ABSTRACT

Simulation models are useful to predict and understand the impact of changes to a manufacturing system. Typical factory simulation models include the parts being manufactured in the factory and the people and resources processing and handling the parts. However, these models do not include equipment or process details, which can affect operational performance such as cycle time and inventory. Separate models are used to evaluate processes and equipment. Thus, it is difficult to evaluate the operational impact of equipment or process changes. However, this information could help factory managers and manufacturing process engineers make better decisions when changing processes or selecting equipment configurations. This paper describes a heterogeneous simulation environment for understanding how equipment changes affect the performance of a wafer fabrication facility. This integrated tool incorporates response surface models that describe process behavior, operational and optimization models of equipment behavior, and a discrete-event simulation model of factory operations. Thus, the tool can measure how process changes and equipment configuration changes change the system performance. We have applied this tool to a specific wafer fab problem.

1 INTRODUCTION

Understanding the operational impact of equipment and process changes typically requires the expertise of multiple engineers and analysts. Each person uses different models to evaluate some segment of the entire manufacturing system. Thus, significant time and effort is needed to gather information to help factory managers and manufacturing process engineers make better decisions when changing processes or selecting equipment configurations.

For instance, experimental approaches to optimizing and controlling manufacturing processes by changing the process parameters has been very successful (see, for example, Stefani et al. 1996). However, process engineers often focus on the process itself and may find it difficult to consider how process parameter changes affect the overall manufacturing system performance. One significant impact of changing process parameters is a change to the time that a process requires, which affects the total lot processing time (the time needed to process all of the wafers). If the process (or a sequence of processes) is performed by a cluster tool, which can process multiple wafers simultaneously, then the impact of the process change on the total lot processing time may be very complex (see, for instance, Herrmann et al. 1999). However, process engineers usually develop response surface models (RSMs) for process rate (like etch rate or deposition rate). Although a higher rate should reduce the total lot processing time, the quantitative relationship is often complex, involving the consequences not only of the process, but also the overhead associated with startup and ending of the process cycle in the tool. Thus, a small change to the process time may change the total lot processing time drastically, or it may not. Process “improvements” that significantly increase the total lot process time and reduce a tool’s capacity (especially if that tool is a bottleneck tool) can seriously degrade
manufacturing system performance by increasing cycle
time and decreasing maximum throughput.

Consider the following, somewhat exaggerated,
scenario. A process engineer wants to modify a particular
semiconductor manufacturing process in order to improve
process performance. Changing the process parameter
values (or recipe) will change the process performance in
various ways. If the change affects the time needed to
perform the process, the process engineer must determine
if the change will affect the equipment’s ability to satisfy
its throughput requirements (wafers processed per day).
Thus, the process engineer calculates the modified process
time and gives that value to an industrial engineer, who
then determines whether the process change is acceptable.
The industrial engineer uses a capacity planning model to
determine, if the proposed process change did occur,
whether the equipment’s utilization would remain at an
acceptable level, since a very high utilization can cause
excessive delays. Such a utilization constraint, although
practical, is a myopic way to avoid potential problems,
since it does not consider benefits that could occur in other
parts of the factory. Also, the process engineer faces a
delay while the industrial engineer does the capacity
analysis. Furthermore, the declaration that a process
change is unacceptable does not give the process engineer
feedback needed to find a change that satisfies everyone’s
requirements.

Consider also the problem of selecting equipment
configurations for a semiconductor wafer fab. A cluster
tool has integrated processing modules linked
mechanically. Typical cluster tools include load locks,
process modules, and a wafer handler. A cluster tool can
process multiple wafers simultaneously. Sequential cluster
tools integrate a sequence of processes, while other tools
have two or more identical modules that are used in
parallel. Hybrid configuration are also possible. Unlike
single-process tools, the complex behavior of a cluster tool
makes analyzing the throughput of different configurations
a difficult task. Adding a second chamber to a tool does
not automatically halve the total lot processing time (or
double the tool capacity). Understanding the impact of
different tool configurations requires an integrated model
that can describe both the tool behavior and the factory
behavior.

This paper describes a heterogeneous simulation
environment (HSE) that integrates a variety of simulation
and analytical models of the manufacturing system and its
components. Specifically, the HSE incorporates response
surface models that describe process behavior, operational
and optimization models of equipment behavior, and a
discrete-event simulation model of factory operations. The
HSE can measure how process changes, equipment
configuration changes, and tool scheduling changes affect
the manufacturing system performance.

Figure 1 shows the functional flow of the HSE. The
HSE Administrator coordinates the user interface and
actions of the analyst with three types of models: process
models, tool models, and a factory model. The user can
modify the parameters of any model, and the Administrator
updates the other models as needed. Thus, the HSE and
the component models, which represent a range of domain-
specific knowledge, allow the user to accomplish modeling
and analysis tasks that would otherwise require multiple
individuals (and more time) to perform.

Figure 2 illustrates some of the relationships between
the models that compose the HSE in our application
(described more in Sections 2, 3, and 4). Note that the
HSE includes a variety of continuous and discrete-event
models. Our HSE framework allows us to replace any
model by another model of the same process or tool. For
example, we could replace a process RSM by a dynamic
process simulation.
The remainder of this paper is organized as follows: Section 2 describes the problem domain. Section 3 discusses the different models that we created for the individual components of the factory. Section 4 describes the heterogeneous simulation environment that integrates the models and supports analysis tools. Section 5 concludes the paper.

2 APPLICATION

Although our approach can be applied to other problems, we have implemented a HSE for a specific “tungsten plug subfactory” that fills tungsten vias (interconnects) on semiconductor devices. This subfactory is just one part of a typical semiconductor wafer fab. The subfactory includes wet clean, liner deposition, and tungsten (W) deposition. The liner deposition step includes titanium physical vapor deposition (Ti PVD) and titanium nitride physical vapor deposition (TiN PVD). The tungsten deposition is a chemical vapor deposition process (W CVD). A simple wet clean tool performs the first process. A cluster tool performs the Ti and TiN deposition processes. A second cluster tool performs the W CVD process. Each cluster tool also has a chamber that performs an orient-and-degas step.

We have chosen to investigate a factory with cluster tools because semiconductor manufacturers are increasingly using cluster tools. Annual sales of cluster tools is projected to increase from $11.2 billion in 1997 to $21.9 billion in 2000 (Semiconductor Business News 1998).

Many workers have developed and described simulation models of semiconductor manufacturing facilities. Because there are too many to list here, we will just mention, as an example, the FabTime Wafer Fab Cycle Time bibliography available at <http://www.FabTime.com/CTBiblio.htm>.

Unlike single-process tools, the complex behavior of a cluster tool makes it a difficult task to determine the relationship between process changes and manufacturing throughput. There are simulation models that describe cluster tool behavior. See, for instance, LeBaron and Pool (1994), Mauer and Schelasin (1994), Atherton et al. (1990), Schruben (1999), Wood (1994). These tools are very useful for evaluating the performance of a specific tool operating under specific conditions. Most discrete event simulation models do not, however, yield insight into how process changes affect cluster tool performance, since they take fixed values for each process step without describing the relationship between process parameters and process step times.

Herrmann et al. (1999) have shown that integrating a process model with a cluster tool simulation can help an engineer determine if a process change will significantly increase the processing time of a lot (a batch of wafers). Their results reveal how the total lot processing time depends upon process parameters, and their analysis tools provide a mechanism to assess this sensitivity both qualitatively and quantitatively.

Pichler et al. (1999) describe an integrated environment for the simulation of VLSI fabrication processes, with a focus on semiconductor device design, and how the processes determine device structure and performance. It does integrate a variety of process simulation tools, but not for evaluating the impact of process changes or tool configurations and algorithms on manufacturing metrics and their tradeoffs against performance.

3 MODELS

The heterogeneous simulation environment incorporates response surface models that describe process dynamics, optimization models of equipment behavior, and a discrete-event simulation model of factory operations. The following sections describe these models in more detail.

3.1 Process Models

In semiconductor manufacturing, as in other manufacturing environments, a manufacturing process is governed by multiple process parameters. When executing the process, the operator (or the computer controlling the process) sets the process parameters to prescribed settings so that the process will run effectively and efficiently. Determining good settings involves many tradeoffs between such things as product quality, product performance, consumables cost, and nominal processing time. Often it is necessary to change the process parameter settings to improve process performance, to enhance yield through improved compatibility with other process steps (i.e., process integration), to restore process performance after a disturbance, or to shift technology design points in accordance with scaling toward more aggressive technology nodes or intermediate steps between nodes.

When attempting to determine new settings, a process engineer may conduct a set of experiments to evaluate how the process parameters affect the process performance. Each experiment may require one or more lots processed under a specific combination of parameter values. In theory, an engineer could conduct an experiment for every possible combination of parameter values. Since there may exist a large number of possible combinations, however, in practice the engineer selects a small subset of the combinations and runs these experiments. Then, using statistical software (like ECHIP), the engineer can construct a response surface model (RSM) that fits the experimental results. The RSM is an empirical (often quadratic) mathematical formula that relates process performance to the process parameter values. (For more information on designing experiments and forming RSMs, see Box & Draper 1987.) The RSM gives the engineer guidance into how the process parameters affect the process performance. The engineer can then select the
new process parameter settings that best meet the process performance goals.

For the subfactory, we constructed response surface models (RSMs) for the relevant processes. These simple formulas relate process performance measures to the process parameters. Specifically we have constructed RSMs for Ti PVD, TiN PVD, and W CVD.

For Ti PVD, a collaborator sent us experimental data from a process demonstration that was intended to analyze the sputtering-induced damage to metal-semiconductor field-effect transistor (MESFET). From this data we constructed an interaction RSM. For TiN PVD, we constructed a RSM from experimental results reported by Hui et al. (1997). The experimenters deposited TiN by Ion Metal Plasma (IMP) PVD for sub 0.25 µm technology.

For the W CVD process, we used an RSM that was based on data collected by Stefani et al. (1996). The deposition rate RSM has the following four process parameters: reactor pressure, deposition temperature, the mole fraction of WF6, and the mole fraction of H2. The output is the average deposition rate (Å/sec). The process is a H2 reduction of WF6, run in an Applied Materials Centura reactor, preceded by a short SiH4 and WF6 nucleation step that deposits a 400 Å seed layer. We have chosen to investigate this simplified process model initially.

Private communication with Stefani confirmed that the process parameters in his model refer to the steady-state growth of W by the hydrogen reduction process involving H2 and WF6, and the formation of the initial seed layer by a SiH4/WF6 process was present as well. The seed layer is illustrative of the fact that real processes entail overhead of several kinds, which add to total process time but not directly through the chemical process behavior. Other examples are the times involved in establishing proper deposition conditions (e.g., pumpdown, gas inlet, heating) and in recovering from them after the deposition is complete (e.g., cooling, pumpdown, venting). In future work we will be considering more complex models which incorporate these additional effects to determine lot processing time and throughput.

Specifically, for the Stefani model, let \( DR \) be the actual deposition rate in Å per second, \( P \) the reactor pressure in torr, and \( T \) the deposition temperature in degrees Kelvin. Then the RSM \( DR(P, T) \) can be expressed as follows (the mole fractions were set to their median values used in the experiments):

\[
DR = 63.55 + 0.4248(P - 80) - 364.6 \left( \frac{1000}{T} - 1.346 \right) \\
- 2.079(P - 80) \left( \frac{1000}{T} - 1.346 \right) \\
+ 1.297 \times 10^{-4}(P - 80)^2 + 945.5 \left( \frac{1000}{T} - 1.346 \right)^2
\]

3.2 Equipment Models

Although researchers have developed techniques for analyzing the behavior of simple cluster tools, our research required the development of new cluster tool models. Herrmann et al. (1999) describe these models in detail, which we summarize here.

To calculate the total time for a cluster tool to process a wafer lot, the cluster tool model includes the process RSM(s) for the process modules on the tool, together with a scheduling model for the entire tool. The total lot processing time in a cluster tool is a function of the lot size, the tool configuration, the wafer handler move times, and the individual process times. The process times are functions of the process parameters, which change the achievable process rate and thus the time required. For simplicity we have assumed that the deposition process time \( D = Th/DR \), where \( Th \) equals the deposition thickness, and \( DR \) is the deposition rate, which is a function of the process parameters (as described in Section 2.1). This simple model neglects subtleties like the nucleation (seed layer) step and equipment overhead before and after the deposition occurs. Ideally we would like to have the deposition process separated into each of these components as separate RSMs, but this is not easy to do with real experiments. For example, there is no way to measure the seed layer deposition thickness independent of bulk thickness since the processes are done sequentially in the same chamber without breaking vacuum.

We have developed four cluster tool models, all implemented as Java programs and compiled into executables. The simplest approach uses a network model for a specified sequence of wafer handler moves. We also have two simulation models. The first simulation model uses a push dispatching rule to select the next wafer handler move, and the second uses a pull dispatching rule to select the next move. The last model is an optimization procedure that can find the best sequence of wafer handler moves. The models require the following input data: the tool configuration (the number of stages and the number of chambers at each stage), the number of wafers in a lot, the wafer handler move time, the process time at each stage, and the tool overhead time. For more information about the optimization procedure, see Herrmann and Nguyen (2000).

3.3 Factory Model

For modeling factory operations, we used Factory Explorer™, a commercial discrete-event simulation software available from Wright, Williams, and Kelly (WWK). The basic factory model consists of a set of Microsoft Excel worksheets that describe the products, the resources, and the sequences of processes that the products must follow. The simulation software converts these factory model spreadsheets into a special file format, runs
the discrete event simulation engine, collects data about the system performance, and generates the results as text files and Microsoft Excel worksheets. In our work, the most important performance measure is the average cycle time of a lot through the factory. We plan to measure cost functions in the future.

The model of the subfactory under consideration includes one product, produced as multi-wafer lots, and three tool groups (wet clean, liner deposition, tungsten deposition). The wet clean is assumed to comprise a single time needed for dip cleaning. The simulation model is a stochastic model and has a random arrival process. Other types of randomness could be included.

Because of the re-entrant nature of semiconductor manufacturing, each wafer lot visits the subfactory of order five times, once for each layer of interconnects. Our subfactory model includes an artificial step that represents the time that the lot spends undergoing other processes between visits to our subfactory. With each visit to our subfactory, the processes are somewhat different, reflecting the fact that the interconnect layer thickness varies with which layer in being manufactured.

4 IMPLEMENTATION

The heterogeneous simulation environment (HSE) integrates the diverse simulation and analytical models described above. This section describes details of our implementation. A key feature of the HSE is its ability to execute models that exist as different types of software. As explained below, the HSE provides a single user interface, which the analyst uses to modify any input data, including the values of process parameters, equipment configurations, and factory data. Using the HSE, the analyst can predict system performance and estimate how the system performance is sensitive to any of these input data. To perform this analysis, the HSE executes each model in turn so that the output from one model can be used in the input for the next model. Finally, the factory simulation model delivers estimates of system performance. Thus, the HSE can measure how process changes and equipment configuration changes affect manufacturing system performance.

4.1 Model Integration

The HSE includes the Administrator, the enhanced factory model, and the cluster tool evaluation models. The Administrator communicates with the models and provides the user interface (see Figure 3). It is written using the Delphi programming language, running in the Windows operating system.

The enhanced factory model is an Excel workbook that contains multiple, inter-related worksheets. These include all of the spreadsheets needed for the basic factory model. One of the additional sheets supports the user interface and lists all of the input variables and the default values. Others implement the process RSMs and contain the necessary input data, formulas, and output data. The cluster tool evaluation models are implemented as described in Section 3.2. The analyst can select which cluster tool evaluation model should be used: the fixed sequence, the push dispatching rule, the pull dispatching rule, or the optimization program.

Figure 3: HSE Interface

Through the Administrator’s user interface, the analyst can view and update any of the input variable values. The Administrator retrieves and stores the values in the enhanced factory model. When the analyst modifies any input data, the Administrator tells the spreadsheets to recalculate, executes the selected cluster tool evaluation model (for each layer), and updates the basic factory model with the results.

4.2 Analysis Tools

Through the user interface, the analyst can tell the Administrator to predict system performance or to estimate how the system performance is sensitive to a specified input variable.

To predict system performance, the Administrator prompts the user to enter a number of parameters that control the simulation: the number of replications, the length (in time) of each replication, and the confidence interval required. The Administrator calls the Factory Explorer simulation engine and passes the required parameters. The Factory Explorer simulation engine performs the desired simulation runs and creates an output file with the system performance in each run. The Administrator reads this output file and calculates the desired confidence interval for the average cycle time. This result is shown to the analyst. Certainly, other performance measures could be collected, but the average cycle time is the most important in our application.
To estimate how the system performance is sensitive to the input variable that the user has selected, the Administrator prompts the user to enter a number of parameters that control the gradient estimation: the number of replications, the length (in time) of each replication, the confidence interval required, and the step size. For a variable whose initial value is \( v \), the analyst can specify a fixed step size \( c \) or a relative step size \( r \).

If the user selects a fixed step size, then the Administrator performs the following steps: the Administrator reduces the variable value to \( v-c \), tells the spreadsheets to recalculate, executes the selected cluster tool evaluation model (for each layer), updates the basic factory model with the results, calls the Factory Explorer simulation engine, and passes the required parameters. After the simulation has created the output file, the Administrator collects the system performance for that point. Then, the Administrator repeats the process after increasing the variable value to \( v+c \). Then, the Administrator can use the results of the second set of simulation runs to calculate a confidence interval for the gradient.

If the user selects a relative step size, the Administrator performs the same set of steps. However, instead of subtracting and then adding \( c \) to the initial value, the Administrator reduces the variable value to \( (1-r)v \) and then increases the original variable value to \( (1+r)v \). This finite differences gradient estimation technique is described more completely in Mellacheruvu et al. (2000).

For example, the analyst can determine how the average lot cycle time is sensitive to the orient-degas (OD) processing time. (Recall that both cluster tools perform this step.) Using a ten percent relative step size, the Administrator can determine a 97.5% confidence interval for the gradient, which is [0.013, 0.017].

5 SUMMARY AND CONCLUSIONS

The HSE integrates a diverse set of simulation and analytical models. An important feature of the HSE is its ability to execute models that exist as different types of software. The HSE provides a single user interface through which the analyst can modify any input data, including the values of process parameters, equipment configurations, and factory data. The HSE can predict system performance and estimate how the system performance is sensitive to any input variable. Thus, the HSE allows the analyst to understand how process changes and equipment configuration changes affect the manufacturing system performance.

These results yield important benefits. First, they demonstrate that, with analysis tools that extend factory simulation models by incorporating process RSMs and equipment models, a process engineer changing the process parameters can quickly determine if the proposed change will significantly increase the average cycle time. If so, it would be prudent to consider a less drastic change. Moreover, the process engineer can understand the system-level impact without requiring any of the industrial engineer’s time. Second, a manager considering whether to purchase a piece of equipment can determine what impact different configurations of the new equipment will have. Third, the industrial engineer can collaborate with the process engineer to see whether changing the scheduling algorithms on a cluster tool would affect system performance.

This new approach, which has not previously been explored, provides a vehicle for direct feedback of manufacturing metrics to process engineers involved in process alterations or tuning. Additionally, the range of possible process recipes that yield acceptable cycle time performance can be identified.

This type of tool will help process engineers understand how process changes affect the system performance. With these results, process engineers can develop better processes, equipment purchasers can make better procurement decisions, and fab managers can improve factory performance.

Heterogeneous simulation environments offer great potential, since they allow an analyst to create more accurate models of complex, inherently heterogeneous systems by reusing existing models that are implemented in different ways. The alternative is to build a large, complex simulation model either as a monolithic model or as a hierarchical model using a single software solution. Scaling the current HSE to a much larger set of processes and models is certainly possible, though it would require significant effort. These difficulties clearly show the need for more research into developing a more flexible integration environment to support heterogeneous simulations.

We plan on enhancing the simulation by adding additional process models to get a broader picture of factory performance. In addition we are developing optimization routines to select the set of equipment that minimizes average cycle time while satisfying throughput and budget constraints. Finally, we are developing more general simulation environments that can interact with other types of factory simulation models so that we can study a broader range of factories.

ACKNOWLEDGMENTS

This research was supported by the Semiconductor Research Corporation and the National Science Foundation under grant DMI 97-13720. We thank our collaborators Michael Fu and Steven Marcus for their useful insights.
REFERENCES


AUTHOR BIOGRAPHIES

JEFFREY W. HERRMANN is an assistant professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the director of the Computer-Integrated Manufacturing Laboratory. He is a member of INFORMS and ASME. He earned his B.S. in applied mathematics from Georgia Institute of Technology. As a National Science Foundation Graduate Research Fellow from 1990 to 1993, he received his Ph.D. in industrial and systems engineering from the University of Florida. He held a post-doctoral research position in the Institute for Systems Research from 1993 to 1995. His current research interests include the design and control of manufacturing systems and the integration of product design and manufacturing system design. His email and web addresses are <jwh2@eng.umd.edu> and <www.isr.umd.edu/~jwh2/jwh2.html>.

BRIAN CONAGHAN is a graduate fellow at the Institute for Systems Research at the University of Maryland working on integrating process models and operational methods. He will receive his M.S. in Electrical Engineering in December of 1999, majoring in Microelectronics. He received a B.S. in Chemical Engineering from Cornell University in 1997. He will be continuing to work on his Ph.D. at the ISR. Prior to coming to Maryland he worked as a co-op at Intel Corp, working on yield analysis and oxide etch issues, and at Advanced Micro Devices Inc. where he worked on wet clean optimization.

LAURENT HENN-LECORDIER is a Research Engineer at the Institute for Systems Research at the University of Maryland, College Park. His research interests are semiconductor process sensing, control, simulation, and optimization. His email and web addresses are <hennlec@isr.umd.edu> and <www.isr.umd.edu/~hennlec>.

PRAVEEN V. MELLACHERUVU is a Graduate Research Assistant for the Institute for Systems Research at the University of Maryland, College Park. He is working towards a M.S. degree in Systems Engineering. He received a B.Tech (1998) degree in Mechanical Engineering from the Indian Institute of Technology, Madras. His email and web addresses are <praveen@wam.umd.edu> and <www.wam.umd.edu/~praveen>.

MANH-QUAN NGUYEN is currently a Graduate Research Assistant of the Institute for Systems Research, University of Maryland College Park. He received the B.S. degree in Mechanical Engineering from the University of
Maryland College Park in May 1999. His current research interest is focusing on cluster tool scheduling.

GARY W. RUBLOFF is currently Director of the Institute for Systems Research and Professor in the Department of Materials and Nuclear Engineering at the University of Maryland, College Park. He received the BA in physics in 1966 from Dartmouth College and the PhD in physics from the University of Chicago in 1971. Following a postdoctoral appointment in physics at Brown University, he was a Research Staff Member and held various management positions in Physical Sciences, Silicon Technology, and Manufacturing Research at IBM Research in Yorktown Heights from 1973-1993. From 1993 to 1996 he was Associate Director of the NSF Engineering Research Center for Advanced Electronic Materials Processing and Professor of Electrical and Computer Engineering at North Carolina State University. His research interests have spanned solid state and surface physics; semiconductor interfaces, materials, and process technology; ultraclean, integrated processing and advanced equipment; and advanced diagnostic techniques. He has published well over 100 papers and holds 15 U.S. patents. He is the founding chairman of the AVS Manufacturing Science and Technology Group, a Fellow of APS and AVS, and active in semiconductor technology roadmapping activities. His current research efforts are aimed at real-time process sensing, dynamic simulation for design optimization, process control, environmentally-conscious manufacturing, manufacturing operations, and manufacturing education and training. His email and web addresses are <rubloff@isr.umd.edu> and <www.isr.umd.edu/~rubloff>.

ROCK Z. SHI is currently pursuing a Ph.D. in the Department of Electrical Engineering at the University of Maryland, College Park, majoring in Microelectronics and minoring in Computer Engineering. From June 1997 to June 1999, he was a graduate research assistant in the Institute for Systems Research at UMCP, while his research was on response surface methodology in integrating product dynamics and process model. From July 1999 till now, he is a full time co-op ASIC design engineer at the ATMELE Corporation Chesapeake Design Center. From August 1995 to May 1997 he studied Biomedical Engineering at Tulane University, New Orleans and received the MS degree in May 1997. During his study at Tulane University, he received Tuition Scholarship and Teaching Assistantship. His master thesis was multichannel coherence spectrum analysis on cognitive EEG and sleep EEG. His email and web addresses are <rshi@glue.umd.edu> and <www.wam.umd.edu/~rshi>.