Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

SIMULATION ANALYSIS ON THE IMPACT OF FURNACE BATCH SIZE INCREASE IN A DEPOSITION LOOP

Boon Ping Gan
Peter Lendermann

Planning and Operation Management Group
Singapore Institute of Manufacturing Technology
71 Nanyang Drive
Singapore 638075, SINGAPORE

Kelvin Paht Te Quek
Bart van der Heijden
Chen Chong Chin
Choon Yap Koh

Industrial Engineering & Planning
Systems on Silicon Manufacturing Co. Pte. Ltd.
70 Pasir Ris Industrial Drive 1
Singapore 519527, SINGAPORE

ABSTRACT

In the dynamic environment of semiconductor manufacturing operations, a bottleneck could be created at the bake furnaces of the deposition loop as capacity expands. Upgrading of the bake furnaces by adding a lot-per-batch in the boat or purchasing a new furnace are two possible solutions to this problem. A simulation model was constructed to assist the decision making, with the behavior of the wet benches (upstream tools) and cluster tools (downstream tools) being modeled in detail. We concluded that a limited number of furnaces upgrade is sufficient to sustain the capacity expansion. But the bottleneck was shifted to an upstream tool, which required the backup tool to be activated to manage the queue. A loading policy that constrains batches to queue at maximum time before loading into the furnaces has to be implemented to balance the efficiency at the furnaces and their downstream tools, without compromising on the cycle time.

1 INTRODUCTION

Semiconductor manufacturing involves very complex processes as wafers need to go through a series of layering, patterning, doping, and heat treatment steps, repeating these steps through the manufacturing process. Reentrant flows, time constraints, varying product mixes, running prototypes and ad-hoc resource breakdowns result in high degree of variability. Analysis of the fab operations with sufficient accuracy is thus not possible with any mathematical models. In general, discrete event simulation can be used to portray this dynamic and high variability behavior of the fab. The simulation model generates artificial history of the fab operations, and it is used to study the impact of different policies or capacity changes to the overall fab performance. Various commercial off-the-shelf simulation packages, such as *AutoSched AP* (Brooks 2001), *Automod* (Brooks 2006), *FlexSim* (FlexSim 2006), and *WITNESS* (Lanner 2006) are available for this purpose. In SIMTech, a proprietary simulation engine for semiconductor manufacturing process modeling has been developed over the years. As compared to the commercial tools, this simulator is more efficient in terms of execution speed and provides high degree of flexibility for customization.

In this paper, we use discrete-event simulation to reveal if upgrading of a bottleneck furnace from 6 to 7-lot per batch is sufficient to cope with the capacity increase, in anticipation of the rising wafer demands. This bake furnace is used in the deposition loop of the metallization process. Our alternative solution to this upgrading would be to purchase a new furnace which costs approximately 10 times more than the upgrading.

Besides, the simulation model is also used to recommend loading policy that needs to be devised due to this upgrading. Our anticipation was that the upgrading will result in lots being queued for longer period of time as now 7-lot of wafers need to be batched instead of 6. The additional lot may have a longer waiting time which could violate the queue time protocol. Such violation will lead to quality issues or rework loading policy.

This paper is organized as follows: In Section 2, we will describe the deposition loop which is the focus of this simulation study. Section 3 focuses on the simulation model development, including challenges in modeling the detailed behavior of wet benches and cluster tools. In Section 4, we present the experimental results, model validation process and analysis of the bake furnace upgrade, including a study of a time-based loading policy. Lastly, we conclude our study with recommended actions and also outline some future developments.

2 THE DEPOSITION LOOP

Figure 1 shows a typical process flow of a deposition loop. It starts from a wet bench where the residue of the photo resists and polymers process are removed through etching and cleaning. The tool is loaded with two lots of the same recipe, dipping into six different baths (solvent and water baths) with different process time.

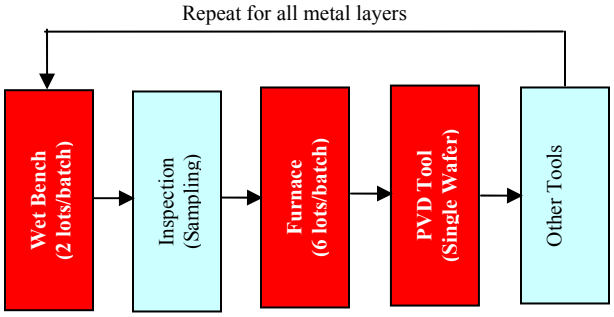


Figure 1: The Deposition Loop

After the etching process, the wafer lots are then sent for inspection and measurement in order to check the consistency of the process. A sampling approach is used for the inspection steps, with rework consideration if the wafers do not meet the specification.

Good wafers are then moved on to the bake furnace to undergo baking process which can be loaded with 6-lot at any one time. It is important that a good batching rule is followed such that the tool is efficiently utilized. Starting the process with a partially filled batch will increase the tool overall utilization while the efficiency suffers. On the other hand, a full batch may result in unnecessary delay to the wafer lots, which in turn result in violation of queue time constraint and cycle time loss. The last step is the glue layers deposition with the Physical Vapor Deposition (PVD) tools. To make use of the PVD tool efficiently, we must ensure that new lot is loaded before the processing wafers are completed, which is as soon as the first chamber becomes available.

After the deposition process, the lot will be sent to other tools for further processing, for example: the via etching process. As these tools are not the bottlenecks, they will not be the focus of this study. The process will repeat the deposition loop depending on the number of metal layers, as illustrated in Figure 1.

Out of the three tools discussed, the bake furnace needs to increase to a 7 lot batch when the fab increases its capacity. Some questions that need to be addressed before a final decision can be made:

1. How many furnaces should be upgraded to support the required capacity expansion?
2. What is the impact of the upgrade to the upstream tools (wet bench) and the downstream tools (PVD)?
3. What is the loading policy? How do we batch the lots? What is the impact of waiting time on batching and hence efficiency?

3 SIMULATION MODEL

To address the questions outlined in Section 2, we created a simulation model for the studied process, using an in-house simulation program. Only the relevant tools were modeled in details for the simulation study, namely the wet bench, furnace, and PVD tools. The rest of the tools were modeled as delays, obtained from the historical data. This was possible as they were not highly utilized, and the overall confidence of the simulation was not compromised.

3.1 Input Release

In this study, we modeled a mixture of different product types (known as technology hereafter), with different number of metal layers. The loading interval of each technology is derived from the demand forecast given by:

$$\text{interval(hours)} = \frac{\text{lot size}}{\text{loading of technology}} * 30 \text{ days} * 24 \text{ hours} \qquad (1)$$

For example, one lot of a technology with demand forecast of 5000 wafers per month will be released at an interval of 3.31 hours using a typical lot size of 23 wafers.

3.2 Wet Benches

The wet bench modeled in this study consists of 6 baths. Batches move from one bath to the next strictly in the arranged bath sequence. There are two alternatives of modeling this behavior. One is to model it's average behavior. But we have ruled out this *approach* as it results in a significant behavior deviation from the actual situation in our model validation exercise. Similar observations were discussed by Jain et al. (1999). This calls for a detailed modeling of the wet benches.

The bath processing time varies depending on the recipes throughput. Such variations will result in efficiency loss and hence need to be addressed by the dispatch rules. The recipes run by the wet benches can further be classified into two groups: clean and dirty. Switching a tool from clean to dirty recipe does not require a chemical change, and thus no setup time is incurred. But the reverse will require a complete chemical change which could incur some hours of non-productive time. Chemical change is also required after the tool has run for a pre-defined number of batches. This is classified as part of the preventive mainte-

nance, which is modeled as tool downtime that will be discussed in Section 3.5.

3.3 Furnaces

The bake furnaces being investigated in this study is much simpler to model than the wet benches and PVD tools as it runs only on one recipe. All lots arriving at the tool can thus be batched together. The only complication in the modeling is on the loading policy. What type of loading policy should we deployed to balance the efficiency and cycle time constraint of the furnaces? We modeled two policies: One is to always load a full batch, another partial batch. In the latter policy, the wafers will be processed when they have reached a predetermined schedule irregardless of their quantities. The efficiency might suffer with this policy, but it complies with the process queue time protocol.

3.4 PVD Tools

The PVD tool is a single wafer processing tool, and the one being used in this study comprises of 4 main chambers. Wafers are moved from one chamber to another, staying at each chamber for a period of time, ranging from half a minute to approximately 3 minutes. Some chambers have duplicates that share the same functionality. Figure 2 shows a schematic diagram of the PVD tool. It has two chambers for type 1 and 4, while only one for type 2 and 3. The "pass" chambers sitting between the two group of chambers are used to move wafers from the first group of chambers to the second group of chambers and vice versa. The sequence of chambers that each wafer will go through is: Loader → 1 → 2 → Pass → 3 → 4 → Pass → Loader. A lot will be unloaded once all the wafers in the lot have gone through the whole process.

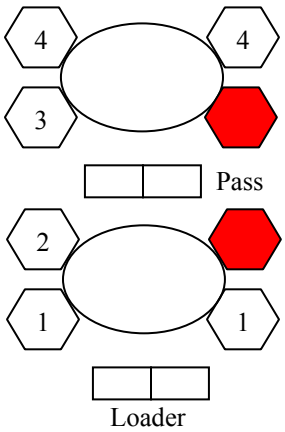


Figure 2: Schematic Diagram for the PVD Tools

The chamber processing time varies depending on recipes. This PVD tool is configured to run 7 different recipes. The effect of mixing fast and slow recipes is less significant in this case as compared to the wet benches. The reason for this is that the PVD tool processes one wafer at a time. Potentially, only the first few wafers are slowed down. Subsequent wafers will be processed at full speed of the recipe. Similar to the wet benches, this detailed behavior of the PVD tool was modeled in our simulation study. It's average behavior is not sufficient to portray it's actual characteristics.

3.5 Downtime Behavior

The random effect that we modeled in this study is the downtime behavior of the tools. We captured the mean time between failure (MTBF) and mean time to repair (MTTR) from a 6-month historical data. The failure includes the preventive maintenance that we perform on the tools on regular basis. Using the MTBF and MTTR, we generate failures using an uniform distribution.

3.6 The Simulation Execution

The simulation runs with wafer lots being released into the deposition loop in a regular interval. Wafer lots arrive at a tool group, a collection of tools with similar capability, and wait to be loaded into next available tool. A dispatch rule is applied to choose the preferred lots if more than one lots is queuing. Table 1 summarizes the number of tools that are available at the respective processing steps in the fab. Take note that there is one backup tool available for the wet bench, to address short term shortages at this tool group. Table 1 summarizes the number of tools that are available for each tool group in the deposition loop.

Table 1: Number of Tools for Each Tool Group

Tool Group	Number of Tools	Number of Backup Tools
Wet Bench	6	1
Furnace	5	0
PVD Tool	9	0

4 EXPERIMENTS

4.1 Model Verification and Validation

The model verification and validation (V&V) process was done iteratively. We begun with modeling the average behavior of the wet benches and the PVD tools. But through the V&V exercise, we found out that this was not sufficient to represent the real situation in the fab. We went back to the drawing board, explored the detailed behavior of the

tools (Section 3) and accordingly improved the granularity of the model. The improved model was then validated with the 6-month historical data; focusing on the tool utilization and downtime. Table 2 shows this correlation of the actual versus the simulation to be less than 5%, hence concluding the validation of the simulation.

Table 2: Correlation of Actual and Simulated Results

Tool Group	Utilization (%)		Down (%)	
	Actual	Simulated	Actual	Simulated
Wet Bench	93.2	96.4	3.9	3.5
Bake Furnace	79.5	81.4	3.8	2.6
PVD Tool	84.2	81.0	18.2	19.0

4.2 Scenarios

To answer the questions listed in Section 2, we ran the first set of experiments, varying the number of furnaces to be upgraded against the fab loading, L with 3K increment each run. With a 3-month simulation warm up period, a simulation run length of two years was deemed sufficient. Upon deciding the number of furnaces to be upgraded, we then experimented the loading policy at the bake furnaces, by varying the waiting time. Table 3 summarizes the parameters for the experiments. The experiments were compared using following metrics:

- Tool group utilization (%):

$$\frac{run\ time}{available\ time} * 100\ \%.$$

- Average queue length (number of lots):

$$\frac{\sum_{i=1}^{n} queue\ wip_{t_i} * (t_i - t_{i-1})}{t_n},$$

where t_i is the time when the queue wip change, and n is the number of times the queue wip changes.

- Wafer per hour:

$$\frac{\sum_{j=1}^{n} \frac{number\ of\ wafers(j)}{max(schedule\ time(j), depart\ time(j-1))}}{n},$$

where j is lot/batch j that is scheduled onto the tool, and n is the total number of batches.

- Cycle Average at a tool group (minutes):

$$\frac{\sum_{k=1}^{n} depart\ time(lot(k)) - arrival\ time(lot(k))}{n},$$

where k is the lot number that is completed by the tool, and n is the total number of lots completed by the tool.

We ran five replications for each experiment, collecting statistics for the above three performance metrics. The average across the 5 runs for these statistics are presented here.

Table 3: Experiment Parameters

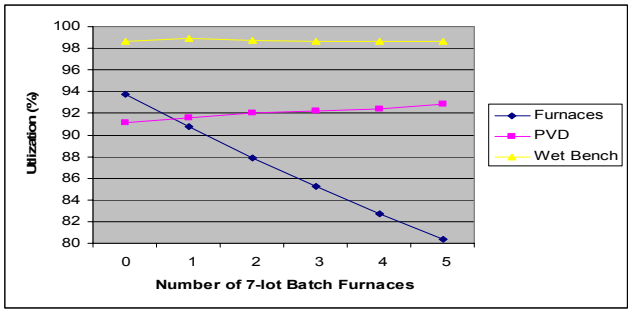
Parameters	Loading		
	L	L+3	L+6
Number of Furnaces Upgrade	0, 1, 2, 3, 4, 5	0, 1, 2, 3, 4, 5	0, 1, 2, 3, 4, 5
After Upgrading *n* furnaces			
Waiting time at Furnace (minutes)	10, 30, 50, 70, ∞	10, 30, 50, 70, ∞	10, 30, 50, 70, ∞

4.2.1 Number of furnace to be upgraded

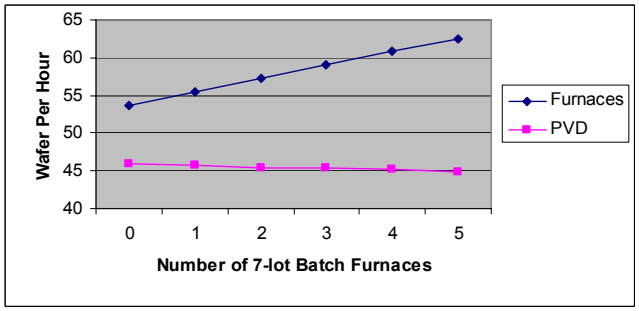
Figure 3 shows the statistics for the simulation, using current loading L, with varying number of furnaces to be upgraded to 7-lot batch, and using full batch loading policy. As more furnaces are upgraded to 7-lot batch, the average queue length at the furnaces (Figure 3(c)) reduces from 7.8 lots to less than 4 lots. This is intuitive as the capacity of the furnaces have increased with the upgrade. Correspondingly, the utilization of the furnaces (Figure 3 (a)) reduces from 94% to 80%. This is accompanied with an improved wafer per hour (WPH), from 54 to 63 wafers per hour.

Figure 4 shows the statistics collected by increasing the loading to L+3. The trends for the three performance metrics remain the same as compare to L loading. The significant difference is that the fab is no longer able to cope with this loading level without the furnace upgrade. As can be seen, the average queue length without furnace upgrade reaches more than 500 lots. At least one furnace needs to be upgraded to bring the average queue length down to below 50 lots. To maintain the utilization below 90%, 4 of the furnaces need to be upgraded. Careful inspection of the average queue length figure (Figure 4(c)) shows that the queue at the wet benches increases to more than 50 lots with 4 furnaces upgrade. The utilization of the wet bench

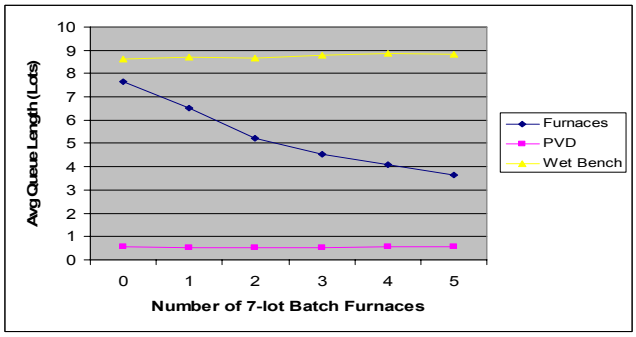
(Figure 4(a)) is closing up to 100%. The backup wet bench needs to be activated to cope with this utilization level.



(a) Utilization



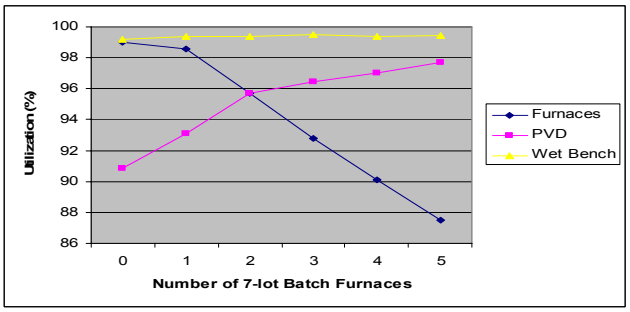
(b) Wafer Per Hour



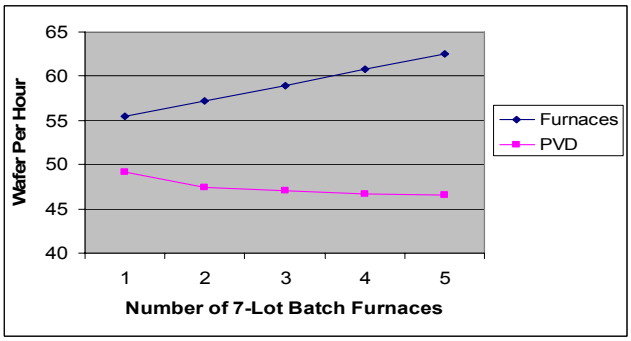
(c) Average Queue

Figure 3: Utilization, Wafer Per Hour, Average Queue Length at L Loading

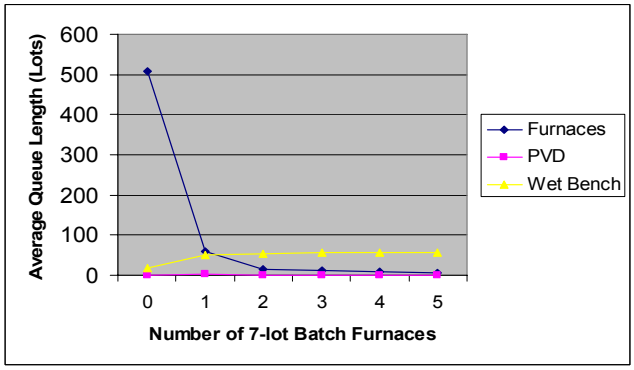
Figure 5 shows the statistics collected with L+3 loading, where the backup of wet bench is being activated. This helps to reduce the queue length of the wet bench down to approximately 5 lots (Figure 5(c)). It does not affect the performance of the furnaces and PVD tools, comparing Figure 4 and 5. From this study, we can conclude that it is a must to upgrade four of the furnaces to 7-lot batch and activate the backup wet bench to cope with L+3 of loading. We increased further the loading on the fab to way above L+6 wafers per month. Simulation results show that going beyond L+6 is not feasible as the wet bench becomes the bottleneck.



(a) Utilization



(b) Wafer Per Hour



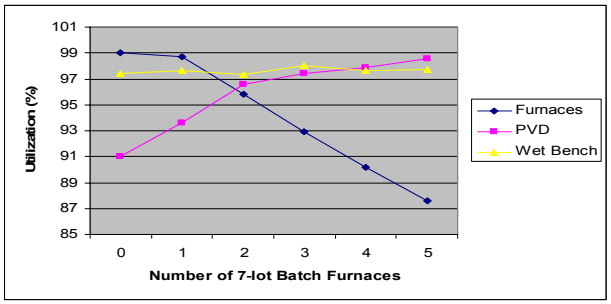
(c) Average Queue Length

Figure 4: Utilization, Wafer Per Hour, Average Queue Length at L+3 Loading

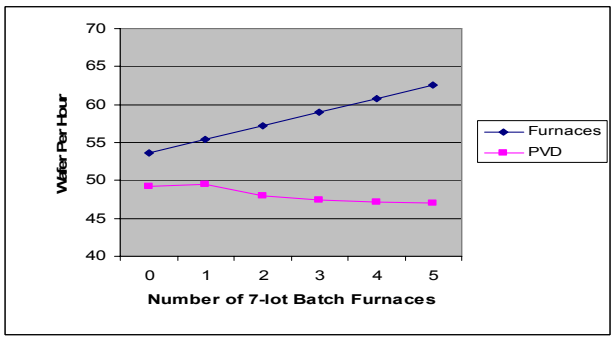
4.2.2 Loading policy at the furnaces

A loading policy was introduced at the furnaces for this simulation study, having four of the furnaces upgraded to 7-lot batch, and the backup wet bench being activated. The loading policy keeps track of the amount of time that a batch has spent waiting to be filled. If the waiting time exceeds the pre-defined value, the batch is loaded into an

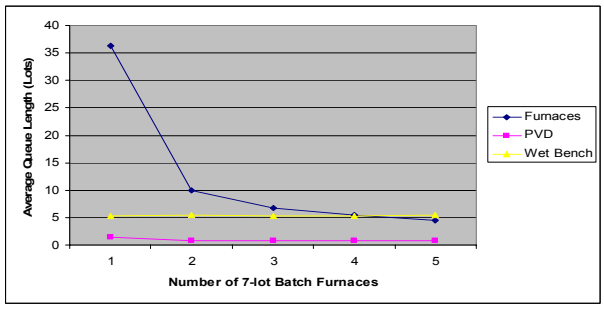
available furnace. On the other hand, if the batch is filled before the waiting time expired, the batch is loaded into an available furnace immediately. This waiting time period is varied from 10, 30, 50 to 70 minutes. At the extreme case, we will always wait for a full batch, which is denoted as Infinity in Figure 6 and 7.



(a) Utilization



(b) Wafer Per Hour



(c) Average Queue

Figure 5: Utilization, Wafer Per Hour, Average Queue Length at L+3 (Backup of Wet Bench Activated)

As shown in Figure 6, the longer the waiting time, the better batching efficiency that we achieve. Our current practice on the line is using a waiting time of approximately 30 minutes. This will result in an average batch size of less than 6.2, a batching efficiency of approximately 90% (average batch size of 6.8 is the maximum). Other implication such as its impact to the cycle average is shown in Figure 7. At 30 minutes waiting time, the cycle average will be approximately 205 minutes and 51 minutes at the furnaces and PVD tools respectively. In fact, the best cycle average was achieved when we always batch full lots. But does this mean that a loading policy that always batch full lot is the best solution?

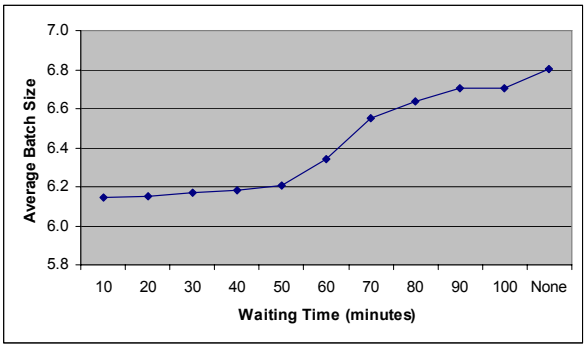
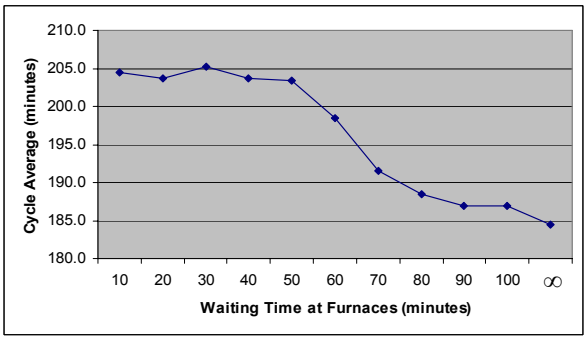
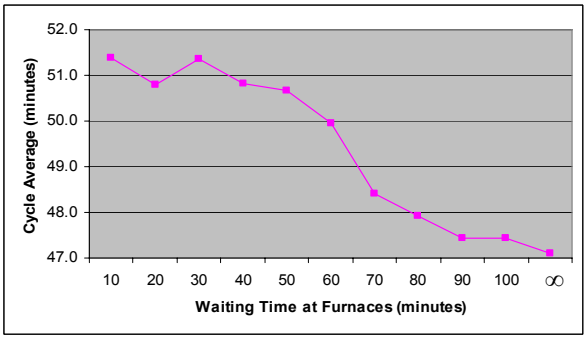


Figure 6: Effects of Waiting Time to the Batching Efficient at the Furnaces

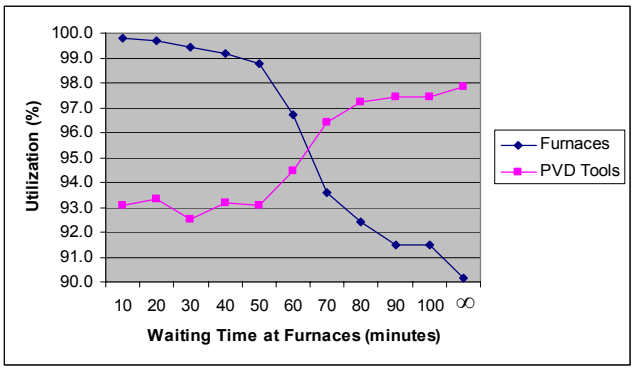


(a) Furnaces

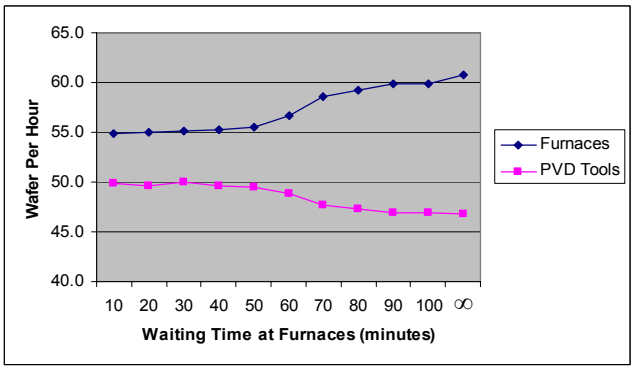


(b) PVD Tools

Figure 7: Effect of Waiting Time to Cycle Average of the Furnace and PVD Tools



(a) Utilization



(b) Wafer Per Hour

Figure 8: Effect of Waiting Time to Utilization and Wafer Per Hour of the Furnaces and PVD Tools

Referring to Figure 8(b), we can see that the wafer per hour of the PVD tools deteriorate relatively more significantly when the waiting time at the furnaces is set above 60 minutes. A lower wafer per hour implies that the PVD tools are not used efficiently, which leads to the increase in the tools' utilization level (Figure 8(a)). To achieve a compromise between the PVD tools' efficiency and the furnaces' batch efficiency, a waiting time of 70 minutes (intersection between the PVD tools' and the furnaces' utilization graph) is chosen. At this value, we will achieve a cycle average of 190 minutes and 47.5 minutes at the furnaces and PVD tools respectively (Figure 7), and a batching efficiency of 97% at the furnaces.

5 CONCLUSIONS

In this paper, we presented our findings in using discrete event simulation to study the impact of furnaces upgrade to the upstream (wet benches) and downstream (PVD tools) tools. Through simulation study, we concluded that our fab was still able to cope with its current loading L without any upgrade. But in order to maintain the utilization level of the furnaces to approximately 90%, upgrading 2 of the furnaces to 7-lot batch was necessary.

When the loading of the fab was increased by an additional 3K L+3, 4 of the furnaces would need upgrading and the backup wet bench was required. A loading policy with waiting time of 70 minutes was chosen to compromise the batching efficiency at the furnaces with the efficiency of the PVD tools. Table 4 summarizes the actions to be taken on the production floor.

Table 4: Actions to be Taken on Production Floor

Loading ('000)	Tool Group		
	Wet Bench	Furnaces	PVD Tools
L	No action	Upgrade two furnaces to 7-lot batch	No action
L+3	Activate one backup	1. Upgrade four furnaces to 7-lot batch 2. Loading policy with waiting time of 70 minutes	No action as waiting time of 70 minutes at the furnaces balances the batching efficiency of furnaces and running efficiency of PVD
L+6			

Through this simulation study, we also learnt that average behavior modeling was not sufficient for some tools, such as the wet benches and PVD tools (cluster tools). It results in a simulation model that does not reflect the real world situation. We overcame this through detailed modeling of these tools . But this poses a great challenge when a large number of tool types are involved. The simulation execution time will increase significantly with detail modeling. A compromise between detailed and average behavior modeling is needed.

Our finding is consistent with the work presented by Peikert, Thoma and Brown (1998) where a simulation model was created for the production area of interest with other fab operations being treated as "black boxes". Though the effort required to build such model was significantly lower, the accuracy of the simulation was not compromised. Pool and Bachrach (2000) compared a detailed and gray (black box modeling) modeling of a fab and made a similar observation where gray modeling gives a reasonably accurate simulation results as compared to detailed modeling. As such, selectively modeling various tools in different level of abstraction is a useful method to perform a quick simulation study.

REFERENCES

Jain S., C.C. Lim, B.P. Gan, Y.H. Low. 1999. Criticality of Detailed Modeling in Semiconductor Supply Chain Simulation. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, 888-896.

Brooks Automation. 2001. AutoSched AP User's Guide.

Brooks Automation. 2006. <http://www.brookssoftware.com/pages/245_automod_overview.cfm> [accessed March 2, 2006].

Flexsim. 2006. <http://www.flexsim.com/software/flexsim> [accessed March 2, 2006].

Lanner. 2006. <http://www.lanner.com/products/ simulation_suite/witness.php> [accessed March 2, 2006].

Peikert, A., Thoma, J., and Brown, S., 1998. A Rapid Modeling Technique for Measurable Improvements in Factory Performance. In *Proceedings of the 1998 Winter Simulation Conference*, Eds: D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, 1011-1015.

Pool M. and R. Bachrach. 2000. Productivity Modeling of Semiconductor Manufacturing Equipment. In *Proceedings of the 2000 Winter Simulation Conference*, Eds: J.A. Joines, R.R. Barton, K. Kang, and P.A. Fishwick, 1423-1427.

AUTHOR BIOGRAPHIES

BOON PING GAN is a Senior Research Engineer with the Planning and Operations Group at the Singapore Institute of Manufacturing Technology. The focus of his research is on the application of distributed simulation technology for supply chain simulation. He received a Bachelor of Applied Science (Hons) in Computer Engineering and a Master of Applied Science from Nanyang Technological University of Singapore in 1995 and 1998, respectively. His research interests are application of parallel and distributed simulation in the manufacturing industry. His email address is <bpgan@SIMTec.a-star.edu.sg>.

PETER LENDERMANN is a Senior Scientist of the Planning and Operations Group at Singapore Institute of Manufacturing Technology. Previously he was a Managing Consultant with agiplan in Germany where his focus was on the areas of supply chain management and production planning. He also worked as a Research Associate at the European Laboratory for Particle Physics CERN in Geneva (Switzerland) and Nagoya University (Japan). He obtained a Diploma in Physics from the University of Munich (Germany), a Doctorate in Applied Physics from Humboldt-University in Berlin (Germany) and a Master in International Economics and Management from Bocconi-University in Milan (Italy). His research interests include modeling and analysis of manufacturing and logistics systems as well as distributed simulation. His email address is <peterl@SIMTech.a-star.edu.sg>.

KELVIN PAHT TE QUEK is a Senior Engineer with the Industrial Engineering and Planning. He is assigned in the capacity section to work on new technology capacity planning and flow delta analysis. He is also involved in productivity section to work the Diffusion Module in the areas of capacity and productivity improvements planning. He received a Bachelor of Mechanical Engineering (Hons) from the University of Birmingham (UK) in 1999. His email address is <path.te.quek@philips.com>.

BART VAN DER HEIJDEN is a Manager with the Industrial Engineering and Planning Department at Systems on Silicon Manufacturing Company Private Limited and he is responsible for productivity improvements. Prior to this, he was manager for diffusion process for SSMC. His previous assignments in Philips include development and manufacturing processes. He received his degree in Chemical Engineering from IHBO 'de Maere' in the Netherlands. His email address is <bart.van.der.heijden@philips.com>

CHEN CHONG CHIN is the Director of Industrial Engineering and Planning Department at Systems on Silicon Manufacturing Company Private Limited. Prior to this, he was with Philips' industrial competence center responsible industrialization of supply chain processes before the subsequent responsibility in global business line/project management. His previous employers include AT&T, Fairchild and Thomson. He holds a Master in Business Administration, University of Brunel (UK) and a Bachelor in Electronics Engineering (Hons), University of Essex (UK). His email address is <cc.chin@philips.com>.

CHOON YAP KOH is a Senior Engineer with Manufacturing Department at Systems on Silicon Manufacturing Company Private Limited. He is involved in daily fab operations at Diffusion and Etching areas, with focus on capacity improvement activities. He received a Bachelor of Mechanical Engineering (Hons) from Nanyang Technological University of Singapore in 2000. His email address is <choon.yap.koh@philips.com>.