Comparing the Economic Impact of Alternative Metrology Methods in Semiconductor Manufacturing

Payman Jula, Costas J. Spanos, *Fellow, IEEE*, and Robert C. Leachman

***Abstract*—Metrology is an essential part of advanced semiconductor manufacturing. It accelerates yield improvement and sustains yield performance at every stage in both new and mature processes. Advances in metrology are needed to achieve challenging industry goals, such as smaller feature sizes and reduced time for introduction of new materials and processes for future technology. To achieve difficult industry goals, it is expected that metrology practices will migrate from offline to inline, and ultimately, to *in situ*. Economic models are needed to study the costs and benefits of introducing new metrology technologies and to compare alternative metrology practices. Several qualitative and quantitative models are presented in this paper to study the elements of revenue and cost associated with different metrology tools and practices. Comparisons between *in situ*, inline and offline metrology systems are made. The cost components of the metrology methods are analyzed and discussed with respect to steady state process control as well as their effect on time to yield. Monte Carlo simulation models are used to study each system under different scenarios.**

***Index Terms*—Continuous-time Markov chain, economics, metrology, semiconductor manufacturing.**

I. INTRODUCTION

HISTORICALLY, semiconductor manufacturers rely on statistical process control (SPC) techniques for maintaining the processes within prescribed specification limits. While semiconductor manufacturing has continued to pursue ever-tightening specifications due to the well-known problems associated with the decreasing feature size, it has also become clear that there is a need for advanced-integrated process control. This approach requires a major shift in operational methods and requires the existence of complex, flexible architectures to meet the above requirements. New metrology tools are introduced as an essential part of these architectures.

Metrology accelerates yield improvement at every stage in both new and mature processes. Appropriate metrology practices can reduce the cost and cycle-time of manufacturing through better characterization of tools and processes. Advances in metrology are needed to achieve difficult industry goals, such as smaller feature sizes and reduced time for introduction of new materials and processes for future technology. To achieve these goals, it is expected that metrology practices will migrate from offline to inline, and ultimately be integrated in the tools (“*in situ*”) [1].

Researchers have concentrated on the economic impact of particular aspects of metrology tools such as the sampling policy [2], [3] and the precision [4]. Dance *et al.* [5] tried to capture the economic behavior of metrology tools through a modified cost of ownership (COO) model. Still there is a need for more comprehensive models to identify elements of cost in complex metrology systems.

Unless convinced otherwise, manufacturers are usually reluctant to adopt major equipment and technology changes because of the short-term uncertainties that arise during the introduction of new technologies. Appropriate metrology models assist the semiconductor manufacturers to assess the costs that drive their businesses and help them in formulating the right operational strategies. The ability to effectively identify cost drivers and manage cost reductions is a competitive advantage for any manufacturer. Therefore, accurate models are needed to study the costs and benefits of introducing new technologies and evaluate different practices. Toward this goal, this paper introduces new analytical models to compare different metrology methods in a litho track system.

Although this study tries to address the economics of metrology systems in a general form, the examples and illustrations are developed for litho track systems. Lithography steps are among the most crucial, and lithography tools are among the most expensive in semiconductor manufacturing. Most of the models offered in this document can easily be modified and extended to other equipment sets and metrology tools.

Fig. 1 shows different metrology methods in a litho track system in terms of the position of the metrology tool in the system. Wafers first enter the track system, where they go through steps such as coating and baking in preparation for the main lithography process (stepper), in which small features are printed on the wafer. After lithography, wafers go through additional steps in the track system, such as post exposure bake (PEB) and development (DE).

The qualities of the features defined during lithography (which in turn depends on the quality of the lithography process) have a direct effect on the quality of the final product. Therefore, we are interested in measuring and controlling the quality of the lithography step. The quality of the process (here the lithography step) is represented by measuring certain quantities on the wafer, such as the critical dimension (CD) of fine printed patterns.

Offline systems, as depicted in Fig. 1(a), have traditionally been practiced by semiconductor manufacturers. In this method,

Manuscript received May 9, 2002; revised July 15, 2002.

P. Jula and R. C. Leachman are with the Industrial Engineering and Operations Research Department, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: payman@ieor.berkeley.edu; leachman@ieor.berkeley.edu).

C. J. Spanos is with the Electrical Engineering and Computer Sciences Department, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: spanos@eecs.berkeley.edu).

Digital Object Identifier 10.1109/TSM.2002.804909

0894-6507/02$17.00 © 2002 IEEE

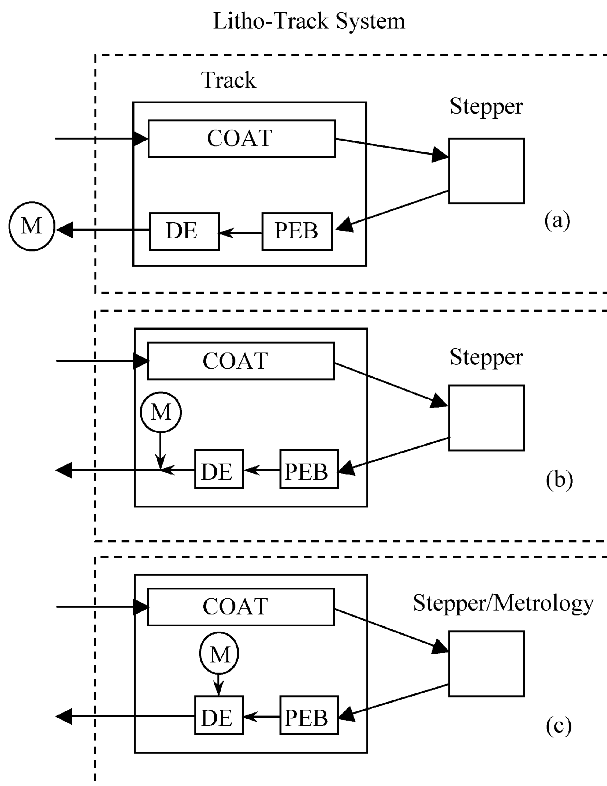


Fig. 1. Different metrology methods applied to a Litho track system: (a) offline, (b) inline, and (c) *in situ*. "M" indicates the position of the metrology tool.

the metrology tool is located after the track system. Wafers are transported to the metrology tool by lots. Lots are then measured by the metrology tool with an appropriate sampling policy. Offline metrology tools are usually accurate and fast, but are also expensive and occupy significant clean room space.

Newer inline systems occupy little footprint in the fab. Their accuracy and speed, however, is generally inferior to offline, though rapidly improving. *In situ* metrology systems are fully integrated and the measurements are done while the wafers are being processed or shortly after the process is completed. *In situ* lithography systems are under development and expected to be introduced with future generations of lithography tools.

To study the elements of cost in the above system, several qualitative and quantitative models are introduced in this paper. In the next section, the major components of the costs and benefits for metrology practices are analyzed and two revenue and cost models are introduced. The effects of metrology methods on revenue during the steady state and the time to maturity are explained. Monte Carlo simulation studies are conducted to compare different scenarios in Section III. First, the results of analytical model are compared to those of simulation model for a simple system. Then, the effects of yield and price structure, control policies, and the precision of metrology tools are examined in a series of scenarios. The results are presented and analyzed for each scenario. Recommendations are provided for each scenario and results are discussed. Conclusions and future avenues of study are explored at the end.

Financing considerations should be addressed along with our models. In this paper, we do not account for the timing of cash flows in our models, or attempt to evaluate the investments in terms of interest rates or discounted returns or tax benefits.

II. Analytical Models of Metrology Methods

In general, since metrology operations are in series with the processes, they reduce the throughput and increase the work in process (WIP) and the cycle time. WIP inventory between a process step and the subsequent inspection is at risk if the process drifts to an undesirable state. Manufacturers have been trying to reduce these risks using different methods such as changing the sampling policies and send-ahead samples.

Simply reducing the number of samples may result in a better cycle time and WIP, but it negatively affects the throughput of good products. Product yields at subsequent steps depend on the quality of information extracted from the metrology data. The quality of information generated from the metrology measurements can be partly characterized by its accuracy, precision and sampling policy.

It is desirable to identify bad products passing through the metrology tool and detect the out of control state of the process as soon as possible. This can be achieved by tightening the acceptance criteria. If, however, these criteria are too tight, then good products may be rejected, or the system may be shut down unnecessarily, resulting in production loss.

Another cause for production loss is the WIP between the process tool and the metrology tool. If the process drifts to an undesirable state, the process keeps manufacturing bad products until they are detected by the metrology tool. All the product in WIP processed since the process went out-of-control needs to be reworked or discarded. A send-ahead (also known as look-ahead) sample method eliminates the WIP risk but reduces the process throughput and utilization. In the send-ahead sampling method, one or more wafers are processed and then submitted for measurement. The remaining wafers in the batch are processed after the measurements are complete, the results are released and the equipment is adjusted.

Therefore, it is also desirable to minimize the WIP in the system. Assuming the same throughput for metrology tools, migrating from offline to inline and *in situ* usually reduces the WIP. In other words, integrated inline and *in situ* metrology operation minimizes the WIP lost with little impact on utilization. However, the feasibility of these approaches and the quality of data collected by inline and *in situ* tools, along with the price tag of these types of equipment, should be considered in making a decision.

A. Overall Equipment Efficiency (OEE)

Overall equipment efficiency (OEE) is one of the most important metrics for measuring equipment performance. OEE is defined as the ratio of the theoretical time needed to produce salable wafers in a given period, divided by the total time in that period [7]. Theoretical time refers to the time required by a machine in perfect working order performing the process specification under ideal conditions.

Since, in this study, we are mainly interested in understanding the differences among metrology practices, we classify the losses in equipment processing time into two main categories.

The first set of losses is associated with the metrology tool, its specifications, and the control policy chosen to detect and improve the bad process. The term "Bad process," in this document, refers to the process that is out of control and produces out-of-spec products; the products that are not conforming to the required specifications set by the fab management. These specifications are those that are measured by the metrology tool. The crosshatched area between OEE and OEE* in Fig. 2 shows the first set of losses. These losses are the focus of this study and will be further explored.

The second set of losses contains any loss that is not captured in the first set. These losses are those that occur regardless of the type of metrology tool and the control policy. Any loss of production due to unavailability of machine, bad utilization of equipment and slow process belongs to this category. The area between the OEE and 100% available time in Fig. 2 shows this set of losses.

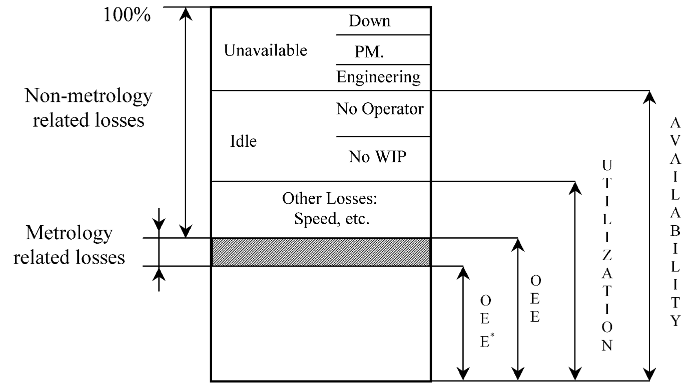


Fig. 2. The concept of OEE.

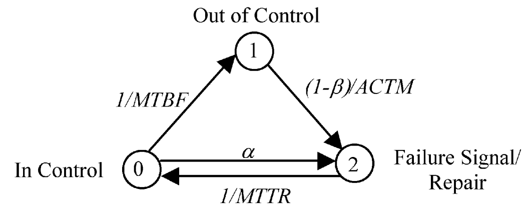


Fig. 3. Continuous-time Markov chain model of a metrology system.

B. A Mathematical Model of Metrology Tools

Assume the main process is up and in the "In Control" state for an exponential amount of time with the mean of mean time between failures (MTBF) of the process. The process goes to the "out-of-control" state and stays in this state until detected by the metrology tool. The quality of information extracted from the metrology measurements can be partly characterized by parameters α and β. The type I error, α, is the probability of rejecting a good product or process. The type II error, β, on the other hand, shows the probability of accepting a bad product or process. The power of metrology, $1-\beta$, is the probability of correctly rejecting a process or product. Accuracy, precision, and sampling policy in metrology are among the factors that affect the quality of information extracted from the metrology tool.

The time that is spent in the out-of-control state by the equipment is proportional to two factors; first, the time required for the results of the metrology tool to become ready, and second, the power of the metrology measurement. It is assumed that the equipment stays in the out-of-control state for an exponential amount of time with the mean of ACTM/$(1-\beta)$, where $(1-\beta)$ is the power of the metrology tool and ACTM is the average cycle time to metrology. ACTM is the response time from the metrology tool, which depends on the amount of WIP between the process and the metrology tool. After the metrology tool gives the signal that the process is out of control, the process is shutdown and the repair starts.

It is assumed that the tool stays in this state, which is called the "Failure Signal/Repair" state, for an exponential amount of time with the mean of the mean time to repair (MTTR). Because of the metrology type I error (α), there is a probability that the metrology tool generates a failure signal even though the process is in the good (in control) state. During any time interval h, in which the process is actually in the good state, the rate at which the equipment will be declared to be in the "Failure Signal/Repair" state is $(\alpha * h)$.

The above system is a description of a continuous-time Markov chain consisting of three states: namely, "In Control," "Out of Control" and "Failure Signal/Repair." Fig. 3 shows this system.

Solving the limiting probability equations of this system [6] result in:

$$P_0 = \frac{1}{1+\frac{\mathrm{ACTM}}{\mathrm{MTBF}(1-\beta)}+\mathrm{MTTR}\left(\alpha+\frac{1}{\mathrm{MTBF}}\right)} \tag{1}$$

$$P_1 = \frac{\mathrm{ACTM}}{\mathrm{MTBF}(1-\beta)\left[1+\frac{\mathrm{ACTM}}{\mathrm{MTBF}(1-\beta)}+\mathrm{MTTR}\left(\alpha+\frac{1}{\mathrm{MTBF}}\right)\right]} \tag{2}$$

where P_0 and P_1 are the long-term probabilities of the process being "in control" and "out of control," respectively.

The process under control produces acceptable products, while the out-of-control process produces bad products that must be reworked. The faster the out-of-control state is detected, the faster the process is calibrated; which limits the amount of required rework. Therefore, the cost of a bad metrology practice is twofold. First, there is the cost due to the lost time of equipment (metrology and litho track), including the expenses of investment in purchasing and installing the machines, maintenance, footprint, etc. The second cost element occurs because of WIP rework, resulting in material, energy and labor costs. These costs are further studied in this section.

C. Revenue Models

Let N_i denote the number of machines of type i that are installed in the factory. Ignoring the requirement that N_i must be an integer, Leachman *et al.* [7] have shown

$$\left(\frac{D}{Y_F}\right) ThPT_i = N_i(\mathrm{OEE}_i^*)(720) \tag{3}$$

where 720 is the number of hours in a month. The left-hand side of this equation expresses the total machine-hours required

to process $w = D/Y_F$ wafers per month; D is the designed output capacity and Y_F is the mature die yield. $ThPT_i$ is the total theoretical process time per wafer (expressed in hours) on equipment type i, considering all process steps performed by that equipment. The right-hand side is the total machine hours that can be devoted to processing (at theoretical rates) considering the achieved equipment efficiency. Assuming a revenue of R_0 for each wafer for the current day, the total revenue per day in the near future can be calculated as

$$\frac{R_0(N_i)\,(\mathrm{OEE}_i^*)\,(24)}{ThPT_i}. \quad (4)$$

Replacing the OEE^* with ($P_0 * \mathrm{OEE}$), where the P_0 is the long run probability of the process being in the good (in-control) state, result in

$$\text{Revenue/Day} = \left[\frac{R_0(N_i)(\mathrm{OEE}_i)(24)}{ThPT_i}\right] \cdot \left[\frac{1}{1+\frac{\mathrm{ACTM}}{\mathrm{MTBF}(1-\beta)}+\mathrm{MTTR}\left(\alpha+\frac{1}{\mathrm{MTBF}}\right)}\right]. \quad (5)$$

As expected, the revenue increases with the decline of α, β, ACTM and MTTR and decreases with the decline of MTBF.

Over the long run, where the price is declining according to a continuous discount factor of γ, the total revenue realized up to time H (expressed in days), assuming zero start-up and production delays, is expressed as

$$\int_0^H \frac{R_0(N_i)\,(\mathrm{OEE}_i^*)\,(24)}{ThPT_i}\, e^{-\gamma t}\, dt = \frac{R_0(N_i)\,(\mathrm{OEE}_i^*)\,(24)}{ThPT_i}\left(\frac{1-e^{-\gamma H}}{\gamma}\right). \quad (6)$$

D. The Effect of Metrology Tools on Ramp-Up

Up to this point, the behavior of metrology tools was considered for mature and stable process technology. However, as depicted in Fig. 4, each process goes through three different phases: development phase where the process is first introduced, the ramp phase where the volume of production is increased, and the mature phase where the process sustains high volume production.

During the development phase, the equipment is installed and an appropriate recipe is applied. In this phase, the process usually does not produce any marketable product. Therefore, this phase is not in our interest. The process starts producing salable products in the ramp phase. In the beginning of this phase, equipment fails more often. After some time, the process is calibrated, the rate of failures declines, and the process becomes mature.

Here, we are interested in studying the effect of the metrology tools on the ramp phase. For simplicity, we approximate the above curve with a step function, where the process has the average ($\mathrm{MTBF}_{\mathrm{low}}$) in the development and ramp phases and jumps to the mature phase ($\mathrm{MTBF}_{\mathrm{high}}$) at time T (Fig. 5).

There are many factors affecting the duration of the ramp phase (T). Studying the behavior of these factors is beyond the scope of this paper. However, it is known that the ramp-up duration, especially at lithography, depends on the knowledge and the experience of the engineers working with the process. Part

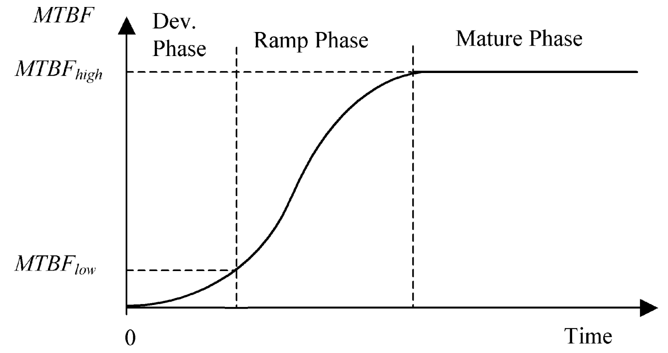


Fig. 4. Different phases of a process life cycle.

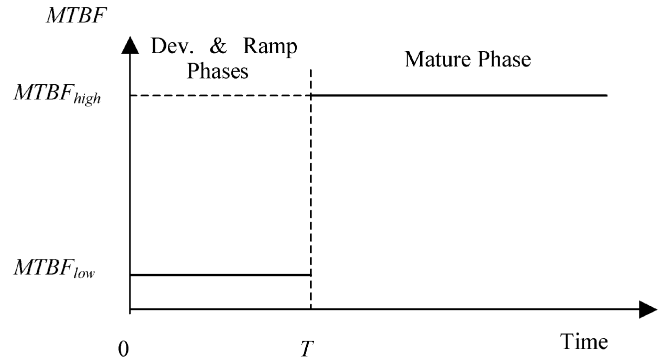


Fig. 5. A simplified process life cycle.

of the experience and knowledge comes from trial and error. Each equipment failure contributes to the knowledge about that equipment/recipe. Here, we assume the time to maturity is a function of the number of detected problems through time. The more problems are found, the more experienced the staff will become. Finally, after k number of trial and errors, the equipment goes to the mature state and the failure rate decreases. We are interested in finding the effect of metrology tools and the control policies on the value of T. Changes of T can then be translated to cost.

The number of required equipment is usually planned for the mature case; therefore, there is some lost revenue due to the unsatisfied demand in the development and ramp phases. Similar to (3), the satisfied demand in development and ramp phase (D_R), assuming the mature die yield, follows

$$\left(\frac{D_R}{Y_F}\right) ThPT_i = (N_i)(\mathrm{OEE}_i)(P_{0R})(720). \quad (7)$$

Here, the P_{0R} is the long-term probability of the process being under control during the development and ramp phases and follows an equation similar to (1). All of the notation in this section concerns the equipment performance in the development and ramp-up phases and is similar to the notation for the mature phase. Using (3) and (7), the unsatisfied demand per month during the development and ramp phases can be calculated as

$$D\left(1-\frac{P_{0R}}{P_0}\right). \quad (8)$$

The duration and the quantity of the lost demand during the ramp period will result in lost revenue during this period.

Considering the continuous-time Markov chain model for the development and ramp phases, therefore, the expected value of T, the elapsed time for k number of repairs, can be calculated as

$$T = (k)(\text{ACTM})/[P_{1R}(1-\beta)]. \tag{9}$$

The total possible revenue during the development and ramp phases, assuming all demands are satisfied, can be expressed as

$$\int_0^T R_0 e^{-\gamma t} \frac{D}{30}\, dt. \tag{10}$$

Here, γ is the continuous discount factor for the exponentially declining sales price. The lost revenue can be calculated as

$$\int_0^T R_0 e^{-\gamma t} \frac{D}{30} \left(1 - \frac{P_{0R}}{P_0}\right) dt. \tag{11}$$

The total lost revenue can be calculated as

$$\Delta R = R_0 \left(\frac{D}{30}\right) \left(1 - \frac{P_{0R}}{P_0}\right) \left(\frac{1 - e^{-\gamma(k.\text{ACTM}/P_{1R}(1-\beta))}}{\gamma}\right). \tag{12}$$

E. Comprehensive Revenue Model

The comprehensive revenue model consists of the combined revenue obtained in the ramp phase and the mature phase. The total revenue obtained in the ramp phase can be expressed as

$$R_0 \left(\frac{D}{30}\right) \left(\frac{P_{0R}}{P_0}\right) \left(\frac{1 - e^{-\gamma(k.\text{ACTM}/[P_{1R}(1-\beta)])}}{\gamma}\right). \tag{13}$$

Given the duration of the mature phase, the total revenue obtained in the mature phase can be calculated by (6). The summation of (6) and (13) should be considered in selecting the metrology setup.

The revenue models are more tailored toward the marketing department's needs versus the manufacturing expenses. In other words, they only consider the incoming cash flow to the company through sales. These models do not consider the outgoing cash flow and the expenses of the company. What if a metrology tool improves revenue, but the price of investment is high? How about the maintenance expenses and labor costs associated with each metrology system? These issues will be addressed by another model, called the cost model, in the following section.

F. The Cost Model of Metrology Methods

Leachman *et al.* [7] expressed the annual expense of a fab as

$$\sum_i \underbrace{(Ce_i + Le_i + Se_i)}_{CM_i} N_i + \underbrace{(Lw + Mw + Sw)}_{CW}.12.w + (Lf + Sf). \tag{14}$$

The first term captures the machine expenses. Ce_i, Le_i, and Se_i are the amortized annual costs due to purchasing, labor, and foot-prints, respectively, per machine of equipment type i. CM_i captures the total amortized annual cost per machine of equipment type i. The second term captures the expenses related to the number of wafers started. Lw, Mw, and Sw are respectively the amortized annual cost due to labor, material, and infrastructure per wafer started. CW is the total amortized annual cost per wafer started. The last term captures the annual fixed cost of manufacturing. Lf and Sf are the fixed labor cost and the fixed space cost, respectively, that are independent of wafer start volume and the number of installed equipment.

Using (1), (3) and (14), the total expenses of the machines per year can then be expressed as

$$\begin{aligned}\text{EPY(Machines)} &= (CM_{\text{litho}})(D)(ThPT_{\text{litho}}) \\ &\quad \cdot \left[\frac{1 + \frac{\text{ACTM}}{\text{MTBF}(1-\beta)} + \text{MTTR}\left(\alpha + \frac{1}{\text{MTBF}}\right)}{720(Y_F)(\text{OEE}_{\text{litho}})}\right] \\ &\quad + (CM_{\text{met}})(N_{\text{met}}) + \sum_{i \in \text{other}} (CM_i) \frac{(D)(ThPT_i)}{(Y_F)(\text{OEE}_i)(720)}.\end{aligned} \tag{15}$$

The "litho" subscript represents the lithography system, which includes the exposure unit and the track line. The first term in (15) captures the effect of metrology in lithography costs through its effective processing time. The second term is the cost associated with the purchase, maintenance and the footprint of metrology devices. The third term captures all other equipment expenses in the fab.

As discussed earlier, different metrology methods generate different amounts of WIP and rework. The rework consumes materials, energy and labor. Furthermore, the mask life, which is considered dependent on the number of exposures, causes the expenses to increase in proportion to the amount of rework. According to our continuous-time Markov chain model, the total out-of-control machine-hours spent processing w_{rl}, the number of wafers in lithography to be reworked, will be:

$$(w_{rl})(ThPT_i) = \left(\frac{P_1}{P_0}\right)(N_i)\left(\text{OEE}_i^*\right)(720). \tag{16}$$

Considering (1)–(3), and (16), the total number of reworked wafers in lithography per month can be calculated based on the total monthly production rate as

$$w_{rl} = \left[\frac{\text{ACTM}}{(1-\beta)\text{MTBF}}\right]\left(\frac{D}{Y_F}\right). \tag{17}$$

The fab total expense per year due to the number of wafers started includes two terms. The first term captures the expenses due to the reworked wafers in lithography steps. These expenses reflect material costs, energy, labor and masks. The second term includes all expenses that are functions of the number of wafers started. All the rework done on the other equipment sets (except lithography) are assumed to belong to this category. Therefore, the total expenses per year due to the number of wafer starts is

$$\begin{aligned}\text{EPY(Wafer started)} &= (CW_{rl})(12)\left(\frac{D}{Y_F}\right) \\ &\quad \cdot \left(\frac{\text{ACTM}}{\text{MTBF}(1-\beta)}\right) + (CW_{\text{other}})(12)w.\end{aligned} \tag{18}$$

The constant terms of (14), Lf and Sf, are assumed to remain unchanged after introducing different metrology methods.

The difference between metrology methods can be calculated according to (15) and (18). This difference can be presented as

$$\begin{aligned}\Delta\text{EPY} =& (CM_{\text{litho}})(D)(ThPT_{\text{litho}})\\ &\cdot \left[\frac{1+\frac{\text{ACTM}}{\text{MTBF}(1-\beta)}+\text{MTTR}\left(\alpha+\frac{1}{\text{MTBF}}\right)}{720(Y_F)(\text{OEE}_{\text{litho}})}\right]\\ &+12(CW_{rl})\frac{(D)(\text{ACTM})}{(Y_F)(\text{MTBF})(1-\beta)}\\ &+(CM_{\text{met}})(N_{\text{met}}).\end{aligned} \tag{19}$$

To choose the best metrology method, manufacturers should consider the elements involved in (19). All the costs associated with acquiring, installing and maintaining the litho track tools should be considered. Special attention should be given to the quality of information extracted from the metrology tools. The failure rate, ease of repair and the position of metrology tool in the system should also be considered.

III. Monte Carlo Simulation Models of Metrology Methods

In previous sections, several analytical models were presented for litho track systems based on some simplifying assumptions. There is still a need to address the issues involved in more complex systems arising in industrial environments. Appropriate models can predict the behavior of these systems under different scenarios and help the decision makers in selecting the best practices in different environments. However, it is very difficult to capture the behavior of complex systems with closed-form mathematical models, similar to those presented in the previous section. As an alternative, we use Monte Carlo (MC) simulation models to study the behavior of more complex systems.

In these models, the results of five 24-hour days with five different initial random seeds are collected for each simulation run. The lithography throughput is considered to be 60 wafers per hour. To accommodate the behavior of a robot in an industrial system, a buffer (with the capacity of one wafer) is considered before and after each station. The revenue generated for each model is then plotted in sets of graphs. Each point in these graphs is based on the information that is statistically collected from $60 \times 24 \times 5 \times 5 = 36\,000$ simulated wafers; each wafer includes 100 dice with individual characteristics. The data are collected after a warm-up period of 50 minutes. SIGMA [8] simulation software was modified and used as a platform for generating the data and collecting the information for these experiments.

The values of the parameters used in these models are either the estimated values in the industry or what experts would expect to see in emerging technologies. The experiments are designed to assist the manufacturers with developing similar models. Decision-makers could develop similar experiments that address their specific needs and accommodate their particular parameter values.

For the center working point, MTBF is 240 minutes and MTTR is set at 20 minutes. For this working point, five samples are selected from each simulated wafer. The CDs of these samples are then measured and the 3σ rule is used for the cutoff line. It is assumed that the results of the offline, inline, and *in situ* metrology are available after approximately 30, 15, and 2 minutes, respectively. The performances of these systems are analyzed with respect to variation in MTBF, MTTR, α, β around the center working point for each of the inline, *in situ* and offline cases. Later in this document, the effect of control policies, yield/revenue structures, the precision of metrology tools and many other parameters are investigated.

The reference of $1000 per chip for 250-nm technology along with the yield/revenue structure of products determines the revenue per chip in these cases. Total revenues on the order of millions of dollars are generated per day in these experiments. Different parameter values would certainly result in different values for revenue. However, readers should keep in mind that the absolute value of revenue is not our interest. We are interested in analyzing the changes in revenue based on the changes in the system. The relative differences will provide us with a better understanding of each system and help us predict the behavior of similar systems in similar working conditions. Therefore, revenues are presented in arbitrary units in this document.

First, a simple model is developed to compare the results of analytical models with those of MC-simulation. The assumptions in this model are consistent with the assumptions under which the analytical models were developed. The second scenario enhances the first scenario by introducing a variance to the process and by considering more realistic structures for the yield and price. In the final scenario, more realistic conditions are introduced to the system. Different random errors are considered for each of the inline, offline and *in situ* tools to capture the different precision associated with each technology. Furthermore, wider and more continuous drifts are considered for the process.

A. Analytical Approach Versus Monte Carlo Simulation

A MC model is designed to verify the accuracy of the results generated from the analytical models presented in the previous section. The assumptions in this model are consistent with the assumptions of exponential failure times and repair times under which the analytical models were developed. The lithography targets a CD of 205 nm at in-control state. It produces bad products with the CD of 225 at out-of-control state. For simplicity, the variance of the process is ignored at this stage; in the next section, the variance will be introduced to the system and its effect will be explored.

Our study [9] shows the consistency between the analytical model and MC-simulation. For example, Fig. 6 shows the effect on revenue from reducing the time between the process and the metrology tool. As shown, both the analytical model and MC-simulation predict a similar pattern. The figure shows an increase in revenue by migrating from offline to inline and *in situ* technology assuming that the same quality of information can be obtained from different metrology tools.

B. The Effects of Process Variation on Revenue

On many products in the semiconductor industry, it is well known that reduction of the critical dimension results in higher

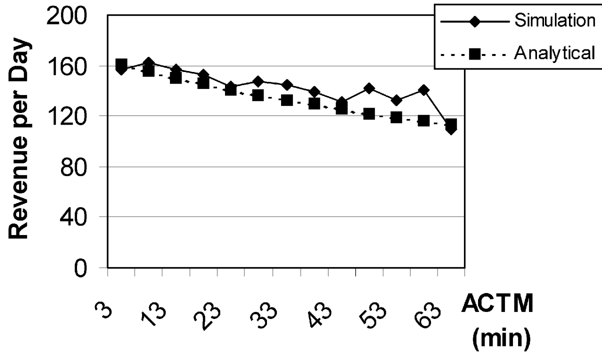


Fig. 6. Revenue per day (arbitrary unit) versus average cycle time to metrology tool (ACTM) for the analytical and simulation models.

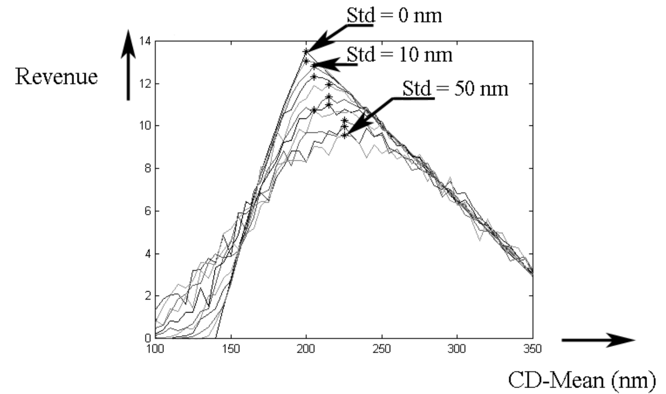


Fig. 8. Revenue (arbitrary unit) versus mean of the process for different standard deviation.

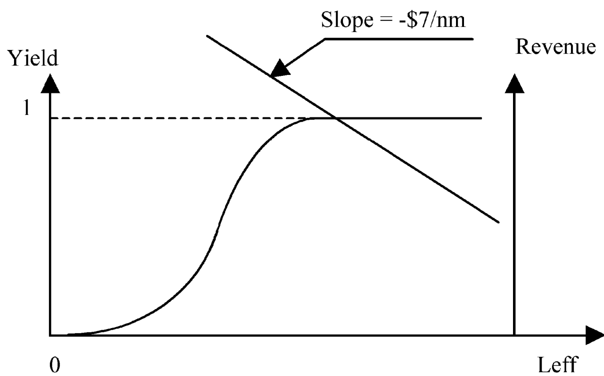


Fig. 7. Relationship between yield, revenue and CD.

revenue. A study by Motorola [10] has estimated an average gain of more than $7 per chip for each nanometer reduction of CD. Therefore, manufacturers tend to reduce the target CD as much as possible. However, reduction of CD may result in downstream manufacturing problems and reduction of yield, which in turn reduces the total revenue (Fig. 7).

Therefore, targeting the right working point is an essential point of the semiconductor business. We want to study the effect of variance on the total revenue and find the best working point for each variance. For simplicity, the yield curve is estimated by a piecewise linear curve, where the die yield is equal to one for CDs more than 200 nm and drops linearly with the decrease of CD until it becomes zero at 140 nm.

Fig. 8 shows the change of revenue versus the change in the targeted mean for different process variance. The star points in this figure show the maximum revenue that can be achieved from the processes with different means but the same standard deviation. Close attention to the behavior of these peaks reveals the reduction of maximum possible revenue with the increase in standard deviation. This indicates that manufacturers should try to minimize the variation in their process to achieve better revenues.

However, due to practical issues there are always uncertainties that cause variation in the process. In these cases, manufactures should choose the best working point for their business. For example, if the standard deviation of a process is 10 nm, with the above assumptions, the best working point would be 205 nm. For the rest of this document, we assume a standard deviation of 10 nm associated with the process, and we try to keep the working point at 205 nm in order to gain the maximum revenue. Assuming revenue of $1000 per chip for a CD of 250 nm will result in a revenue of $1315 per chip for a CD of 205 (assuming the $7/nm decline rate). Another negative effect of variance on revenue is due to the risk involved in the quality of information extracted from the product measurements.

Consider a process with a standard deviation of 10 nm that is targeted to work at 205 nm but it may go to the bad state of 225 nm after a random time with the distribution $N(\mathrm{MTBF}, \mathrm{MTBF}/10)$. (In this document $N(\mu, \sigma)$ notates a normal distribution with mean μ and standard deviation σ.) The process stays in the bad state until detected by the metrology tool. The shutdown/repair signal is generated when the average of the CDs measured from the sample points exceeds the cutoff line threshold. The process is then shut down and all the bad products in WIP are sent to rework. The process will be back in the good state after a random repair time with distribution $N(MTTR, MTTR/10)$. The $7/nm decline rate is observed in this case and there is no revenue for the products with CDs more than 220 nm, reflecting tight specifications set by management. Fig. 9 shows the in-control and out-of-control cases.

Changing the number of sample points taken from each wafer and adjusting the cutoff line of the control policy affects the type I and type II errors. Suppose Z_ζ represents the point on the Standard Normal distribution $N(0, 1)$ with the probability of upper tail equal to ζ. Then the following equations hold

$$\frac{X-205}{\sigma/\sqrt{n}} = \frac{\Delta}{\sigma/\sqrt{n}} = Z_{(1-\alpha)}$$
$$\frac{X-225}{\sigma/\sqrt{n}} = \frac{20-\Delta}{\sigma/\sqrt{n}} = Z_\beta. \tag{20}$$

Here, n is the number of sample points in each wafer, X is the CD of the cutoff point, and Δ is the cutoff distance from the mean as shown in Fig. 9. Assuming the standard deviation of 10 nm ($\sigma = 10$) results in the number of sample points

$$n = \frac{\left(Z_{(1-\alpha)} - Z_\beta\right)^2}{4}. \tag{21}$$

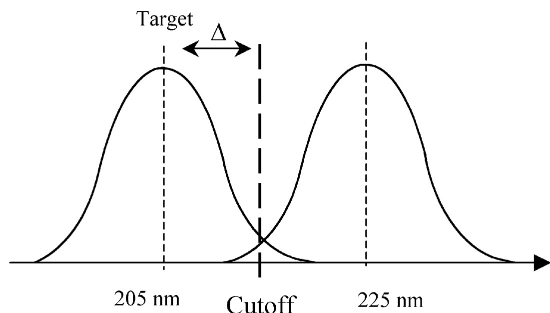


Fig. 9. In-control versus out-of-control.

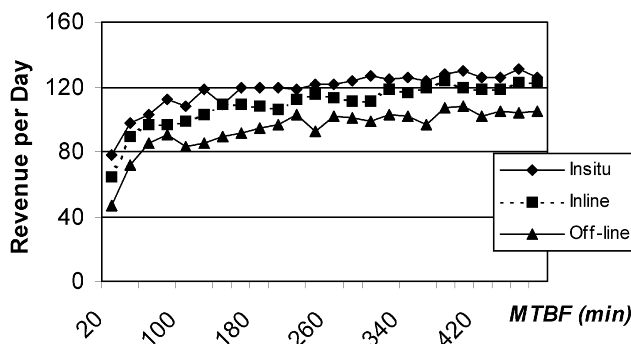


Fig. 10. Daily revenue (arbitrary unit) versus different values of MTBF.

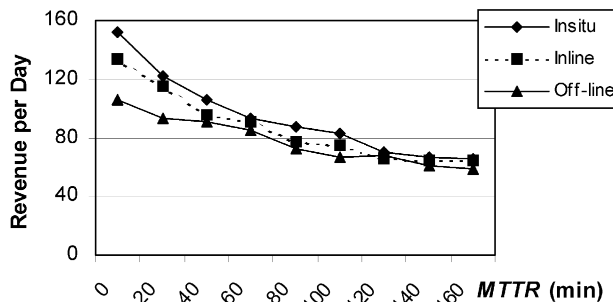


Fig. 11. Daily revenue (arbitrary unit) versus different values of MTTR.

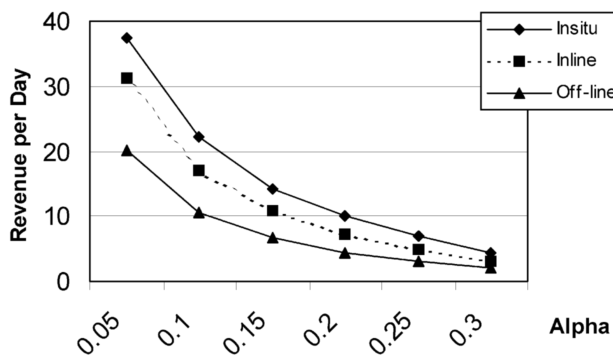


Fig. 12. Changes of revenue (arbitrary unit) versus changes in type I error.

To obtain the desired α and β, first, the number of sample points (n) is calculated and rounded to the closest integer. Then, the value of Δ is obtained. Δ and n, together, will specify new values for α and β, which are very close to the desired values. For example, the standard 5 sample points along with the 3σ ($\Delta = 13.41$ nm) rule for cutoff results in the values of 0.0013 and 0.0702 for α and β, respectively. Whenever possible, these values are set for the working point of the model.

Fig. 10 shows the importance of MTBF in generating revenue. It should be noted that changes in MTBF, especially at low values, have a greater impact on revenue than changes of equal magnitude when MTBF has high values. Here, the *in situ* method shows better revenue than the inline and offline techniques; which in turns confirms the value of measurement response time when other things are equal.

The change in the revenue of this system versus the changes in MTTR is depicted in Fig. 11. As expected, it shows decreasing revenue with increasing MTTR. Furthermore, the rate of change of the revenue decreases with the increase of MTTR. Similar to previous charts, the *in situ* method here outperforms the inline and offline methods.

Fig. 12 captures the effect of type I error (α) on the revenue. As α increases, more repair/shutdown signals are generated by the metrology tool, which results in frequent shutdowns of the system and therefore, increasing production loss. Our study shows that in this case the revenue generated per wafer is not very sensitive to the changes in α.

Type II error (β), on the other hand, has a noticeable effect on the quality of the products. By increasing β, more bad product is produced while the process is considered to be in control. In this case, total revenue per day is not very sensitive to changes in β, and it shows only slight decline with increasing β. In this experiment, the low price decline rate and the fact that the bad process state is constant and very close to the good state contribute to the changes of revenue based on changes of α and β. The price structure plays an important role in the decision to choose the metrology method. We have encountered some examples [9] where choosing offline is superior to inline and *in situ* even when the same quality of information is obtained from different metrology tools. Price structures in these cases were such that producing bad product was justified versus shutting down the system. If customers are willing to pay premium prices for bad product and/or the price of shutting down the system is very high, the manufacturers may prefer to stay with offline metrology.

C. The Effect of Metrology Precision

Previous models considered the precision of metrology tools embedded in α and β. Here, we introduce the precision of metrology tools as a parameter of the system and study its effect on system performance. Furthermore, in this case, the changes in the targeted mean may occur gradually. In other words, there are an infinite number of out-of-control states in the system. New market observation reveals a steeper line plotting revenue as a function of changes in CD. The price decline rate also has been modified to accommodate recent changes in the market.

In this scenario, the process is targeted to work at 205 nm but it may drift after a random time according to the distribution $N(\mathrm{MTBF}, \mathrm{MTBF}/10)$. The new working point can be

anywhere in the [0, +30] nm neighborhood of the previous working point. The process keeps changing the working points (getting worse and worse) until it is detected by the metrology tool. Here, we introduce noise to each measured point with a distribution of $N(0, Std_ERR)$. This noise models the precision of the metrology tool. The standard deviations of the noise for the central working point are set to the values of 3 nm, 2 nm, and 0.5 nm respectively for *in situ*, inline and offline metrology tools.

The shutdown/repair signal is generated when the average of the CDs measured from the sample points exceeds the cutoff line threshold. The process is then shut down and all of the bad product in WIP is sent to rework. The process will be back under control after a random repair time with distribution $N(\mathrm{MTTR}, \mathrm{MTTR}/10)$. Our observation of the current advanced logic market shows an approximate of $10/nm drop for the price of each chip. The $10/nm price drop is implemented in this case and there is no revenue for the products with more than 220-nm CDs. Changes in revenue versus changes in MTBF and MTTR in this scenario are similar to those in the previous case. In this scenario, *in situ* still outperforms the inline and offline methods. Revenue gaps, however, are narrower in this case due to differences in metrology precision. Fig. 13 shows the effect of the metrology precision on the total generated revenue. The chart reveals that the revenue is very sensitive to changes in precision. This chart can be used to justify the migration from offline to inline and *in situ* based on the precision achieved by the different technologies. It can be used further to identify the break-even points of each of these technologies. For example, according to Fig. 13, migrating from offline to *in situ* is justified when the offline and *in situ*, respectively, have precisions of 0.5 and 3 nm, but it cannot be justified if the *in situ* precision is worse than 6 nm.

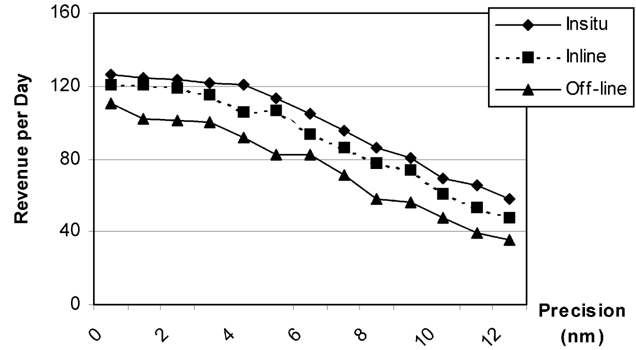


Fig. 13. Revenue per day (arbitrary unit) versus the metrology precision.

IV. Summary and Conclusion

In this paper, we have provided a framework for the economic analysis of metrology tools in semiconductor manufacturing. In order to study the elements of revenue and cost in semiconductor manufacturing due to different metrology tools and practices, several qualitative and quantitative models were presented in this paper. The differences between *in situ*, inline and offline metrology systems were analyzed. The proposed models should be modified and adjusted to address the practical issues in particular industrial environments.

A framework was suggested for the steady state case based on a continuous-time Markov chain model. Based on this framework, two analytical models were developed. The first model emphasizes the revenue generated from each system. This model was extended to include the revenue loss due to the delayed time-to-yield in ramp up phases. The second model focuses on system costs. The cost model estimates the expenses of the manufacturing system to satisfy certain demands for a long period of time. These analytical models present the important factors that affect the performance of the system and capture the most important relationships among these factors. To study more complex systems, Monte Carlo simulation models were generated. Different price and yield structures were implemented in these scenarios. Many complexities were introduced into these systems and the results were analyzed.

All of these models confirm the importance of selecting appropriate metrology tools and methods. The revenue and cost of these systems are very sensitive to metrology tool specifications, metrology structure, and even the price structure and the control policies. In most situations, especially when the process is newly introduced and the failure rate is high, the *in situ* metrology outperforms the inline and offline methods.

To make the decision to migrate from offline to inline or to *in situ*, many factors should be taken into account. The expenses of purchasing, installing and long-term maintenance costs along with the footprint, labor and material costs should be considered. The quality of products and the revenue associated with each technique should be studied. Both the market situation and the control policy play important roles in the decision-making process. *In situ* and inline metrology provide better response times than offline metrology, and are very useful especially in ramp-up phases. However, the quality of information extracted from these methods should be comparable to that from offline methods. Future technologies will reduce this gap and it is expected that *in situ* metrology will become the main trend in the semiconductor industry.

There are not many publications about the economic aspects of metrology tools in semiconductor manufacturing. More studies should be conducted to capture the effect of different factors on metrology tools in different environments. The analytical models presented in this document should be further enhanced to address more practical issues. The accuracy and the sensitivity of these models should be assessed in industrial environments and adjustments should be made accordingly. Hybrid methods, with the combination of inline, *in situ* and offline methods, should be studied. One approach may be to emphasize inline metrology with tight specifications during the introduction of a process and gradually change the emphasis to offline as the process matures. In general, effective and efficient algorithms should be developed for data acquisition and control of metrology systems.

REFERENCES

[1] International Technology Roadmap for Semiconductors (2001). [Online]. Available: http://www.public.itrs.net.

[2] T. Raz, Y. T. Herer, and A. Grosfeld-Nir, "Economic optimization of off-line inspection," *IIE Trans.*, vol. 32, no. 3, pp. 205–217, March 2000.

[3] R. Williams, D. Gudmundsson, K. Monahan, and J. G. Shanthikumar, "Optimized sample planning for wafer defect inspection," in *1999 Proc. IEEE Int. Symp. Semiconductor Manufacturing Conf.*, pp. 43–46.

[4] K. Tang and H. Schneider, "Selection of the optimal inspection precision level for a complete inspection plan," *J. Quality Technol.*, vol. 20, no. 3, pp. 153–156, July 1988.

[5] D. L. Dane, T. DiFloria, and D. W. Jimenez, "Modeling the cost of ownership of assembly and inspection," *IEEE Trans. Comp. Packag. Technol.*, vol. 19, no. 1, pp. 57–60, Jan. 1996.

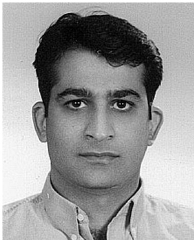
[6] S. M. Ross, *Introduction to Probability Models*, 7th ed. New York: Academic, 2000.

[7] R. C. Leachman, J. Plummer, and N. Sato-Misawa, *Understanding fab economics*. Berkeley: Competitive Semiconductor Manufacturing (CSM) Program Publication, Univ. of California, 1999, vol. CSM-47.

[8] L. W. Schruben, *Graphical Simulation Modeling and Analysis: Using SIGMA for Window*. Danvers, MA: Boyd & Fraser, 1995.

[9] P. Jula, "The economic impact of metrology methods in semiconductor manufacturing," Dept. of Elect. Eng. and Comput. Sci., Univ. of California, Berkeley, M.S. Project Report, 2001.

[10] D. Gerold, R. Hershey, K. McBrayer, and J. Sturtevant, *Run-to-Run Control Benefits to Photolithography*. Lake Tahoe, NV: Sematech AEC/APC, Sept. 1997.



Payman Jula received the B.S. degree in electrical engineering from Tehran University, Iran, in 1992 and the M.S. degree in industrial engineering from Western Michigan University, Kalamazoo, MI, in 1996. He has been with the University of California at Berkeley since 1996, where he completed the Management of Technology program in Haas business school (1998), and received the M.S. degree in electrical engineering and computer science in 2001.

He is currently a researcher at the Competitive Semiconductor Manufacturing (CSM) group at U.C. Berkeley and a Ph.D. candidate at the department of Industrial Engineering and Operations Research.

Mr. Jula's research interests include the economical analysis, simulation, production planning and scheduling, and supply chain management in the semiconductor industry. He has consulted number of companies in these areas.

Costas J. Spanos (S'79–M'81–SM'96–F'00) was born in 1957 in Piraeus, Greece. He received the electrical engineering diploma (with honors) from the National Technical University, Athens, Greece, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1981 and 1985, respectively.

From June 1985 to July 1988, he was with the advanced CAD development group of Digital Equipment Corporation in Hudson MA, where he worked on the statistical characterization, simulation and diagnosis of VLSI processes. In 1988, he joined the faculty at the Department of Electrical Engineering and Computer Sciences of the University of California, Berkeley, where he is now a Professor. He has been the Director of the Berkeley Microfabrication Laboratory from 1992 to 2000. His research interests include the development of flexible manufacturing systems, the application of statistical analysis in the design and fabrication of integrated circujits, and the development and deployment of novel sensors and computer-aided techniques in semiconductor manufacturing.

Dr. Spanos was the editor of the IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING from 1991 to 1994. He has published more than 100 refereed publications and has received best paper awards in 1992, 1997, and 2001. He has served in the technical committees of the IEEE Symposium on VLSI Technology, the International Semiconductor Manufacturing Sciences Symposium, the Advanced Semiconductor Manufacturing Symposium and the International Workshop on Statistical Metrology.

Robert C. Leachman received the A.B. degree in mathematics and physics, the M.S. degree in operations research and the Ph.D. degree in operations research, all from the University of California at Berkeley (U. C. Berkeley).

He is a Professor of Industrial Engineering and Operations Research at the University of California at Berkeley, where he serves as Director of the Competitive Semiconductor Manufacturing (CSM) Program. He is the author of more than 50 technical publications concerning production operations management and productivity improvement, and he has been a consultant in these areas to many corporations. He has been a member of the U. C. Berkeley faculty since 1979.

In 1981, Dr. Leachman was the winner of the Nicholson Prize from the Operations Research Society of America. In 1995, Dr. Leachman was the winner of the Franz Edelman Award Competition sponsored by the Institute for Operations Research and the Management Sciences (INFORMS), recognizing his work to design and implement automated production planning systems in the semiconductor industry. In 2001, he was the runner-up in the Franz Edelman Award Competition, recognizing his work for automated floor scheduling and cycle time management in the semiconductor industry. The Edelman Award is the highest accolade from INFORMS, given annually recognizing outstanding practice of the management sciences.