Interdisciplinary Research on  
Modeling and Scheduling of Semiconductor Manufacturing Operations  

Donald Collins, Ph.D., Manufacturing Engineering Technology,  
Forouzan Golshani, Ph.D., Computer Science,  
Frank Hoppensteadt, Ph.D., Math and Electrical Engineering,  
Christian Ringhofer, Ph.D., Math,  
Jennie SI, Ph.D., Electrical Engineering,  
Kostas Tsakalis, Ph.D., Electrical Engineering  
Arizona State University  

Abstract  

This paper will describe the ongoing interdisciplinary research work of a group of faculty and students working within the Systems Science and Engineering Research Center at Arizona State University (SSERC) in collaboration with INTEL and Motorola.  

One of the strongest driving forces in the economy of most developed countries is manufacturing. In the United States, one of the most important components of this driving force is the manufacturing of semiconductors in fabrication facilities (FAB). Not surprisingly, tremendous efforts have been expended to reduce the art of semiconductor manufacturing to a science.  

While the improvements in scale and yield of semiconductor manufacturing has been spectacular and are well known, the improvement in operational methods of process flow and process scheduling have not been as impressive. The most obvious reason for this is the inherent physical/chemical complexity of these processes. Semiconductor manufacturing processes are hundreds of steps long, and factories typically operate multiple processes with each one running many product devices to serve a dynamic marketplace. Process flows are re-entrant and the work in process can be subject to rework. Finite equipment and human resources exhibit a wide range of variability of availability and performance. There are batching or set-up considerations on machines that process by wafer, by lot, or by batch.  

One practical consequence of all of this complexity is that decisions must be made on very different time-scales. The smallest time-scale being every few minutes during floor operations, usually about resource allocation - which step/lot combination to run next, which down machine to repair next, and so on. The medium time-scale of every few days or weeks, responding to market changes. To the largest time-scale of every few months concerning process changes, design improvements, etc. The integration of all of these operational decisions defines the performance of the factory. The key difficulty in making these decisions lies in the intricacy by which they affect the long- and short-term factory performance, as well as each other.
Given the economics of FAB performance, these decisions should be based on sound systems models built on fundamental principals. Unfortunately, these important decisions are currently based on a little local data and rough rules of thumb. The main goals of this research project were:

- To investigate novel applications and employ proven tools from the theory of dynamical systems and from control theory to the factory-wide resource scheduling (decision) problem and product release generation in semiconductor manufacturing, and
- To develop a model for university-industry dialog in the semiconductor manufacturing area by having industry personnel and interdisciplinary faculty and students working as active participants in the research.

The Systems Science and Engineering Research Center

SSERC is made up of faculty and students from across many departments and three colleges, College of Engineering and Applied Sciences, College of Liberal Arts and Sciences and the College of Technology and Applied Sciences. The objective of the Center is to foster interdisciplinary research work for faculty and students. The faculty meets at least once per week in a scheduled seminar series where individuals and groups present their latest research. An outside speaker is brought in from either industry or another university several times per semester. The Center also sponsors a 1 to 2 day mini-conference per academic year on topics of specific interest to its members. This has been an excellent model to introduce faculty and students to new ideas for interdisciplinary collaboration. The small group of faculty listed here is an example of this collaboration. Each member brings his/her own expertise and when integrated makes the total more than just the sum of the individuals.

This interdisciplinary research encompasses hierarchical mathematical and stochastic simulation modeling for semiconductor manufacturing, from the release of raw wafers at the start through the completion and shipment of the devices. The implementation of “optimal scheduling policies” has recently been recognized as an important and challenging problem in the Semiconductor Industry.\textsuperscript{2,7,18,20,45,46} The competitive operation of modern fabrication processes requires the development of precise 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 of an optimal scheduling policy remains highly nontrivial, involving the solution of a constrained optimization problem with respect to often-conflicting objectives while any admissible policy must possess certain robustness properties in the presence of uncertainty. This research has developed hierarchical simulation modeling software tools that can be used to manage variabilities arising on the FAB floor\textsuperscript{3,4,5,6,7,8,10,11,15,18,23,24,25,26,31,38,43}. These tools will enable managers and operators to avoid work in progress (WIP) build-ups caused by these variabilities. We investigated the integration of several levels of FAB flow product-release and resource-scheduling algorithms integrated with machine process controllers, including predictive controllers in real time (see Figure 1.) with the aid of discrete event simulation.

Information Systems Management

Intelligent information systems are at the heart of control, processing and optimization of
Our goals for the information support systems integrated with the FAB Manufacturing Execution System (MES) such as PROMIS® or WorkStream® and Master Database system such SAP® or Oracle® include provisions for a knowledge-based, multi-level integrated architecture for information flow/processing across all stages of manufacturing from oxide growth chamber simulations through controller and supervisory levels to the enterprise level.

Specifically, in the context of semiconductor manufacturing, data analysis and data mining are the basis for quality control assurance, prediction of time to failure, raw material needs, inventory control, executive report generation, etc. In the context of our multi-level information system architecture, data analysis and filtering must take place at all levels. Basically, raw data is modeled into a summative form, and then passed on for use at another level. As information travels up the hierarchy, more refined numbers and semantics are included.

The research involved the development of effective simulation models of existing FABs. Present plans are to extend beyond the simulation and include effective interaction with the FABs relational databases (e.g., Motorola’s PROMIS® and Intel’s WorkStream®). In particular, bi-directional interfaces for all controllers, schedulers, and release generators are being investigated (Figure 1.). There are four levels in the hierarchy of a FAB integrated simulation model. The lowest level (level 1) comprises machine process controllers; next (level 2) predictive controllers; third (level 3) resource schedulers, and finally (level 4) the product release generator. The long-range objective of this research is to assist FAB Managers in maintaining a stable FAB with maximum utilization of resources and with minimum inventory in an acceptable time frame.

Our research was based on past and existing multidisciplinary ASU/Intel team(s) and ASU/Motorola team(s) organized through the SSERC working on operational methods for semiconductor manufacturing. The ASU/Intel team(s) have been working on an Intel provided abstract problem of the FAB process and the ASU/Motorola team(s) have been working in several FABs solving multiple scale problems from the atomistic scale of Chemical Vapor Deposition (CVD) to machine level controllers in diffusion and to predictive controllers for Reactive Ion Etch (RIE) process to resource scheduling and product release generators of actual FABs.

A specific problem motivated by these predictive maintenance controllers is to design scheduling and release policies that maintain the highest possible performance in the presence of regular maintenance operations or, in general, predictable equipment failures. It is clear that for this to be possible, the scheduling algorithm should exchange information with the low-level controllers and in turn with the release policy generator. In this respect, our goal was to investigate the type of information that needs to be communicated. For example, based on our recent work on uncertainty characterization in system identification, the "in-spec" operation of the real-time controllers can be described with relatively simple and computable measures.

While such procedures can help in predicting potential equipment failures, the use of this information by the scheduling and release policies remains unclear. During our recent work we
observed that this problem could be formulated as a time-varying control problem, for which control theory offers a variety of tools. Our plan was to investigate the use of Model-Based compensatory design techniques to develop and test scheduling algorithms that provide "near-optimal" performance with respect to large-but-low-frequency perturbations, such as predictable equipment failures and emergency maintenance.

A Mini-FAB Model

Intel has provided the basic definition of the abstract problem in an operational format. One facet has been the specification of a FAB model (see Figure 2.). The Intel design goals for this model were:

- to include all of the relevant features that make semiconductor manufacturing operations difficult, and
- to structure the model so that the features could be included incrementally and could be increased (decreased) in magnitude of impact.

The features included are re-entrant flow, multiple products including test wafers, equipment sets with variable availability from both preventive and emergency maintenance, equipment sets that require different setups for different process steps and products, equipment sets that batch, technicians that have overlapping training and floor responsibility for loading and repairing machines, a multi-cell layout, a finite capacity material handling system with transporters and
stockers, and multiple performance metrics including both volume and speed. Another facet has been the details of how Intel would run the modeled FAB given the current operational methods used across Intel FABs. This model and these metrics have been encoded into a simulation-based test-bed using the Extend® simulation software to provide realistic evaluation of new ideas against Intel’s current practice.

Figure 2. Schematic diagram of the Intel 5 machines 6 steps process flows

Technical Summary of the Research

Fundamentally, the problem of scheduling of semiconductor manufacturing operations and product release is a problem of resource allocation in the presence of uncertainty. This can be viewed as a systems/controls problem where resources or manipulated variables include the release of material into the FAB, production machines, transportation systems, and operators’ etc. Uncertainty enters in several aspects of machine and work availability, e.g., machine failures, operator reliability, etc. In an attempt to re-cast this problem in a more tractable form, simplified descriptions of the FAB are frequently employed (e.g., flow models)1,16 while much of its variability is attributed to generic uncertainty. Such a description is useful in extracting gross properties of the manufacturing system, but its applicability is limited. In general and even in the absence of stochastic perturbations or modeling errors, the manufacturing systems under consideration cannot achieve the performance suggested by simple flow models. The principal obstacles are the nonlinear dynamics caused by the re-entrant nature, batching, and the generally nontrivial relations between the processing, transport, and loading times. Following simulations of an abstraction of an Intel FAB, there is strong evidence19,37 that the dynamics of re-entrant manufacturing systems behaves chaotically, see also1,16. Moreover, these effects become more pronounced and critical as the manufacturing system is pushed to operate near capacity. For this reason, we believe that an in-depth analysis of the nonlinear dynamics involved in manufacturing systems will be crucial in the design of effective scheduling and control systems. We propose a hierarchical approach10,43 based on the previously mentioned time-scale decomposition between short- and long-term objectives. We will model the fast time-scale where decisions are made at the machine level, through detailed models that describe the nonlinear system dynamics (“inner-loop” or “low-level” controller). The objective of the inner-loop controller is to follow, as well as possible, commands issued by the high-level (outer-loop) controller. Operating on a fast time-scale, it can take into account the process nonlinear dynamics and define operating trajectories that trade-off variability minimization, cycle-time minimization and throughput maximization in a locally optimal way. Effectively, the inner-loop
controller allows the approximation of the system by a simpler aggregate model for which the computation of long-term optimal scheduling strategies can be performed in a systematic and efficient manner. The high-level decisions evolve at a slow time-scale (e.g., daily or weekly) and are based on aggregate models aimed to optimize the overall system performance, compensating for infrequent or low frequency variabilities.

The study of control-oriented modeling involves the model development of manufacturing systems for control purposes, using techniques/ideas from linear and nonlinear robust control theory. Here, model aggregation and characterization of the resulting modeling error will be central in assessing the fidelity of the predicted closed-loop performance.

Motorola has provided Ringhofer, et al, FAB data for real time control of Chemical Vapor Deposition (CVD) processes. CVD processes a mixture of reacting gases together with a wafer in a reaction chamber, depositing a thin film on the surface. This process is one of the fundamental processing steps in semiconductor manufacturing. A full process simulator has been developed to enable control of the film thickness and its morphology from external reactor set points and allow for the extraction of parameters for higher level controllers. To simulate simultaneously in real time the gas flow in the reactor and the evolution of the wafer surface involves extremely challenging computational problems, which are amenable to solutions using parallel and distributed computer architectures.

Motorola has provided Si, et al, a large amount of production data for the Reactive Ion Etching (RIE) process to study ways to monitor and control the process at the tool level. We will systematically study the functional correlation between in-situ sensor measurements, post process measurements and optical emission end-point signals. One of the critical signals for this study is the optical emission end-point trace, since in current engineering practice this signal is used as the sole input to determine the end point of etching processes, such as polysilicon, metal and oxide etches. This optical emission trace has embedded in it information about the plasma, the effect of the plasma on the wafer, and consequently, the final wafer attributes.

There were two specific aims included in the RIE research. 1) To achieve tight real time control over the etch time to reduce final wafer variances in remaining and critical dimensions. 2) To provide real time correlation analysis for "out of norm" optical emission signals in order to determine their causes for either warnings or control. To achieve the above goals, we have developed predictive models for the optical emission traces from real time in-situ measurements, an optimizer to determine the main etch time that leads to minimum variance in final wafer dimension, and an off-line analysis relating in-situ sensors to post measurements. Si, et al performed a systematic analysis of existing historical data to differentiate between "normal" from "abnormal" in-situ patterns entirely based on existing/available production data from Motorola.

Study of the fundamental performance limitations

An important problem at this level was to establish the theoretical performance limitations that are due to re-entry, batching and set-up times. For this purpose, discrete optimization techniques
were used, reminiscent of statistical mechanics and simulated annealing, to map out boundaries of the “feasible region.”

Design of inner- and outer-loop scheduling policies

Here, the results of the previous analysis will be integrated with long-term optimization/control policies in a practical methodology for the design of scheduling policies. This involved tools from optimal, nonlinear and time-varying control theory\textsuperscript{23,39,45}.

Summary

The researchers investigated the novel application of proven tools from modern control theory to the factory-wide semiconductor fabrication problem. Adopting a hierarchical approach, their objective was to provide a general and systematic methodology for the design of long-term, e.g., daily, scheduling and release policies.

In particular, the main focal point of this research revolved around the relation between low-level control (e.g., real-time) variability and high-level control (scheduling) performance. As shown by Flores-Goday, et al, in our recent work (and well understood in practice, Little’s Law and Kingman’s Formula), tight low-level control can provide a significant improvement of the variability in scheduling operations \textsuperscript{3,4,5,6,7,8,10,21,25,26,42}. It was also evident, however, that a suitable design of the scheduler is necessary in order to take full advantage of the tight low-level control.

At the second level of research, we looked into real time process scheduling problems. As is well known, the job-shop problem is one of the NP-complete problems. Various scheduling strategies such as FIFO, etc., have been investigated by Collins, et al, at a Motorola FAB\textsuperscript{36,38,39,40,41,42,43}. The results were a 33.9\% reduction in cycle-time over a period of six months.

The present focuses on real time scheduling using heuristic rules \textsuperscript{3,4,5,6,7,8,19,21,22,23,24,26,28,29,30,31,32,33,34,35,37,38,39,40,41,42,43}. Aside from the traditional constraint optimization approaches, and heuristics, we are presently investigating with a Ph.D. Student and a Masters Student a computational tool based on learning. Various schedulers are taken as input to the "intelligent" scheduler; a pattern analyzer is first used to screen these inputs in order to figure out the correlation’s among various input patterns. Another intelligent agent will then be used to judge whether the performances of the known schedulers. A positive reinforcement is released for good performance and the associated pattern is associated closer to the schedule, otherwise a penalty will be given. By such an approach, we believe the scheduler will learn to be smarter and smarter. We have conducted some basic research in this form of learning. Our closed loop system has demonstrated its improved performance over learning. Once the system is trained, delivering a schedule should be automatic.

Results

The overall goal of this team of researchers will be to develop control policies that can be evaluated by means of simulations and, eventually, results in implementations in real FABs.
22. Kumar, S. and Kumar, P.R., The last buffer first serve priority policy is stable for stochastic re-entrant lines, Technical report, Coordinated Systems Laboratory, University of Illinois, Urbana, IL, received 1994.
25. Limanond, S., Si, J., and Tseng, YL., "Production data based optimal etch time control design for a reactive ion etching process" IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING, 12: (1) 140-147 FEB 1999.


Biographical Information

DONALD W. COLLINS
Donald Collins is a Professor of Manufacturing Engineering Technology, at Arizona State University East. He received his M.S. and Ph.D. in Electrical Engineering and Computer Science from the University of Illinois, Chicago. Dr. Collins is an international consultant to semiconductor companies throughout the US and Japan.

FOROUZAN GOLSHANI
Forouzan Golshani is a Professor of Computer Science at Arizona State University. Dr. Golshani received a B.S. in Industrial Engineering from Arya Mehr University of Technology, Tehran, Iran, in 1976, a M.S. and Ph.D. in Computer Science in 1979 and 1982, respectively, from Warwick University, England. He has authored 120 papers and nine US patents.

FRANK HOPPENSTEADT
Frank Hoppensteadt is a Professor with a joint appointment in Mathematics and Electrical Engineering at Arizona State University. He is also the Director of Systems Science and Engineering Research Center. Dr. Hoppensteadt received a B.S. in Mathematics and Physics from Butler University in 1960, and a Ph.D. in Mathematics from the University of Wisconsin in 1965. Dr. Hoppensteadt is author and co-author of many books and several patents.

CHRISTIAN RINGHOFER
Christian Ringhofer, is a Professor in Mathematics at Arizona State University. Dr. Ringhofer received a B.S. in Mathematics from the Technical University of Vienna in 1978, a M.S. in Computer Science in 1980 from Technical University of Vienna, and a Ph.D. in 1981 from the Technical University of Vienna, Austria.

JENNIE SI
Jennie Si is an Associate Professor of Electrical Engineering at Arizona State University. Dr. Si received a B.S. in 1985 from Tsinghua University, an M.S. in 1988 from Tsinghua University as well, and a Ph.D. from the University of Notre Dame in 1991.

KOSTAS TSAKALIS
Kostas Tsakalis is an Associate Professor of Electrical Engineering at Arizona State University. Dr. Tsakalis received a B.S. in Ch.E. from NTUA, Greece, in 1983, M.S. in Ch.E. in 1984, E.E. in 1985, and a Ph.D.in E.E. in 1988, all from USC.