THE SIMULATION OF INTEGRATED TOOL PERFORMANCE
IN SEMICONDUCTOR MANUFACTURING

John L Mauer
Geer Mountain Software Corp.
South Kent, Ct 06785

Roland E. A. Schelasin
IBM Technology Products
Essex Junction, Vt 05452

ABSTRACT

Integrated tools in semiconductor manufacturing have become increasingly complex; the throughput and cycle time are no longer easily related to the individual process times. Detailed simulation models of these tools provide a means to evaluate the various performance characteristics. Further, proper planning for integrated tools requires flexible simulation models in the hands of line engineers themselves.

1 INTRODUCTION

Semiconductor manufacturing equipment used to process a batch of silicon wafers through just one process operation. Because the production of electronic chips now requires hundreds of these process operations, many sequential operations have been combined in integrated tools, commonly known as cluster tools. The use of such cluster tools in semiconductor manufacturing has increased rapidly over the last few years. (Newboe 1990, McNab 1990, Bergendahl et. al. 1990, Larrabee 1991, Wood et. al. 1991, Mauer et. al. 1992). With the improvement of wafer handling equipment and the development of reliable automation, not only have simple sequential operations been combined but also parallel processing has been introduced at bottleneck steps.

These tools are sufficiently complex that a manufacturing engineer can no longer predict the tool's performance from simple measurements of process times. The engineer must either take a manufacturing tool off-line to measure its response or guess from existing data in order to get values for cycle time and throughput. Detailed simulation models of these integrated tools provide a better means of checking tool capacity especially when process, tool, or manufacturing line changes take place (Mauer et. al.).

In particular, this paper describes the discrete event simulation models of two tools: a photolithography cluster used to pattern thin films on wafers, and a parallel processing wet bench used primarily to clean wafers. These models are built using commercial simulation software, ProModel, and run on a PC.

Indeed, the typical manufacturing engineer has sufficient computing power on his desk to run these models himself, thus significantly reducing the learning time. A simulation expert is no longer needed to provide information, but rather to provide a flexible reusable model built as a sophisticated calculator. Our goal was to provide line engineers with their own flexible simulation models.

2 ASYNCHRONOUS INTEGRATED TOOLS

An integrated tool, or cluster, is made of three principal parts: process chambers or modules, a central robot(s), and the input/output handling system. The process modules can handle either a whole cassette of wafers, such as the chemical tank of a wet bench, or single wafers, such as the bake module in a photolithography cluster. The robot must service each module and the input/output area according to a specified algorithm.

If the integrated tool can handle only one cassette of wafers at a time, the job flow through the tool is synchronous. The throughput of the tool is still inversely related to the cycle time. However, the cycle time may not be directly related to the wafer process times; the robot may not be able to keep up creating internal queues. In this case, simulation is needed to predict capacity of the tool for changes of process times, process sequence, or module configuration.

If the integrated tool can handle multiple cassettes, each with an independent arrival time, the calculation of tool parameters is non-trivial. Because of this asynchronous behavior, the throughput is no longer inversely related to the cycle time. The jobs interfere with each other. Traditional tool planning methods are

inadequate. In this case, simulation is required not only for changes in the tool but also for changes in line loading and process sequence.

As in a complete manufacturing line, where any one tool can restrict line capacity, any one component of an integrated tool can limit the overall tool performance. The mix of chamber assignments and numbers of parallel modules must be studied carefully to generate a cluster configuration that will optimize overall capacity. Even scheduling within a cluster, wherein the tool software determines the robot response to work requests, can severely limit tool capacity.

In light of this discussion, detailed simulation, rather than experimentation on a real tool, provides an ideal means of study and optimization. These models must provide sufficient flexibility to allow variation of component configuration, robot speeds, scheduling algorithms, and other cluster parameters as appropriate in order to satisfy the real user, our line engineer.

3 FLEXIBLE SIMULATION MODELS

The design of flexible simulation models is somewhat different from the design of simulation models with multiple but well determined routings. An entity or part, (in these models, a silicon wafer or cassette of wafers), must be able to travel from one location to any other, with normal tool restrictions, so that the user can specify the routing at run time. The process routing becomes an attribute of the part, and is assigned to it from a routing array either at entry into the simulation or step by step during the simulation. Since the process times are also variable, these become an attribute of the part as well. While the routing specifies process steps, the configuration of the tool specifies where modules for each process step are physically located within the tool. Thus the internal logic must tie together process and physical location while running. In ProModel, the routing and configuration arrays are initially filled with default values which are replaced, at run time, with the data specified by the user.

The internal design of the model is further complicated by the scheduling of robot response. Each semiconductor tool has a different algorithm for robot service; for some tools, this algorithm is a parameter as changeable as process times. In order to account for all variations of robot scheduling, the demand logic for the robot is contained in hidden locations throughout the models. Processed parts, upon leaving real process locations in the model, typically go to such hidden locations in order to test the robot algorithm. This also requires that statistics for process locations be calculated

independent of the normal statistical package available in the simulation software.

One further note on robot scheduling bears witness to a problem still faced by designers of integrated tools: gridlock! In tools where parts can revisit process modules, a gridlock situation can develop whereby two or more parts demand access to the other's locations. Unless the tool vendor has provided a buffer location and the software to use it, the tool will hang. And some tools do. In spite of the model builder's desire to provide efficient gridlock removal, the tool must be modeled according to the tool vendor's design.

In contrast to the complexity of the internal design of the model, the external appearance needs to be simpler. The user is a line engineer who will not patiently edit text files let alone debug models made unworkable by his changes. For the models described here, the model interface is the PMI interface to ProModel, which provides a scrollable screen for parameter input prior to running a model. Yet even this interface, originally provided to ease modeler experimentation, does not yet provide sufficient ease of use. Still, the availability of simulation models at all is a welcome relief for most manufacturing engineers.

4 PHOTOLITHOGRAPHY CLUSTER

Photolithography in semiconductor manufacturing is the process of coating wafers with a thin film of light sensitive material, exposing that material to a pattern of light, and developing that pattern such that holes are left in the thin film. Subsequent processing effects the underlying wafer. Photolithography clusters tend to be very modular.
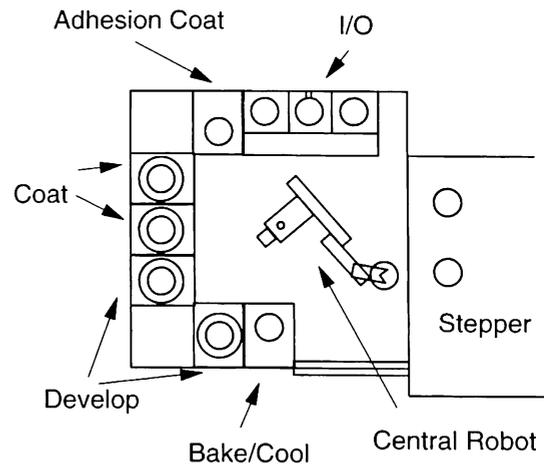

Figure 1: schematic of integrated photo cluster

thruput (wafers/hr)

normalized process time

cycle time (minutes)
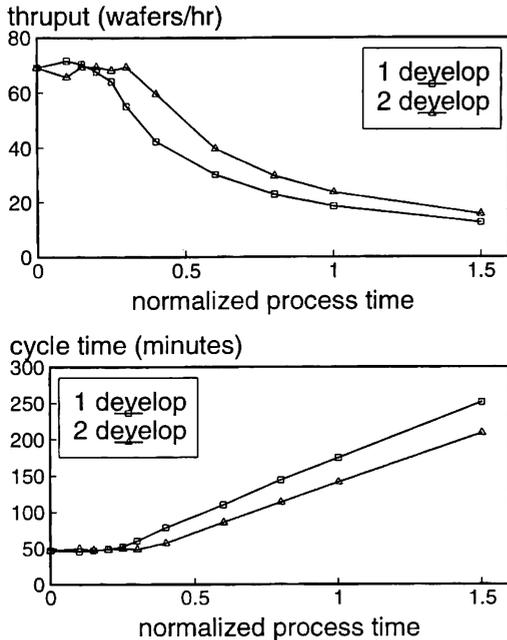

normalized process time

Figure 2: capacity of photo cluster

In particular, we modeled a photo cluster with a central multi-axis robot. The modules are arranged about the robot in a C shape with an optical camera called a stepper occupying the remaining side as shown in Figure 1. One module location is reserved for physical access into the tool and two for cassette input/output and wafer storage. The remaining six locations can be used for processing which includes vertical stacking of process modules. Although the cluster schematic shows one specific layout, both the tool and the model allow the user to assign different process modules to any location, and to create module stacks where system constraints allow it.

The process routing for each batch of wafers can specify any sequence through the system; the model uses backup modules as the real system does. In this way, various configurations were tested for validation of the model both at the equipment vendor and in IBM manufacturing lines. The results were within +/-5%.

The model found immediate application on a manufacturing line where, for a given process routing, the modules of the photo cluster was gating delivery of wafers to the stepper. One additional cluster/stepper would have been required to obtain the needed improvement in the capacity of the photo sector. However, simulation of the existing tools showed that the addition of one develop module to each existing cluster would provide the necessary increase in capacity, as shown in Figure 2. This solution was the most effective, saving both time and money.

5 INTEGRATED WET BENCH

A wet bench in semiconductor manufacturing is a series of tanks filled with chemicals or water rinses designed to clean or etch wafers. An integrated wet bench is shown in Figure 3. In practice the tool can be made larger or smaller by design, but the initial simulation model was done for this configuration.

The process tanks are divided into 4 chemical processing tanks in the rear and 4 quick dump rinse tanks in the front of the system. The chemical tanks may be designated as safe or unsafe meaning that the cassette of wafers can be left in the tank for extended times or must be removed at a specific time, respectively. A central robot, operating off the front of the system, delivers the cassettes from tank to tank, to the final rinse, to the dryer, and from/to the load/unload area. In general the robot is forced to stay with cassettes in unsafe tanks if the service time for the next robot operation is too long, but may leave cassettes in safe tanks. The robot is forced to wash itself in a special arm wash after each exposure to chemicals.

The chemicals each have a lifetime before they need to be dumped and replaced automatically; some lifetimes are as short as 90 minutes with replacement taking 20-30 minutes due to bath temperature stabilization.

The load/unload area has 4 chemically inert cassettes for processing. In this area, a separate mechanism transfers the wafers from the traveling cassette to a chemically inert cassette or the reverse when the processing is finished. In this way, 4 cassettes of wafers may be independently processed in an asynchronous manner.

The model was validated against 8 tools in an IBM manufacturing line, and then used to investigate the wet bench tool group for a new semiconductor manufacturing line.
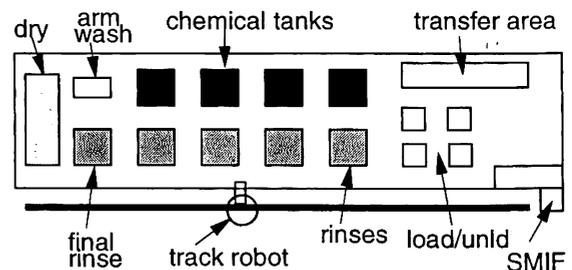

Figure 3: schematic of integrated wet bench

As an example of this use, we chose a tool designed for wafer cleaning. In this case, some of the tanks contain a chemical with a very short lifetime. The quick dump rinses, the final rinse, and the dryer were set to process for representative times, although the rinses are actually variable in practice. The arrival of the runs to the tool was exponentially distributed, with a mean arrival time of slightly greater than one cassette per hour. If we allowed only one job to be processed in the tool at a time, then the average cycle time was 108 minutes. This cycle time represents the raw process time since other jobs did not interfere, but it does include the average effects of the chemical down times.

As we increase the number of jobs allowed in the tool at one time, the cycle time increases but the overall time of the run in the line decreases as shown in Figure 4. The net effect of parallel processing is the reduction of long external queue times by the introduction of short internal queue times. The thruput of the tool also increases with parallel processing to nearly one cassette per hour. Because we chose a flow of cassettes greater than the capacity of the tool, the queue times are a function of the allowed capacity of the WIP station, taken as 10 cassettes maximum. With this constraint, the actual run time of the job goes from 11 times the raw process time to 6.5 with increasing parallel processing.
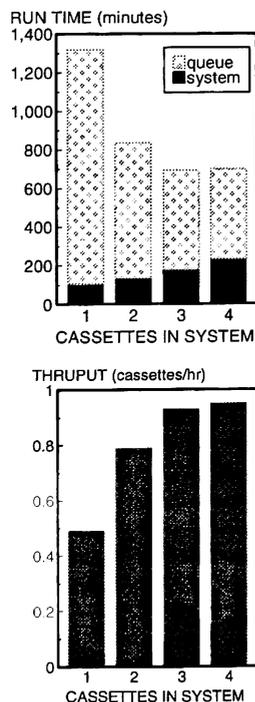
6 CONCLUSION

The increasing integration of semiconductor process tools has led to a need for tool simulation in order to calculate the thruput and the cycle time. With central robot handling and parallel processing of wafers the cycle time is no longer a linear function of the process time. The asynchronous processing of cassettes of wafers decouples the normal relationship between thruput and cycle times. The cycle time in the tool now depends on line loading. Furthermore, some of the external queue time is replaced by shorter internal queue times leading to shorter run times as well as higher thruput.

Flexible simulation models, as described here, provide the understanding necessary to optimize the use of such integrated tools. By putting these models in the hands of equipment and process engineers, we have reduced the time necessary to improve the use of integrated semiconductor tools. The results of such models have enhanced the status of simulation as an engineering tool.

REFERENCES

Bergendahl, Albert S., David V. Horak, Paul E. Bakeman and David J. Miller, 1990. "Cluster Tools, Part 2: 16Mb DRAM Processing", *Semiconductor International*, September

Larrabee, Graydon B., 1991. "The Intelligent Microelectronics Factory of the Future", in *Proceedings of IEEE/SEMI Int'l Semiconductor Manufacturing Science Symposium*, Burlingame, CA, May, p30-34

Mauer, John L., Roland E. A. Schelasin and Peter J. Miller, "Analysis of Cluster Tool Performance in Semiconductor Manufacturing" *13th International Electronics Manufacturing Technology Symposium*, September 1992, p129

McNab, T. Kieth, 1990. "Cluster Tools, Part 1: Emerging Processes", *Semiconductor International*, August

McNab, T. Kieth, 1990. "Cluster Tools, Part 3: Technical Issues", *Semiconductor International*, October

Newboe, B., 1990. "Cluster Tools: A Process Solution?", *Semiconductor International*, July

Figure 4: capacity of integrated wet bench

Wood, Samuel C. and Krishna C. Saraswat, 1991. "Modeling the Performance of Cluster-Based Fabs", in *Proceedings of IEEE/SEMI Int'l Semiconductor Manufacturing Science Symposium*, Burlingame CA, May, p8-14

AUTHOR BIOGRAPHIES

John Mauer received an S.B. degree in Physics from MIT in 1967 and a Ph.D in Engineering and Applied Science from Yale University in 1972. Since joining IBM in 1973, he has worked on electron beam systems for semiconductor lithography, plasma processing and semiconductor device process development. Other projects include semiconductor line modeling, yield analysis, planarization, and device isolation. From, 1987 to 1991, he worked on focused ion beam applications and x-ray lithography. Dr. Mauer is currently President of Geer Mountain Software Corp. and works on applications of simulation technology.

Roland Schelasin has a Bachelor's degree in Mechanical Engineering and a Master's degree in Engineering Management from Brigham Young University. He is currently working as a Staff Engineer/Scientist at IBM's semiconductor manufacturing site in Burlington, Vermont. His work experience includes clean room project management, strategic and tactical fabricator planning, competitive analysis, as well as capacity modeling and simulation of integrated semiconductor equipment.