Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

ASSESSMENT OF POTENTIAL GAINS IN PRODUCTIVITY DUE TO PROACTIVE RETICLE MANAGEMENT USING DISCRETE EVENT SIMULATION

Sungmin Park
John Fowler Ph.D.
Matt Carlyle Ph.D.

Arizona State University
Tempe, AZ 85287, U.S.A.

Matt Hickie

Motorola MOS 12
Chandler, AZ 85224, U.S.A.

ABSTRACT

Photolithography is often the constraining equipment in semiconductor wafer fabrication plants due to the number of times the product must process through it. Modern day photolithography is performed on a cluster tool that is a combination of a stepper and track. It is obvious that the combined availability of the cluster tool is critical to throughput, but what is not so obvious is the throughput restriction from a secondary constraint known as a reticle. Every layer of a product needs a unique reticle for processing. Setup issues arising from the requirement of reticles affects productivity of photolithography and the entire wafer fabrication line if photolithography is the bottleneck. Efficient management of reticles (with regard to setup and storage on a stepper) based on current system status provides a strategic and tactical advantage. In this paper, a SLAM discrete event simulation model is to mimic the setup and storage of reticles. This enables the collection of information that can used to identify potential gains in tactical reticle management. The simulation model will be explained in detail along with output results for the tactical issues. The relationship between the simulation and the network flow model for proactive reticle management will also be discussed.

1 INTRODUCTION

Complexity in the manufacturing process for semiconductor fabrication plants causes many headaches in the world of scheduling. Without an easy solution for many of the problems, one important area to focus on for improvement is the fabrication plant bottleneck. The definition of a bottleneck is “any resource whose capacity is equal to or less than the demand placed upon it” (Goldratt 1986). The throughput of a manufacturing line is directly related to the bottleneck’s throughput.

Photolithography is often found (and sometimes planned) to be a bottleneck because the equipment is very expensive and the process is performed multiple times on product. This reentrant flow is typical in semiconductor wafer fabrication processing.

If photolithography is the bottleneck, then any improvement in the efficiency of this operation will help the fab achieve a higher throughput rate. Much effort has gone into increasing the availability of the photolithography tool. This tool is often a combined track and stepper clustertool. The focus on availability helps increase tool throughput, but there is still a secondary constraint known as a reticle that can limit tool utilization, regardless of increased tool availability.

Reticles are required to be on the tool to run product through the photolithography machine. The reticle constraint is a permanent resource constraint, which means it is required on the machine for the complete processing of the product. Reticles are also product and layer dependent which means that the quantity of required reticles for a production fab will depend on the number of layers per product and the number of products. This typically results in hundreds if not thousands of required reticles for a production wafer fab. Because of this, the reticle storage and setup issues are critical to reducing cycle time and increasing throughput. This paper will detail the reticle management issues faced by a production wafer fab. It will also detail a SLAM discrete event simulation model and how it is use to evaluate strategic reticle load scenarios. It will also explain what simulation data can be fed into a network flow model for the purpose of providing a methodology to proactively manage reticles.

2 PHOTOLITHOGRAPHY

The Photolithography process is a 3 step process that first consists of coating the wafer uniformly with a material known as photoresist. After coat comes the transferring of the reticle pattern into the photoresist by projecting light through the reticle while it is placed over the wafer. This second process, referred to as exposing, is repeated or stepped over all circuits on the wafer. The third and final process after coating and exposing is developing the exposed photoresist. This process removes the polymerized sections of the photoresist from the previously exposed wafer. After this, the photolithography process for the layer is complete and the wafers will continue for further processing on other equipment.

The reticle required during the exposure process is a piece of glass with an enlarged pattern on it. A different reticle is required for every layer of a product's flow. Each time a layer is run, the reticle for that layer of that product must be setup on the tool. Typically more advanced logic products require 15-30 layers to produce. This means 15-30 different reticles are required to be setup for a specific product. This number multiplied by the different products being manufactured at a fab, quickly cause the number of required reticles and setups to be very large.

3 RETICLE MANAGEMENT ISSUES

To better understand the scope of the constraining reticle resource and hence, the model being used to simulate photolithography, it is worth detailing the different requirements and constraints that the reticles must adhere to in the wafer fabrication plant. One main issue to evaluate is how a wafer plant's product diversity drives the quantity of reticles required in production. This will be explained with a generic example of a medium size, microcontroller wafer fabrication plant. The number of reticles in a plant will highlight the issue of reticle storage. It will also highlight the issue of reticle setup and reticle inspection issues. After explaining these, the strategic and tactical issues for minimizing the effect of reticle fixture constraints will be illustrated.

3.1 Reticle Requirements in a Wafer Fabrication Plant

The reticles for a certain *product A* are typically referred to as reticle *set A*. The reticles inside *set A* are typically referred to by the layer number. For this work, the layers will be numbered numerically from 1 to the maximum layer for the product. However, this is typically not the layer numbering method used by a wafer fab.

A conservative estimate for an average size production fab would be that it has the capability to produce 80 different products. Again, each product will require a unique set of reticles to produce each layer of the product. For this case, we will assume that there is only one set of reticles for each product (no duplicate reticle sets).

Each of the 80 different products will require a certain number of layers. To keep things simple in understanding the quantity of reticles required, an average of the number of layers for the 80 products will be estimated. In this case, the estimate will be that the average number of layers per product is 17 for the 80 different products. The multiplication of 80 and 17 gives us 1360 different, unique reticles that need to be stored in this average fab example.

3.2 Reticle Storage

Reticle storage is accomplished through two methods. The first method is storage on the actual photolithography machines. Most modern photolithography machines have some amount of storage located on the actual machine. Storage on the machine is something that is beneficial in two ways. First, reticle storage on a machine is considered a safe place to put a reticle for storage because it is a self-contained clean space where the reticle will be protected from the any human handling and particles in the manufacturing environment. Second, reticle storage on the machine means that no human interaction is required to load that reticle on the machine, which results in no human required setup.

A general configuration for a machine with reticle storage is that is has 6-14 slots for reticles. For this example, it will be estimated the manufacturing line has 20 photolithography machines that store 10 reticles each. This results in 200 total storage slots on the machines. Unfortunately, subtracting 200 from the required 1360 reticles still causes a shortage of 1160 storage slots. These 1160 reticles are typically stored in the wafer fabrication clean room in what is referred to as a reticle stocker, which is the second method of storing a reticle. This stocker is a self-contained, clean environment electronic closet from which reticles can be stored and from which reticles can be retrieved.

Another constraint for photolithography can occur depending upon the method of storing the reticles in the machine. The general assumption for storage on a machine would be that any reticle could be stored on any of the storage slots. In this case, the reticles are free to be put in any slot and the photolithography machine will acquire that reticle when it is time to run the wafers for that reticle. This is referred to as the *free reticle assignment* case.

Another storage method is that each reticle must be programmed on each machine to be located in a specific storage slot. When the machine is ready to process wafers for a specific reticle, it determines the slot number for the associated reticle and it utilizes the reticle in that slot only. This is a bit rudimentary, and the more advanced machines will scan the barcode of the reticle in the storage slot number to check if it matches the required reticle id. This is referred to as the *fixed reticle assignment* case. With this case, there is an obvious need to balance the reticle layers assigned to the reticle storage slots to avoid having some reticle storage slots over utilized while other slots are under utilized.

3.3 Inspection and Setup

From the above information, it is evident that there is not enough storage in the actual photolithography tools for all of the reticles needed to run products in a wafer fab. This results in moving of reticles from the stocker to the machine, from the machine to the stocker, and from a machine to another machine. Since the reticle contains the pattern that is exposed onto the wafer, it is very important that the reticle be clean (i.e. have no particles on it) and that the pattern on the reticle be unaltered. Unfortunately, the need to move the reticles for setup on machines creates possible opportunities for contamination to get on the reticle and possible opportunities for damage to the reticle pattern from electro-static discharge (ESD) or general wear and tear from physical handling.

To ensure reticle quality, reticles can be inspected for particles and pattern integrity after being transported from storage to a machine and from a machine to another machine. This inspection unfortunately takes time and resources to perform, and if not done ahead of time, can cost cycle time at the photo process because material will have to wait while the reticle is inspected and then physically loaded into the storage slot. It is apparent that the less moves a reticle experiences, the less inspections and the less setups it will require. The benefit from this is increased throughput from the tool, less cycle time for the material being processed, and fewer resources to inspect the reticles.

3.4 Strategic and Tactical Issues

Strategically, the largest opportunity for improvement is identifying when to purchase an extra set of reticles for a product whose demand makes up a large percentage of the plants products. Reticles are not inexpensive fixtures. A single reticle costs anywhere from five to ten thousand dollars. A whole set of reticles for the example product with 17 layers would therefore cost around 85-170 thousand dollars. This is costly enough to make having too many reticles a waste of money. Conversely, not having enough reticles could limit the capacity of the photolithography area. For a manufacturing line it is tragic to let something that represents only a small fraction of the around 5 million dollar cost of the photolithography tool, limit capacity and cause increases in product cycle time. There exists a trade off between spending money on extra reticle sets and making sure there are ample reticle sets available to not increase product cycle time. Discovering the relationship between start rate product mix and cycle time through based on a set number of reticles provides a strategic manufacturing advantage and therefore, a potential gain.

Tactically there is gain related to reducing the cycle time lost to reticle setups. A machine that has the ability to store “X” amount of reticles could avoid some setups on reticles if they remain in the storage on the tool. Specific to this, a large opportunity exists for the fixed reticle assignment case. Assigning specific reticles to a slot, highlights the need to make sure the assignment is appropriate for the product mix. Another opportunity is in the area of load leveling of the product waiting for reticles in storage on a machine. It is very probable that one machine could end up with a disproportionate amount of lots in queue waiting for the reticles stored on that tool while another machine could be idle because it does not have the correct reticles stored in it. The potential here is to figure out a method to avoid this imbalance and at the same time minimize setups.

4 SIMULATION MODEL

The primary reason to use a discrete event simulation for such analysis of this system is because the reentrant behavior of the photolithography process conceptually has the same context of the recursive logic of discrete event procedures. This discrete event simulation model was built for two purposes. The first is to provide a tool to evaluate the performance measures of the photolithography process that are associated with current management of reticles. Specifically, we are interested in utilization, setup time, and product cycle time. The second purpose is to utilize the future production data available from the simulation event calendar to show potential improvements in the photolithography performance measures by using proactive reticle management. This specifically focuses on the tactical reticle balancing issue.

4.1 Model Assumptions

The simulation model was built with the following assumptions. Wafers are simulated from start of processing to completion for the product flow. The wafers are processed and transported in groups called lots. For simplicity, each lot consists of the same number of wafers, and only the photolithography processing steps are modeled with capacity constraints. The other processing steps are modeled simply as exponentially distributed cycle time delays.

For the capacitated photolithography step, the decision was to model the processing time of one lot of any product at any layer as a 1 hour constant. This assumption does not quite match reality as different products have different photolithography requirements that affect the cycle time. Two of these requirements are the number of steps needed to expose the wafer and the exposure energy. Regardless, it was decided that generalizing the cycle time to a fixed time per lot would not adversely affect the model with regard to the reticle management issues being investigated.

There were 3 exponential delay times identified for the processing steps other than the photolithography step. The first delay is the time the lot takes upon entry of the lot into the system to reach the first photolithography step. The second delay is the time it takes after completing the first photolithography step until the next photolithography step. This delay represents the time between photo layers for the whole flow of the product. The third delay is the time it takes from the last photolithography step to the completion of the product. Each reentrant delay is modeled as an exponential distribution. The above cycle time delays were created as variables to be assigned at the product family level. Table 1 shows the delays used for each product in this model.

The wafer product portfolio was modeled to match real world wafer fab conditions. The plant was modeled to be able to process 2 distinctly different technologies. This is very often the case as a fab is ramped up with one product technology, and as time passes, new technologies emerge that must be run in the fab.

Each of the above 2 technologies is broken down into 4 families. This gives 8 different product families for the model. Each product is representative of differences within the technology. For instance, one difference might be in the number of metal layers for the product. These differences will cause the number of photo layers to change between the families and this is why each family has a different required number of reticles.

Within each of the 8 families, there are 12 different products. Each product in a family has the same number of photo layers and the same delay times. What differs among the products in the family is that the products within a family each require different reticle sets. This means that each layer of each product in the family needs a unique reticle to be processed at a layer. The difference in the reticles would be analogous to different performance characteristics for the product required by a broad range of customers.

In summary, the product portfolio for the simulation model totals 8 families of 12 products each. Thus total number of products being produced at this plant is 96. This is not unrealistic for a production fab. Each of the 96 products has a certain number of reticles required to process it through the fab. The numbers of layers per product family along with the delays are shown in Table 1.

Table 1: Product Portfolio

TECHNOLOGY	FAMILY	Number of Products	# Photo Layers	Photo Proc Time (min)	Delay 1 (hour)	Delay 2 (hour)	Delay 3 (hour)
A	1	12	16	60	36	48	24
A	2	12	17	60	36	48	24
A	3	12	18	60	36	48	24
A	4	12	19	60	36	48	24
B	5	12	24	60	24	52	16
B	6	12	25	60	24	52	16
B	7	12	26	60	24	52	16
B	8	12	27	60	24	52	16

All 8 families and all 12 products in each family are generated with an equal probability in the ARRIVAL event. The time between creation for consecutive lots is exponentially distributed at 105 minutes.

The model includes 16 photolithography tools. Each tool is identical to the next tool. Each can store the 12 reticles on the machine including the reticle being utilized, and for this research, the assumption is that the reticle storage on the tool is fixed assignment. All products can be run on all photolithography tools. The tools are also modeled to be available 100% of the time. While adding variation to availability is more realistic, it also complicates the investigation into the prediction of gains due to proactive reticle management. For this reason, it was not included. Table 2 shows the equipment detail used in the simulation.

Table 2: Equipment Data

# Photo Equipment	Processing Time	# of Slots
16	60 minutes	12

4.2 Model Logic

Figures 1 and 2 show the PSUEDO SLAM II network diagram for the discrete event simulation. Figure 1 shows the higher level overview of the simulation, which displays the primary method of implementing the discrete event call procedures. Figure 2 has 3 different segments. Each segment details the event calls from the higher level Figure 1. Segments A, B, and C show the detail behind the FORTRAN code of the Arrival, Reticle Setup, and End of Service Events respectively. Even though the simulation actually uses the event orientation of SLAM, the diagram only represents an equivalent SLAM II network diagram.

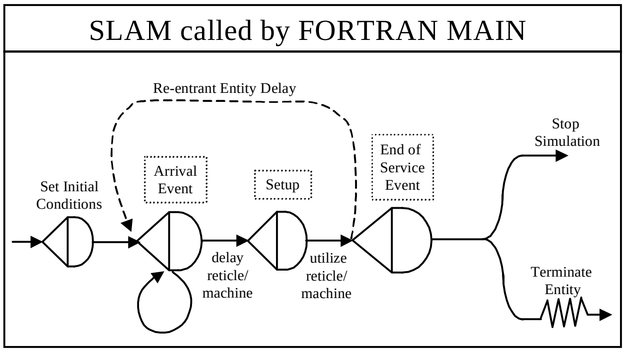

Figure 1: FORTRAN MAIN

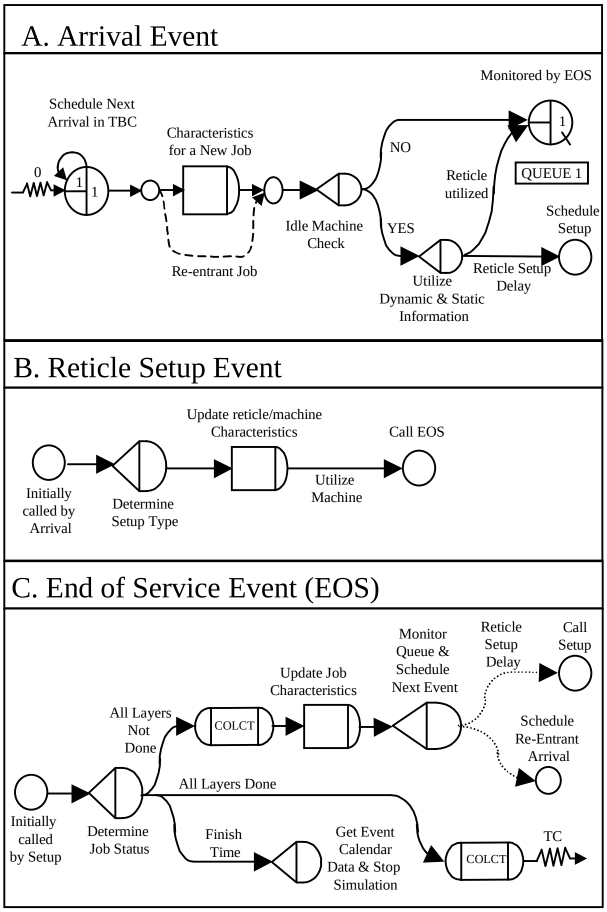

Figure 2: Service Events

4.2.1 Arrival Subroutine

In more detail, the ARRIVAL subroutine (Figure 2) is where the entities are created with a defined time between creation. After creation of an entity, there will be an assignment of some attributes to the entity. These attributes are only assigned upon creation of an entity. The attributes assigned in ARRIVAL are shown in Table 3. If the ARRIVAL subroutine is being called from the END OF SERVICE subroutine, then the initial attribute assignment is skipped because it was already completed when the entity was created.

Table 3: Simulation Entity Description

Entity Attribute	Description
1	Creation Time
2	Arrival Time For Each Layer
3	Family Identification
4	Product Identification
5	Current Layer Identification
6	Machine Currently Serving Entity
7	Reticle ID Required for Current Layer
8	Slot Number on Machine for Required Reticle ID
9	Sequence Dependent Setup Delay
10	Origin Source for Required Reticle ID
11	Dynamic Reticle Information Table

After assignment of attributes or the ARRIVAL event was called from the END OF SERVICE event, the ARRIVAL subroutine checks to see if there are any idle (non-utilized) machines at this time. If there is no idle machine, then the lot is filed in the queue. If one idle machine is detected, the simulation will check the machine arrays determine if there are any other idle machines.

In the case where there is only one idle machine, the subroutine will ascertain the reticle identification that is in the assigned slot number for the layer of the entity being evaluated in the ARRIVAL subroutine. After determining the reticle id, the slot number, the simulation investigates the location of the reticle id required for the lot. It will first check to see if the required reticle is on the idle machine in the assigned slot. If the reticle is already located on the machine, the simulation will assign a delay time of 0 for the setup time delay and then call the SETUP event. The no time delay is appropriate because the reticle is on the machine and nothing has to be done by a human to inspect or setup the reticle.

If the reticle is not on the machine, the appropriate slot for all other machines are checked for the reticle. If it is in one of the other machines, the simulation will evaluate the reticle to determine if it is being utilized at that time. If it is not utilized, the simulation will move the reticle and assign a delay of 30 minutes for the setup and then call the SETUP event. This time represents the human time it takes to remove the reticle from the machine, inspect the reticle for particles, and install it into the idle machine. This delay time is referred to as the machine to machine delay time and is the same for all reticles moved from one machine to another.

If the reticle is utilized on another machine, the simulation will not run that lot. In this case, the lot will be place in the queue to be evaluated for selection again by the END OF SERVICE event. This must occur because there is only one reticle for the layer of that product, and it is not desirable to leave a machine idle because it is waiting for a reticle that is being utilized on another machine.

If the reticle is not on any of the other machines, the reticle must be in the stocker. If the reticle is in the stocker, it is known to be idle. In this case, the simulation will assign a delay of 15 minutes for the setup of the reticle. This represents the time it takes to remove it from the stocker, and install the reticle into the idle machine. This delay time is referred to as the stocker to machine delay and is the same for all reticles being moved from the stocker to any machine.

If there is more than 1 idle machine, the simulation will essentially perform the same checks as in the single idle machine case except it must check multiple machines for the reticle, and it must determine which idle machine will be processing the lot.

4.2.2 Setup Subroutine

The SETUP subroutine (Figure 2) is called upon completion of the ARRIVAL subroutine or upon completion of the END OF SERVICE subroutine when a condition exists. This condition is that the END OF SERVICE has detected a lot in queue that can be run on the now idle photolithography tool.

The ARRIVAL subroutine actually determines the type of delay to be taken by the SETUP subroutine. Table 4 shows the 3 different setup types that can occur in the simulation. The SETUP call is used to adjust the machine and reticle utilization numbers to correctly report what part of the resources utilization was due to setup and what was due to processing. SETUP will always call the END OF SERVICE subroutine upon its completion.

Table 4: Setup Types

Reticle Move Type	Steps	Time (minutes)
Machine to Machine	Inspection + Setup	30
Stocker to Machine	Setup	15
Stays on Machine	No Action	0

4.2.3 End of Service Subroutine

The END OF SERVICE subroutine (Figure 2) starts with evaluating whether or not the current time is greater than or equal to the ending simulation time. If this is true, the simulation is halted. If not true, the simulation evaluates whether the current lot has just finished its last layer (completed). If this is true, the data is collected on the entity and it is terminated from the system. Table 5 shows the data collected upon termination of an entity from the system.

If the lot is not completed with all processing, then intermediate data is collected on the entity and the attributes of the entity are updated. Table 6 shows the intermediate data collected. After collection of the data, the entity has its attributes updated to represent the required reticle for the next layer. The updated attributes that are changed are shown in Table 7.

Table 5: Collected Data at Entity Termination

Data Collection	Description
Time In System	Time from creation to completion for entity
Time Per Layer	Time from last layer to current layer completion
Family Time in System	Creation time to completion for FAMILY of Product

Table 6: Collected Data at Layer Completion

Data Collection	Description
Time Between Layer	Time from last layer to current layer

Table 7: Attribute Update After Layer Completion

Attribute	Description	Action
2	Arrival time for each	Set to current
5	Current layer	Increment to next
6	Machine current serving	Set to 0
7	Reticle ID for current layer	Set to current
8	Machine slot # for reticle	Set to slot for current reticle
9	Sequence dependent setup	Set to 0
10	Origin source for reticle	Set to 0
11	Dynamic reticle information	Set to 0

Upon completion of the END OF SERVICE subroutine the lot will have been assigned a re-entrant time to go back into the ARRIVAL subroutine. In addition to this, a photo tool will be idle. The simulation will go back in and evaluate whether or not there is a lot in the queue that the machine can process. If it fails to find a lot to run, the subroutine will keep monitoring the queue.

4.3 Data Structures

In the building of this simulation model, it is important to detail the data arrays and reticle index structure because it explains how the logic is used to monitor the reticle performance measures for the fixed assignment case. The first data structure is used to identify the slot assignment for each reticle of a product. Every product that is in the simulation must have an assigned slot for every reticle that is required to make the product. The number of reticles defines the number of layers for the product. This data structure is static because it is read upon initialization of the simulation and does not change. This data structure is implemented with a 2 dimensional array. The row index corresponds to family identity *m* and the column index matches to layer number *n*. Figure 3 shows a general skeleton of the static information table with *m* products each requiring *n* layers. Each cell of the static information table keeps the slot number that is pertinent to the given family and layer.

$$\begin{array}{c|ccccc}
ARRAY_{(F1,\, n1)} & dd_{1,1} & dd_{1,2} & \cdots & dd_{1,n_1-1} & dd_{1,n_1} \\
ARRAY_{(F2,\, n2)} & dd_{2,1} & dd_{2,2} & \cdots & dd_{2,n_2-1} & dd_{2,n_2} \\
ARRAY_{(F3,\, n3)} & dd_{3,1} & dd_{3,2} & \cdots & dd_{3,n_3-1} & dd_{3,n_3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
ARRAY_{(Fm-1,\, nm-1)} & dd_{m-1,1} & dd_{m-1,2} & \cdots & dd_{m-1,n_{m-1}-1} & dd_{m-1,n_{m-1}} \\
ARRAY_{(Fm,\, nm)} & dd_{m,1} & dd_{m,2} & \cdots & dd_{m,n_m-1} & dd_{m,n_m}
\end{array}$$

Figure 3: Static Product Reticle Slot Assignment Array

Another data structure is used to keep track of the where reticles are currently residing. It is referred to as the slot reticle location array. This structure is the needed because during the simulation run, reticles continually move between machines and the storage stocker as needed.

When a reticle moves into a machine, it must first refer to the static reticle slot assignment table to get the reticle's assigned slot number. It then must put this reticle into that slot, but to do this correctly, it must know what reticle currently resides in the slot. Then, it must remove that reticle from the slot and put in the new reticle.

To accomplish this, a 2 dimensional dynamic information table is used. The row index *i* represents the slot number on the machines and the column index *j* represents the machine number for the slots. A reticle stored in slot *i* and machine *j* is identified array(*i*, *j*) and the index of the reticle is stored in the array position. The frame of this dynamic information table, array(*i*,*j*), is shown in Figure 4.

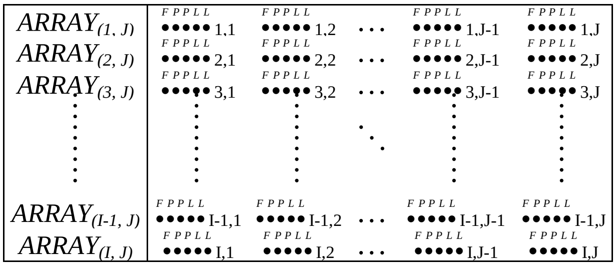

Figure 4: Dynamic Slot Reticle Location Array

There is a 3rd array referred to as the utilized reticle array which is a 1 dimensional array that contains the list of reticles currently being utilized on the actual photolithography tool. The simulation must check this array before it pulls reticles out of photo tools. The rows of this array correspond to the number of photolithography tools, *j*. Reticles will be stored on a tool, but only 1 reticle per tool can actually be used for processing. If the reticle in the slot is currently being utilized, the array will contain the index of that reticle. If it is utilized, then the simulation is not allowed to remove the reticle and replace it with the needed reticle.

The reticle index is another data structure that is important to discuss. It is indispensable to devise a rule to index reticles based on their uniqueness. Internally, the simulation model uses the 5 digits indexing rule such as $\overset{F}{1}\overset{P}{2}\overset{P}{3}\overset{L}{4}\overset{L}{5}$. The first digit represents a family identification. The second and the third digit correspond to product identification. The fourth and the fifth digit identify a layer that will use the indexed reticle. This indexing scheme provides the needed information in the name of the reticle to help make decisions based on the coded logic.

5 MODEL RESULTS

It is reasonable to hypothesize that the photolithography process will not reach steady state quickly due to the re-entrant feature and long product cycle times. Hence, the simulation model needs to have comparably long warm-up period to avoid initialization bias. The model is executed with a run length 1,000,000 minutes that is approximate 2 years. The first half of it is truncated. The total number of simulation runs is 50.

The simulation model was used to evaluate the strategic reticle issues that develop due to different constraints. The first simulation run involved evaluating the machine utilization given a mix that started each one of the 96 products in equal quantity. Figure 5, shows the average machine utilization of each machine for the run. The dotted line illustrates an unbalance, derived from a certain machine selection rule that forced the lower number tool to be utilized first. The created a scenario that favored some machines over others.. The issue was resolved by searching idle machines with random starting position, this unbalance is resolved like the balanced solid line.

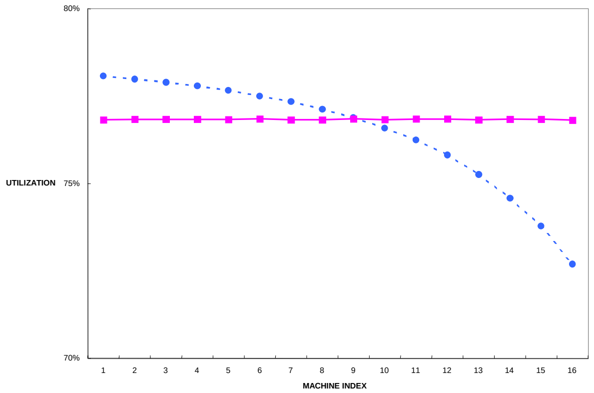

Figure 5: Machine Utilization

After changing the equipment selection logic, the simulation was evaluated for slot utilization. Figure 6 shows the slot utilization results from the runs. Since each machine has the same slot assignment policy for the reticles, the slot utilization can be averaged across all machines. The operational slot assignment policy, directly impacts this performance measure. Therefore, slot balance must be achieved base on the product mix running or the utilization may start to hinder throughput on the tool.

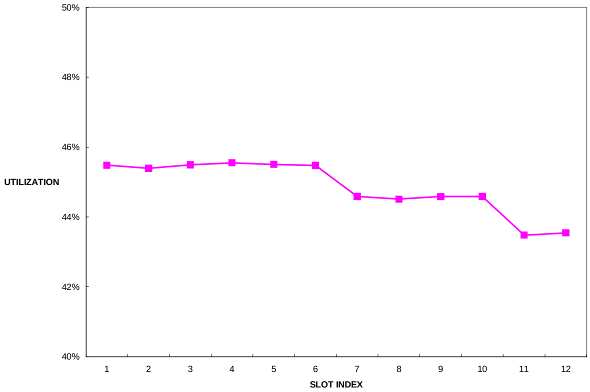

Figure 6: Slot Utilization

Another area to investigate with the simulation is the relationship between cycle time and product mix. Specifically, does the cycle time of the product in the system increase, as the product mix becomes more homogenous? When the product mix is very diversified, the main issue seems to be the need for many reticles and associated setups required. In this case, the utilization of the reticle is very low and the required setup high. In the homogenous product mix case, the utilization of the reticle starts to increase and the number of setups start to decrease. It is thought that the increased utilization of the reticles will start to cause lots to have to wait for reticles and not machine availability. This waiting for the dual constraint causes increased time in system for the product.

To test this case the simulation was run at 3 different mix ratios. Mix 1 was 8 families with 12 products each evenly released. Mix 2 was 8 families with 8 products evenly released. Mix 3 was 8 families with 4 products evenly released. Mix 1 required 2064 reticles while mix 2 and mix 3 required 1376 and 688 for respectively. Mix 2 is a 33% reduction in the required reticles and mix 3 is a 66% reduction. The evaluation of the output shows that mix 2 resulted in a product cycle time increase of 1.65%. Mix 3 resulted in a product cycle time increase of 4.81%. Figure 7 displays the cycle time lost due to the homogenous mix.

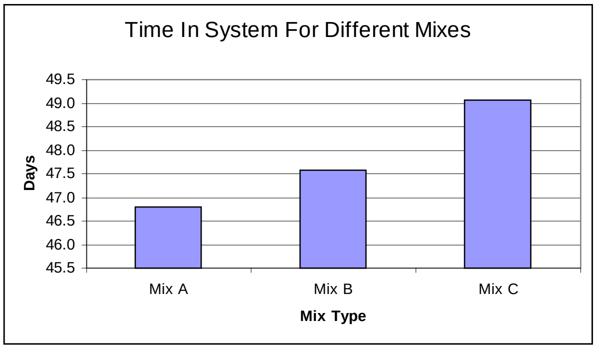

Figure 7: TIS for Mixes

6 FUTURE EVENT INFORMATION

Knowing the future events to be scheduled in simulation provides specific opportunity if the data can be analyzed. Utilizing this simulation, at identified times, data can be written out of the future event calendar and the data structures that can then be fed into a specifically designed network flow model. An example of this network flow model can be found in a paper presented at the *Proceedings of the 1999 International Conference on Semiconductor Manufacturing Operational Modeling and Simulation* by Hickie, Fowler, and Carlyle titled "Photolithography Reticle Management Through Network Analysis".

This network flow model requires knowing what lots are currently in inventory along with the predicted arrival times of future lots. It also must know what reticle is required for all lots in inventory along with all lots to be arriving. In addition to this, the location of all required reticles for the lots must be known. The final piece of information is the contents of the dynamic slot reticle array. This would give the reticle indexes currently in storage on all photolithography machines

This information can be fed into the network flow model. The model takes the information and solves the network as a minimum setup/maximum flow problem. The resultant output is unfortunately not a direct reticle assignment, but it can be used to establish a minimum setup reticle assignment with a heuristic. With a new reticle assignment, the data assignment can be fed back into the simulation and the simulation can continue to run. This method can be repeated between the simulation and the network flow. Consequently, this relationship could result in the enhancement of system performances due to less inspection and setup on reticles.

7 CONCLUSION

This research mainly focused on the building of the discrete event simulation model for the reentrant photolithography process. It discussed current reticle management issues that hinder performance of the photolithography tool. Based on the simulation model, output analysis on the system performance measures was carried out for various scenarios. The primary output analyses are summarized as follows.

- Random machine selection rule was implemented into the simulation to balance out the machine utilization.
- The slot utilization was imbalance was investigated for the fixed reticle assignment policy.

- For data transaction generated in the course of reticle management, 2 key data structures were developed.
- Future event information, from the simulation was detailed to explain how it can be fed into a network flow model to provide a proactive method of determining reticle setup.

BIBLIOGRAPHY

Goldratt, E. M. And Cox, J., The Goal, North River Press, Inc.,New York, 1986

Matt Hickie, John W. Fowler, and W. Matthew Carlyle. 1999. "Photolithography Reticle Management Through Network Analysis". In *Proceedings of the 1999 International Conference on Semiconductor Manufacturing Operational Modeling and Simulation*, ed. John W. Fowler, Jeffery K. Cochran, and Courtland M. Hilton, 127-132. Sponsored by The Society For Computer Simulation International, San Diego, CA, U.S.A.

AUTHOR BIOGRAPHIES

SUNGMIN PARK is a graduate student pursuing a Ph.D. in Industrial Engineering at Arizona State University. His research interests include simulation output analysis on single system and applied statistics.

MATT HICKIE is a graduate student pursuing a Masters degree in Industrial Engineering at Arizona State University. He is working for Motorola at Chandler, Arizona as a Manufacturing Systems Engineering Manager. His research interests include optimization and scheduling on semiconductor manufacturing system.

JOHN W. FOWLER is an Assistant Professor of Industrial Engineering at Arizona State University. He holds a Ph.D. from Texas A&M University. His research interests include manufacturing systems analysis, discrete event simulation, and applied operations research.

W. MATTHEW CARLYLE is an Assistant Professor of Industrial Engineering at Arizona State University. He holds a Ph.D. from Stanford University. His research interests include operations research, computer aided processes, optimization, and mathematical modeling.