INTELLIGENT SIMULATION-BASED LOT SCHEDULING OF PHOTOLITHOGRAPHY TOOLSETS IN A WAFER FABRICATION FACILITY

Amr Arisha Paul Young

School of Mechanical and Manufacturing Engineering Dublin City University Glasnevin, Dublin 9, Ireland.

ABSTRACT

Scheduling of a semiconductor manufacturing facility is one of the most complex tasks encountered. Confronted with a high technology product market, semiconductor manufacturing is increasingly more dynamic and competitive in the introduction of new products in shorter time intervals. Photolithography, being one of the processes repeated often, is a fabrication bottleneck. Lot scheduling within photolithography is a challenging activity where substantial improvements in factory performance can be made. The proposed scheduling methodology integrates two common approaches, simulation and artificial intelligence. Using detailed simulation modeling within a structured modeling method, a comprehensive model to characterize the photolithography process was developed. An artificial intelligence scheduler was then developed and integrated with the model with the goal of reducing Work-In-Process (WIP), setup time, and throughput time. The results have shown a significant improvement in lot cycle time as well as tool utilization, improved the throughput time by an average of 15% and is currently in use for scheduling the photolithography process.

1 INTRODUCTION

Semiconductor manufacturing is one of the most complex manufacturing processes in the world. Scheduling of wafer fabrication facilities is among the most challenging planning activities encountered these days due to random yields and rework, complex product flow, time-critical operations, batching, simultaneous resource possession, and rapidly changing products and technologies.

 The competitive operation of modern fabrication (FAB) processes requires the development of precise models and rules for allocating the available resources within the FAB so as to optimize the production performance. Although simply stated, such an objective is elusive, primarily due to the size and complexity of modern FABs. The determination to have better scheduling policies remains highly nontrivial, involving the solution of constrained problems within bottleneck processes (e.g. photolithography) with respect to often-conflicting objectives while any admissible policy must posses certain robustness properties in the presence of uncertainty.

 The interdisciplinary research in scheduling of semiconductor manufacturing encompasses mathematical models (Ignizio 2004), stochastic modeling for semiconductor manufacturing (Hunter *et al*. 2002; Collins 2002), and simulation modeling (Ignizio 2002; Arisha 2003) for different applications within the wafer fabrication facilities. Simulation provides an effective tool for defining the path from competitive concept to real world solutions (Nayani and Mollaghasemi 1998). The use of simulation within dynamic manufacturing systems provides the only method to study new and existent complex interactions for which analytic or static models provide at best a low fidelity model with corresponding low accuracy. For example, the tools in the photolithography process are extremely expensive and hence, the risk attached to perform experimentation within the real systems is very high.

 The hybrid photolithography model presented in this paper was developed as a hierarchical simulation model, which includes the variabilities arising on the FAB floor, with an integrated neural network scheduler. The primary objective is to provide the managers and planning staff with an intelligent scheduler to improve photolithography area performance and to reduce WIP build-ups caused by variabilities and other constraints.

 In this paper a brief description of the photolithography process, Section 2, is followed by a detailed review of the scheduling problem for photolithography in Section 3. Having defined the objectives of the simulation in Section 4, the development of a modeling approach to address the issue is then described in Section 5, which also includes the integration of Artificial Intelligence (AI) techniques for dynamic lot scheduling. The techniques and results confirming the validity of the model are outlined in Section 6 before the results of a study carried out in a major semiconductor facility are presented in Section 7.

2 PHOTOLITHOGRAPHY PROCESS

Wafer fabrication is the most technologically complex and capital intensive stage of semiconductor manufacture. It involves the processing of silicon wafers to create the semiconductor devices in the wafer and build up the layers of conductors and dielectric on top that provide the complex interconnection between devices. Hundreds of operations are required to build a complex component such as a microprocessor. The main areas in wafer fabrication are shown in Figure 1. Photolithography is the most complex operation, requiring the greatest precision.

Figure 1: Wafer Fabrication Main Processes

 During the process the circuit pattern is transferred from a mask onto a photosensitive polymer and finally replicates the pattern in the underlying layer. The object of this process is the accurate and precise definition of a three-dimensional pattern on a semiconductor substrate. The basic photolithographic sequence is shown in Figure 2. Wafers move through the FAB in homogenous lots held in special containers. Typically, a lot to be processed goes through a coating operation, where the wafers are coated with a photo-resistant substance. The lot is then moved to the exposure operation where the patterns are projected on the wafers. The exposed wafers are moved to the developing operations. Once these steps are completed, the lot typically is moved to post-photolithography analytical operations. The amount of metrology is dependent on the product and the layer being processed.

3 PHOTOLITHOGRAPHY SCHEDULING PROBLEM

Photolithography is usually the bottleneck process with the most expensive equipment in a wafer FAB (Akcalt *et. Al*

Figure 2: Simplified Diagram of Typical Photolithography Process Flow

2001). Being one of the processes that is repeated the most during fabrication, any improvement in photolithography will consequently improve the overall performance. Lot scheduling is mainly based on the allocation of available tools over time to meet a set of performance criteria. Typically the scheduling problem involves a set of lots (different products/layers) to be processed, where each lot requires a particular set of operations/processes to be completed.

The scheduling of the photolithography area is a very difficult activity due to two main issues; complexity and variability. The process builds the required layers with such critical dimensions that it also needs complex metrology procedure to ensure the quality of the outcome. The expense of photolithography tooling is such that manufacturers cannot afford to buy more than the minimum number of tools and use the existing ones as long as possible to reduce costs, consequently many non-identical parallel tools can be found in the floor. The process flow is re-entrant and even more dynamic within photolithography area than other areas of the production system. In addition, the process is sensitive to product/layer changes with associated setup times. There are many sources of variability within the process such as high product-mix, lot priority issues, lack of formal lot scheduling rules within the floor. Maintenance including preventive maintenance, random yields, and labor dedication is also a crucial issue. Moreover, lack of a prior information about future lots for processing mean that scheduling must be realtime, increasing the complexity.

These result in a conservative operating policy with;

- low overall performance,
- low tool utilization/ high cost,
- more "Work In Process" inventory build,
- delay in delivery of orders,
- increase in throughput time per lot, and
- increase in tool cycle time.

 Two key issues for scheduling had to be established before the problem could be addresses, the qualifying matrix (Section 3.1) and the lot selection criteria (Section 3.2).

3.1 Qualifying Matrix

The factory cannot replace older equipment as long it is still functioning and the replacement period is not due, which means the performance of each tool in the group is unique. The manufacturing team uses a qualifying matrix (QM), updated periodically based on manufacturing policies, that defines which tool is capable of processing each layer. For example, the manufacturing team always assigns the hard/complex layers to new tools as the older tools may not be capable of achieving the required quality in a timely manner. Table 2 illustrates a sample of the qualifying matrix showing the tools and the layers on which they are able to perform. A similar table could also be drawn up with regard to product, but in this work it was assumed that all the tools can process a qualified layer on any product. In actual production more than 10 layers and over 20 tools are involved in the toolset represented by the model developed.

3.2 Selection Criteria

The manufacturing team had significant input in assigning a short list of the major constraints on the flow of lots through photolithography process. The schedule generated for a manufacturing run is highly dependent on the particular criteria used in the scheduling process. There are several criteria – most are dynamic – that will affect the selection of a particular tool to process an incoming lot. These criteria can be either process-oriented or wafer-oriented:

- A. Process oriented criteria relate to the equipment itself such as technology, maintenance, .. etc.
- B. Wafer oriented criteria relate to the lot information such as product, layer, .. etc.

Figure 3 shows the breakdown of criteria informally used by the production engineers to schedule the lots.

The lot may visit the photolithography process a number of times in order to build the required layers, increasing the model complexity and the level of variability. In addition each photolithography tool uses at least 13 operations to complete each layer. Hence, for example, if there are 10

Figure 3: Lot Scheduling Criteria

layers for 6 different products to be processed on 5 tools, that means there is almost $(60!)^5$ combinations to consider within the schedule. While the number of combination may be reduced due to constraints, there is still no possibility of performing an exhaustive set of experiments to find the optimum schedule. Indeed, explicit enumeration of such a problem requires too much time to be considered as an option (Arisha *et al*. 2002). Adding more tools is impractical on two counts, the capital cost is such that only the minimum number of tools can be installed and more tools increase the size of the scheduling problem. Therefore, there is an immense need for a powerful decision support system to minimize production cost and increase productivity throughout the existing toolsets.

4 MODEL OBJECTIVES

The intelligent scheduling model developed in this work has the following objectives:

- Characterize photolithography tools both individually and as a toolset.
- Examine the impact of various production plans on the performance of the photolithography toolset.
- Provide manufacturing/production engineering staff with a robust risk assessment tool for lot scheduling.
- Develop an efficient multi-criteria scheduling model.
- Demonstrate the feasibility of integrating simulation based models and AI techniques to provide effective scheduling.

 To do this, the model must provide the user with different performance measures such as average tool utilization, throughput time, WIP in front of tools, number of mask changes, and tool cycle time. This output allows the production control staff to understand the load distribution in the toolset for different production scenarios. A number of parameters are used as inputs to the model to define the scenarios and include product mix, output demand, scheduled maintenance interval, tool buffer capacity and the qualification matrix.

5 SCHEDULING METHODOLOGY

The main phases of building a simulation model have been discussed in many references (e.g. Law 2003; Banks *et. al* 2001). Table 1 shows the main phases of the model development including the integration phase for the Neural Network (NN) module and the design of experiments.

 Data collection is a crucial step and hence time must be spent to ensure the validity of data. The required data is stored in an associated database rather than hard-coded into the model. The information gathered included, but was not limited to, equipment run rates, initial setup times, mask change setup time, equipment loading and unloading times, material handling times, and equipment qualification based on layer (see Section 3.1). Maintenance has been classified into; preventive maintenance which occurs at constant frequencies and unscheduled breakdowns (common on these tools). The model has specified probability distributions for mean time to failure as well as the mean time to repair based on available historic data. These assumptions were documented within the conceptual modeling phase. Prior manufacturing experience and lot loading within the floor was obtained from production engineers in to order to set up the model framework.

 In order to make effective use of simulation in manufacturing systems, it is often helpful to develop a simple, intuitive model that describes the subsystem elements and the relationships among the elements in the simulation model. This project used the Integrated computer aided manufacturing DEFinition (IDEF0) to gain an understanding of the complex systems on a heuristic basis. IDEF0 was selected as it offers a structured top-down approach, is simple to use and provides a good means of describing the functional processes within a manufacturing environment. The next step involved the 'art of modeling' as products, process, equipment, constraints, assumptions and objectives together with all the related information were analyzed and prepared for coding. The simulation model was coded from the conceptual model, Figure 4, using commercial event based simulation software (Extend 2003).

5.1 Selection Criteria Evaluation

For decision making in the model some weight must be assigned to each of the selection criteria, based on their importance, before the scheduler can be run to minimize particular process measures. The following order of evaluating these criteria was established in close consultation with the manufacturing team.

- Qualifying Matrix
- Maintenance (Scheduled $&$ Unscheduled)
- Tool status
- Buffer status
- Lot priority
- Previous layer & product
- Previous layer
- Previous product

This means the selection of the best tool to process a lot will be made by evaluating a score for each tool based on evaluation of the above criteria in order. First, the selected tool must qualify to build the layer required. Second, it should not be busy or in maintenance, either preventive or unscheduled. Once the tool is identified as available scores are determined by the buffer status in front of the tool, the priority of the lot and the properties of the previous lot processed by the tool to reduce tooling reconfiguration.

Arisha and Young

Figure 4: Schematic Diagram of the Lot Flow in the Model

5.2 Selection Algorithm

The model uses a weighted-score approach for evaluating the possible alternatives in the following manner:

Assume an incoming lot O_{ij} has to process layer '*j*' for product '*i*', where $i = 1, 2, ..., n_i$, and $j = 1, 2, 3, ..., n_j$. The scheduling problem is to assign a specific tool to process this lot. The tool with highest score is the optimum for the selected criteria, as shown in the formula below;

$$
Max\big|S_m\big|_{m=1}^{m=N_{QT}}
$$

where, S_m is the score of tool '*m*' and N_{QT} is the number of qualified tools for this layer.

The score of tool '*m*' can be calculated based on the following equations:

$$
S_m = \sum_{c=1}^n \sum_{r=1}^r K_c \pi_r
$$

where, πr is the weight of the selection criterion 'r', (e.g. r = Maintenance), n is the number of selection criteria, and Kc is binary variable $(0 -1)$ to set if the criterion is applied (1) or not (0). The neural network is used to set the weights assigned to each criteria.

5.3 Neural Network Module

A neural network was constructed to map the scheduling selection criteria, lot sequence, priority setting, etc. to a schedule rule that gives the best performance measure under the imposed constraints. The NN has been designed and trained in conjunction with simulation outputs as shown in Figure 5.

The outputs are the weights of the selection criteria to be used in the simulation to provide the best expected per-

Figure 5: NN Module Integrated with Simulation Model

formance, when measured by a particular performance measure (e.g. throughput time), for the system to process a given sequence of lots arriving at the photolithography toolset. Hence, the number of output nodes corresponds to the number of selection criteria used. The neural network was trained using various production scenarios, the training goal being to match model throughput time with existing experimental data (Figure 6).

Figure 6: Neural Network Training Output

 The neural network outputs also provide a ranking of the relative importance of each selection criterion (Figure 7). While not shown here, the weights and importance were found to vary with the particular optimization measure used, in line with expectations.

Figure 7: NN Module Output

5.4 Model Constraints

The developed model is limited by certain assumptions which are in line with practical operation of the toolset.

- 1. The maximum number of steps in any operation is 13.
- 2. The maximum number of tools is 30.
- 3. The maximum number of layers to be processed is 13.
- 4. An operator is always available.
- 5. The qualification matrix is two-dimensional.
- 6. The photolithography masks are always available.

The performance measure used here, throughput time, is the elapsed time between the lot first entering the toolset and completion of the photolithography process on the last layer.

6 VERIFICATION AND VALIDATION

The strength of decisions made on the basis of simulation is a direct function of the validity of the output data. It is evident that validation must therefore be an integral part of building any simulation model, right from input data collection through model development to output data analysis. The goal of the approach undertaken for this model, outlined below, was to verify that the outputs from the model were valid and directly useful in the FAB. A number of approaches were combined to confirm the status of the model.

The first verification approach could be called, in quality assurance terms, an 'internal audit'. The software used for simulation produces a trace file, which consists of detailed output representing the step-by-step progress of the model over time, allowing detection of subtle errors. The trace file showed that some stations had higher utilization values than would be expected. These numbers do not appear directly in the overall model output, but would influence the results. To ensure that these times were not overlooked they were checked by people other than the modeler to confirm that the correct logic was followed for each step.

 The overall output from the model was then checked for reasonableness, similar to an 'external audit', by production staff. Finally, the most definitive test established that the simulation output data closely resembled the data from the actual system. A set of different run parameters from the factory floor were provided and simulated to ensure that the throughput time and cycle time levels were close to the actual values. This confirms the belief that the logic and assumptions in the model are correct. Figure 8 shows the comparison of model output with the actual data from the FAB for seven different scenarios and shows a maximum deviation of 4.9% (considerably better than other models run under the same scenarios).

Figure 8: Simulation vs. Actual Data

7 RESULTS

Having developed a model which closely simulates the performance of the actual system, experiments can be carried out using the model and used to optimize the performance of the factory without interrupting the production flow. In this study, the experiments were conducted based on orthogonal design and Taguchi methods (Arisha 2003). The input parameters (factors) and their variations (levels) as well as the performance measures were determined by the manufacturing teams (Table 3).

 Each level contains a value based on the recommendations from manufacturing personnel. For example, the stepper buffer capacity for a tool is selected as one of the following four sizes: 1, 3, 5, or 8. The scheduling criteria has only two levels which switch the intelligent-agent on or off, allowing an evaluation to be made of the effect of operator knowledge/experience on the performance of the system. Five different dispatching rules (e.g. First Come First Serve (FCFS), Wafer with Highest Layer Number First (W-HLNF)) were considered.

 The Taguchi method allows fewer experiments to be conducted while still obtaining the statistical significance and the near optimum levels for each factor. The experimental procedure is beyond the scope of this paper but further detail can be found in (Arisha *et al*. 2003).

Based on the Analysis of Variance (ANOVA) Table 4, the main control parameters (i.e. SC, PM and DR) have a statistically significant impact on the throughput time. In contrast, parameters such as wafer starts and stepper buffer capacity which are often adjusted on the shop floor to improve performance, are not seen to be statistically significant.

Table 4: Analysis of Variance Matrix

Factor	DOF	SSB	SSB/DOF	F_{cal}
SC		0.621	0.621	28.852
PM		0.329	0.082	3.821
WS		0.068	0.017	0.791
B	$\mathbf{3}$	0.107	0.036	1.654
DR		2.129	0.532	24.714
Error	15	(0.323)	(0.022)	
Total	31	4.47758		

The results suggest that experimentation should focus attention on the alternatives available for the product-mix and selection criteria, and only then the other parameters, to improve the global shop performance.

More production scenarios under various production demands and product mix have been conducted to compare the performance of the system under standard operation with the predicted performance using the intelligent scheduler. Figure 9 shows a consistent reduction in throughput time when using the intelligent lot scheduling. WIP build in front of the tools has also been reduced, Figure 10. It is worth mentioning that the number of mask changes required to complete the production is also reduced, saving an average of 18% in set-up times for the tools.

These model results were promising enough for the manufacturing staff to implement the model for scheduling in the FAB itself.

8 CONCLUSIONS

Scheduling within the semiconductor industry is a very challenging activity. From manufacturers and researchers'

Figure 9: Comparison of Throughput Times (TPT) with and without Intelligent Photolithography Scheduling (IPS) for Different Production Scenarios

Figure 10: WIP Reduction due to Intelligent Scheduling

observations, the complexity of this activity will increase in the future with:

- Increased global competition.
- Variations in customer demands.
- Decreasing product life cycle.
- Rapid changes in technologies.
- constraints (e.g. technological, quality, and production).

The simulation developed here has been effectively used in scheduling of these complex processes in the photolithography area. Developing such effective models incorporating all the process details, operating details, and manufacturing procedure details for scheduling is extremely complex. Well-thought out hybrid models based on simulation and neural networks can be used to predict and examine the performance of the photolithography process as well as the impact of various production parameters on that performance. A good simulation model provides not only numerical measures of system performance, but provides insight into system performance (Carson 2003). The new model provides a number of interesting insights into the performance benefits from a tacit understanding of system behavior. As one would expect the greatest benefit is obtained from improvements at the lot throughput time and average WIP reduction.

Applying Intelligent Scheduling has a significant effect on improving the lot distribution across the tools taking into consideration that uniform lot distribution across the tools is impossible due to the high variability in the system (e.g. qualifying matrix, product-mix, and unscheduled maintenance). In addition, the number of mask changes required to complete a specific production order is also reduced compared to the situation where lots are assigned manually on the shop floor.

ACKNOWLEDGMENTS

The authors wish to thank the industrial partner for their dedicated support during the project. Special appreciation is expressed to *Mr. Ger Ryan*, *Mr. Mike O'Dwyer*, and *Mr. Frank Turpin.*

REFERENCES

- Akcalt, E., K. Nemoto, and R. Uzsoy 2001. Cycle-Time improvements for Photolithography process in semiconductor manufacturing. *IEEE Trans. Semiconductor Manufacturing* 14(1), 48-56.
- Arisha, A. 2003. Intelligent Shop Scheduling for Semiconductor Manufacturing. Doctoral dissertation, School of Mechanical and Manufacturing Engineering, Dublin City University, Dublin, Ireland.
- Arisha, A. P. Young, and M. El Baradie. 2003. Evaluation of Scheduling Strategies on the Performance of a Flexible Manufacturing Cell - A Simulation Study. In *Proceedings of the 32nd ICC&IE Conference*, Limerick, Ireland, 700 –706.
- Arisha, A., P. Young, and M. El Baradie 2002. Flow Shop Scheduling Problem: A Computational Study. *Proceedings of the Sixth International Conference on Production Engineering and Design for Development (PEDD6)*, Cairo, Egypt, 543 – 557.
- Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2001. *Discrete Event Simulation*, Third Edition, Prentice-Hall, Upper Saddle River, N.J.
- Carson, John S. 2003. Introduction to modeling and simulation. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P.J. Sanchez, D. Ferrin, and D.J. Morrice, 7–13. Piscataway, NJ: IEEE
- Collins, Donald. 2002. Interdisciplinary Research on Modeling and Scheduling of Semiconductor Manufacturing Operations. Available online via <www.acadzinc. com/AASEEInterdisciplinarySSERC.htm> [accessed Sept 09, 2002].
- EXTEND v 6.0 simulation software, **</mark>** inmaginethatinc.com>, [accessed May 18, 2004].
- Hunter, J., D. Delp, D. Collins, and J. Si. 2002. Undersanding a Semiconductor Process Using a Full-Scale Model. *IEEE Trans. Semiconductor Manufacturing* 15(2), 285-289.
- Ignizio, J. 2002. A generic toolset simulation template. *Intel Corporation*, Feb. 2002*.*
- Ignizio, J. 2004. Optimal maintenance headcount allocation: an application of Chebyshev Goal Programming. *Int. J. Prod. Res.* 42(1), 201-210.
- Law, Averill. 2003. How to conduct a successful simulation study. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P.J. Sanchez, D. Ferrin, and D.J. Morrice, 66 – 70. Piscataway, NJ: IEEE
- Nanyani, N. and M. Mollaghasemi 1998. Validation and verification of the simulation model of a photolithography process in semiconductor manufacturing. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D.J. Mederios, E.F. Watson, J.S. Carson, and M.S. Manivannan, 1017–1022. Piscataway, NJ: IEEE

AUTHOR BIOGRAPHIES

AMR ARISHA, Ph.D., B.Sc. is a senior researcher with the Industrial and Manufacturing Engineering Research group in Dublin City University, sponsored by Intel-Ireland. He received his Ph.D. in Manufacturing Engineering from Dublin City University. Dr. Arisha has published over 12 journal and international conference articles in the area of manufacturing systems simulation and scheduling. His research interests include analysis and optimization of manufacturing systems using stochastic simulation modeling, and Artificial Intelligence techniques. His research includes project management, virtual manufacturing, Taguchi methods for experiments, and quality management. He is a member in IIE, IMECH, IEI, ESE, and ASME. His e-mail address is <amr.arisha@dcu.ie> and his web address is http://webpages.dcu.ie/~arishaa/

PAUL YOUNG, B.A. B.A.I., Ph.D., is a senior lecturer in the School of Mechanical and Manufacturing Engineering in Dublin City University (DCU) and Head of the Industrial and Manufacturing Engineering (IME) Research group. He joined DCU in 1998 having spent three years in University College Dublin (UCD) as a postdoctoral researcher in Manufacturing Science. He spent 4 years as a project engineer consulting in automotive NVH with ESTECH Corporation, (Yokohama, Japan) following his PhD studies in Machine Tool Vibration in Trinity College Dublin. His research interests include simulation and optimization of manufacturing systems as well as measurement, simulation and optimization of the performance of mechanical systems. His e-mail address is <Paul.Young@dcu.ie> and his web address is <http://webpages.dcu.ie/~youngpa>.