IIE Transactions (2002) **34**, 191–209

Using in-line equipment condition and yield information for maintenance scheduling and dispatching in semiconductor wafer fabs

THOMAS W. SLOAN[1]* and J. GEORGE SHANTHIKUMAR[2]

[1]*Department of Management, School of Business Administration, University of Miami, Coral Gables, FL 33124, USA*
E-mail: tsloan@miami.edu
[2]*Department of Industrial Engineering and Operations Research and Walter A. Haas School of Business, University of California at Berkeley, Berkeley, CA 94720, USA*
E-mail: shanthikumar@ieor.berkeley.edu

Received January 2000 and accepted July 2001

Yield is one of the most important measures of manufacturing performance in the semiconductor industry, and equipment condition plays a critical role in determining yield. Researchers and practitioners alike have traditionally treated the problems of equipment maintenance scheduling and production dispatching independently, ignoring how equipment condition may affect different product types or families in different ways. This paper addresses the problem of how to schedule maintenance and production for a multiple-product, multiple-stage production system. The problem is based on the situation found in semiconductor wafer fabrication where the equipment condition deteriorates over time, and this condition affects the yield of the production process. We extend a recently developed Markov decision process model of a single-stage system to account for the fact that semiconductor wafers have multiple layers and thus make repeated visits to each workstation. We then propose a methodology by which the single-stage results can be applied in a multi-stage system. Using a simulation model of a four-station wafer fab, we test the policies generated by the model against a variety of other maintenance and dispatching policy combinations. The results indicate that our method provides substantial improvements over traditional methods and performs better as the diversity of the product set increases. In the scenarios examined, the reward earned using the policies from the combined production and maintenance scheduling method was an average of more than 70% higher than the reward earned using other policy combinations such as a fixed-state maintenance policy and a first-come, first-serve dispatching policy.

1. Introduction

Semiconductor wafer fabrication facilities (fabs) often manufacture a wide variety of products. Using the same equipment set, one fab can produce memory chips, logic devices, and microprocessors. *Yield*, defined as the fraction of working devices that emerge from the manufacturing process, is the most important performance measure for many fabs. Particulate contamination within the process equipment is major source of yield loss (Borden and Larson, 1989), and the level of equipment contamination will affect the yield of different product types differently.

In recent years, a great deal of scientific and engineering effort has been invested in developing ways to assess equipment condition in real time. Several researchers have explored how in-line data can be collected and used to predict end-of-line yields (Nurani *et al.*, 1996; Lee, 1997; Cunningham and MacKinnon, 1998; Nurani *et al.*, 1998; Segal *et al.*, 1999). Other researchers have studied how to use devices such as *in situ* particle monitors, which indicate particle levels in the equipment during production, to improve cleaning schedules and increase yields (Borden and Larson, 1989; Peters, 1992; Hunter and Nguyen, 1993; Takahashi and Daugherty, 1996). In all of this work, production decisions such as the release of new work into the fab and lot dispatching are not considered.

Another stream of research has been directed toward linking yield information with lot-sizing decisions, i.e., the decision of *how many* units to feed into a production process given that some yield loss is expected. Yano and Lee (1995) review over 120 articles related to lot-sizing models with random yield. Much of the research in this area has focused on multiple-stage, single-product

---

* Corresponding author

0740-817X

systems; examples include the works by Lee and Yano (1988), Gong and Matsuo (1990), and Wein (1992). Examples of research on multiple-stage, multiple-product systems include the works by Akella *et al.* (1992) and Tang (1992), Gong and Matsuo (1997), Shen (1997), and Chitchachornvanich (1998). These models account for yield loss, but they do not explicitly link the process or equipment condition to the yield loss.

Most research that explicitly links equipment condition to yield is focused on single-product, single-stage systems. The classic Economic Manufacturing Quantity (EMQ) model has been extended in various ways to account for changing equipment condition and inspection policies; examples include the works by Rosenblatt and Lee (1986), Porteus (1986, 1990), Lee and Rosenblatt (1987, 1989), Lee and Park (1991), and Makis and Fung (1998). Since only one product is produced, the question of dispatching, or *which* product to process next, is not considered.

The first research to incorporate yield information in dispatching decisions is the work by Cunningham (1995). The author demonstrates how using yield information can increase throughput, reduce flow time, and improve yield predictability. However, the process yield is not explicitly connected with the equipment condition. In summary, most research related to the problem under study does not explicitly link equipment condition to yield loss. The models that do make this connection are models of single-product systems, and therefore they do not address the issue of dispatching.

Sloan and Shanthikumar (2000) develop a model that simultaneously determines maintenance and production schedules for a single-stage, multiple-product system. This model is the first of a multi-product system in which equipment condition is explicitly linked to yield loss and in which yield information is used in dispatching decisions, i.e., used proactively for decision making rather than only for predictive purposes. While this model reveals some important lessons in the effort to bridge the gap between maintenance and production scheduling, it is not clear how the results would be implemented in a complex, multi-stage system. Indeed, Yano and Lee (1995) point out the difficulty in solving analytical models of multi-product, multi-stage systems and highlight the need for "heuristic solution procedures that are computationally inexpensive and easy to implement" (Yano and Lee, 1995, p. 331).

With these issues in mind, the current paper extends this earlier work in two ways. First, we extend the model presented by Sloan and Shanthikumar (2000) to account for the fact that semiconductor wafers have multiple layers and thus make repeated visits to each workstation. Second, we develop a methodology by which the single-stage results can be applied in a multi-stage system. Using a simulation model of a four-station wafer fab, we test the policies generated by the new model against a variety of other maintenance and dispatching policy combinations. We examine four scenarios, varying the number of products, the number of layers per product, and the expected yields. The results indicate that the proposed method provides substantial improvements over traditional methods and performs better as the diversity of the product set increases. In the scenarios examined, the reward earned using the policies from the combined production and maintenance scheduling method was an average of more than 70% higher than the reward earned using other policy combinations such as a fixed-state maintenance policy and a first-come, first-serve dispatching policy.

2. Analytical model

2.1. *Problem statement*

We consider the problem of determining the production and maintenance schedules for a multiple-product, multiple-stage system. (The term "product" could potentially include to several product types within the same technology family. For example, several 128 Megabit DRAM products made using the same technology would be in the same family and treated as one product in our model.) At each stage, there is a machine whose condition deteriorates over time, and the condition affects the yield of different product types differently. After observing the machine condition, we must decide whether to stop production and clean the machine or to continue producing. If production is picked, we must also decide which product to process. As mentioned above, the machine condition could be gauged by using an *in situ* particle monitor that detects the number of particles in a piece of equipment while it is operating. This is effectively a continuous measure, but for modeling purposes, we divide the continuum into discrete segments, or *states* (this is discussed in greater detail in Section 6). In addition, we model time in terms of discrete periods.

Let X_m^t denote the state of machine m in period t, where $t = 1, 2, \dots$ Each machine can be in any one of $M + 1$ states, indexed by $i = 0, 1, \dots, M$, where state 0 denotes the best possible condition and state M denotes the worst possible condition. (We have assumed that all machines have the same number of states simply for notational convenience; the model could easily address the case in which each machine has a different number of states.) We can produce any one of K products, indexed by $k = 1, 2, \dots, K$, and each product requires n_k visits to each workstation. We sometimes refer to products as *wafers* and to visits as *layers*. Let a_m^t denote the action taken in period t with respect to machine m. Possible actions include producing one of the product types and cleaning the machine. The cleaning action is denoted as action $K + 1$. Define $\beta_m(i, k, l)$ as the expected yield of product k, layer l

when machine m is in state i. We assume that the yield for all products decreases as the machine condition deteriorates, i.e., that $\beta_m(i,k,l)$ decreases in i for each k and l; however, the magnitude of the change may be different for different products. The profit earned for each unit of product k is R_k, and the cost to clean machine m is C_m, independent of the state. This last assumption is based on the fact that the main cost of cleaning stems from the disruption of production and the time that it takes to disassemble and re-assemble the equipment rather than from the level of contamination.

We assume that changes in the machine state depend only on the current state and the action taken. That is, the machine state changes according to the transition probabilities

$$\begin{aligned} &P_m(j\,|\,i,a)\\ &\quad = \Pr\{X_m^{t+1} = j\,|\,X_m^t = i, a_m^t = a; X_m^{t-1}, a_m^{t-1}; \ldots; X_m^1, a_m^1\},\\ &\quad = \Pr\{X_m^{t+1} = j\,|\,X_m^t = i, a_m^t = a\}. \end{aligned}$$

Under these conditions, $\{X_m^t, t \geq 1\}$ forms a Markov chain, and the problem can be treated as a Markov decision process.

2.2. *Solution procedure*

We wish to extend the single-stage model developed by Sloan and Shanthikumar (2000) to a multiple-stage environment in which products make multiple visits to each workstation. Even with a small number of products, visits, and workstations, the dimensionality of such a problem is quite large. Thus, we propose a heuristic approach, treating each workstation in isolation and using the single-stage model to determine the production and maintenance schedules for each workstation. While such policies may not be optimal in a theoretical sense, we expect them to provide substantial benefits over policies that ignore yield altogether. The following procedure will be used:

1. For each machine, determine a simple maintenance policy (i.e., cleaning threshold).
2. Determine the expected yields for the entire production process based on the simple maintenance policy and a First-Come, First-Serve (FCFS) dispatching policy.
3. For each machine, determine the "optimal" production and maintenance policy using the model below, assuming that all other stations use the simple maintenance policy and an FCFS dispatching policy.

A *policy* is a decision rule that prescribes an action for each state, and our objective is to determine a policy that maximizes the long-run expected average profit. Define $x_m(i,a)$ as the proportion of time action a is taken when machine m is in state i. Possible actions include cleaning the machine, denoted as action $K+1$, and producing one of the products. We treat each layer of each product as a unique action, so the "produce" action is written as k, l, denoting layer l of product k.

As described above, we first determine a simple maintenance policy for each machine, ignoring the yield differences between the products. Since all products are treated the same way, the problem is reduced to choosing between "produce" and "clean" for each state. We simply need to find the point at which it is more profitable to stop and clean the machine than to continue producing. The simple maintenance policy for machine m is determined by solving the linear program below. (Some additional technical assumptions needed to ensure the existence of an optimal policy are included in Appendix A.)

$$\text{maximize} \sum_i \sum_a x_m(i,a) r_m(i,a), \tag{1}$$

$$\text{subject to} \sum_a x_m(i,a) = \sum_i \sum_a x_m(i,a) P_m(j\,|\,i,a) \quad \text{for all } j, \tag{2}$$

$$\sum_i \sum_a x_m(i,a) = 1, \quad \text{and} \tag{3}$$

$$x_m(i,a) \geq 0 \quad \text{for all } i, a, \tag{4}$$

where $r_m(i,a)$ is the reward for taking action a when machine m is in state i. This type of linear program, first proposed by Manne (1960), is a standard method for solving Markov decision processes. Equation (2) represents the state balance equations for the Markov chain that governs the machine state transitions. Recall that the $x_m(i,a)$'s refer to proportions of time, and Equation (3) ensures that the proportions sum to one. Equation (4) requires that all of the proportions are non-negative.

Suppose that we need to meet some production requirements in the sense that we want γ_k to be the fraction of total good output made up of product k, where $\sum_k \gamma_k = 1$. For this part of the solution procedure, we assume that a First-Come, First-Serve (FCFS) scheduling policy is used. In the long-run, an FCFS policy would be equivalent to choosing products randomly according to the production ratios. Since product k requires n_k layers, the "produce" decision is equivalent to producing layer l of product k with probability γ_k/n_k. When the machine is in state i, the expected reward for the "produce" decision is $\sum_k (\gamma_k/n_k) R_k \beta_m(i,k,l)$. The "clean" decision costs C_m, independent of the machine state. We solve the linear program for each machine to determine its simple maintenance policy. The solution tells us the proportion of time the machine spends producing and the proportion of time that it is being cleaned when the simple maintenance policy is employed.

Now we can determine the "optimal" production and maintenance policy for one machine assuming that a simple maintenance policy and an FCFS dispatching policy are used at all other machines. We do this by solving the above linear program with a modified objective function and some additional constraints. Define $Y_m(k,l)$ as the average yield of product k, layer l on machine m when a simple maintenance policy is employed (a different policy is used for each machine). The future expected yield of a wafer is simply the product of the average yields for all future steps (all remaining layers at each machine). Thus, for producing layers $l = 1, 2, \ldots, n_k$ of products $k = 1, 2, \ldots, K$, the reward is

$$\sum_{i=0}^{M}\sum_{k=1}^{K}\sum_{l=1}^{n_k}\left(x_m(i,k,l)R_k\beta_m(i,k,l)\prod_{n\neq l}Y_m(k,n)\prod_{s\neq m}\prod_{n}Y_s(k,n)\right),$$

and for cleaning, the reward is

$$-C_m\sum_{i=0}^{M}x_m(i,K+1).$$

We use these values in the objective function (1). In addition, constraints are needed to ensure that sufficient quantities of the intermediate layers are produced:

$$\sum_i x_m(i,k,l+1) = \sum_i x_m(i,k,l), \quad l = 1,2,\ldots,n_k-1; \quad k = 1,2,\ldots,K.$$

We must also ensure that the finished products are produced in the proper proportions by including the following constraints.

$$\frac{\sum_i x_m(i,k,l)\beta_m(i,k,l)}{\sum_k\sum_l\sum_i x_m(i,k,l)\beta_m(i,k,l)} = \frac{\gamma_k}{n_k}, \quad l = 1,2,\ldots,n_k;\ k = 1,2,\ldots,K,$$

where γ_k is the proportion of (final) product k required.

We denote the optimal production and maintenance policy for machine m as $\pi_m^*(i,a)$. The optimal policy may be randomized, i.e., we may choose actions according to some probability distribution, where $\pi_m^*(i,a)$ is the probability that action a is taken given that machine m is in state i. The optimal policy is determined as follows:

Step 1. Find an optimal solution, $\mathbf{x}_m^* = [x_m^*(i,a)]$. Define the set $S^* = \{i : x_m^*(i,a) > 0, \text{for some } a\}$.

Step 2. For $i \in S^*$, $\pi_m^*(i,a) = x_m^*(i,a)/\sum_a x_m^*(i,a)$; for $i \notin S^*$, $\pi_m^*(i,a) = 1$ for some arbitrary a.

Computationally, the proposed method is not particularly demanding. To determine a production and maintenance policy for a particular machine, one must first solve the linear program specified by Equations (1) through (4) for each machine in the system. Since there are $\sum_k n_k$ possible production actions and one possible maintenance action in each state, each of these initial linear programs has $M \times \left(\sum_k n_k + 1\right)$ variables and $M + 1$ constraints. Next, the expanded linear program outlined above must be solved, and this problem will have $\sum_k n_k - K$ additional constraints for the intermediate layers and $\sum_k n_k$ additional constraints for the finished products. In Section 5, we examine a fab that produces 10 products, each with 20 layers. Two machines, each with five states, exhibit the deterioration pattern of interest. To determine the optimal policy, we solve two linear programs each with 1005 variables and six constraints. The expanded linear program has 1005 variables and 402 constraints. The solution of all of these problems takes only a few seconds using CPLEX on a DEC 8400 workstation. Thus, computing time should not be an issue even for problems involving a large number of products and/or layers per product.

The model implicitly assumes that all products are available at each decision epoch. We would expect this assumption to be close to reality in high-volume, make-to-stock fabs, but it may not be realistic for some environments. In fact, actually implementing the optimal policy may require high levels of Work-In-Process (WIP) inventory which would in turn lead to long flow times. On the other hand, Cunningham (1995) demonstrated that the flow time for *good* chips could be reduced by giving priority to wafers with high yield. In other words, while the overall flow time may increase, the yield is higher, so the average flow time for a working device decreases. For some fabs, the cost of long flow times might outweigh the benefit of increased yields.

2.3. *Monotone optimal policy*

Under some additional assumptions, it can be shown that a monotone optimal policy exists. In the semiconductor industry, leading-edge products typically have very small minimum feature sizes or line widths. A device with a smaller minimum feature size has more circuitry per unit area than a device with a larger minimum feature size, and, other factors being equal, is worth more. However, devices with smaller minimum feature sizes are also more sensitive to the equipment condition, and, therefore, have lower yields than devices with larger minimum feature sizes. To model this phenomenon, we rank the products in descending order of sensitivity to the machine state. A product with a lower index is *more* sensitive and has a higher unit profit than a product with a higher index. We also assume that as the condition of the machine deteriorates, the expected profit decreases less for less-sensitive products than for more-sensitive products. If these assumptions hold, then a monotone optimal policy will exist; that is, we will produce more-sensitive products when the machine condition is good, and as the machine condition deteriorates, we will produce less-sensitive products. Eventually, the machine condition will reach the cleaning threshold, and we will stop and clean the machine. Such policies are desirable because they are intuitive and easy to implement. The details are discussed in Appendix A.

3. Simulation model

Semiconductor manufacturing is an extremely complex process that involves hundreds of steps and requires repeated visits to the same set of workstations. (For a description of the wafer fabrication process refer to Uzsoy *et al.* (1992).) Owing to the complexity and repetitive nature of the manufacturing process, discrete-event computer simulation models have become a popular tool to study wafer fabrication lines. Many researchers have employed simulations to compare the performance of different scheduling policies (Glassey and Resende, 1988; Wein, 1988; Cunningham, 1995; Glassey *et al.*, 1996). Below we describe the details of the simulation model used to compare different combinations of maintenance and scheduling policies.

3.1. *Workstations*

Our fab model consists of four workstations. Each station represents one of the primary process steps in the fabrication process: oxidation/deposition, photolithography, etch, and ion implant/diffusion. The oxidation/deposition and etch stations exhibit the deterioration process described above, and we assume that all of the yield loss will occur during these operations. While additional yield losses will surely occur at other stations in a real fab, our focus is on the effect of yield loss that can be avoided through changes in the dispatching and maintenance policies.

Both the deposition and etch stations have a finite number of machine states. A high state number indicates a worse condition than a low state number. As the machine is used, the condition deteriorates. The transition from state to state is governed by a Markov chain with known transition probabilities, as described in the previous section. In all of the experiments, we model the machine deterioration as a five-state Markov chain. The actual parameter values (e.g., transition probabilities and yields) are different for each station.

As mentioned previously, the yield depends on the product (k), layer (l), machine (m), and machine condition (i). Note that we use the term *yield* to refer to die yield, i.e., the number of good chips on a wafer divided by the total number of chips. Furthermore, we assume that there is no line yield loss, i.e., all of the wafers started in the fab complete the entire production process.

We model the yield at the deposition and etch stations using a beta probability distribution. The beta distribution is often used when actual data are not available (Law and Kelton, 1991). In addition, we know that the yield values must be between zero and one, the same range as the beta distribution. We generated different yield matrices for the deposition and etch stations using the beta pseudo-random number generation procedure described by Bratley *et al.* (1987). We assumed that the yield was the same for each layer of a given product, i.e., $\beta_m(i,k,1) = \beta_m(i,k,2) = \cdots = \beta_m(i,k,n_k)$ for each i and k, but different yields for each layer could easily be incorporated. More details on the yields are discussed in Section 4.3.

We have defined the time units in the simulation – as in the analytical model – in terms of *periods*. One period is required to process each wafer at each station. In reality, it may take longer to process a given product type at a particular station. Even the same product type may have different processing times at the same workstation, depending on its stage of completion. For example, many products require "high current" and "low current" implant operations, and the high current implant takes up to eight times as long as the low current implant. While the simulation model has the capability to incorporate different processing times for different products and workstations, we felt that this would add complexity without shedding any additional light on the questions of interest.

3.2. *Dispatching mechanism*

The heart of the simulation model is the dispatching mechanism. It is the bridge between the analytical model and the actual decision making on the shop floor. Our goal was to find a dispatching policy that could improve performance but would still be relatively easy to implement.

The analytical model implicitly assumes that all layers of all products are available at each decision epoch. The output of the model is a policy that indicates when to clean the machine and when to produce, including which product (and layer) to produce in each state. In practice, however, the specified product may not be available or more than one of the specified product may be available. Thus, we interpret the output of the analytical model as a list of *candidates*. If machine m is currently in state i, layer l of product k is a candidate if $\pi^*_m(i,k,l) > 0$. In principle, one must also specify "secondary" rules, i.e., rules to use when more than one candidate is available and rules to use when no candidates are available. Unless otherwise specified, FCFS is the secondary rule employed. This is an important feature that distinguishes this model from the single-stage model of Sloan and Shanthikumar (2000).

3.3. *Release of new work*

We chose to use a closed-loop release policy, i.e., we release new work into the fab once the level of work remaining drops below a certain threshold. This approach is commonly used in the semiconductor industry. Define W as the total amount of work to be released each time. The release policy is straightforward: when the total amount of work in the system (in terms of layers) drops below a certain threshold, release W layers of new work. For each experiment (described in the next section), we had to determine an appropriate value for W and an appropriate work release threshold. We want to end up with γ_k proportion of *good* wafers of product k, and each wafer of product k requires n_k layers. The proportion of new product k wafers to release is defined as

$$\gamma'_k = \frac{n_k \gamma_k / \overline{Y}_k}{\sum_k n_k \gamma_k / \overline{Y}_k},$$

where $\overline{Y}_k$ is the observed yield of product k thus far. For each k, release $\lceil W\gamma'_k/n_k \rceil$ wafers, where $\lceil x \rceil$ is the smallest integer greater than or equal to x.

The choice of the threshold and W will have an important effect on the results. A high threshold and/or large value of W will result in high WIP and long flow times but give us more flexibility with the dispatching decision, i.e., bring us closer to the policy generated by the analytical model. Fab managers would need to consider this trade-off when choosing the release threshold.

4. Experiments

4.1. *Overview*

The purpose of the simulation study is to explore how production and maintenance schedules that incorporate equipment condition and yield information can affect fab performance. We do this by comparing the performance of the combined production and maintenance schedules generated using the proposed analytical model with the performance of combinations of "traditional" production and maintenance policies. Sloan and Shanthikumar (2000) demonstrated that for single-stage systems, the policies from the analytical model have the biggest effect when many products are produced and when the production is spread out among the products. Thus, we would expect benefits for fabs that:

- Produce multiple-product types.
- Produce a relatively large quantity of each product type.
- Do not place high emphasis on due dates (e.g., produce-to-stock environment).
- Place high emphasis on yield (as opposed to flow time, for example).

With these things in mind, we designed experiments to determine which combinations of dispatching rules and cleaning policies are the most effective and which factors have the biggest influence on performance. We examine four different cases or fabs: a four-product fab in which each product has four layers; a four-product fab in which each product has 10 layers; a four-product fab in which the products have 20, 15, 15, and 12 layers, respectively; and a 10 product fab in which each product has 20 layers.

For each fab, we conduct two sets of experiments. In Experiment 1, all of the conditions necessary to guarantee a monotone optimal policy are met (refer to Section 2.3). In Experiment 2, these conditions are not necessarily met.

4.2. *Dispatching rules and maintenance policies*

Our objective is to maximize the total average long-run profit while maintaining a certain level of output of each product. We assume that there are no due dates and that the setup time is independent of the product and of the sequence of production. The main focus of this study is to compare the performance of a variety of simple, myopic rules. That is, we want to find rules that are relatively easy to determine and easy to implement. We consider the following dispatching policies, divided into categories based on the information that they use:

Simple (non-yield)

Arrival time

- FCFS (First-Come, First-Serve): Select the lot that arrived first at *this* station.
- LCFS (Last-Come, First-Serve): Select the lot that arrived last at *this* station.
- FIS (first in shop): Select the lot that arrived first in the *shop*.

Processing time

- SPT (Shortest Processing Time): Select the lot that has the shortest processing time.
- LPT (Longest Processing Time): Select the lot that has the longest processing time.
- SRPT (Shortest Remaining Processing Time): Select the lot that has the shortest remaining processing time.
- LRPT (Longest Remaining Processing Time): Select the lot that has the longest remaining processing time.

Reward

- VAL (value): Select the lot that has the highest value (i.e., unit profit).

Yield-based

- CYLD (current yield): Select the lot that has the highest current, cumulative yield.

- FRWD (future reward): Select the lot that has the highest expected future reward, i.e., unit profit times cumulative yield so far times expected yield of remaining steps. The expected yield of future steps is computed based on the actual yields observed at those stages thus far during the simulation.

We consider the following equipment maintenance policies:

- ST (fixed-state): A repair threshold (state) is determined based on the expected equipment deterioration. At the beginning of each period, the machine state is observed. If the state is at or above the threshold, the machine is cleaned.
- FT (fixed-time): A repair threshold (state) is determined based on the expected equipment deterioration. We then compute the expected time to reach the threshold, say T periods, given that the machine starts in state 0. The machine is then cleaned every T time periods. Suppose that the repair threshold is state $\hat{\imath}$. Since we have assumed Markov deterioration, $T = 1/p_{\hat{\imath}}$, where $p_{\hat{\imath}}$ is the steady-state proportion of time that the machine spends in state $\hat{\imath}$ (Wolff, 1989).
- FN (fixed-number): Similar to the fixed-time policy. A repair threshold (state) is determined based on the expected equipment deterioration. We then compute the expected number of periods to reach the threshold, say N time periods, given that the machine starts in state 0. Each product requires one time period to process, so we clean the machine after every N units have been produced. If the machine is never idle, then this policy will be identical to the fixed-time policy.

For each simulation run, we must specify a dispatching rule and a maintenance policy. Our benchmark is a first-come, first-serve dispatching rule and a simple threshold policy. We have observed these policies being used in several fabs. Our goal is to investigate the performance differences between various combinations of the scheduling and maintenance policies described above and the policy generated by the analytical model. We test 16 combinations of the dispatching policies and maintenance policies described above for each scenario outlined below.

4.3. *Summary of factors and levels*

As mentioned above, we will examine four fabs. Below we summarize the parameters of the simulation model and report the values used in the simulation runs. For each fab, we examine several different combinations of scheduling and maintenance policies.

- *Number of products*: Fabs 1 through 3 produce four products, and Fab 4 produces 10 products.
- *Layers per product*: The products have four layers each in Fab 1; 10 layers each in Fab 2; 20, 15, 15, and 12 layers each, respectively in Fab 3; and 20 layers each in Fab 4.
- *Output proportions*: For all scenarios, the output is evenly distributed among all products. Thus, $\gamma_k = 0.25$ for each k in Fabs 1 through 3, and $\gamma_k = 0.10$ for each k in Fab 4.
- *Unit profit*: In Fabs 1 through 3, the profit per unit (in dollars) is 800, 400, 200, and 100, respectively, for products 1 through 4. For Fab 4, the profit per unit (in dollars) for is 1000, 900, 800, 700, 600, 500, 400, 300, 200, and 100, respectively, for products 1 through 10.
- *Equipment deterioration*: For all scenarios, we assume "slow" deterioration for the deposition station and "medium" deterioration for the etch station. The transition probability matrices are reported in Appendix B.
- *Cleaning cost*: For all scenarios, the cost to clean the deposition station is \$10, and the cost to clean the etch station is \$20.
- *Yield*: We generated 10 different yield matrices for each station and each experiment from beta probability distributions, using a different random number seed for each matrix. The means and variances of the distributions used to generate the per layer yields are reported in Table 1. The values generated for a particular matrix were sorted to ensure that the model assumptions discussed in Section 2 were met (e.g., yields tend to be lower as the equipment condition gets worse). We assume that for a given product the yield is the same for each layer, but different per layer yields could easily be incorporated. Fabs 1 through 3, which are all four-product fabs, have the same per layer yields for a given product and a given random number seed, but the final yields will be different because the number of layers per product differs for each fab. The same procedure was used to generate yields for Fab 4. Table 2 reports the mean and variance of final expected yields for each experiment and each fab. Appendix B lists a representative final yield matrix for each fab model and each experiment.

Table 1. Base means and variances of per layer yields for simulation experiments

Exp.	*Oxi./Deposition*		*Etch*		*Monotone policy*
	Mean yield	*Variance (% of mean)*	*Mean yield*	*Variance (% of mean)*	
1a	0.950	10	0.900	20	yes
1b	0.975	10	0.950	20	yes
2a	0.950	10	0.900	20	no
2b	0.975	10	0.950	20	no

Table 2. Expected means and variances of final yields for simulation experiments

		Exp. 1a	*Exp. 1b*	*Exp. 2a*	*Exp. 2b*
Fab 1	Mean yield	0.764	0.867	0.587	0.766
	Variance	0.036	0.010	0.043	0.015
Fab 2	Mean yield	0.556	0.718	0.306	0.535
	Variance	0.105	0.042	0.108	0.058
Fab 3	Mean yield	0.452	0.585	0.194	0.406
	Variance	0.150	0.091	0.137	0.094
Fab 4	Mean yield	0.694	0.769	0.122	0.308
	Variance	0.121	0.075	0.130	0.111

Test runs were performed for each scenario to determine appropriate run lengths. For a given yield level (matrix), we expect the average reward to converge to a limiting value after some time, so we examined the average reward at intervals of 10 000 time units and concluded that the limiting value had been reached if the change from the last inspection point was less than 0.10%. The run length for Fabs 1 and 2 was 2.5 million time units. For Fab 3, a run length of 5 million time units was used, and the run length for Fab 4 was 10 million time units.

5. Results

For each fab and for each experiment, we tested 16 combinations of the dispatching policies and maintenance policies as described in Section 4.2. Recall that the base policy is an FCFS dispatching policy with a fixed-state maintenance policy. Tables 3–5 report the *highest* percentage improvement (as compared to the base policy) for each policy combination category for each fab. Detailed results for all 16 policy combinations for each fab are reported in Appendix C.

The results suggest a number of important lessons. First, the performance of the fixed-time and fixed-number maintenance policies is extremely poor compared to the fixed-state policy. Recall that when using a fixed-state policy, one observes the machine state at each decision epoch, whereas the fixed-time and fixed-number policies are based on the expected deterioration. Lack of real-time information about the equipment condition dramatically reduces yields. While this may seem obvious, some fab managers have been reluctant to invest in the equipment and training needed to implement such cleaning policies. Additional implementation issues are discussed in Section 6.

The second lesson revealed is that dispatching decisions have an important effect on yield and that it is important to match products and machine states. In other words, if different products are affected differently by equipment contamination levels, then these differences should be accounted for in shop-floor scheduling decisions. The

Table 3. Summary results of simulation experiments for Fab 1

Policy		*Exp. 2a*	*Exp. 1a*	*Exp. 2b*	*Exp. 1b*
Dispatch	*Clean*				
Combined	Combined	16.2	12.8	7.4	4.7
Simple (non-yield)	Fixed-state	0.0	0.1	0.0	0.0
Yield-based	Fixed-state	0.1	0.0	0.0	0.1
FCFS	Fixed-number	−81.5	−80.4	−80.6	−80.4

Percent improvement over base policy.

Table 4. Summary results of simulation experiments for Fab 2

Policy		*Exp. 2a*	*Exp. 1a*	*Exp. 2b*	*Exp. 1b*
Dispatch	*Clean*				
Combined	Combined	73.4	43.0	23.2	15.0
Simple (non-yield)	Fixed-state	0.0	0.1	0.1	0.1
Yield-based	Fixed-state	0.3	0.7	0.1	0.2
FCFS	Fixed-number	−140.6	−116.6	−113.4	−108.8

Percent improvement over base policy.

Table 5. Summary results of simulation experiments for Fab 3

Policy		*Exp. 2a*	*Exp. 1a*	*Exp. 2b*	*Exp. 1b*
Dispatch	*Clean*				
Combined	Combined	477.3	168.5	48.4	38.1
Simple (non-yield)	Fixed-state	0.4	0.4	0.1	0.2
Yield-based	Fixed-state	1.2	0.6	0.8	0.8
FCFS	Fixed-number	−605.1	−230.9	−151.0	−139.9

Percent improvement over base policy.

Table 6. Summary results of simulation experiments for Fab 4

Policy		*Exp. 2a*	*Exp. 1a*	*Exp. 2b*	*Exp. 1b*
Dispatch	*Clean*				
Combined	Combined	64.3	57.2	83.5	35.2
Simple (non-yield)	Fixed-state	0.1	0.3	0.0	0.1
Yield-based	Fixed-state	0.2	−0.2	0.1	0.7
FCFS	Fixed-number	−65.9	−147.2	−182.2	−120.2

Percent improvement over base policy.

results from the combined policies generated by the analytical model are substantially better than the results from the simple policies. The improvement ranges from 4.7% (Fab 1, Exp. 1b) to over 400% (Fab 3, Exp. 2a). In most cases, yield-based policies provide a slight improvement over non-yield policies, but the combined policies are still far superior. This result is particularly interesting given that the maintenance policies paired with the combined, simple, and yield-based dispatching policies all have the *same* cleaning threshold. This result supports the findings of the earlier study of single-stage systems by Sloan and Shanthikumar (2000).

It is interesting to note how the magnitude of improvement provided by the combined policies changes under different conditions. Referring to yield values in Tables 1 and 2, we see that the expected yields decrease in the order Exp. 1b > Exp. 2b > Exp. 1a > Exp. 2a. Referring to Tables 3 through 6, the level of improvement provided by the combined policies increases in the opposite order in all but one instance (Fab 4, Exp. 2b). The pattern is illustrated in Fig. 1, which shows the average improvement provided by the Comb./FCFS policies as compared to the base policy for all experiments. This result suggests that since yields are bounded by zero and one there is less opportunity for the combined policies to improve upon the simple policies when products have high expected yields. Put differently, the dispatching decisions will have a smaller influence as the mean yield increases.

The differences between the Experiment 1 and 2 results is somewhat counter-intuitive at first glance. We would expect that the combined policies in Experiment 1, which imposes more structure on the yield values, would provide more improvement than those in Experiment 2. But the results indicate just the opposite. We believe that this is an artifact of the method used to generate the experimental yield values. As revealed in Table 2, the expected yields for Experiment 2 are lower that those for Experiment 1. This fact suggests that what matters most is the yield differences between different states for a given product rather than yield differences between products. Whether or not the assumptions necessary for the monotone optimal policy are met, the combined policies still provide sizable improvements over the simple policies, and the fixed-state maintenance policy is still far superior to the fixed-number and fixed-time policies.

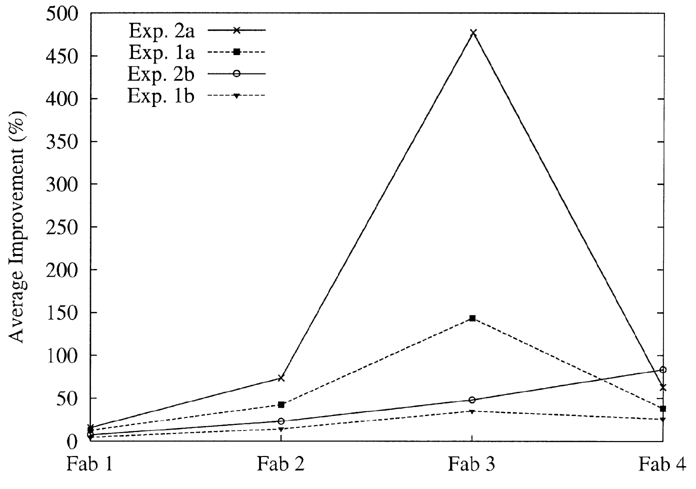

Fig. 1. Average improvement provided by Comb./FCFS policies.

Finally, the most encouraging result is that substantial, consistent improvements can be made by applying the analytical model output in a heuristic fashion. For example, using the combined policy along with an FCFS secondary policy (Comb./FCFS in the tables) is almost always as good as the best combined policy. In other words, the secondary policy does not make a significant difference in performance. Determining the optimal combined policy is straightforward once the yields have been characterized (see Section 6), and minimal effort is needed to apply the policy.

6. Implementation issues

The success of the proposed methodology depends heavily on the availability of fairly detailed information about equipment condition and yield. How would fabs go about gathering this information? First, one must have the ability to observe the machine state. Many researchers have demonstrated how in-line equipment information can be used to assist in process control (May and Spanos, 1993; Bunkofske *et al.*, 1996; Edgar *et al.*, 1999). As discussed above, we have proposed the use of *in situ* particle monitors (ISPMs) to gauge the machine state. Other researchers have demonstrated the effectiveness of using real-time ISPM data in assessing equipment contamination levels, improving equipment cleaning schedules, and increasing yields (Borden and Larson, 1989; Peters, 1992; Hunter and Nguyen, 1993; Takahashi and Daugherty, 1996). While an ISPM provides an essentially continuous measure of state, the particle count could be broken down into discrete states for use in the analytical model, as illustrated in Fig. 2.

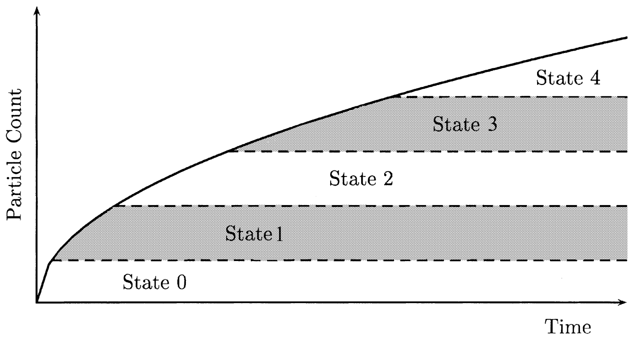

Fig. 2. Translating particle count into discrete states.

Finally, and perhaps most importantly, detailed product yield information must be acquired. The information needed by the analytical model, namely the $\beta(i, k, l)$ values for each product and each layer, is not likely to be readily available in most production fabs. However, ISPMs and the like are becoming more and more prevalent, and tremendous efforts have been made to link particle counts and other in-line measurements to yield, often on a per layer level of detail (Cunningham, 1995; Nurani *et al.*, 1996, 1998; Lee, 1997; Cunningham and MacKinnon, 1998; Segal *et al.*, 1999). Initial yield estimates could be developed using the techniques described in these references, and the yield values would be updated as more data are collected and analyzed. Furthermore, once the relationship between yield and machine state is characterized for a few product families, this information could be used as a starting point for related product families, i.e., those with similar designs and/or circuit dimensions. Thus, "perfect" information about yield is not required initially to benefit from the in-line measurements currently available. Even using simple policies such as, "Do not run Product 1 when the particle count is greater than X," could help increase yields.

The current study demonstrates the value of using in-line process and yield information proactively to improve yield rather than simply for predictive purposes. It is our hope that this lesson will provide added incentive to develop the required technologies and expertise to collect and analyze more detailed data. Clearly, the implementation details need to be adapted to the specific needs and data-collecting capabilities of particular fabs.

7. Conclusions

The purpose of this paper was to examine the effects of using in-line equipment condition and yield information for equipment maintenance scheduling and dispatching. The problem was motivated by an application is semiconductor manufacturing where particulate contamination has a substantial effect on yield, and different products types or families are affected to different degrees by such contamination. We extended a recently developed Markov decision process model to simultaneously determine the equipment maintenance and production schedules for each stage of a multiple-stage, multiple-product production system with the objective of maximizing the long-run expected average profit. A simulation model of a four-station semiconductor wafer fab was used to compare the performance of policies generated by the model with that of other policies commonly used in practice. We examined many different scenarios, varying factors such as the number of products, the number of layers per product, and the expected yields.

Several conclusions may be drawn from this study. First, substantial gains can be made simply by obtaining accurate information about the equipment condition. Fixed-state maintenance policies that use real-time information about equipment condition are much better than fixed-time and fixed-number policies that rely only on expected deterioration patterns. Second, incorporating yield information into the simple policies – e.g., processing the product with the highest attained yield first – provides only a slight improvement over a simple first-come, first-serve policy. Third, we found that the combined policies generated by the analytical model were consistently superior to any simple policy, even when the same maintenance policy was used. In other words, the dispatching policy can have a big impact on final yields and profits. The degree of improvement increases as the mean yield decreases and as the number of layers per product increases. A greater diversity of products gives the combined policies more opportunities to match the products to the appropriate machine states, resulting in higher yields than the simple dispatching policies. In the scenarios studied, the improvement provided by the combined policies ranged from 4.7% to more than 400%, with an average of more than 70%. In an industry as competitive as semiconductor manufacturing, even small improvements in yield can mean the difference between long-term success and failure.

The results of the simulation study are encouraging. We have seen that the benefits achieved in the single-stage environment can be extended to a multiple-stage system. Furthermore, the policies are easy to implement once the analytical model has been solved. More work needs to be done on expanding the simulation model to include more "real world" complexities such as batch operations, equipment failures (which would allow for so-called "opportunistic" maintenance) and sequence-dependent setups. In addition, including multiple machines at each station and allowing multiple repair actions would also be of interest. Future research will be directed toward studying these issues.

Appendices

Appendix A: Optimal policies (from Section 2)

A.1. Existence of an optimal policy

Several assumptions are needed to ensure the existence of an optimal solution to the optimization problem in Section 2. We assume that rewards are bounded and that the cost to clean the machine m, denoted as C_m, does not depend on the machine condition. The primary cost of cleaning is due to the disruption of production caused by taking the equipment apart and putting it back together, so the level of contamination will not significantly change the cost to clean the equipment. Therefore, it seems that a constant cleaning cost is appropriate in this context, but a variable cleaning cost could easily be incorporated. We assume that as the machine condition deteriorates, the yield decreases, and the yield for all products is zero when the machine is in the worst state. (The terms *decreasing* and *increasing* are used in the non-strict sense.) Cleaning can be initiated from any state, and this action is successful with probability one. These assumptions are summarized as follows:

Assumption 1: $|R_a| < \infty$ for all a.
Assumption 2: $\beta_m(i,a)$ decreases in i for all a.
Assumption 3: $\beta_m(M,a) = 0$ for $a \neq K+1$.
Assumption 4: $\beta_m(i,K+1) = 1$ for all i.

We assume that once in state M, the machine cannot leave this state unless it is cleaned. If production is chosen, the state transition is not affected by the choice of which product to produce. This assumption reflects the operations under consideration where we do not expect the deterioration to be influenced by product type, because processing one product type does generate more particles than any other. In other contexts, this assumption may not hold, and the formulation would have to be modified accordingly. We also assume that cleaning the machine returns it to state 0 with probability one. These conditions are expressed as

Assumption 5:

$$P_m(j\,|\,i,a) = \begin{cases} P_m(j\,|\,i), & \text{for } a = 1,2,\ldots,K, \\ 1, & \text{for } a = K+1, j = 0, \\ 0, & \text{for } a = K+1, j \neq 0. \end{cases}$$

We assume that as the machine condition deteriorates, it is more likely that the machine will go to a worse state than a better state. We can express this as

Assumption 6:
For each l, $\sum_{j=l}^{M} P_m(j\,|\,i)$ increase in i.

If these assumptions hold, a stationary, average-reward optimal policy will exist (Sloan and Shanthikumar, 2000).

A.2. Existence of an monotone policy

Additional assumptions are needed to prove the existence of a monotone optimal policy. Specifically, we assume that a product with a lower index is *more* sensitive and is worth more than a product with a higher index. This can be written as

Assumption 7: R_a decreases in a.
Assumption 8: $\beta_m(i,a)$ increases in a for all i.

Table A1. Final yields for Fab 1

Exp.	*Oxi./Dep.*	*Etch*
1a	$\begin{bmatrix} 0.9606 & 0.9801 & 0.9900 & 1.0000 \\ 0.8295 & 0.8877 & 0.9327 & 0.9494 \\ 0.8122 & 0.8619 & 0.9132 & 0.9143 \\ 0.7938 & 0.8284 & 0.8896 & 0.9103 \end{bmatrix}$	$\begin{bmatrix} 0.9224 & 0.9606 & 0.9801 & 1.0000 \\ 0.6022 & 0.8110 & 0.8779 & 0.9610 \\ 0.5677 & 0.7472 & 0.7936 & 0.9315 \\ 0.4935 & 0.5837 & 0.7274 & 0.8824 \end{bmatrix}$
1b	$\begin{bmatrix} 0.9606 & 0.9801 & 0.9900 & 1.0000 \\ 0.9338 & 0.9374 & 0.9516 & 0.9735 \\ 0.8780 & 0.9151 & 0.9339 & 0.9675 \\ 0.8443 & 0.8861 & 0.9200 & 0.9646 \end{bmatrix}$	$\begin{bmatrix} 0.9224 & 0.9606 & 0.9801 & 1.0000 \\ 0.8508 & 0.9533 & 0.9754 & 0.9998 \\ 0.7911 & 0.9114 & 0.9707 & 0.9944 \\ 0.7596 & 0.8463 & 0.9583 & 0.9849 \end{bmatrix}$
2a	$\begin{bmatrix} 0.9336 & 0.8406 & 0.8253 & 0.8724 \\ 0.9050 & 0.8277 & 0.8036 & 0.7979 \\ 0.7958 & 0.7575 & 0.7618 & 0.7192 \\ 0.7677 & 0.7173 & 0.7263 & 0.6997 \end{bmatrix}$	$\begin{bmatrix} 0.9160 & 0.7412 & 0.6911 & 0.8381 \\ 0.5984 & 0.6989 & 0.6217 & 0.4730 \\ 0.5237 & 0.5012 & 0.6046 & 0.4413 \\ 0.4392 & 0.4496 & 0.5104 & 0.4239 \end{bmatrix}$
2b	$\begin{bmatrix} 0.9739 & 0.9230 & 0.9120 & 0.9767 \\ 0.9605 & 0.9138 & 0.8947 & 0.9430 \\ 0.8880 & 0.8488 & 0.8898 & 0.8301 \\ 0.8603 & 0.8078 & 0.8538 & 0.7970 \end{bmatrix}$	$\begin{bmatrix} 0.9573 & 0.8633 & 0.8349 & 0.9164 \\ 0.7799 & 0.8394 & 0.7941 & 0.6999 \\ 0.7332 & 0.7186 & 0.7837 & 0.6783 \\ 0.6769 & 0.6840 & 0.7246 & 0.6662 \end{bmatrix}$

Table A2. Final yields for Fab 2

Exp.	*Oxi./Dep.*	*Etch*
1a	$\begin{bmatrix} 0.9044 & 0.9511 & 0.9753 & 1.0000 \\ 0.6266 & 0.7424 & 0.8402 & 0.8783 \\ 0.5944 & 0.6897 & 0.7970 & 0.7993 \\ 0.5614 & 0.6246 & 0.7464 & 0.7906 \end{bmatrix}$	$\begin{bmatrix} 0.8171 & 0.9044 & 0.9511 & 1.0000 \\ 0.2815 & 0.5923 & 0.7222 & 0.9053 \\ 0.2428 & 0.4826 & 0.5611 & 0.8374 \\ 0.1711 & 0.2603 & 0.4512 & 0.7315 \end{bmatrix}$
1b	$\begin{bmatrix} 0.9044 & 0.9511 & 0.9753 & 1.0000 \\ 0.8427 & 0.8508 & 0.8833 & 0.9352 \\ 0.7223 & 0.8010 & 0.8428 & 0.9208 \\ 0.6550 & 0.7391 & 0.8119 & 0.9137 \end{bmatrix}$	$\begin{bmatrix} 0.8171 & 0.9044 & 0.9511 & 1.0000 \\ 0.6677 & 0.8872 & 0.9397 & 0.9996 \\ 0.5566 & 0.7930 & 0.9284 & 0.9860 \\ 0.5029 & 0.6589 & 0.8991 & 0.9627 \end{bmatrix}$
2a	$\begin{bmatrix} 0.8421 & 0.6479 & 0.6187 & 0.7109 \\ 0.7791 & 0.6233 & 0.5789 & 0.5688 \\ 0.5650 & 0.4993 & 0.5065 & 0.4387 \\ 0.5164 & 0.4357 & 0.4495 & 0.4095 \end{bmatrix}$	$\begin{bmatrix} 0.8030 & 0.4730 & 0.3970 & 0.6430 \\ 0.2770 & 0.4084 & 0.3048 & 0.1539 \\ 0.1985 & 0.1778 & 0.2842 & 0.1294 \\ 0.1278 & 0.1356 & 0.1861 & 0.1170 \end{bmatrix}$
2b	$\begin{bmatrix} 0.9360 & 0.8184 & 0.7943 & 0.9427 \\ 0.9042 & 0.7982 & 0.7572 & 0.8636 \\ 0.7430 & 0.6638 & 0.7469 & 0.6279 \\ 0.6866 & 0.5865 & 0.6736 & 0.5671 \end{bmatrix}$	$\begin{bmatrix} 0.8967 & 0.6926 & 0.6368 & 0.8038 \\ 0.5372 & 0.6455 & 0.5619 & 0.4098 \\ 0.4603 & 0.4377 & 0.5438 & 0.3790 \\ 0.3769 & 0.3870 & 0.4469 & 0.3623 \end{bmatrix}$

Table A3. Final yields for Fab 3

Exp.	*Oxi./Dep.*	*Etch*
1a	$\begin{bmatrix} 0.8179 & 0.9276 & 0.9631 & 1.0000 \\ 0.3926 & 0.6397 & 0.7701 & 0.8558 \\ 0.3533 & 0.5728 & 0.7115 & 0.7643 \\ 0.3151 & 0.4936 & 0.6448 & 0.7543 \end{bmatrix}$	$\begin{bmatrix} 0.6676 & 0.8601 & 0.9276 & 1.0000 \\ 0.3728 & 0.6984 & 0.7813 & 0.9783 \\ 0.3309 & 0.6169 & 0.7277 & 0.9300 \\ 0.2018 & 0.4024 & 0.6760 & 0.8508 \end{bmatrix}$
1b	$\begin{bmatrix} 0.8179 & 0.9276 & 0.9631 & 1.0000 \\ 0.7102 & 0.7848 & 0.8301 & 0.9227 \\ 0.5217 & 0.7169 & 0.7737 & 0.9057 \\ 0.4290 & 0.6354 & 0.7316 & 0.8974 \end{bmatrix}$	$\begin{bmatrix} 0.6676 & 0.8601 & 0.9276 & 1.0000 \\ 0.4459 & 0.8357 & 0.9110 & 0.9995 \\ 0.3098 & 0.7062 & 0.8946 & 0.9832 \\ 0.2529 & 0.5349 & 0.8525 & 0.9554 \end{bmatrix}$
2a	$\begin{bmatrix} 0.7091 & 0.5215 & 0.4867 & 0.6640 \\ 0.6070 & 0.4921 & 0.4404 & 0.5081 \\ 0.3193 & 0.3528 & 0.3605 & 0.3721 \\ 0.2667 & 0.2876 & 0.3014 & 0.3425 \end{bmatrix}$	$\begin{bmatrix} 0.6449 & 0.3253 & 0.2501 & 0.5886 \\ 0.0767 & 0.2610 & 0.1682 & 0.1058 \\ 0.0394 & 0.0750 & 0.1515 & 0.0860 \\ 0.0163 & 0.0499 & 0.0803 & 0.0762 \end{bmatrix}$
2b	$\begin{bmatrix} 0.8762 & 0.7404 & 0.7079 & 0.9317 \\ 0.8176 & 0.7131 & 0.6589 & 0.8386 \\ 0.5520 & 0.5409 & 0.6455 & 0.5721 \\ 0.4714 & 0.4491 & 0.5529 & 0.5062 \end{bmatrix}$	$\begin{bmatrix} 0.8040 & 0.5764 & 0.5082 & 0.7695 \\ 0.2886 & 0.5186 & 0.4212 & 0.3428 \\ 0.2119 & 0.2896 & 0.4010 & 0.3121 \\ 0.1421 & 0.2408 & 0.2988 & 0.2957 \end{bmatrix}$

We also assume that as the condition of the machine deteriorates, the expected profit decreases less for less-sensitive products than for more-sensitive products. This can be expressed as

Assumption 9:

$$\begin{aligned} &R_{a-1}\beta_m(i-1,a-1) - R_{a-1}\beta_m(i,a-1) \\ &\geq R_a\beta_m(i-1,a) - R_a\beta_m(i,a) \geq 0, \end{aligned}$$

for $a = 2, 3, \ldots, K$ and $i = 1, 2, \ldots, M-1$. If these assumptions hold, it can be shown that a monotone optimal policy exists (Sloan and Shanthikumar, 2000).

Appendix B: Data for simulation experiments

The machine state transition probabilities for "slow" and "med" deterioration are

$$\boldsymbol{P}_{\text{slow}} = \begin{bmatrix} 0.9 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.9 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.9 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.9 & 0.1 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$\boldsymbol{P}_{\text{med}} = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

The yield matrices were generated from beta distributions with (per layer) means and variances listed in Table 1. The beta distribution is useful in this case because we do not have actual yield data, but clearly the values must be between zero and one. For Experiment 1, we fixed base yields for each product (i.e., state 0 yield) and then generated the yields for other states, discarding those that were larger than the base yield. We then tested to see if the assumptions needed to ensure a monotone policy had been met. If not, we discarded the entire matrix and started over.

For Fabs 1 through 3, we generated 10 yield matrices for Experiment 1 and 10 yield matrices for Experiment 2. For Fab 4, which has 10 products, we generated 10 more yield matrices for each experiment. Tables A1–A4 report the final yield values for a representative yield matrix for the oxi./dep. and etch stations of each fab for Experiments 1 and 2 (each column represents one product). Both the oxi./dep. and etch stations have five states, and the yield is zero in the worst state for all products and layers. That is, $\beta_m(M, k, l) = 0$ for each k and each l, so it is not necessary to report the final yields for state M.

Appendix C: Detailed results of simulation experiments

For each fab, we tested 16 combinations of dispatching and maintenance policies. The policies are divided into four categories: *combined* policies generated using the analytical model, *simple* (*non-yield*) dispatching policies with a *fixed-state* threshold maintenance policy, *yield-based* dispatching policies with a *fixed-state* threshold maintenance policy, and *FCFS* dispatching with a *fixed-number* or *fixed-time* maintenance policy. The base policy

Table A4. Final yields for Fab 4

Exp.	*Oxi./Dep.*	*Etch*
1a	$\begin{bmatrix} 0.8179 & 0.9046 & 0.9512 & 0.9753 & 0.9876 & 0.9938 & 0.9969 & 0.9984 & 0.9992 & 1.0000 \\ 0.7091 & 0.7916 & 0.8650 & 0.9427 & 0.9813 & 0.9903 & 0.9948 & 0.9964 & 0.9980 & 0.9994 \\ 0.6070 & 0.6977 & 0.7643 & 0.8261 & 0.8459 & 0.9060 & 0.9140 & 0.9234 & 0.9613 & 0.9697 \\ 0.3193 & 0.6666 & 0.7438 & 0.8171 & 0.8404 & 0.8993 & 0.9078 & 0.9167 & 0.9608 & 0.9690 \end{bmatrix}$	$\begin{bmatrix} 0.6676 & 0.8179 & 0.9046 & 0.9512 & 0.9753 & 0.9876 & 0.9938 & 0.9969 & 0.9984 & 1.0000 \\ 0.6449 & 0.7981 & 0.8980 & 0.9481 & 0.9719 & 0.9858 & 0.9924 & 0.9966 & 0.9980 & 0.9994 \\ 0.0767 & 0.1766 & 0.5366 & 0.5432 & 0.9503 & 0.9782 & 0.9896 & 0.9951 & 0.9977 & 0.9990 \\ 0.0394 & 0.1198 & 0.5020 & 0.5422 & 0.9488 & 0.9766 & 0.9892 & 0.9949 & 0.9976 & 0.9989 \end{bmatrix}$
1b	$\begin{bmatrix} 0.8179 & 0.9046 & 0.9512 & 0.9753 & 0.9876 & 0.9938 & 0.9969 & 0.9984 & 0.9992 & 1.0000 \\ 0.8176 & 0.9044 & 0.9510 & 0.9752 & 0.9875 & 0.9937 & 0.9968 & 0.9984 & 0.9992 & 1.0000 \\ 0.5520 & 0.6060 & 0.7300 & 0.8425 & 0.9014 & 0.9611 & 0.9715 & 0.9815 & 0.9864 & 0.9902 \\ 0.4714 & 0.5236 & 0.6648 & 0.8226 & 0.8893 & 0.9471 & 0.9699 & 0.9812 & 0.9862 & 0.9899 \end{bmatrix}$	$\begin{bmatrix} 0.6676 & 0.8179 & 0.9046 & 0.9512 & 0.9753 & 0.9876 & 0.9938 & 0.9969 & 0.9984 & 1.0000 \\ 0.2886 & 0.4056 & 0.5743 & 0.8620 & 0.9268 & 0.9444 & 0.9928 & 0.9958 & 0.9978 & 0.9999 \\ 0.2119 & 0.3157 & 0.5274 & 0.8083 & 0.8831 & 0.9323 & 0.9921 & 0.9954 & 0.9978 & 0.9998 \\ 0.1421 & 0.2957 & 0.5042 & 0.7846 & 0.8571 & 0.9194 & 0.9820 & 0.9859 & 0.9967 & 0.9992 \end{bmatrix}$
2a	$\begin{bmatrix} 0.7669 & 0.6623 & 0.6601 & 0.8154 & 0.3586 & 0.7199 & 0.6025 & 0.9272 & 0.7282 & 0.2973 \\ 0.5884 & 0.6604 & 0.3654 & 0.6612 & 0.3353 & 0.4425 & 0.3544 & 0.3308 & 0.6469 & 0.2762 \\ 0.3484 & 0.3282 & 0.2126 & 0.4324 & 0.1468 & 0.4018 & 0.3034 & 0.2259 & 0.4426 & 0.2654 \\ 0.2709 & 0.1589 & 0.2085 & 0.3982 & 0.1452 & 0.1888 & 0.2709 & 0.2023 & 0.1852 & 0.2432 \end{bmatrix}$	$\begin{bmatrix} 0.8944 & 0.7506 & 0.9375 & 0.9996 & 0.2701 & 0.1898 & 0.6354 & 0.9911 & 0.2704 & 0.3023 \\ 0.6054 & 0.7474 & 0.7467 & 0.2489 & 0.1218 & 0.1161 & 0.0416 & 0.8498 & 0.0582 & 0.0932 \\ 0.1085 & 0.0856 & 0.1311 & 0.1836 & 0.0932 & 0.0629 & 0.0238 & 0.0882 & 0.0445 & 0.0893 \\ 0.0130 & 0.0196 & 0.0205 & 0.0122 & 0.0121 & 0.0162 & 0.0184 & 0.0157 & 0.0293 & 0.0548 \end{bmatrix}$
2b	$\begin{bmatrix} 0.8762 & 0.6698 & 0.6309 & 0.8887 & 0.9493 & 0.8668 & 0.6434 & 0.8420 & 0.8720 & 0.9452 \\ 0.8176 & 0.6371 & 0.5733 & 0.7458 & 0.6775 & 0.7211 & 0.6011 & 0.8373 & 0.6927 & 0.7412 \\ 0.5520 & 0.4407 & 0.5579 & 0.3942 & 0.6469 & 0.4616 & 0.5525 & 0.7141 & 0.3806 & 0.6818 \\ 0.4714 & 0.3440 & 0.4538 & 0.3216 & 0.5837 & 0.4173 & 0.5028 & 0.6261 & 0.3234 & 0.5506 \end{bmatrix}$	$\begin{bmatrix} 0.8040 & 0.4797 & 0.4056 & 0.6461 & 0.9846 & 0.8936 & 0.4283 & 0.8426 & 0.1623 & 0.5042 \\ 0.2886 & 0.4167 & 0.3157 & 0.1679 & 0.4953 & 0.5901 & 0.3561 & 0.5743 & 0.1441 & 0.4404 \\ 0.2119 & 0.1916 & 0.2957 & 0.1436 & 0.4347 & 0.2050 & 0.2892 & 0.5274 & 0.1341 & 0.2869 \\ 0.1421 & 0.1498 & 0.1997 & 0.1312 & 0.3301 & 0.1787 & 0.2373 & 0.3971 & 0.1226 & 0.1960 \end{bmatrix}$

Table A5. Detailed results of Fab 1 simulations

Oxi./Dep.		*Etch*		*Exp. 1a*			*Exp. 1b*			*Exp. 2a*			*Exp. 2b*		
Dispatch[†]	*Clean*[‡]	*Dispatch*[†]	*Clean*[‡]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]
Comb./FCFS	Comb.	Comb./FCFS	Comb.	66.12	1.60	12.8	72.32	1.04	4.6	51.49	4.54	15.8	65.26	2.99	7.4
Comb./FRWD	Comb.	Comb./FRWD	Comb.	63.79	1.80	8.8	71.48	1.02	3.4	49.70	4.54	11.8	63.52	2.78	4.5
Comb./VAL	Comb.	Comb./VAL	Comb.	63.77	1.80	8.8	71.53	1.02	3.5	48.70	3.89	9.6	63.36	2.71	4.3
Comb./CYLD	Comb.	Comb./CYLD	Comb.	65.29	1.79	11.4	72.36	1.04	4.7	51.66	4.78	16.2	64.92	2.89	6.8
FCFS	ST(4)	Comb./FCFS	Comb.	64.35	1.84	9.8	71.13	1.22	2.9	50.10	4.48	12.7	63.56	2.98	4.6
Comb.	Comb.	FCFS	ST(4)	59.79	1.70	2.0	70.00	1.41	1.3	46.40	4.47	4.4	62.08	2.79	2.1
FCFS	ST(4)	FCFS	ST(4)	58.63	1.74	–	69.12	1.53	–	44.45	3.87	–	60.77	2.82	–
LCFS	ST(4)	LCFS	ST(4)	58.55	1.76	−0.1	69.14	1.53	0.0	44.37	3.87	−0.2	60.76	2.81	0.0
FIS	ST(4)	FIS	ST(4)	58.63	1.74	0.0	69.12	1.53	0.0	44.45	3.87	0.0	60.77	2.82	0.0
SRPT	ST(4)	SRPT	ST(4)	58.57	1.77	−0.1	69.14	1.53	0.0	44.40	3.88	−0.1	60.77	2.82	0.0
LRPT	ST(4)	LRPT	ST(4)	58.60	1.75	0.0	69.12	1.52	0.0	44.45	3.88	0.0	60.76	2.81	0.0
VAL	ST(4)	VAL	ST(4)	58.67	1.75	0.1	69.15	1.52	0.0	44.48	3.87	0.1	60.79	2.81	0.0
CYLD	ST(4)	CYLD	ST(4)	58.45	1.75	−0.3	69.04	1.54	−0.1	44.51	3.87	0.1	60.79	2.81	0.0
FRWD	ST(4)	FRWD	ST(4)	58.65	1.73	0.0	69.16	1.53	0.1	44.40	3.89	−0.1	60.78	2.81	0.0
FCFS	FT(40)	FCFS	FT(8)	11.24	0.44	−80.8	13.38	0.29	−80.6	8.11	0.82	−81.8	11.55	0.58	−81.0
FCFS	FN(40)	FCFS	FN(8)	11.49	0.39	−80.4	13.57	0.31	−80.4	8.21	0.86	−81.5	11.80	0.61	−80.6

[†] *Dispatch* refers to the dispatching policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the dispatching policies, refer to Section 4.2.

[‡] *Clean* refers to the cleaning policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the cleaning policies, refer to Section 4.2.

[**] *Profit* refers to the average profit (in dollars) earned for this policy.

[*] *Diff.* refers to the percentage difference in average profit for this policy as compared to the base policy.

Table A6. Detailed results of Fab 2 simulations

Oxi./Dep.		*Etch*		*Exp. 1a*			*Exp. 1b*			*Exp. 2a*			*Exp. 2b*		
Dispatch[†]	*Clean*[‡]	*Dispatch*[†]	*Clean*[‡]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]
Comb./FCFS	Comb.	Comb./FCFS	Comb.	16.70	1.41	42.3	22.80	0.88	14.4	9.26	2.56	73.4	17.51	2.24	23.2
Comb./FRWD	Comb.	Comb./FRWD	Comb.	15.87	0.98	35.2	21.94	0.83	10.1	8.25	2.50	54.5	16.30	1.94	14.7
Comb./VAL	Comb.	Comb./VAL	Comb.	15.58	1.07	32.8	22.07	0.78	10.7	7.59	1.90	42.0	16.23	1.84	14.2
Comb./CYLD	Comb.	Comb./CYLD	Comb.	16.78	1.47	43.0	22.91	0.88	15.0	9.16	2.68	71.6	17.34	2.16	22.0
FCFS	ST(4)	Comb./FCFS	Comb.	16.07	1.67	36.9	22.08	0.98	10.8	8.55	2.47	60.1	16.68	2.12	17.3
Comb.	Comb.	FCFS	ST(4)	12.26	1.34	4.4	20.47	1.22	2.7	5.83	1.74	9.2	14.89	1.84	4.8
FCFS	ST(4)	FCFS	ST(4)	11.74	1.34	–	19.93	1.29	–	5.34	1.64	–	14.21	1.83	–
LCFS	ST(4)	LCFS	ST(4)	11.67	1.34	−0.6	19.93	1.28	0.0	5.30	1.63	−0.8	14.20	1.82	−0.1
FIS	ST(4)	FIS	ST(4)	11.74	1.34	0.0	19.93	1.29	0.0	5.34	1.64	0.0	14.21	1.83	0.0
SRPT	ST(4)	SRPT	ST(4)	11.68	1.34	−0.4	19.92	1.28	0.0	5.30	1.64	−0.8	14.21	1.83	0.0
LRPT	ST(4)	LRPT	ST(4)	11.73	1.34	0.0	19.93	1.29	0.0	5.33	1.64	−0.1	14.21	1.83	0.0
VAL	ST(4)	VAL	ST(4)	11.75	1.34	0.1	19.94	1.29	0.1	5.34	1.64	0.0	14.22	1.82	0.1
CYLD	ST(4)	CYLD	ST(4)	11.69	1.34	−0.4	19.89	1.29	−0.2	5.35	1.64	0.3	14.23	1.83	0.1
FRWD	ST(4)	FRWD	ST(4)	11.82	1.33	0.7	19.96	1.28	0.2	5.30	1.63	−0.7	14.22	1.82	0.0
FCFS	FT(40)	FCFS	FT(8)	−2.00	0.04	−117.0	−1.80	0.04	−109.0	−2.19	0.05	−141.1	−1.96	0.05	−113.8
FCFS	FN(40)	FCFS	FN(8)	−1.95	0.04	−116.6	−1.75	0.03	−108.8	−2.17	0.06	−140.6	−1.91	0.06	−113.4

[†] *Dispatch* refers to the dispatching policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the dispatching policies, refer to Section 4.2.

[‡] *Clean* refers to the cleaning policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the cleaning policies, refer to Section 4.2.

[**] *Profit* refers to the average profit (in dollars) earned for this policy.

[*] *Diff.* refers to the percentage difference in average profit for this policy as compared to the base policy.

Table A7. Detailed results of Fab 3 simulations

Oxi./Dep.		*Etch*		*Exp. 1a*			*Exp. 1b*			*Exp. 2a*			*Exp. 2b*		
Dispatch[†]	*Clean*[‡]	*Dispatch*[†]	*Clean*[‡]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]	*Profit*[**]	±95%	*Diff.*[*]
Comb./FCFS	Comb.	Comb./FCFS	Comb.	4.50	1.03	143.3	8.20	0.47	35.1	1.30	1.62	477.3	7.04	1.97	48.1
Comb./FRWD	Comb.	Comb./FRWD	Comb.	4.95	0.80	167.1	8.33	0.59	37.1	0.65	1.31	288.4	6.10	1.77	28.5
Comb./VAL	Comb.	Comb./VAL	Comb.	4.97	0.91	168.6	8.38	0.57	38.1	0.75	1.46	315.7	6.11	1.77	28.5
Comb./CYLD	Comb.	Comb./CYLD	Comb.	4.56	0.88	146.1	7.98	0.58	31.4	1.19	1.57	443.2	6.83	1.88	43.8
FCFS	ST(4)	Comb./FCFS	Comb.	3.62	0.93	95.5	7.40	0.53	21.9	1.12	1.59	425.0	6.70	1.77	41.1
Comb.	Comb.	FCFS	ST(4)	2.12	0.90	14.2	6.50	0.63	7.1	−0.16	1.04	54.6	5.07	1.67	6.7
FCFS	ST(4)	FCFS	ST(4)	1.85	0.88	–	6.07	0.61	–	−0.35	0.95	–	4.75	1.59	–
LCFS	ST(4)	LCFS	ST(4)	1.83	0.89	−1.3	6.07	0.61	0.0	−0.36	0.94	−5.3	4.74	1.59	−0.2
FIS	ST(4)	FIS	ST(4)	1.85	0.88	0.1	6.07	0.61	0.0	−0.35	0.95	−0.5	4.75	1.59	0.0
SRPT	ST(4)	SRPT	ST(4)	1.82	0.88	−1.6	6.04	0.61	−0.6	−0.36	0.94	−4.1	4.73	1.59	−0.4
LRPT	ST(4)	LRPT	ST(4)	1.85	0.88	0.0	6.07	0.61	0.0	−0.35	0.95	−0.7	4.75	1.59	0.0
VAL	ST(4)	VAL	ST(4)	1.86	0.88	0.4	6.08	0.60	0.2	−0.34	0.95	0.4	4.75	1.59	0.1
CYLD	ST(4)	CYLD	ST(4)	1.84	0.87	−0.5	6.06	0.61	−0.2	−0.34	0.95	1.2	4.76	1.60	0.1
FRWD	ST(4)	FRWD	ST(4)	1.86	0.89	0.6	6.12	0.60	0.8	−0.35	0.94	−1.9	4.76	1.58	0.1
FCFS	FT(40)	FCFS	FT(8)	−2.43	0.01	−231.3	−2.42	0.01	−139.9	−2.44	0.00	−607.0	−2.43	0.01	−151.2
FCFS	FN(40)	FCFS	FN(8)	−2.42	0.01	−230.9	−2.42	0.01	−139.9	−2.44	0.00	−605.1	−2.42	0.01	−151.0

[†] *Dispatch* refers to the dispatching policy used. “Comb.” refers to the combined production and maintenance policies generated by the analytical model. For details on the dispatching policies, refer to Section 4.2.

[‡] *Clean* refers to the cleaning policy used. “Comb.” refers to the combined production and maintenance policies generated by the analytical model. For details on the cleaning policies, refer to Section 4.2.

[**] *Profit* refers to the average profit (in dollars) earned for this policy.

[*] *Diff.* refers to the percentage difference in average profit for this policy as compared to the base policy.

Table A8. Detailed results of Fab 4 simulations

Oxi./Dep.		*Etch*		*Exp. 1a*			*Exp. 1b*			*Exp. 2a*			*Exp. 2b*		
Dispatch[†]	*Clean*[‡]	*Dispatch*[†]	*Clean*[‡]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]	*Profit*[**]	*±95%*	*Diff.*[*]
Comb./FCFS	Comb.	Comb./FCFS	Comb.	7.07	2.32	38.0	15.02	0.95	26.0	−0.55	0.54	62.9	5.42	0.98	83.5
Comb./FRWD	Comb.	Comb./FRWD	Comb.	7.99	2.66	55.9	15.52	0.96	30.2	−0.78	0.42	47.2	4.86	0.91	64.8
Comb./VAL	Comb.	Comb./VAL	Comb.	8.06	2.40	57.2	16.12	0.88	35.2	−0.71	0.47	51.5	5.02	0.89	69.9
Comb./CYLD	Comb.	Comb./CYLD	Comb.	6.73	2.28	31.3	13.98	0.78	17.3	−0.53	0.50	64.3	5.37	0.95	81.9
FCFS	ST(4)	Comb./FCFS	Comb.	6.58	2.14	28.4	14.44	1.02	21.1	−0.73	0.44	50.5	4.94	0.93	67.4
Comb.	Comb.	FCFS	ST(4)	5.40	2.09	5.3	12.26	0.97	2.9	−1.37	0.22	7.1	3.31	0.62	12.1
FCFS	ST(4)	FCFS	ST(4)	5.13	1.92	–	11.92	1.00	–	−1.47	0.20	–	2.95	0.59	–
LCFS	ST(4)	LCFS	ST(4)	5.06	1.93	−1.4	11.90	1.00	−0.1	−1.48	0.20	−0.7	2.94	0.59	−0.3
FIS	ST(4)	FIS	ST(4)	5.13	1.92	0.0	11.92	1.00	0.0	−1.47	0.20	0.0	2.95	0.59	0.0
SRPT	ST(4)	SRPT	ST(4)	5.05	1.92	−1.4	11.86	1.00	−0.5	−1.47	0.20	−0.2	2.94	0.59	−0.6
LRPT	ST(4)	LRPT	ST(4)	5.12	1.91	−0.1	11.92	1.00	0.0	−1.47	0.21	0.3	2.95	0.60	0.0
VAL	ST(4)	VAL	ST(4)	5.14	1.92	0.3	11.93	1.00	0.1	−1.47	0.20	0.1	2.95	0.60	0.0
CYLD	ST(4)	CYLD	ST(4)	5.11	1.91	−0.3	11.90	1.01	−0.2	−1.47	0.20	0.2	2.96	0.60	0.1
FRWD	ST(4)	FRWD	ST(4)	5.12	1.95	−0.2	12.01	1.00	0.7	−1.48	0.20	−0.3	2.95	0.59	−0.1
FCFS	FT(40)	FCFS	FT(8)	−2.42	0.00	−147.3	−2.42	0.00	−120.3	−2.44	0.00	−66.0	−2.43	0.00	−182.4
FCFS	FN(40)	FCFS	FN(8)	−2.42	0.00	−147.0	−2.41	0.01	−120.2	−2.44	0.00	−65.9	−2.43	0.00	−182.2

[†] *Dispatch* refers to the dispatching policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the dispatching policies, refer to Section 4.2.

[‡] *Clean* refers to the cleaning policy used. "Comb." refers to the combined production and maintenance policies generated by the analytical model. For details on the cleaning policies, refer to Section 4.2.

[**] *Profit* refers to the average profit (in dollars) earned for this policy.

[*] *Diff.* refers to the percentage difference in average profit for this policy as compared to the base policy.

is an FCFS dispatching policy with a fixed-state maintenance policy. The Tables A5–A8 on the preceding pages report the percentage improvement for each combination (as compared to the base policy) for each fab and each experiment.

The abbreviation *Comb.* is used to refer to the combined production and maintenance policies generated by the analytical model. When implementing the combined policies, secondary rules are needed when no candidates are available and when more than one candidate is available. The secondary policy used is listed for each of the *Comb.* policies. For example, *Comb./FRWD* indicates that an FRWD secondary policy is used.

The notation ST(i) refers to a fixed-state maintenance policy with threshold i. The threshold is determined using the method described in Section 2.2. The notation FT(T) refers to a fixed-time maintenance policy that calls for repair every T time units, and FN(N) refers to a policy that cleans after N units are produced. The fixed-time and number policies are determined based on the fixed-state threshold. If i is the threshold, then on average it will take $1/p_{\hat{\imath}}$ periods to reach state $\hat{\imath}$, where $p_{\hat{\imath}}$ is the steady-state proportion of time that the machine spends in state $\hat{\imath}$. In each case, we round *down* so that $T = \lfloor 1/p_{\hat{\imath}} - 1 \rfloor$, where $\lfloor x \rfloor$ is the largest integer less than or equal to x. Rounding down means that we are, on average, cleaning earlier than needed. The same procedure is used to determine the value of N for the fixed-number policies.

Acknowledgements

The authors thank Robert Leachman for comments and suggestions on an earlier version of this paper. This work was supported in part by the Alfred P. Sloan Foundation grant for the study on "Competitive Semiconductor Manufacturing."

References

Akella, R., Rajagopalan, S. and Singh, M.R. (1992) Part dispatch in random yield multistage flexible test systems for printed circuit boards. *Operations Research*, **40**, 776–789.

Borden, P.G. and Larson, L.A. (1989) Benefits of real-time, *in situ* particle monitoring in production medium current implantation. *IEEE Transactions on Semiconductor Manufacturing*, **2**, 141–145.

Bratley, P., Fox, B. and Schrage, L. (1987) *A Guide to Simulation*, 2nd edn, Springer-Verlag, New York.

Bunkofske, R.J., Pascoe, N.T., Colt, J.Z. and Smit, M.W. (1996) Real-time process monitoring, in *Proceedings of the 1996 7th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, IEEE Press, Piscataway, NJ, pp. 382–390.

Chitchachornvanich, T. (1998) Optimizing semiconductor fabrication scheduling in the face of uncertainty, Ph.D. thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720.

Cunningham, S.P. (1995) Applications of spatial statistics to semiconductor wafer defects. Ph.D. thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720.

Cunningham, S.P. and MacKinnon, S. (1998) Statistical methods for visual defect metrology. *IEEE Transactions on Semiconductor Manufacturing*, **11**, 48–53.

Edgar, T.F., Cambell, W.J. and Bode, C. (1999) Model-based control in microelectronics manufacturing, in *Proceedings of the IEEE Conference on Decision and Control*, IEEE Press, Piscataway, NJ, pp. 4185–4191.

Glassey, C.R. and Resende, M.G.C. (1988) Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, **1**, 36–46.

Glassey, C.R., Shanthikumar, J.G. and Seshadri, S. (1996) Linear control rules for production control of semiconductor fabs. *IEEE Transactions on Semiconductor Manufacturing*, **9**, 536–549.

Gong, L. and Matsuo, H. (1990) Stabilizing work-in-process and smoothing production in a production system with random yield. Technical report, Graduate School of Business, University of Texas, Austin, TX 78712.

Gong, L. and Matsuo, H. (1997) Control policy for a manufacturing system with random yield and rework. *Journal of Optimization Theory and Applications*, **95**, 149–175.

Hunter, J. and Nguyen, H.K. (1993) *In-situ* particle monitoring reduces wafer defects. *Semiconductor International*, **16**, 80–84.

Law, A.M. and Kelton, W.D. (1991) *Simulation Modeling and Analysis*, 2nd edn, McGraw-Hill, New York.

Lee, F. (1997) Advanced yield enhancement: integrated yield analysis, in *Proceedings of the 1997 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, IEEE Press, Piscataway, NJ, pp. 67–75.

Lee, H.L. and Rosenblatt, M.J. (1987) Simultaneous determination of production cycle and inspection schedules in a production system. *Management Science*, **33**, 1125–1136.

Lee, H.L. and Rosenblatt, M.J. (1989) A production and maintenance planning model with restoration cost dependent on detection delay. *IIE Transactions*, **21**, 368–375.

Lee, H.L. and Yano, C.A. (1988) Production control in multistage systems with variable yield losses. *Operations Research*, **36**, 269–278.

Lee, J.S. and Park, K.S. (1991) Joint determination of production cycle and inspection intervals in a deteriorating production system. *Journal of the Operational Research Society*, **42**, 775–783.

Makis, V. and Fung, J. (1998) An EMQ model with inspections and random machine failures. *Journal of the Operational Research Society*, **49**, 66–76.

Manne, A. (1960) Linear programming and sequential decisions. *Management Science*, **6**, 259–267.

May, G.S. and Spanos, C.J. (1993) Automated malfunction diagnosis of semiconductor fabrication equipment: a plasma etch application. *IEEE Transactions on Semiconductor Manufacturing*, **6**, 28–40.

Nurani, R.K., Akella, R. and Strojwas, A.J. (1996) In-line defect sampling methodology in yield management: an integrated framework. *IEEE Transactions on Semiconductor Manufacturing*, **9**, 506–517.

Nurani, R.K., Strojwas, A.J., Maly, W.P., Ouyang, C., Shindo, W., Akella, R., McIntyre, M.G. and Derrett, J. (1998) In-line yield prediction methodologies using patterned wafer inspection information. *IEEE Transactions on Semiconductor Manufacturing*, **11**, 40–47.

Peters, L. (1992) Twenty good reasons to use *in situ* particle monitors. *Semiconductor International*, **15**, 52–57.

Porteus, E.L. (1986) Optimal lot sizing, process quality improvement and setup cost reduction. *Operations Research*, **34**, 137–144.

Porteus, E.L. (1990) The impact of inspection delay on process and inspection lot sizing. *Management Science*, **36**, 999–1007.

Rosenblatt, M.J. and Lee, H.L. (1986) Economic production cycles with imperfect production processes. *IIE Transactions*, **18**, 48–54.

Segal, J., Gordon, A., Sajoto, D., Duffy, B. and Kumar, M. (1999) Framework for extracting defect density information for yield modeling from in-line defect inspection for real-time prediction of random defect limited yields, in *Proceedings of the IEEE International Symposium on Semiconductor Manufacturing Conference*, IEEE Press, Piscataway, NJ, pp. 403–406.

Shen, Y. (1997) Stochastic wafer fabrication scheduling. Ph.D. thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720.

Sloan, T. and Shanthikumar, J.G. (2000) Combined production and maintenance scheduling for a multiple-product, single-machine production system. *Production and Operations Management*, **9**, 379–399.

Takahashi, K.M. and Daugherty, J.E. (1996) Current capabilities and limitations of *in situ* particle monitors in silicon processing equipment. *Journal of Vacuum Science and Technology, A: Vacuum, Surfaces, and Films*, **14**, 2983–2993.

Tang, C.S. (1992) Controlling inventories in an acyclic assembly system. *Management Science*, **38**, 743–750.

Uzsoy, R., Lee, C.-Y. and Martin-Vega, L.A. (1992) A review of production planning and scheduling models in the semiconductor industry part I: system characteristics, performance evaluation and production planning. *IIE Transactions*, **24**, 47–60.

Wein, L.M. (1988) Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, **1**, 115–130.

Wein, A.S. (1992) Random yield, rework and scrap in a multistage batch manufacturing environment. *Operations Research*, **40**, 551–563.

Wolff, R.W. (1989) *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, NJ.

Yano, C.A. and Lee, H.L. (1995) Lot sizing with random yields: a review. *Operations Research*, **43**, 311–334.

Biographies

Thomas W. Sloan is an Assistant Professor of Management at the University of Miami. He received a Ph.D. and an M.S. from the University of California, Berkeley and a B.B.A. from the University of Texas at Austin. His research interests include stochastic models of manufacturing systems, production scheduling, and manufacturing performance measurement. His previous work has appeared in such journals as *International Journal of Production Research* and *Production and Operations Management*.

J. George Shanthikumar received a B.Sc. degree in Mechanical Engineering from the University of Sri Lanka in 1972 and the M.A.Sc. and Ph.D. degrees in Industrial Engineering from the University of Toronto, Ontario, Canada, in 1977 and 1979, respectively. He is Professor of Industrial Engineering and Operations Research (in the College of Engineering) and Manufacturing and Information Technology (in the Walter A. Haas School of Business) at the University of California, Berkeley. His research interests are in production and service systems modeling and analysis, queueing theory, reliability, scheduling, stochastic processes, simulation, and supply chain management. He has written or written jointly over 250 technical papers on these topics. He is a co-author (with John A. Buzacott) of the book *Stochastic Models of Manufacturing Systems* and a co-author (with Moshe Shaked) of the book *Stochastic Orders and Their Applications*. He is (or was) a member of the editorial boards of the *IIE Transactions on Design and Manufacturing*, *International Journal of Flexible Manufacturing Systems*, *Journal of Discrete Event Dynamic Systems*, *Operations Research*, *Operations Research Letters*, *OPSEARCH*, *Probability in the Engineering and Informational Sciences*, *Production and Operations Management*, and *Queueing Systems*: *Theory and Applications*.