Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

AUTOMATED RETICLE HANDLING: A COMPARISON OF DISTRIBUTED AND CENTRALIZED RETICLE STORAGE AND TRANSPORT

Anne M. Murray

Hudson Valley Research Park
2070 Route 52 (M/S 56A)
IBM Microelectronics
Hopewell Junction, NY 12603, U.S.A.

David J. Miller

1000 River Road
M/S 967K
IBM Microelectronics
Essex Junction, VT 05452, U.S.A.

ABSTRACT

The implementation of Automated Material Handling Systems (AMHS) in 300mm semiconductor facilities provides the opportunity to realize significant benefits in fabricator productivity and performance. The leverage associated with automated reticle delivery to photolithography process tools may be less apparent than a fab-wide AMHS. However, a high product mix environment requires the tracking, storage and transportation of thousands of reticles to successfully process wafers on photolithography tools. The failure to deliver reticles in an accurate and timely manner will negate many of the competitive advantages associated with automated wafer handling. Implementing an automated reticle management system (ARMS) requires an evolution from traditional reticle storage and management methodologies. In this paper, we review the application of simulation analysis to explore centralized versus distributed reticle storage and handling alternatives for an overall ARMS strategy.

1 INTRODUCTION

The cost of a 300mm fab can exceed $2 billion. Technology development costs can add hundreds of millions of dollars to the total investment cost of a new product in a leading edge facility. It is critical to maximize the return on that investment. Several papers (McIlvaine 1999; Rust, Wright, and Shopbell 2002; Pillai et al. 1999) have documented the ability of an Automated Material Handling System (AMHS) to maximize equipment productivity throughout a fabricator with timely and accurate wafer delivery.

However, in the critical photolithography sectors, a wafer AMHS only addresses a portion of the material handling requirement. The correct reticle must be available in the desired photolithography tool with proper timing relative to the lot arrival. In order to ensure high throughput in photolithography areas, reticle deliveries must be planned and executed in parallel with automated lot deliveries. It becomes increasingly difficult to meet this requirement as the number of reticles that are controlled and delivered grows (Murray and Miller 2002, Lambson, Choudhury, and Davis 1996).

IBM's initial 300mm fab was designed to handle a high product mix environment with potentially tens of thousands of unique reticles to support a wide diversity of customers and applications (Campbell, Rohan, and MacNair 1999). As part of the fab design process, it was necessary to consider an automated solution to the reticle management problem. The objective was to define a cost-effective Automated Reticle Management System (ARMS) that provided the ability to track and instantly locate reticles throughout the facility, and then deliver them to the photolithography tools to maximize wafer processing.

Simulation was employed to quantify the effect of various reticle management options, including storage, transport and delivery considerations (Miller, 1994). The results played a valuable role in making knowledgeable ARMS design, layout and implementation decisions. In this paper, we discuss some of the simulations employed to help us define and quantify potential reticle management solutions. We conclude with a description of the ARMS solution final design.

2 APPROACH

There are a number of considerations associated with a reticle management solution. In order to assess the cost effectiveness of potential ARMS solutions, one must define, analyze and quantify the various options and solutions available and answer those questions including:

- How many reticles must be managed (stored and tracked) to meet fab requirements for volumes, product and part number mix, etc.?
- How many reticles must be transported to and from photolithography tools in a given time period?
- Where should reticle storage be placed, given the constraints of available space within (or outside)

the facility and fab throughput and cycle time requirements?

- How should reticles be stored: bare glass, single reticle per carrier, or multiple reticles per carrier? What are the implications of each (e.g., empty carrier management, kitting, etc.) on throughput and cycle time?
- What performance metrics can be achieved by the various potential solutions and which provides the best overall cost of ownership result for the fab considering investment costs, operating costs, capacity, cycle time, etc.?

Simulation was one of the primary methods used to help analyze and answer a number of these questions, including:

- Analysis of a tool-centralized approach to determine reticle transfer rates. Inputs included assumptions concerning reticle exchange frequency per wafer pass, lot pass or carrier pass; photolithography tool OEE; average lot size; and average number of reticles per lot
- Analysis of distributed reticle storage versus centralized storage options
- Analysis of bare reticle stockers versus pod stockers
- Analysis of single reticle pod strategy versus multi-reticle pod to determine effect on ARMS traffic loading
- Iterative sizings to determine ‘optimum’ number of vehicles, delivery track paths, number of stockers, number of storage bins, number of interface ports, and other equipment solutions
- Ongoing analysis of reticle management optimization via RTD rule simulation and analysis.

3 DISTRIBUTED VERSUS CENTRALIZED RETICLE STORAGE SIMULATION

This paper reviews the analysis of distributed reticle storage (that is, reticle storage co-located in photolithography bays) versus centralized (that is, storage primarily in one remote area) storage and handling options. A centralized, remote storage option is attractive because it frees up additional floor space in the photolithography bays, allowing the installation of additional photolithography equipment. However, there are concerns that a remote storage solution may not be able to provide reticles in a timely manner, thus starving the photolithography tools and negating any advantage associated with additional tools. A distributed storage option has the advantage of reticles being close to the photolithography tools and thus minimizes the time to transfer reticles to a tool. However, the distributed reticle storage solution requires more process floor space in the photolithography bay than the centralized solution.

We also briefly discuss our analysis of the storage of reticles in pods versus bare reticle storage. There are a number of considerations in this analysis including storage density, reticle protection, and the requirement to load and unload pods for transport and storage. We focus on the third issue in this discussion because of the implications to delivery time and throughputs. This is not meant to reduce the importance of other factors in the analysis, however they are typically outside the scope of simulation analysis.

3.1 Assumptions

Throughout the development of the ARMS strategy and solution, 300mm fab planning assumptions were applied along with potential uplifts to bound the system requirements. However, regardless of mix, the throughput of the photolithography tools ultimately determines how quickly reticles can be delivered and exchanged at the tools. Thus, we employed a photolithography tool centric approach to define the reticle exchange rate at the tool and the required ARMS cycle time. Using a scenario that a different reticle delivery was needed at the tool for each lot delivery provided one upper bound on our ARMS system’s required delivery rate. Other cases including multiple part numbers per FOUP and multiple reticles per wafer exposure were also considered.

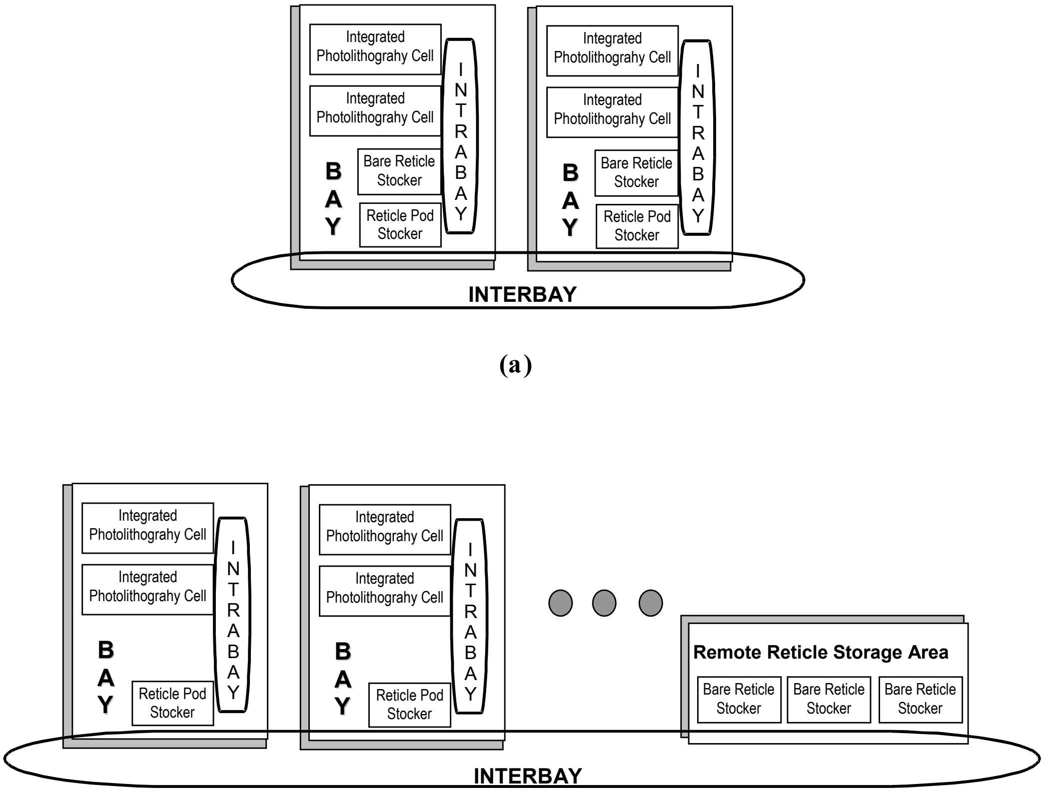
The tool’s required reticle exchange rate was determined using the tool’s throughput OEE, average lot size, average part numbers per lot, and average number of reticles per layer. After an individual tool’s reticle exchange rate was determined, then this rate was used across a bay of tools and propagated through the ARMS components. For one analysis, we determined a rate of 5 incoming reticles per hour per tool. Using this exchange rate, we then determined the required move rate for the entire ARMS system.

We then proceeded to compare reticle storage options given reticle delivery requirements. We could either place high-density reticle storage in a remote, but centralized location at one end of the building with limited capacity reticle pod stockers within each photolithography bay, or we could place the high-density reticle storage with the reticle pod stockers within each photolithography bay (see Figure 1).

In both systems, the reticle was transported in a SEMI E-109 Reticle SMIF pod. There was one critical detail in this layout that worked to the advantage of the ARMS system: like photolithography tools were grouped together in each bay. Thus only a small percentage of reticles would be exchanged between the bays.

3.2 Simulation Description

An AutoMod ^{TM} (version 10) model of the ARMS system was developed for a performance analysis of each of the two storage strategies. Most of the basic constructs within the AutoMod^{TM} system were used to define the ARMS system layout and components. Proprietary reticle transport system dispatching logic and vehicle control



(a)

(b)

Figure 1: (a) Distributed and (b) Centralized Automated Reticle Management Systems

logic was employed to model the specific behaviors of the reticle transport system. The model incorporated previously verified vehicle and stocker logic that included precise descriptions of operation including speeds, interfaces, and control logic.

Inputs to the model included transfer requests, physical track layout including travel paths and distances, and available resources including vehicles, stockers and interface ports. The reticle pod movements (or transfer requests) were exponentially distributed. A delivery was initiated only if the target equipments' loadport was available. The model monitored the availability of the loadports on both reticle stockers and photolithography tools and randomly chose the destination loadport from those available. These loadports had equal probability of being selected.

The possible reticle pod source and destination movement combinations varied for the two different storage strategies considered. For the centralized remote storage option, the possible movements included:

- Remote Storage Center Bare and/or Pod Stockers to/from Bay Pod Stockers
- Bay Pod Stocker to/from Photolithography Tool
- Bay Pod Stocker to/from Bay Pod Stocker in Another Bay.

The busier travel routes were between the Remote Storage Center Stockers and the Bay Stockers and between the Bay Pod Stockers and the photolithography tools. The between bay moves were limited.

The possible movement combinations for the distributed system where reticles were stored primarily in the photolithography bays were fairly more complicated and included:

- Remote Storage Center Bare Stocker and/or Pod Stocker to/from Bay Pod Stockers (for reticle introduction or removal to or from the fabricator)
- Bay Pod Stocker to/from Photolithography Tool
- Bay Pod Stocker to/from Bay Bare Reticle Stocker to Photolithography Tool
- Bay Pod Stocker to/from Bay Pod Stocker in Another Bay.

The busiest travel route was between the Bay Pod Stocker and the Photolithography tool. Since reticle storage was co-located with like photolithography tools, a small percentage of moves were generated between bays and between the remote storage center and the bays. One early finding was the advantage of using a unidirectional flow through bare reticle stockers to minimize the likelihood of deadlock at the bare reticle stockers' two loadports. That is, by allowing reticles to move directly from the bare reticle stocker to the photolithography tool but not in reverse, minimized the contention at the bare reticle stockers loadports.

A combination of standard AutoMod$^{\text{TM}}$ model performance statistics and custom statistics were used to characterize the models' performance. The standard metrics used were:

- Vehicle utilization
- Stocker cycle time and utilization
- Stocker component utilization (e.g.,loadport conveyor).

The custom metrics used were:

- Number of reticle transfers made per loop
- Number of reticle deliveries made to photolithography tools in each bay and the associated cycle time (mean, maximum, minimum, standard deviation)
- Interbay (Bay-to-Bay) Transfer Time (mean, maximum, minimum, standard deviation)
- Pod Pick-Up Latency. Time between pod retrieval request and vehicle pickup. (mean, maximum, minimum, standard deviation).

3.3 Simulation Results and Discussion

For each simulation run, there was a 2 hour startup period to reach a steady state condition. After that point, data were collected for at least 10 hours for each run. A number of simulation runs were performed to review an array of potential alternative layouts, track routes, storage capacities, and stocker locations. These were summarized into a comparison of the best performing solutions for the two primary options (see Table 1).

An initial glance at these summary statistics shows that the distributed system is superior to the centralized solution by approximately 10% on the key metric of mean total delivery time. Additional details reveal further advantages of the distributed system.

In the distributed system, 88.7% of the reticles resided in the correct lithography bay when required for processing. In contrast, only 20.8% of reticles were located in the correct photolithography bay in the centralized case.

This difference drove significantly higher volumes on the interbay transport portion of the centralized system, to the point where the remote storage option case required a dedicated interbay system with dual-pod-capacity overhead vehicles (OHV's) to attain required throughputs. The distributed system, however, was able to utilize a single-pod-capacity overhead hoist (OHT) vehicle for both interbay and intrabay transport, simplifying the overall system including control, maintenance and support benefits.

The frequency of interbay moves associated with the centralized system contributed an average of 3.32 minutes per reticle delivery and required that reticles be stored in reticle pod stockers in the photolithography bays for pick up and delivery to achieve desired overall delivery times. The distributed solution was able to utilize dense bare reticle storage in the photolithography bays to maximize the number of reticles in a minimum footprint and still outperform the centralized system, since a transfer of a reticle into a pod could be accomplished in less time than an interbay move.

So while the average interbay delivery time was lower using the dual capacity vehicles in the centralized system (4.2 vs. 5.6 minutes), a much higher percentage of reticles require interbay delivery (79.2% to 11.3%) and thus the resulting impact on average total reticle delivery time to the photolithography tools was over 5X higher than the distributed system (3.32 minutes to 0.63 minutes). Additionally, the centralized interbay system required significantly more (2.5X) single-pod-capacity OHT vehicles to provide comparable delivery times of the dual-pod-capacity system. However, more OHT vehicles in the centralized case caused considerable traffic congestion and an unacceptable degradation of the interbay delivery times in the centralized ARMS system.

The intrabay delivery statistics showed that, while the average centralized case delivery times were better than that for the distributed case (3.6 minutes versus 5.7 minutes), the difference was associated with the large quantity of linked moves involving the bare reticle stocker. Specifically, the transfer time in the distributed case incorporated the movement of an empty pod from the pod stocker to the bare reticle stocker, the retrieval of the reticle into the pod, and then the transfer of the reticle (now in a pod) to the final destination photolithography tool.

From a pragmatic perspective, the centralized case relies heavily on the ability of the reticle management system

Table 1: A Summary of Simulation Results for the Centralized and Distributed Automated Reticle Management Systems

Model Type	**Mean Interbay Delivery Time (Minutes)**	**Deliveries Requiring Interbay Moves (%)**	**Mean Interbay Time Across all Reticle Deliveries (Minutes)**	**Interbay Vehicle Utilization (%)**	**Interbay Vehicle Count**	**Intrabay Mean Delivery Time (Minutes)**	**Intrabay Vehicle Utilization (%)**	**Intrabay Vehicle Count**	**Mean Total Delivery Time (Minutes)**	**Bay Pod Stocker Utilization (%)**
Centralized	4.2	79.2	3.32	47	15 (OHV)	3.6	55	5 (OHT)	6.92	72
Distributed	5.6	11.3	0.63	44	12 (OHT)	5.7	57	7 (OHT)	6.33	52.6

to effectively use the storage space of the reticle pod stockers by pre-kitting reticles to eliminate the time required to transfer a reticle into a pod. Without this assumption, the centralized solution advantage of delivery time (from the photolithography tool's view) increases to unacceptable levels. The distributed case assumed the use of bare reticle stockers in the flow and thus the pre-kitting reticles into the pod stockers were not a requirement to achieve the performance illustrated in the simulation model. In fact, pre-kitting reticles, will improve the performance of the distributed system. Using these simulation results, their associated analyses, facility and operational considerations, and overall cost of ownership, delivery time and throughput capabilities, the recommendation was to implement a distributed ARMS system to manage and transport over 10,000 reticles.

Our system incorporates three methods of reticle storage: an internal buffer in the photolithography tool for immediate-use reticles; reticle pod stockers for storing not only empty pods but reticles to be used soon for the bay; and bare reticle storage for high-density long-term reticle storage. The reticles are transported to photolithography tools and stockers via an overhead transport system (OHT) that leverages the AMHS OHT vehicle platform. The reticle OHT system is completely independent of the wafer OHT system. The coordination of reticle and wafer deliveries and the interaction of the different types of reticle storage is orchestrated via a host-level, real-time dispatching and scheduling system that encodes the operational rules of the facility.

4 CONCLUSION

The capital investment in photolithography equipment demands that they be fully utilized. We suggest that this cannot be realized in an automated high-product mix 300mm facility without an automated reticle management system (ARMS) to coordinate reticle delivery with the automated wafer delivery system. Using an AutoMod^{TM} simulation environment, we were able to examine centralized and distributed reticle storage and transport designs. The results demonstrated that a distributed solution would provide shorter average delivery times to photolithography tools, would support higher capacity handling throughputs by avoiding the bottleneck inherent in a centralized area design, and would allow a more cost effective solution in terms of transport systems and storage solutions. Based on the results of the simulation analysis, we have elected to implement a distributed ARMS system. Our distributed ARMS system has located bare reticle stockers and pod stockers in situ with the photolithography equipment.

The next important phase of simulation analysis will examine how to optimize the three forms of reticle storage we employ (internal reticle buffer of the photolithography tool, reticle pod stocker, and bare reticle stocker) in balance with the reticle handling system. Conceptually, the photolithography tool's internal buffer should be reserved for reticles that are to be used in the immediate future, the reticle pod stocker for reticles being "pre-staged" for use within a time window of lesser immediacy, and the bare reticle stocker for long-term reticle storage. Our simulation analysis will enable us to determine the extent of the pre-stating time window for the reticle pod stocker and the timing required for managing the introduction and withdrawal of reticles in photolithography tool's buffer. The complexity of reticle scheduling is driven by the fab business model, manufacturing operational rules and reticle usage constraints; however, we must also take into account how these scheduling "rules" influence the effectiveness of the reticle handling system.

ACKNOWLEDGMENTS

The authors would like to recognize the contributions of Masahiro Ikeda of Murata Automated Systems, Inc. for building the AutoMod^{TM} models described in this paper. Additionally, we would like to thank the IBM IMD Automation Team for their stimulating discussions, guidance, and providing a sounding board for this work.

REFERENCES

Campbell, P.L., D. Rohan, and E. MacNair. 1999. A Model of a 300mm Wafer Fabrication Line. *In Proceedings of the 1999 Winter Simulation Conference*.

Lambson, C., M. Choudhury, and R. Davis. 1996. Automated Reticle Transport and Stepper Loading. *Solid State Technology*, October, 1996: 97-105.

McIlvaine, B. 1999. Automation is Essential for Next-Generation Fab. *Semiconductor Business Weekly*. July 1999.

Miller, D. 1994. The Role of Simulation in Semiconductor Logistics. *In the Proceedings of the 1994 Winter Simulation Conference*: 885-891.

Murray, A., D. Miller. 2002. The Next Step in 300mm Automation: Automated Reticle Handling in Photolithography. *In the Proceedings of the 2002 Brooks-PRI Automation Worldwide Automation & Performance Symposium*, October, 2002.

Pillai, D., T. Quinn, K. Kryder, and D. Charlson. 1999. Integration of 300mm Fab layouts and Material Handling Automation. *In the Proceedings of the 1999 Winter Simulation Conference*: 23-26.

Rust, K., R. Wright and M. Shopbell. 2002 Comparative Analysis of 300mm Automated Material Handling Systems (AMHS). *In Proceedings of the International Conference on Modeling and Analysis of Semiconductor Manufacturing*, April 10-12, 2002. 240-245.

AUTHOR BIOGRAPHIES

ANNE M. MURRAY is the project leader of IBM's 300mm Reticle Automation System in East Fishkill, NY. She has been with the 300mm Automation group for the past 4 years. Prior to her work with IBM, Anne received a Ph.D. in Electrical and Computer Engineering at Carnegie Mellon University in Pittsburgh, PA for her work in telemanipulation systems. She may be contacted via email at `ammurray@us.ibm.com`.

DAVID J. MILLER is a Senior Technical Staff Member for IBM Microelectronics. His interests include semiconductor factory automation, material handling systems and discrete simulation. He received a BS in Mathematics from Georgetown University in Washington, DC, an MS in Computer Science from the University of Vermont in Burlington, VT, and an MS in Manufacturing Systems Engineering from Lehigh University in Bethlehem, PA. He may be contacted via email at `davemill@us.ibm.com`.