Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Discrete-event simulation of wafer fabrication facility

Karen A. Pitts
David Sarnoff Research Center
Subsidiary of SRI International
Princeton, N.J. 08543-5300

ABSTRACT

Using discrete-event simulation (process interaction approach with SIMAN), we modeled the product flow through a research and experimentation facility for wafer fabrication. The model was verified using past history; it predicts typical cycle times and average operations/day closely. Experiments were run with the model to isolate bottlenecks, to determine how best to alleviate the bottlenecks, and to determine the effect of changing system parameters, such as people, equipment, loading, failure rates on equipment, and hours of operation (add a second shift).

1. INTRODUCTION

1.1 Problem Description. The IC (Integrated Circuit) Center is a facility for research and experimentation in wafer fabrication at David Sarnoff Research Center. The IC Center was established in 1968 and from the beginning users have complained work takes too long. IC Center personnel feel they are doing the best they can given a one-shift operation, no redundancy on many critical pieces of equipment, and limited people resources. Various management strategies, as well as advice from outside consultants, have been tried in the past, with little improvement in cycle times, especially for complete device lots.

The following model was developed to analyze what parameters were critical to cycle time and to provide management with a tool to decide how to spend money to improve the Center. For this model, I used the process-interaction approach. The specific software is SIMAN (SIMulation ANalysis).

1.2 Model Overview. Product flow through the IC Center is a complex queueing system, where lots wait at various places for resources, either people or machines. Specifically, the IC Center is a complex job-shop, where arriving lots require the various process areas, in a variety of orders and with differing number of individual process steps (some lots require three steps, some 500). An integrated circuit is formed by repeating a core of steps: growing or depositing a layer, patterning the layer with photolithography, and transfering the photoresist pattern to the layer by etching. The process areas, or work stations, in the model are furnace (grow an oxide), poly-nitride (deposit poly or nitride), other depositions (glass and metal), photo (pattern the layer), ion implant (implant the layer), plasma etch (etch the layer with a plasma), wet etch (etch with a chemical bath), testing, and a miscellaneous step, which represents items such as special experiments, adding controls, information about the experiment, and inspecting a lot; the miscellaneous step is called "tracker" since it is typically performed by wafer trackers (people responsible for expediting the movement of lots through the IC Center).

In addition to the job-shop model of routing lots to various work stations, I modeled machine breakdowns, people absences (illness or vacation), and a one-shift operation with a break for lunch.

A block diagram of the simulation is below.

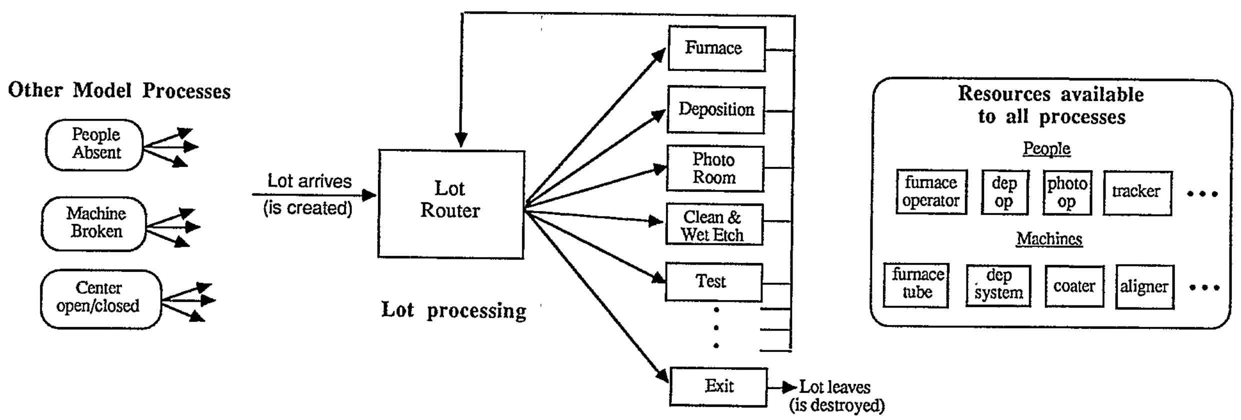

Figure 1. Block Diagram Of Simulation Model Of IC Center

The model is sizable: 2000 SIMAN blocks, 80 experimental parameters, 13 stations, and 26 attributes for each lot.

2. SUMMARY

• The model has been verified (compared to actual operation). The model matches some parameters, such as average operations/day and average completion times closely. The model fails to predict the lots (especially shorter lots) which have long cycle times. The model also underpredicts work in progress.

• Several model changes to improve verification were made, mostly adding more detail or making the model match actual operation more closely. None of these changes improved model predictions.

• The current system has two bottleneck areas, photo, especially the stepper, and plasma. Both of these areas are typically overloaded in a wafer fab since steppers cost $1,000,000 and plasma machines (and you need several) run $500,000 to $1,000,000. In addition, the model predicts occasional problems in the furnace area. Adding a two machines, a plasma etcher and a stepper, and two operators, plasma and stepper, allows substantially more capability.

• The sensitivity of the bottlenecks to changes in parameters is as follows. The stepper is close to being overwhelmed; an increase in the lots to the stepper or an increase in the complexity of stepper lots will cause queue times to explode. Adding a plasma operator significantly reduces cycle time as well as reducing queue times in plasma. Adding a stripper (a plasma etcher to strip photoresist) significantly reduces stripper time, but degrades times for the other plasma machines.

• Several other parameters were tested for sensitivity. Adding more wafer trackers (who do about 65% of the processing steps) does not reduce queue times or cycle times. Changes in many of the processing times had little impact. Decreasing the number of lots in the system does not improve cycle time. Adding redundancy in equipment and people beyond the areas listed above does not improve cycle time. Vastly reducing failure rates helps a little. *By far the biggest improvement comes from increasing the number of hours the center is in operation!*

• Even though SIMAN is marketed heavily for PC's, it is capable of handling models of this size, although I was constantly increasing SIMAN array dimensions. New parameter settings took about one hour to evaluate on a VAX 8800. The most useful SIMAN commands were SEQUENCE, ROUTE, STATION, and SYNONYM. I found the lack of IF-THEN-ELSE frustrating; my code branched and branched and branched.

3. MODEL DESCRIPTION

The model resources for the IC Center model are people and machines. The people and machines for the work areas (modeled as work stations) are:

Furnace:	operator, repair person, furnaces, poly-nitride systems
Deposition:	operator, repair person, dep systems
Photo:	stepper operator, scanner operator, photo repair person, primer, coater, aligner, developer, postbake oven
Plasma:	operator, repair person, plasma machines
Ion Implant:	operator, repair person, implanters
Clean/ wet etch:	wafer trackers, clean/test/ microscope repair, cleaning stations
Test:	test equipment (trackers)
Tracker:	microscopes (trackers)
System:	system repair person

The model of the current operation has forty-eight machines and twenty-seven people.

Processes and their entities are (1) lot creation, routing, and processing; (2) machine breakdowns; (3) people absences; and (4) providing for shifts (close the center down over lunch and overnight), which I've called daily scheduling.

3.1 Entities. The entities are lots, machine breakdown, people absences, and the scheduler, respectively. Entities representing lots are temporary entities: they are created when a lot arrives to the system, remain while the lot is processed, and are destroyed when the processing is complete. The last three categories, machine breakdowns, people absences, and the scheduler are permanent entities; they are created at the start of the simulation and remain for the duration of the simulation. There is one entity for each person (so they can be sick or go on vacation) and for each machine (so the machine can break).

Typical attributes for a lot include lot number, current process step, total number of process steps, which tracker is assigned to the lot, whether special equipment is required, and so on. Attributes for machine breakdowns include which failure distribution and downtime distribution is used during the simulation, time of the next breakdown, repair time, and so on.

3.2 Processes. For all lot processing, an entity (a lot) will arrive at a station, seize both a machine and an operator, be processed, and move on to the next station. For people absences, the absence entity typically preempts the person, and delays for the absence. For machine breakdowns, a machine is preempted, then a repair person is seized to complete the repair. The daily scheduler uses the SIMAN command ALTER to make operators available for work at the start of work periods, and unavailable over lunch and overnight.

3.3 SIMAN Comments. More complex (or new) SIMAN commands which I used heavily are SEQUENCE, ROUTE, and STATION, which are especially useful for lot routing and processing, and SYNONYM, because I don't like variable names such as A(1) and X(22). I did have some problems, which yielded the following recommendations:

1) use ALTER in *only one* process for any resource
2) minimize the use of PREEMPT
3) keep your SYNONYMS short
4) limit the number of branches for any BRANCH block to 10

Systems Modeling recommended 2) and 4) to me when I started tripping over system bugs.

4. VALIDATION AND VERIFICATION

The model has been tested extensively for validation (does the model do what we want it to?). At this time there are no known bugs. It does have some idiosyncracies, but there are no serious problems. In verification (does it match reality?), I checked for transient behavior as well as matching actual operation. Since SIMAN allows me to load the model up at the beginning, the model reached a steady state very quickly. For verification, I used two years of previous lot processing history.

4.1 Cycle Times. The table below compares actual cycle times to cycle times predicted by the model. Twenty-five percent indicates 25% of the lots are completed in the given time; 50% corresponds to half the lots completed in the given time, 75% to three-fourths, and so on. Note that the 25% and 50% match closely, although the model tends to predict on the low side. The times in the 75% to 90% range are frequently not even close, typically underpredicted by several weeks. We tried to correct this by changing the model (occasionally adding long times for processing to mimic when something goes wrong), adding a logical sequence of steps, and updating failure rates. None of the changes helped. Histograms of the first three groups follow in Figure 2.

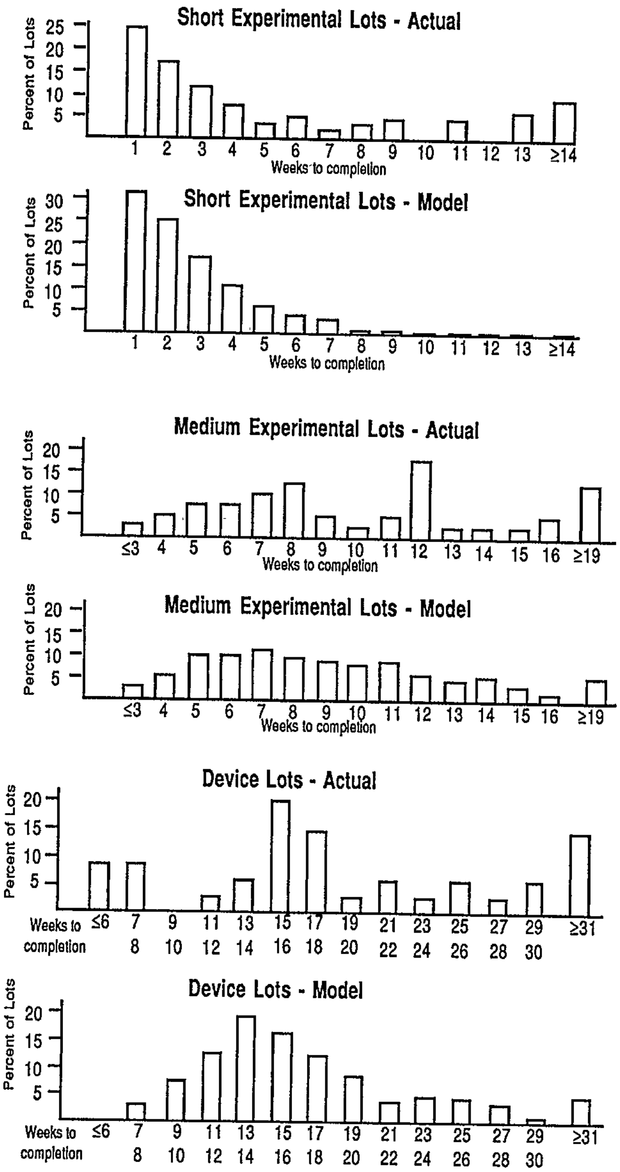

Figure 2. Histogram Comparison Of Cycle Times: Model Vs. Actual

Table 1: Cycle Time Comparison

Type of lot (number of steps)	Percentage (of lots completed)	Actual Completion (in weeks)	Model Prediction (in weeks)	Model error (in weeks)
Short experimental	25	1	1	0
(3-55 steps)	50	3	2	-1
(64% of the lots)	75	8	4	-4
	85	11	5	-6
Medium experimental	25	7	6	-1
(30-90 steps)	50	9	9	0
(18% of the lots)	75	12	12	0
	90	20	15	-5
Device lots	25	13	13	0
(93-167 steps)	50	17	15	-2
(15% of the lots)	75	25	20	-5
	90	31	26	-5

4.2 Number Of Operations/Day. Another measure which correlates well is average operations/day. The table below compares actual to model predictions. I have more work going through photo and less work going through trackers; the other numbers are very close.

Table 2: Average Operations/Day

Step or area	Actual	Model
Furnace	7-8	7
Deposition	2.5-3	2.9
Photo	9-11	15
Plasma etch	5.5-6.6	5.5
Implant	1.7	1.7
Clean/wet etch	7-9	7.5
Test	3	2.9
Tracker (misc.)	21-27	20.9

4.3 Work In Progress. We don't have good numbers on the typical number of active lots in the center, but the general feeling is the numbers of active lots are typically between forty and sixty. The model configuration which I used for validation has an average of thirty-five lots in the system, with a minimum of thirteen and a maximum of sixty-seven.

5. MODEL EXPERIMENTATION

In all experimentation, for each parameter combination I simulated two years of operation, with five independent simulations (ten total years of simulated time). I then compared the averages from the five simulations for one set of conditions to the averages from the five simulations for other sets of conditions. Statistical significance was determined using general linear models and Tukey's means comparison using SAS's PROC GLM.

5.1 Add People To Bottlenecks (Photo And Plasma). I compared two plasma operators to one plasma operator and four photo operators to three photo operators (add a substitute). The additional person at plasma significantly reduces cycle time; the person at photo does not (the stepper queue problem is equipment, not operator availability). Figure 3)a and 3)b contain the box plots; the horizontal axis displays the change (add a person, '3 operators' vs. '4 operators'); the vertical axis displays the measure (time in queue, 'Hours in Stepper Queue').

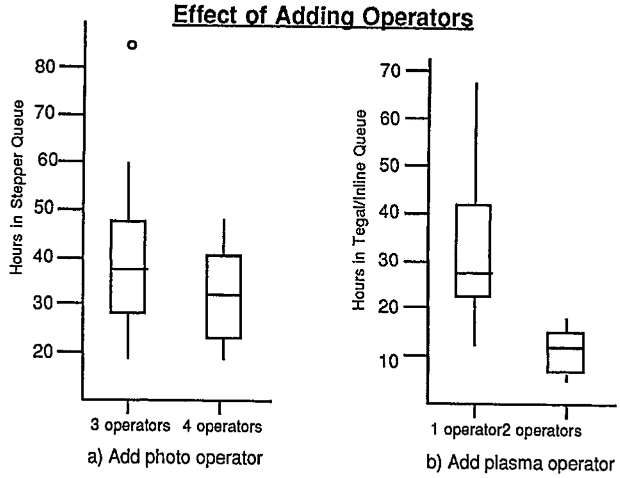

Figure 3: Add People To Bottleneck Areas

5.2 Add Both People And Equipment To Bottlenecks. Adding a stepper, stripper (plasma etcher), stepper operator, and a plasma operator significantly reduces device cycle time from 11.6 weeks to 10.5 weeks (given the level of variability from simulation run to simulation run, device cycle times should differ by one week to be statistically significant). Queue waiting times for the stepper and the stripper are drastically reduced; times for other machines in the same areas, however, are increased. See Figure 4 and 5 for box plots. Again, the horizontal axis is the change to the system; the vertical axis is the measure for comparison. Adding people <u>and</u> machines produces a marginal improvement in cycle time, although the station queue times are improved. The real benefit to additional people and machines comes with adding more load to the center. The current system cannot handle many more lots entering the center without cycle time degradation.

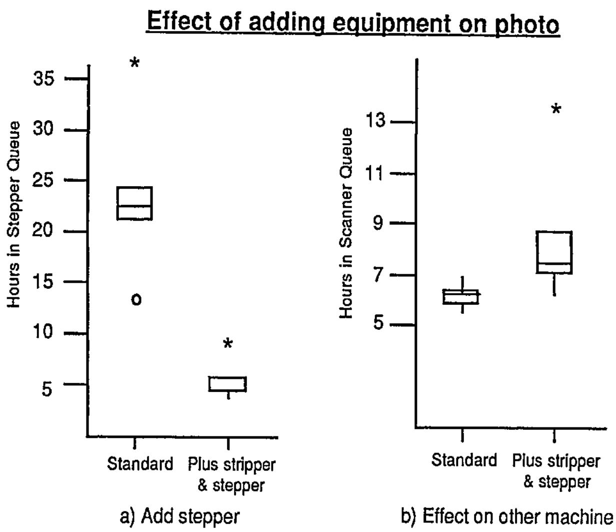

Figure 4: Add Equipment To Photo

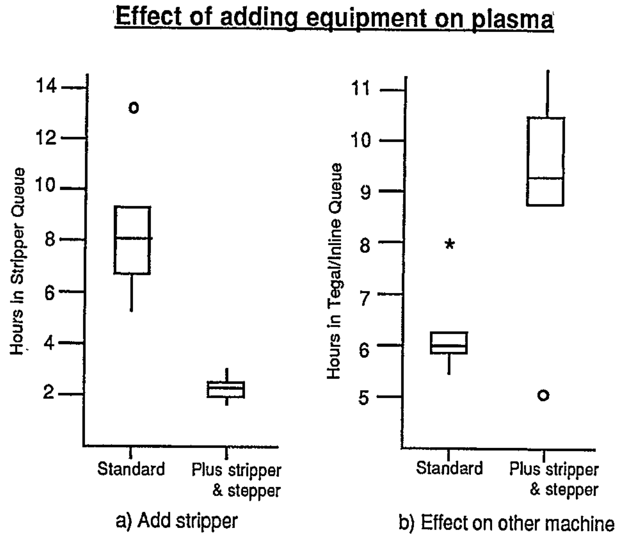

Figure 5: Add Equipment To Plasma

5.3 Does More Redundancy Help? After adding the two people and two machines to alleviate the bottlenecks, I checked whether more redundancy helped. I added more people (eight additional), more machines (twenty-one additional), and both more people and machines. The additional people helped more than the equipment, neither made much difference. *The way to get shorter cycle times for device lots is not a room twice the size.* Figure 6 illustrates the effect of redundancy.

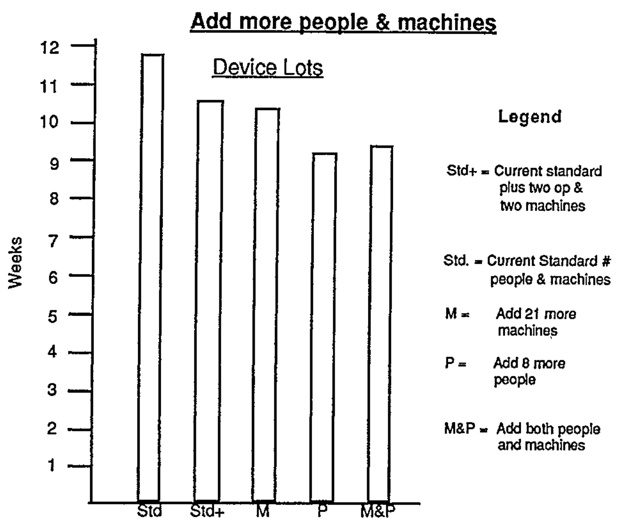

Figure 6: The Effect Of Redundancy

5.4 Change Load In The Center. With the current configuration, increasing the number of lots arriving lengthens cycle times, especially device lots. Decreasing the load does <u>not</u> improve cycle times significantly. With the addition of a stripper, plasma operator, stepper, and stepper operator, cycle time doesn't degrade as quickly when load on the center increases. Changes in cycle time are graphed in Figure 7; the horizontal axis is the proportion of current load, i.e., .5 corresponds to half as many lots arriving, 1 corresponds to current load, 2 corresponds to twice as many lots arriving. The box plot for changing load for a system with additional people and equipment is located in Figure 8, with increases in load on the horizontal axis and device cycle time in weeks on the vertical axis.

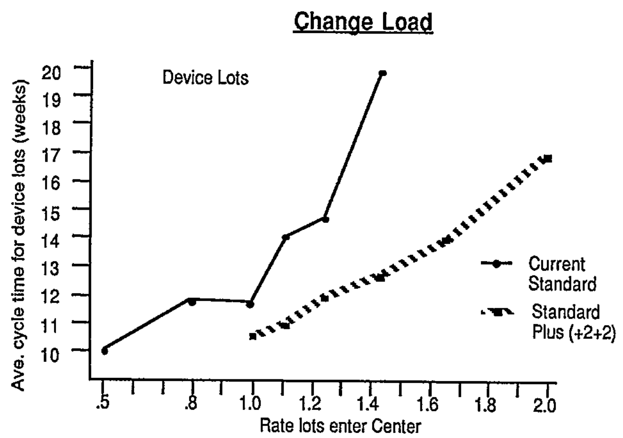

Figure 7. Effect Of Increasing Or Reducing Load In The Center

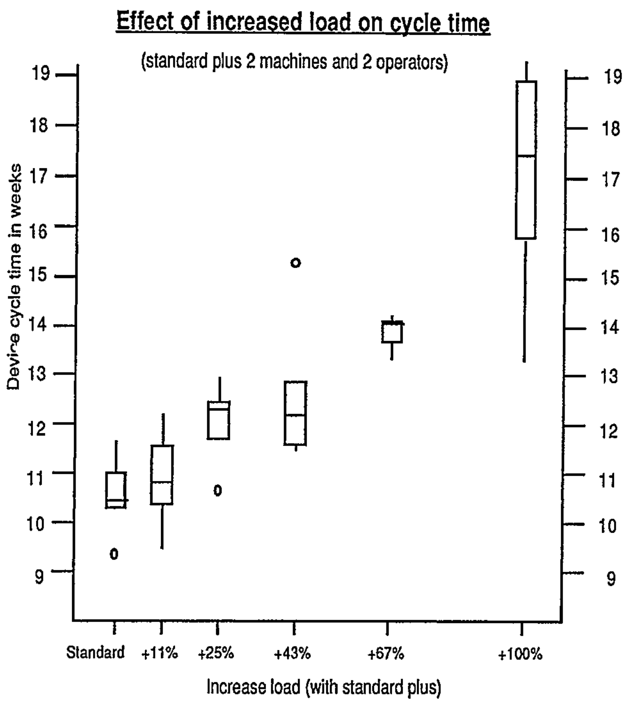

Figure 8: Box Plot Of Cycle Time With Additional Load

5.5 Change Failure Rates. Simplistically, I increased or decreased failure rates on all equipment. One might want to pursue the effect of selectively increasing or reducing failure rates. Where we have redundancy (and no bottlenecks), changes in failure rates will probably have little effect. For reduced failure rates, the device cycle time improves from 11.6 weeks to 9.7 weeks; for increased failure rates, device cycle times explode. Figure 9 contains the box plot, with changes in the failure rates on the horizontal axis and device cycle time (in weeks) on the vertical axis.

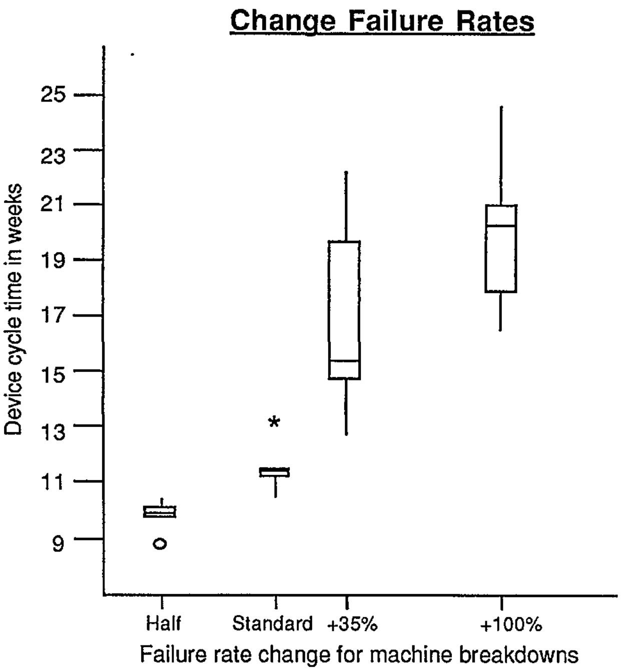

Figure 9: Effect Of Changes In Failure Rates

5.6 Lengthen The Work Day. To check how sensitive cycle time is to the number of working hours, I gradually lengthened the work day from a one-shift operation to a three-shift operation. The results are illustrated in Figure 10, with increases in shift operations on the horizontal axis and device cycle time (in weeks) on the vertical axis. The averages are below.

All cycle times are statistically different from each other except difference between 66% increase and two shifts. The minimum significant difference for device lots is about one week. For improving cycle time, nothing beats lengthening the hours the center is in operation. Figure 10 contains a graph of various improvements: reduce failure rates, reduce load, add many people and machines, lengthen the hours. The y-axis of Figure 10 is proportion improvement over current cycle time for device lots, e.g. .20 corresponds to a 20% improvement in cycle times.

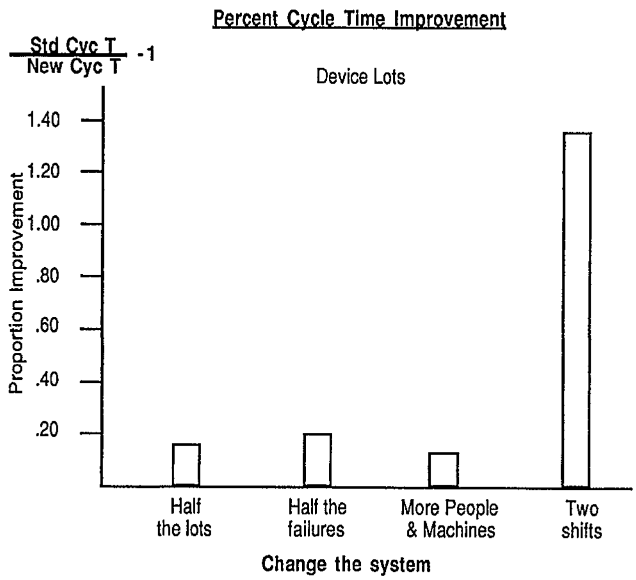

Figure 10. Comparison Of Several Changes

5.7 Add load to second shift. Since adding hours shows clear benefits, I investigated how much load a two-shift operation could handle. If we double the load with a second shift, do we get the same cycle times as we have in a one-shift operation with the current load? As you can see in the Table 4 and Figure 11, a two-shift operation has considerably more capability for handling lots; recall that current cycle time for device lots is 11.6 weeks. Minimum statistical differences are again about one week for device lots.

Table 3: Impact of Lengthening Work Day

Measure	One shift	33% inc.	66% inc.	Two shifts	Three shifts
Device cycle time in weeks	11.6	7.5	5.6	4.9	2.6
Stepper queue time (hours)	23.7	9.0	4.1	3.2	≤ 1 hr
Plasmatherm queue time (hours)	13.5	9.7	6.2	4.3	≤ 1 hr
Stripper queue time (hours)	8.3	4.2	1.8	1.8	≤ 1 hr
Poly/nitride (hours)	12.4	8.4	5.6	4.0	≤ 1 hr

Table 4: Add Load To Two-shift Operation

Load on Center	Cycle Time	
Current load	4.7	|
Add 25%	5.0	|
Add 50%	5.2	|
Add 75%	5.4	|
Add 100%	5.4	|
Add 150%	7.9	
Add 200%	9.5	

Averages connected by a line are not statistically different. Adding 100% is the same as twice the current load (twice the number of lots enter the Center each week). Adding 200% is the same as three times the current load (triple the number of lots enter the Center each week). With a two-shift operation, triple the load is not as bad as current cycle times. There are clear advantages to adding a second shift.

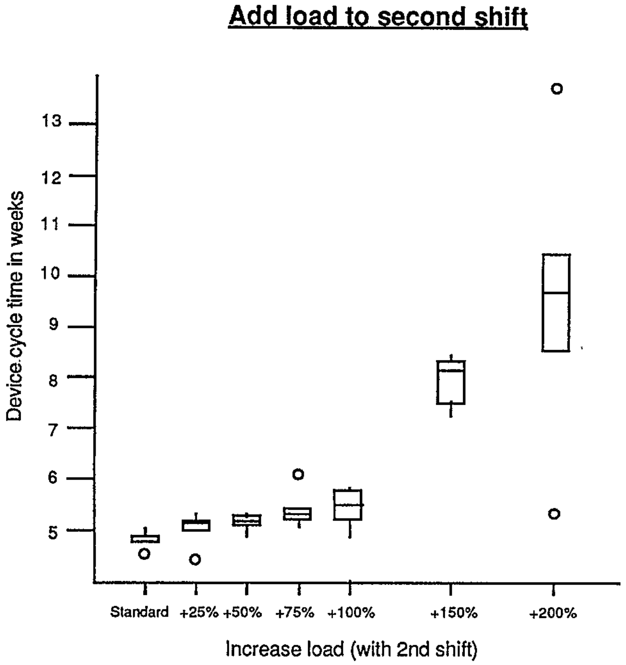

Figure 11: Increase Load With Two-shift Operation

ACKNOWLEDGEMENTS

Many people in the IC Center contributed to collecting the data and developing the model. I would like to especially thank Mike Leahy and Doris Flatley for their time and patience, as well as Harold, Justine, Verne, Regina, and Bill, who showed me the facility and explained their work areas. I would also like to thank Rod Borchardt, system administrator for the VAX where SIMAN is installed, who is always helpful in solving system problems, of which I've had plenty.

REFERENCES

Fishman, George (1973). *Concepts and Methods in Discrete Event Digital Simulation*, Wiley

Fishman, George (1975-1976). *Concepts in Discrete Event Digital Simulation*, manuscript and class notes

Pegden, C. Dennis (1986). *Introduction of SIMAN with Version 3.0 Enhancements*, Systems Modeling Corporation

BIOGRAPHY

KAREN A. PITTS has been a member of the technical staff at the David Sarnoff Research Center since July, 1976. There she has worked on a variety of operations research and statistical projects in consumer electronics, solid state processing, and telecommunications; her interests include discrete-event simulation, experimental design, statistical graphics, and subjective viewer tests. She received a B.A. in mathematics from St. Olaf College in 1974 and an M.S. in Operations Research and Systems Analysis from Univerisity of North Carolina at Chapel Hill in 1976. Ms. Pitts is a member of ASA and ASQC.

David Sarnoff Research Center
Subsidiary of SRI International
CN 5300
Princeton, N.J. 08543-5300
(609) 734-3045