different Sequences

chamber is used, wafers would have to wait at Chamber A until the previous wafer clears the PVD chambers and exits to Chamber B. To clarify, consider the wafer sequence displayed in Figure 5.10.

1. Start from loadlock
2. Chamber E
3. Chamber A
4. Chamber 3
5. Chamber 2
6. Chamber 1
7. Chamber 2
8. Chamber B
9. Return to loadlock

Figure 5.10 – Wafer Sequence using 2 ARC Layers

As shown, each wafer requires two separate ARC layers. Using the baseline machine illustrated by Figures 1.2 and 5.2, the wafer chamber sequence for the 2 ARC layer metal would be: E, A, 3, 2, 1, 2, B. However, the wafers would become "jammed" using this sequence, as eventually both wafers in Chamber 3 and 1 must wait for Chamber 2, while the wafer in Chamber 2 must wait for Chamber 1 to clear. Fortunately, the Endura 5500 control software is sophisticated enough to prevent this occurance. In this case, the subsequent wafer would be held at Chamber A until the previous wafer passed through Chamber 2 twice. Necessarily, then, the preferred configuration for processing wafers was the baseline configuration, plus adding a second ARC chamber in position 4. Thus, differences between layers seems to prohibit the gains anticipated by the single preferred configuration discussed earlier.

Chamber 5

Applied Materials recently introduced Chamber 5, which is mounted on the mainframe between Chambers 2 and 3. Due to space restrictions, a PVD type chamber cannot be added. Only a chamber comparable in size to the degas/orient chamber may be placed in this position. Applied Materials intends this chamber to perform a maintenance function. A "dummy" (non-product) wafer would sit within Chamber 5, and be moved into a PVD chamber to serve as a protective cover for the wafer pedestal during an automated cleaning operation. This automated activity would replace the requirement of introducing dummy wafers from outside the machine for this purpose. Another, although unintended, function of this chamber could be to serve as an in process buffer. Indeed, the use of this chamber as a buffer allows a sequence with two ARC layers to be run using only one ARC chamber. This would allow Chamber 4 to be used as a redundant conductor deposition chamber, thereby gaining the cycle time benefits mentioned earlier. The wafer sequence required for the two ARC layer stack is shown in Figure 5.11. Differences are highlighted in italics. It is assumed that Applied Materials will develop addressing methods and sequencing accessibility similar to the existing Chambers 1 through 4. Notice that in this sequence both Chambers 4 and 5 are used as buffer chambers to avoid "jamming" the wafers. This introduces the potential quality problem of cross-contaminating the Chamber 4 conductor chamber with particles from other chambers. Another, more subtle problem is displayed in Figure 5.12, which displays the utilization of each process chamber. Note that the utilization of Chambers 4 and 5 differs. This means that not every wafer undergoes the same processing: some use Chamber 4, some use Chamber 5, and some use neither. Figure 5.13 shows this clearly. This creates processing problems because the wafers have differing time/temperature profiles. The duplication of results of present processing recipes would have to be demonstrated using the new sequence. Additionally, if this equipment is used in future processes, it would be unknown whether development

of a desirable two ARC layer or two barrier layer stack would be undermined by these differing wafer histories.

1. Start from loadlock
2. Chamber E
3. Chamber A
4. Chamber 3
5. Chamber 2
6. Chamber 1
7. *Chamber 2, 4, or 5*
8. Chamber 2
9. Chamber B
10. Return to loadlock

Figure 5.11 – Wafer Sequence using 2 ARC Layers using Chamber 5

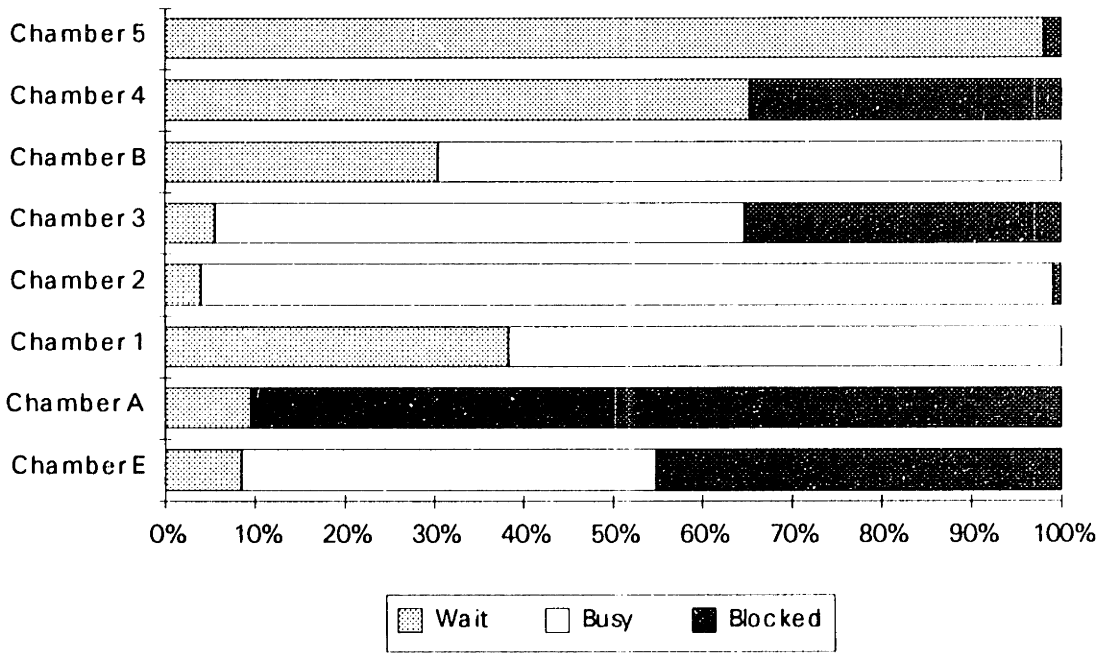

Figure 5.12 – Chamber Utilization using Chamber 5 and Rough Model

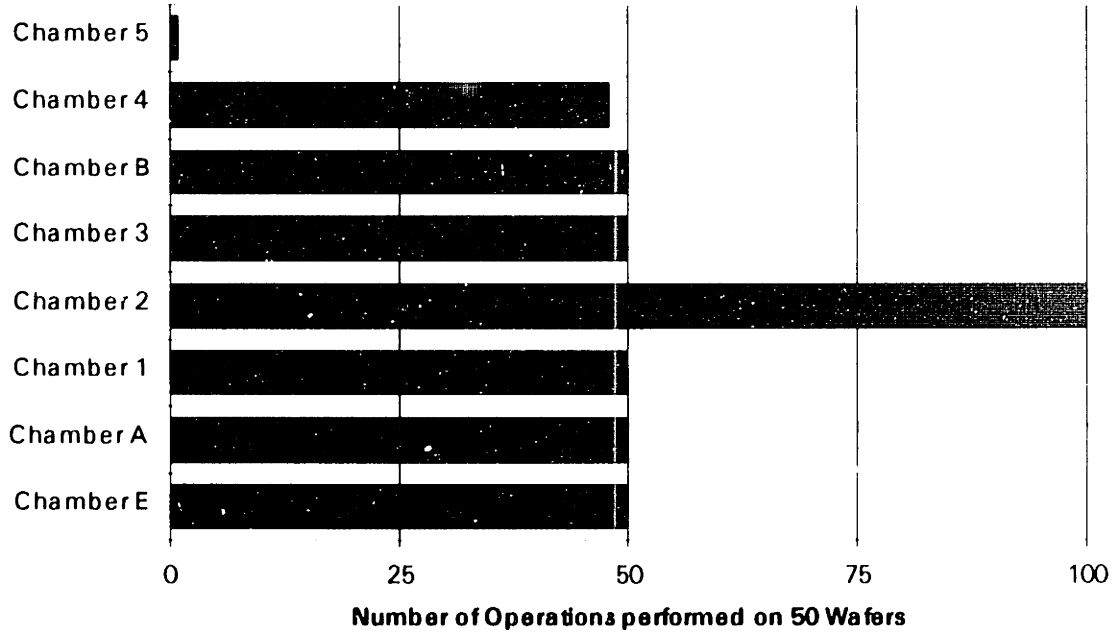

Figure 5.13 – Chamber Operations using Chamber 5 and Rough Model

Developing the Detailed Model

The rough model has yielded two intriguing hypotheses: the throughput enhancing value of redundant conductor chambers, and the feasibility of using Chamber 5 as an intra-machine buffer. However, the rough model can only hint at the scale of cycle time savings offered by a second conductor chamber or the extent of difference in wafer processing time profiles inferred by the use of Chamber 5. Thus, a more accurate model is required to quantify these values. The "detailed" model developed must reflect delays not included in the rough model. These delays include chamber delay time caused by a busy robot, robot transfer times between process chambers, and mechanical delays caused by wafer pedestal movement. Including mechanical delay within the chamber is a simple matter, obtained by merely appending additional time to the chamber recipe time requirements. Reflecting the robots, however, involves replicating a portion of the robot control logic

within the model. An example of the contrast of the logic required for the rough and detailed models is shown in Appendix B.

As another comparison, development time for the detailed model took two additional weeks, compared to two days for the original rough model. However, the detailed model now reflects actual cycle times to within +/- 3%.

Now the hypotheses of the rough model may be validated. The simulation runs of the sequences in Figures 5.6, 5.7 and 5.8 are repeated using the detailed model. The net results are shown in Figures 5.14 and 5.15. This time, one lot, or 25 wafers is used. Load lock pumping and venting times are ignored. Additionally, since the cycle times are "real" the corresponding wafer throughput rates are also tabulated.

These results are similar but not identical to the rough model. The addition of conductor chambers yields smaller returns due to overall larger cycle times per wafer and the additional delays associated with the robots. Conceptually, the biggest change is the relative merit of adding a second adhesion chamber. The simple model shows a cycle time decrease of only 2% of the base time, while the detailed model shows a 10% decrease. Obviously, the buffer and transfer robots are integral parts of the machine and resources that must be included to obtain accurate comparisons.

These results show that the optimal strategy for increasing throughput is to dedicate one configuration for metal layers and one configuration for adhesion layers (barrier layers perform equally well on either machine). The improvement in metal cycle times for this strategy results in an 8% weighted cycle time improvement for the entire wafer.

The detailed model can be complicated further by the addition of Chamber 5. The two ARC layer metal stack is run again, and the rough model results in Figures 5.12 and 5.13 are repeated for the detailed model, and displayed in Figures 5.16 and 5.17.

Configuration (#degas/orient, #conductor)	1,1	1,2	2,1	2,2	3,1	3,2
Metal 1 Cycle Time (minutes)	61.1	61.1	47.6	47.6	47.6	47.6
Metal 1 Throughput (wafers/hour)	24.6	24.6	31.5	31.5	31.5	31.5
Throughput increase (vs. standard configuration)	0%	0%	28%	28%	28%	28%

Configuration (#degas/orient, #conductor)	1,1	1,2	2,1	2,2	3,1	3,2
Metal 2,3 Cycle Time (minutes)	61.1	61.1	54.8	48.0	54.8	48.0
Metal 2,3 Throughput (wafers/hour)	24.6	24.6	27.4	31.3	27.4	31.3
Throughput increase (vs. standard configuration)	0%	0%	11%	27%	11%	27%

Configuration (#degas/orient, #conductor)	1,1	1,2	2,1	2,2	3,1	3,2
Metal 4 Cycle Time (minutes)	84.8	84.8	76.2	49.6	76.2	49.3
Metal 4 Throughput (wafers/hour)	17.7	17.7	19.7	30.2	19.7	30.4
Throughput increase (vs. standard configuration)	0%	0%	11%	71%	11%	72%

Planned Configuration

Preferred Configuration

Figure 5.14 – Metal Cycle Times for Various Chamber Configurations

Configuration (#degas/orient, #etch, #ARC, #conductor)	1,1,1,1	1,2,1,1	2,1,1,1	2,1,2,1	2,2,1,1	2,2,2,1
Adhesion Cycle Time (minutes)	72.2	72.9	72.9	72.9	57.0	49.7
Adhesion Throughput (wafers/hour)	20.8	20.6	20.6	20.6	26.3	30.2
Throughput increase (vs. standard configuration)	0%	-1%	-1%	-1%	27%	45%

Configuration (#degas/orient, #etch, #ARC, #conductor)	1,1,1,1	1,2,1,1	2,1,1,1	2,1,2,1	2,2,1,1	2,2,2,1
Barrier Cycle Time (minutes)	43.9	43.9	43.9	43.9	43.9	43.9
Barrier Throughput (wafers/hour)	34.2	34.2	34.2	34.2	34.2	34.2
Throughput increase (vs. standard configuration)	0%	0%	0%	0%	0%	0%

Planned Configuration Preferred Configuration

Figure 5.15 – Adhesion/Barrier Cycle Times for Various Chamber Configurations

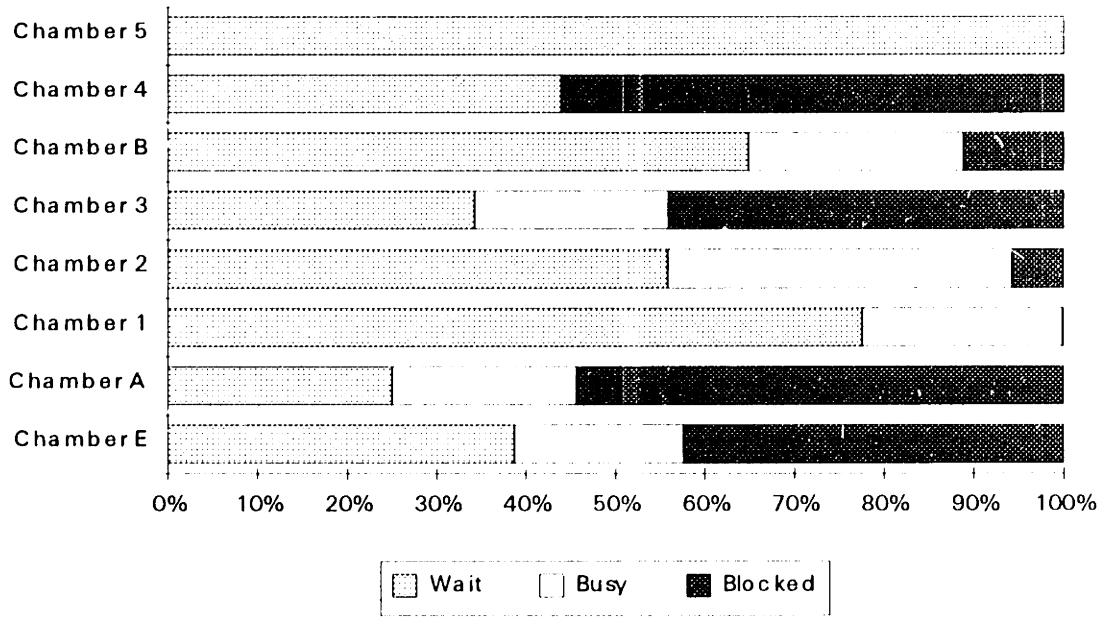

Figure 5.16 – Chamber Utilization using Chamber 5 and Detailed Model

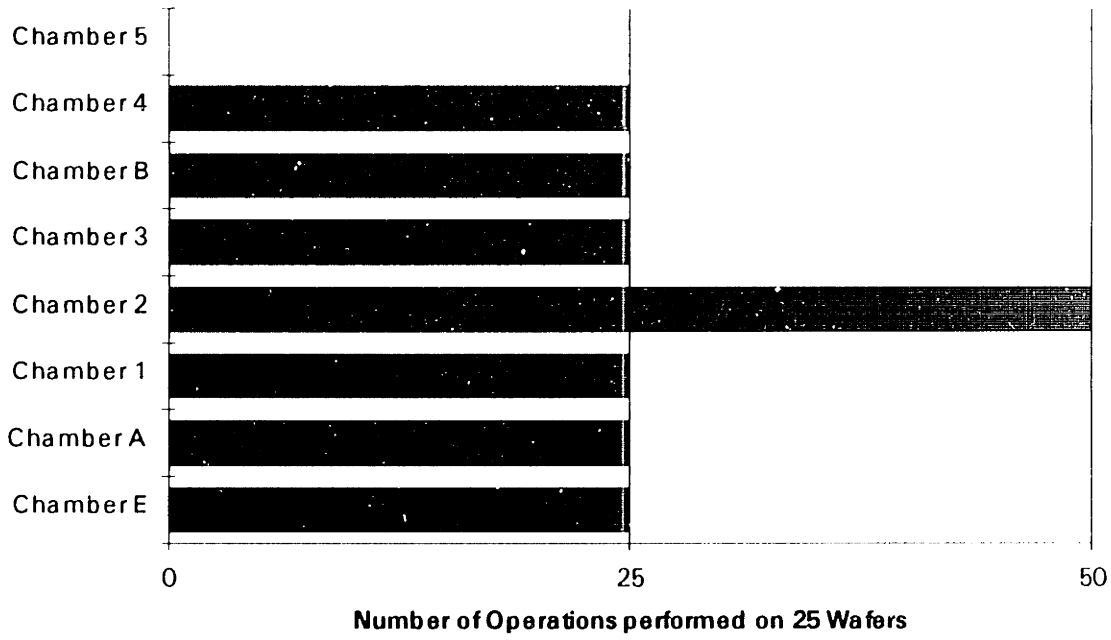

Figure 5.17 – Chamber Operations using Chamber 5 and Detailed Model

Chamber 5 within the Detailed Model

Note that the utilization of each chamber and number of operations in each chamber is consistent. In fact, Chamber 5 is not used – only Chamber 4 serves as a buffer (the software has treated Chamber 4 as the default in this case). This infers that not only are the wafer processing profiles identical, but that the need for two buffer chambers, and hence the risk of cross-contamination, is avoided. Chamber 5 alone may serve as the buffer. Furthermore, the gain on the two single ARC layered metal stacks approaches 31%, while the time penalty for using only one ARC chamber on the two ARC layered stacks is only 13%. This improvement in metal layer processing results in a net improvement of 8% for the present generation fabrication sequences.

Recommendations for Machine Configuration

For the next generation process, an ideal mix of machines would be equal numbers of the metal configuration (including two conductor chambers) and the adhesion configuration (including two anti-reflective coating chambers). Note in Figure 5.15 that the adhesion configuration also has a conductor chamber. Although slower than its dedicated counterpart, this gives the adhesion configuration the ability to run metals, unlike the existing "TNN" configuration, which uses only three PVD chambers: one barrier and two ARC chambers. The adhesion tool would thus be able to help process conductor layers. This benefit is difficult to quantify in capacity planning stages, but could be highly beneficial if operations downstream of metal layers in the completed fab are discovered to be the plant bottleneck. In this case, processing conductor layers takes precedence in order to establish a continuous inventory in front of the bottleneck, to ensure its maximum utilization. Each machine being able to run each layer increases the plant's resilience to Murphy's Law, as the effect of unexpected breakdowns are less pronounced. Thus, the recommended mix is both fast and flexible. The metal configuration also conforms to the ideal configuration for minimizing maintenance downtime discovered in Chapter 4. In general, the redundant chambers of both the metal and adhesion configurations increases the value of operating the machine in handicapped mode.

Comments on Modeling and Simulation

The rough model was built quickly, but as its name implies, gave rough results. However, that step in the modeling process was valuable in that it allowed quick testing of ideas without a high investment in time. Once the rough model indicated distinct gains were possible, it became necessary and worthwhile to spend the time to construct a more detailed and more accurate model. The model also produced results that were impossible

to achieve using static spreadsheet models. Nevertheless, it is important to notice that the use of constraint theory gave valuable first-order indications of where gains could be made, and eliminated the need for many experimental models.

Appendix A – Typical Target/Shield Change Sequence

Chamber 4 Target/Shield Change 7/20/93

Event	*Time (min.)*
CEPT RPR status	2
Retrieve gaskets	1
Turn off RF/DC breaker	2
Switch AMAT to remote	1
Log life, pressure on PM sheet	1
Vent chamber	8
Remove shield	0.5
IPA chamber	2.5
Dispose of shield	2
Install pedestal	2
Install shield	2
IPA ceramic ring	1
IPA/re-install o-ring	0.5
Replace ceramic ring	0.5
Pedestal to process position	0.5
Align shield/pedestal/wafer	1
Install perimeter bolts	2
Install AFS cover/cable	2
Check pedestal/ground R	3
Disconnect P101 cable	1
Start pumpdown	3
Set plasma conditions	1
Create logical wafer	0.5
Start AFS	0.5
Reduce gas flow	1
Remove target foil	1
Unscrew target bolts	4
Remove target	1
Wipedown/treat o-ring	1
Install target	5
Install o-ring	1
IPA target	1
Fill target cooling water	2
Cover target with foil	1
AFS complete	14
Disconnect RF cable	1
Chamber venting	8.5

Remove AFS and wafer	1
IPA shield	1
Close chamber	1
Pump down	8
Start leak check on RGA	2
Leak check	4
Start pump/purge	1
End repair - CEPT	1
Reset target life	1
Pump/purge complete	50
Run RGA	1
Bakeout	240
Post bake	30
Particle pre-read	3
Heater reduced/on	1
Load target burn-in	1
Enter RPR status - CEPT	0.5
Burn-in complete	75
Remove from RPR - CEPT	5
Particle post-read	3
Measure (4-11pt) uniformity	10
Measure (1-121pt) burn-in uniformity	14
Normal - CEPT	1
total minutes	536.5
total hours	8.94

Appendix B – Logic Required for Rough and Detailed Models

Example 1: **Degas/Orient chamber for AMAT Endura 5500**

Rough model Output Logic:

```
PUSH TO ROUTE
```

Detailed model Output Logic:

```
IF ISTATE(cha) = 1 AND (TYPE = met_1 OR TYPE = met_2 OR TYPE = met_3
    OR TYPE = met_4)
    PUSH TO ROUTE
ELSEIF pc_used = 0 AND ISTATE(cha) = 1 AND (TYPE = adh_1 OR adh_2
    OR adh_3 OR adh_4)
    PUSH TO ROUTE
ELSEIF pc_used = 1 AND num_pcii = 1 AND ISTATE((pcii(1)) = 1
    PUSH TO bfr
ELSEIF pc_used = 1 AND num_pcii = 2 AND (ISTATE((pcii(1)) = 1
    OR ISTATE((pcii(2)) = 1)
    PUSH TO bfr
ELSE
    WAIT
ENDIF
```

Example 2: Transfer chamber robot for AMAT Endura 5500

Rough Model Robot Output Logic:

- none -

Accurate Model Robot Output Logic:

```
IF TYPE = barrier
        IF seq = 6 AND num_barrier = 1 AND ISTATE(barrier(1)) = 1
                PUSH TO ROUTE
        ELSEIF seq = 6 AND num_barrier = 2 AND (ISTATE(barrier(1)) = 1
        OR ISTATE(barrier(2)) = 1))
                PUSH TO ROUTE
        ELSEIF seq = 8 AND ISTATE(chb)) = 1
                PUSH TO ROUTE
        ENDIF
ELSEIF (TYPE = adh_1 OR TYPE = adh_2 OR TYPE = adh_3 OR TYPE = adh_4)
        IF seq = 8 AND num_barrier = 1 AND ISTATE(barrier(1)) = 1
                PUSH TO ROUTE
        ELSEIF seq = 8 AND num_barrier = 2 AND (ISTATE(barrier(1)) = 1 OR
        ISTATE(barrier(2)) = 1)
                PUSH TO ROUTE
        IF seq = 10 AND num_arc = 1 AND ISTATE(arc(1)) = 1
                PUSH TO ROUTE
        ELSEIF seq = 10 AND num_arc = 2 AND (ISTATE(arc(1)) = 1 OR
ISTATE(arc(2)) = 1)
                PUSH TO ROUTE
        ELSEIF seq = 12 AND ISTATE(chb)) = 1
                PUSH TO ROUTE
        ENDIF
ELSEIF (TYPE = met_1 OR TYPE = met_2 OR TYPE = met_3 OR TYPE = met_4)
        IF seq = 6 AND num_barrier = 1 AND ISTATE(barrier(1)) = 1
                PUSH TO ROUTE
        ELSEIF seq = 6 AND num_barrier = 2 AND (ISTATE(barrier(1)) = 1 OR
        ISTATE(barrier(2)) = 1)
                PUSH TO ROUTE
        IF seq = 8 AND num_conductor = 1 AND ISTATE(conductor(1)) = 1
                PUSH TO ROUTE
        ELSEIF seq = 8 AND num_conductor = 2 AND (ISTATE(conductor(1)) = 1 OR
        ISTATE(conductor(2)) = 1)
```

```
            PUSH TO ROUTE
        IF seq = 10 AND num_arc = 1 AND ISTATE(arc(1)) = 1
            PUSH TO ROUTE
        ELSEIF seq = 10 AND num_arc = 2 AND (ISTATE(arc(1)) = 1 OR
ISTATE(arc(2)) = 1)
            PUSH TO ROUTE
        ELSEIF seq = 12 AND ISTATE(chb)) = 1
            PUSH TO ROUTE
        ENDIF
ELSEIF
        WAIT
ENDIF
```

Detailed Model Actions on Start:

```
IF path(seq-1) = cha
        IF path(seq+1) = barrier
            xfr_ct = 23
        ELSEIF path(seq+1) = arc
            xfr_ct = 25
        ELSEIF path(seq+1) = conductor
            xfr_ct = 25
        ELSE
            xfr_ct = 24
        ENDIF
ELSEIF path(seq-1) = barrier
        IF path(seq+1) = arc
            xfr_ct = 23
        ELSEIF path(seq+1) = conductor
            xfr_ct = 24
        ELSEIF path(seq+1) = chb
            xfr_ct = 25
        ELSE
            xfr_ct = 24
        ENDIF
ELSEIF path(seq-1) = arc
        IF path(seq+1) = barrier
            xfr_ct = 23
        ELSEIF path(seq+1) = conductor
            xfr_ct = 23
        ELSEIF path(seq+1) = chb
            xfr_ct = 25
```

```
        ELSE
                xfr_ct = 24
        ENDIF
ELSEIF path(seq-1) = conductor
        IF path(seq+1) = barrier
                xfr_ct = 25
        ELSEIF path(seq+1) = arc
                xfr_ct = 23
        ELSEIF path(seq+1) = chb
                xfr_ct = 25
        ELSE
                xfr_ct = 24
        ENDIF
ENDIF

seq = seq + 1
```