Production Planning & Control,
Vol. 17, No. 7, October 2006, 742–754

Taylor & Francis
Taylor & Francis Group

Improving preventive maintenance scheduling in semiconductor fabrication facilities

A. CRESPO MARQUEZ*†, J. N. D. GUPTA‡ and J. P. IGNIZIO§

†Department of Industrial Management, School of Engineering,
University of Seville, Camino de los Descubrimientos s/n, 41092 Sevilla, Spain
‡Department of Accounting and Information Systems, College of Administrative
Science, The University of Alabama in Huntsville, Huntsville, AL 35899, USA
§Intel Corporation, Fab 11X ME, Mail Stop RR7-332, 4100 Sara Road,
Rio Rancho, NM 87124, USA

The complexity of modern semiconductor manufacturing processes and the need for realistic considerations when modelling their short term availability and reliability make it very difficult to use analytic methods. In this paper, we review the process of preventive maintenance planning and shop level maintenance scheduling in semiconductor fabrication facilities (fabs). We show that the use of Monte Carlo continuous time simulation modelling to improve preventive maintenance scheduling allows the assessment of alternative scheduling policies that could be implemented dynamically on the shop floor. Using a simulation model, we compare and discuss the benefits of different scheduling policies on the status of current manufacturing tools and several operating conditions of the wafers production flow. To do so, we estimate measures of performance by treating simulation results as a series of realistic experiments and using statistical inference to identify reasonable confidence intervals.

Keywords: System simulation; Preventive maintenance scheduling; Semiconductor wafer manufacturing

1. Introduction

In a semiconductor wafer production facility, process tool availability determines factory capacity and serves to drive factory performance in terms of outputs, inventory, cycle time, and WIP velocity (Ignizio 2004). The availability of a production system (EN 13306:2001) is its ability to be in a state to perform a required function under given conditions at a given instant of time or during a given time interval, assuming that the required external resources are provided. In order to ensure a certain level of system availability over time at a specified cost, the best possible preventive maintenance (PM) plans need to be designed, scheduled, and implemented.

A wafer fabrication is a process in which a single crystal semiconductor in the form of a rod or boule is transformed by cutting, grinding, polishing and cleaning into a circular wafer (with the desired diameter) used in semiconductor manufacturing. The process of manufacturing chips typically consists of more than a hundred steps. In the beginning, a thin film of oxide is formed or deposited on the surface of the wafer in a process called oxidation. Then, a process called photolithography is used to transfer the desired pattern onto the surface of the oxidised silicon wafer. During the photolithography process, photo-masks are used by equipment known as steppers to project light through a photo-mask and a high-aperture lens. The intensity of the light casts the pattern on the photo-mask.

*Corresponding author. Email: adolfo.crespo@esi.us.es

Production Planning and Control
ISSN 0953–7287 print/ISSN 1366–5871 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/09537280600901525

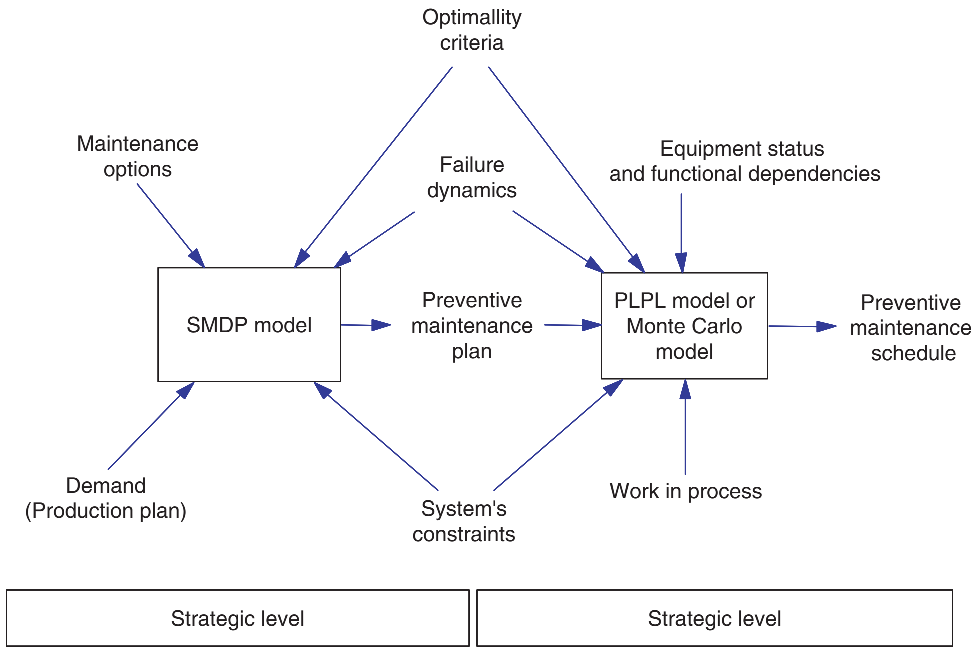


Figure 1. Preventive maintenance scheduling process.

In the subsequent etching process, portions of the oxide surface under the pattern are dissolved away. Finally, in a process called doping, impurities are introduced onto the exposed surface to form device elements such as the source and drain of a transistor.

The design of a preventive maintenance plan needs to take into account such factors as (1) the production plan, (2) the tool's failure dynamics, (3) the operating conditions of the process and (4) different possible maintenance actions and their consequences according to required investments in instrumentation, diagnostic, and repair tools (Crespo Márquez and Sánchez Herguedas 2002). The generic process of preparing a preventive maintenance plan is shown in figure 1. Studies of repairable systems for finite time periods have shown that semi-Markovian decision processes (SMDPs) are useful to solve the preventive maintenance design problem at the strategic level (see, for instance, Gerstbakh 1976, 1977, 2000, Becker 1999, Papazoglou 2000, Abboud 2001). Semi-Markovian models offer a good trade-off between the complexity of the formulation (mathematical format and data requirements) and results provided in terms of detailed production systems behaviour replication. When the preventive maintenance plan design problem is modelled as a SMDP, dynamic programming can help find an optimal solution (Bellman 1957, Howard 1960, Scarf 1997, Gerstbakh 2000, Campbell and Jardine 2001).

Once the optimal preventive maintenance plan is determined, the tactical and process level problem is to schedule short-term preventive maintenance activities. For this problem, we consider the following primary factors:

1. The preventive maintenance plan.
2. The status of the production system (state variables: WIPs).
3. The tool operating condition.
4. The possible functional dependencies among tool and tool components.
5. The system failure dynamics.

For semiconductor wafer facilities, this problem has been approached using mixed integer linear programming (MILP) models (Yao *et al.* 2004). These MILP models have a planning horizon shorter than the time between two successive PM activities of a manufacturing tool.

In using MILP models for preparing PM schedules, it is assumed that the set of maintenance activities to be carried out on a tool is known (maintenance plan). It is also assumed that the start times of each activity are known. This is done by minimising an objective function based on various cost elements. However, in practice, these assumptions rarely hold and hence MILP models do not necessarily lead to workable PM schedules. In semiconductor wafer manufacturing, the production flow variability generally increases the cycle time and the work-in-process inventory levels (Hopp and Spearman, 1996). This lowers the performance of a manufacturing system. Therefore, it is desirable to reduce the variability of the production flows. Such ideas have been recently applied by different researchers when dealing with the problem of finding optimal maintenance schedules (see, for instance, Hopp *et al*. 1989, Simon and Hopp 1995, Van Der Duyn Schouten and Vanneste 1995, Hsu 1999, Liu and Cao 1999).

In this paper we propose to solve the shop floor-level PM scheduling problem using a Monte Carlo (stochastic) simulation (Dekker and Groenendijk 1995). We pay particular attention to the assessment of advanced maintenance scheduling policies such as those involving functional dependencies and work-in-process levels. We also benchmark the results of different maintenance policies considering throughput of the wafer manufacturing process, and not the availability of individual tools, as the variable to optimise. Each scheduling policy is defined by a set of parameter values and we will then limit the space of parameter values (grid simulation) to explore in our work by considering the technical constraints of the problem. Initial grid simulation results are plotted as sample results for a certain random number generators seeds sets. After that, we simulate for different seeds of the random numbers in our experiments in order to see consistency of our initial results. Finally we present the conclusions associated with each experiment.

The rest of the paper is organised as follows: section 2 is dedicated to the continuous time Monte Carlo modelling of a semiconductor manufacturing tool set's preventive maintenance activities and the description of several configuration examples used in the simulation study. In section 3 we present and discuss results of the simulation study. Finally, section 4 concludes the paper with a summary of our findings and directions for future research.

2. Simulation modelling of the PM scheduling problem

Monte Carlo simulation is the generation of certain random and discrete events in a computer model to create a realistic time frame scenario of the system. The simulation is carried out in the computer and estimates are made for the desired measures of performance (Hoyland and Rausand 2004). The simulation is actually a series of realistic experiments where statistical inference is used to estimate confidence intervals for the performance metrics. In general, the events can be simulated either with variable time increments (discrete event simulation; a similar work using discrete event simulation can be found in Charles *et al*. 2003), or with fix time increments, at equidistant points of time (continuous time simulation; see Charles *et al*. 2003).

In this paper, we use the continuous time simulation technique, which evaluates the system's state at the end of every constant time interval (Δt), records the new system's state and collects the statistics of interest. Then the time is incremented another Δt, and so on. As a simulation tool, we use the VENSIM simulation environment (Ventana Systems®), which has special features to assist in easy Monte Carlo type simulation experiments, and to provide confidence interval estimations. This method allows us to consider various relevant aspects of systems operation (such as K-out-of-N, redundancies, functional dependencies or component repair priorities) which can hardly be captured by analytical models (Marseguerra and Zio 2000, 2002).

To describe a generic continuous time stochastic model for a tool's maintenance the following list of notations is used. However, this variable list could be later subscripted according to the number of tools in a tool-set.

Tool status information related variables

A_t	Tool availability (1 available, 0 unavailable) at t.
AA_t	All tools available (1 yes, 0 no) at t.
CA_t	Decrease in tool's age due to corrective maintenance action in t.
IFS_t	In-front stock status in t.
LC_t	Time when the last corrective maintenance, for a tool in t, started.
LP_t	Time when the last preventive maintenance, for a tool in t, started.
$\lambda(T_t)$	Failure rate of the system in t.
MB_t	Maintenance backlogged (1 yes, 0 no) at t.
PA_t	Decrease in tool's age due to preventive maintenance action in t.
$PTB_{t,i}$	Time that a PM action is being in backlog in t.
RN_t	Random number within the interval $(0, 1)$, generated in t.
RM_t	Maintenance released (1 yes, 0 no) in period t.
SM_t	Sheduled maintenance (1 yes, 0 no) in period t.
T_t	Tool's age in t.
TI_t	Increase of system's age in period t.
TO_t	Decrease of system's age in period t.
$TBI_{t,i}$	Time increase (1 yes, 0 no) of a PM action in backlog in period t.
$TBD_{t,i}$	Time decrease (1 yes, 0 no) of a PM action in backlog in period t.

Model parameters

CT	Average time of a corrective maintenance action.
MxW_i	Maximum desired in-front stock level to release maintenance.

- n Minimum age of the tool to do preventive maintenance actions.
- N Maximum age of the tool to do preventive maintenance actions.
- PT Average time of a preventive maintenance action.

2.1 *Modelling tool's age*

We use the developments in Crespo *et al.* (2003) to describe the age of the system as follows:

The process requires first to model the age of the system (T_t):

$$T_t = T_{t-1} + TI_t - TO_t \tag{1}$$

We assume that age will increase when the tool is available, i.e. we assume that available means 'running', neither idling nor standing-by, therefore:

$$TI_t = A_t \tag{2}$$

and age will decrease when the system is maintained:

$$TO_t = \begin{cases} PA_t, & \text{if } PA_t <> 0 < \text{ and } CA_t <> 0 \\ PA_t + CA_t, & \text{otherwise} \end{cases} \tag{3}$$

$$CA_t = \begin{cases} T_t, & \text{if } \lambda(T_t) \geq RN_t \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where RN_t is a random number generated for every t within the range $(0, 1)$, $\lambda(T_t)$ is the failure rate of the system, and CA_t and PA_t are decreases in the system's age as a consequence of the corrective and preventive maintenance actions respectively. Notice how according to equation (4) CA_t is not equal to zero at the time a failure occurs. Also according to equation (1) notice that we assume that after a failure or after a preventive maintenance, the age of the system is set to zero.

2.2 *Modelling tool availability*

The conditions of a tool that make it unavailable will be the corrective or preventive maintenance that is being carried out. In order to model availability, we therefore need to keep track of the moment in time when the last failure or preventive maintenance took place, and the time that those activities lasted. Then, following Crespo *et al.* (2003), we can model availability as in (5) below:

$$A_t = \begin{cases} 1 - (\text{Pulse}(LC_t, CT, t) + \text{Pulse}(LP_t, PT, t)), & \text{if } LC_t > 0 \text{ or } LP_t > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Notice that when $t = 0$, $LC_t = LP_t = 0$ (LC_t and LP_t are the times when the last corrective preventive maintenance started for a tool in t, respectively). The function Pulse is defined as follows:

$$\text{Pulse }(a, b, t) = \begin{cases} 1, & a < t < a + b \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

For instance, imagine that we have a tool-set with two tools ($i = 1, 2$), and both of them need to be in operating conditions in order to preventively maintain one of them, then the condition to fulfil would be $AA_t = 1$, where AA_t is defined as follows:

$$AA_t = \prod_{i=1}^{i=2} A_{t,i} \quad \text{with} = 1, 2. \tag{7}$$

2.3 *Modelling maintenance activities backlog*

Although a tool's age or a wafer's yield may indicate starting a preventive maintenance action, it may be beneficial to leave the tool functioning while another tool is down (under corrective or preventive maintenance). Therefore, it is necessary to model the possible backlog of maintenance activities, i.e. activities which are due and waiting to be carried out by the maintenance department. This concept of maintenance backlog will be very practical when considering a feasible time range for PM actions to be carried out on a tool. We model the backlogged activities and the time they are backlogged as follows:

$$SM_{t,i} = \begin{cases} 1, & \frac{t_i}{n} = \text{Int}\left(\frac{t_i}{n}\right) \quad \text{and} \quad t_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$$MB_{t,i} = MB_{t-1,i} + SM_{t,i} - RM_{t,i} \tag{9}$$

In our model, maintenance actions will be scheduled according to (8) when $T_t \geq n$. The backlog of maintenance activities is formalised as in (9), where $RM_{t,i}$ represents the PM activity released (its value is 1 when the activity is released). A PM activity will only be released when a certain condition which characterises the preventive maintenance policy (defined later in this section) is met. The time a PM scheduled action is backlogged is being modelled to use it as a control variable and also to use it later as a performance measure. It is defined as follows:

$$PTB_{t,i} = PTB_{t-1,i} + TBI_{t,i} - TBD_{t,i} \tag{10}$$

$$TBI_{t,i} = \begin{cases} 1, & RM_{t,i} = 0 \quad \text{and} \quad MB_{t,I} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$TBD_{t,i} = PTB_{t,i}\ RM_{t,i} \tag{12}$$

where $TBI_{t,i}$ and $TBD_{t,i}$ are increases and decreases in the time for which PM activity is backlogged, respectively. Equation (11) expresses how backlog time increases when the activity is not released while equation (12) formalises decreases in time when PM is released. Therefore, this will later set the backlogged time to 0 in equation (10).

2.4 *Modelling preventive maintenance policies*

In this section, we model the way PM actions are released. The following policy options are considered: age based maintenance, age and availability based maintenance and age and in-front buffer maintenance. When a preventive activity is released, we will record this time (in LP_t) to allow downtime modelling as explained previously in equation (5). Notice that, in this example, we constrain the time that a backlogged activity can be backlogged so it will be released before a new preventive maintenance is scheduled. Also, regardless of the PM policy, (13) and (1) will later set up tool's age to zero.

$$PA_{t,i} = T_t RM_{t,i} \tag{13}$$

2.4.1 Age based maintenance policy. In the age based maintenance policy, we assume that the tool is preventively maintained when it reaches a certain number of periods of time N without a failure. Otherwise, it is correctively maintained at the failure time (see equation (14)).

$$RM_{t,i} = \begin{cases} 1, & \text{if } T_t \geq N \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

2.4.2 Age and availability based maintenance policy. In this policy, PM activity is released when both tools are available and $N \geq T_t \geq n$. In case maintenance is overdue, the activity is released. The formulation of this policy is in equation (15):

$$RM_{t,i} = \begin{cases} PTB_{t,i} < N - n & \begin{cases} 1, & (SM_{t,i} = 1 \text{ or } MB_{t,i} = 1) \\ & \text{and } AA_t = 1 \\ 0, & \text{otherwise} \end{cases} \\ PTB_{t,i} \geq N - n, & 1 \end{cases} \tag{15}$$

Notice that when trying to release a backlogged activity, the first thing we do is check whether the time at which the PM scheduled action has been backlogged exceeds the time limits. In case we do not exceed the time limits, the maintenance is released only if both tools are OK (i.e. $AA_t = 1$). In order to facilitate understanding of this PM policy, figure 2 depicts two simple cases of the scheduling and release of PM activities in tool no. 1, located within a tool-set with two tools (T1 and T2), when tool no. 2 is being maintained. The circles in figure 2 denote the times which triggers backlog, scheduling and release of the PM activity.

2.4.3 Age and in-front buffer maintenance policy. Figure 3 shows some examples and explanation of cases when PM activities are released based on the tool's age and in-front inventory status (for two tools in parallel in the same tool-set processing the same operation).

As we mentioned earlier, maintenance policies to explore in this paper are based not only on current manufacturing tools status but also on several operating conditions of the wafers production flow. One of these conditions is the WIP status. It seems reasonable that low in-front stock status could be a desirable condition to release maintenance activities since queuing phenomena could be reduced during tools preventive downtime.

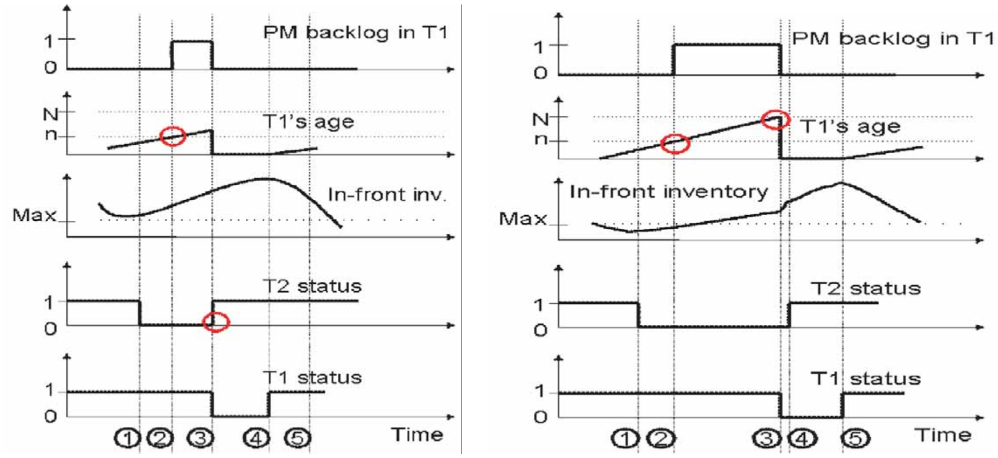
$$RM_{t,i} = \begin{cases} PTB_{t,i} < N - n & \begin{cases} 1, & (SM_{t,i} = 1 \text{ or } MB_{t,i} = 1) \\ & \text{and } IFS_t \leq MxW_i \\ 0, & \text{otherwise} \end{cases} \\ PTB_{t,i} \geq N - n, & 1 \end{cases} \tag{16}$$

Equations (15) and (16) determine when the RM (release maintenance) variable has a value of 0 or 1. For instance, in equation (16), notice that maintenance is released ($RM = 1$) when: (1) The tool's age is within the maintenance time window ($n < PTB_t < N$), there is a maintenance already backlogged or just scheduled for this time period, and all tools are available; (2) we cannot backlog a maintenance action any longer (because we run out the feasible time window to backlog maintenance ($N - n$)).

Notice that these equations reproduce possible preventive maintenance scheduling policies on the shop floor.

2.5 *Specific wafer production flow scenarios*

In order to apply the above maintenance scheduling concepts, the following two simple production flow scenarios are used in our simulation experiments. The production flow scenario 1 (in figure 4) consists of two tools sets ($TSet_1$ and $TSet_2$) each one with only one tool ($Tool_1Tset_1$ and $Tool_2Tset_2$). The production flow scenario 2 (in figure 5) consists of two tools sets



Case 1, sequence of events (see figure 1 - left) :

(1) T2 fails
(2) PM action is backlogged because T1's age reaches n periods but T2 is down.
(3) PM action is released as soon as T2 is up after maintenance and n≤T1's age ≤N.
(4) T1 is up again after the PM, both tools are working and in-front inventory decreases. Notice how from (1) to (4) in-front inventory will tend to increase since only one tool is working.

Case 2, sequence of events (see figure 1 - right) :

(1) T2 fails, inventory will tend to increase.
(2) PM action is backlogged, because T1's age reaches n periods and T2 is down. T1's age reached N periods and Pm action is released even with T2 down.
(3) T1 is up again after maintenance, T2 is still down.
(4) T2 is up again. Notice how from (3) to (4) the inventory would tend to increase at a higher rate than from (1) to (3) once both tools are down.
(5) T1 is up again after its PM.
Inventory would decrease after this event once both tools would be working.

Figure 2. Cases for age and availability based PM policy. Cases 1 (left) and 2 (right).

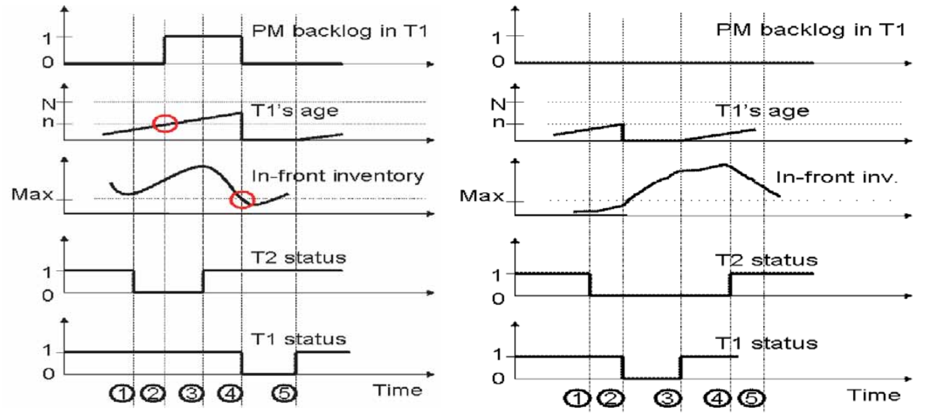
(TSet1 and TSet2), the first one containing two tools (Tool1Tset1 and Tool2Tset1) and the second one with only one tool (Tool3Tset2). TSet1 performs operation 1 and TSet2 operation 2.

In figures 4 and 5, the variable PR T*i* TS*j* denotes the production rate of Tool *i* in TSet *j*. The variable T*i* TS*j* TcomR denotes the production completion rate of Tool *i* in TSet *j*. In-front stockers (TSet *i* in-front Stocker) will be this time capacity constrained (to 200 wafers, which is a short number in real fabs but adjust reasonable top accumulations for the wafer flows we are considering), and reaching this stockers constraint will provisionally stop the inflow of wafers to those stockers. Tools production capacity will be also constrained to the tools processing speed, which will be according to table 1.

In all scenarios, arrival rate of wafers (AR_t in figures 6 and 7) is assumed to be exponential ($\lambda = 0.428$ wf/min) and failure rate of all tools in all cases is assumed to be uniformly distributed ($\mu = 0.000145$ f/min) within the interval of 'time since last maintenance' (age T_t) for which the tool is in operation. Even when preventive maintenance is momentarily delayed, we assume that the same failure rate is applicable. (A selection of parameters values was carried out according to real maintenance and repair database records obtained from semiconductor fabs. Initial conditions and simulation horizon were selected in order to test the different policies ensuring that enough number of maintenance actions would be scheduled and released for each simulation.)

For each machine, we will assume that preventive maintenance actions are planned to take place every 1440 minutes. Age intervals to explore alternative maintenance scheduling policies will go from $n = 1440$ to $N = (n + 1400) = 2880$ min (i.e. there is a range of 1400 min for a possible PM delay in order to meet more favourable system operational conditions to carry out the PM). With this condition, no more than one maintenance action can be backlogged at any time.

Time required to accomplish a PM is assumed to be 200 minutes while the unscheduled maintenance (UM) will require 800 minutes. This will be for all tools and scenarios to simulate. Some readers may argue that this time should also be random; we, however, assume here



Case 1, sequence of events (see figure 1 - left) :

(1) T2 fails
(2) PM action is backlogged because T1's age reaches n periods but in-front inventory is over maximum value.
(3) T2 is up after maintenance, n≤T1's age ≤N, but in-front inventory is still over the maximum, therefore PM maintenance action keeps backlogged. Notice how from (1) to (3) in-front inventory will tend to increase since there is always just one tool working.
(4) T2 is up, n≤T1's age ≤N and in-front inventory is now under the maximum, therefore PM maintenance action is released in T1. Notice how from (3) to (4) the inventory will tend to decrease since both tools are working. After (4) inventory would again increase once only T2 would be working.

Case 2, sequence of events (see figure 1 - right) :

(1) T2 fails, inventory will tend to increase.
(2) PM action is released, not backlogged, because T1's age reaches n periods and in-front inventory is below the maximum value. Both tools are down and inventory will increase at a higher rate.
(3) T1 is up again after maintenance, T2 is still down.
(4) T2 is up and both tools are again working. Notice how from (3) to (4) the inventory would tend to increase at a similar rate than from (1) to (2). After (4) inventory would decrease once both tools would be working.

Figure 3. Cases for age and in-front buffer based PM Policy. Cases 1 (left) and 2 (right).

that maintainability programmes have taken them to variability ranges where this assumption is reasonable.

In all cases, initial conditions of the system are assumed to be as follows:

- Initial number of wafers in TSet 1 in-front stocker = 5 wafers.
- Initial number of wafers in TSet 2 in-front stocker = 5 wafers.
- Initial operational condition for all tools = 'idling'.
- Initial 'time since last maintenance' (age T_0) of the tools 0, 50, and 100 minutes for tool 1, tool 2, and tool 3 respectively.

In our simulations, as we argued in the introduction of the paper, the overall objective against which we will compare maintenance scheduling policies will be to maximise the throughput of the wafer manufacturing process (total output) and not the availability of individual tools.

Finally, the simulation horizon will be a total of 40 000 minutes in all cases. Age based and age and in-front buffer maintenance scheduling policies are tested in scenario 1, while in scenario 2 we also test the age and availability policy for tools in $TSet_1$.

3. Simulation results

In this section, we discuss the results obtained from our simulation experiments for each scenario. We also summarise the managerial implications of these results.

3.1 *Results for scenario 1*

Sample results considering the same set of seeds for the pseudo-random number generation in the model are presented in Figure 6. These results deal with the arrival of wafers and the failure appearance stochastic mechanism. A total of 400 simulation runs were made for maintenance policy classes, grids that could be characterised as Class Max1, Max2. Such a maintenance scheduling class would release maintenance of Tool i when, within their planned age window [1440, 2840], the in-front stocker i is below Max i value. For this example,

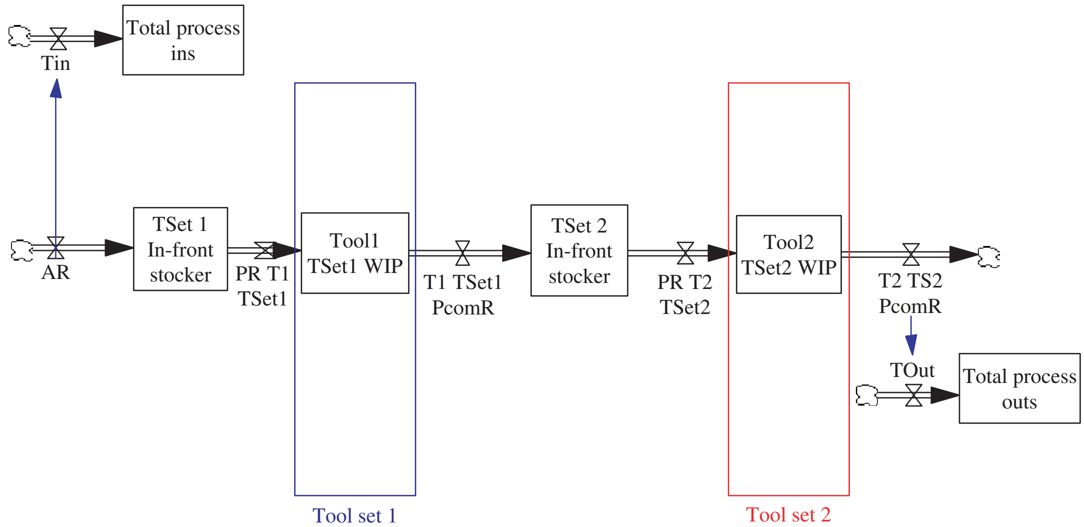


Figure 4. Production flow configuration for scenario 1.

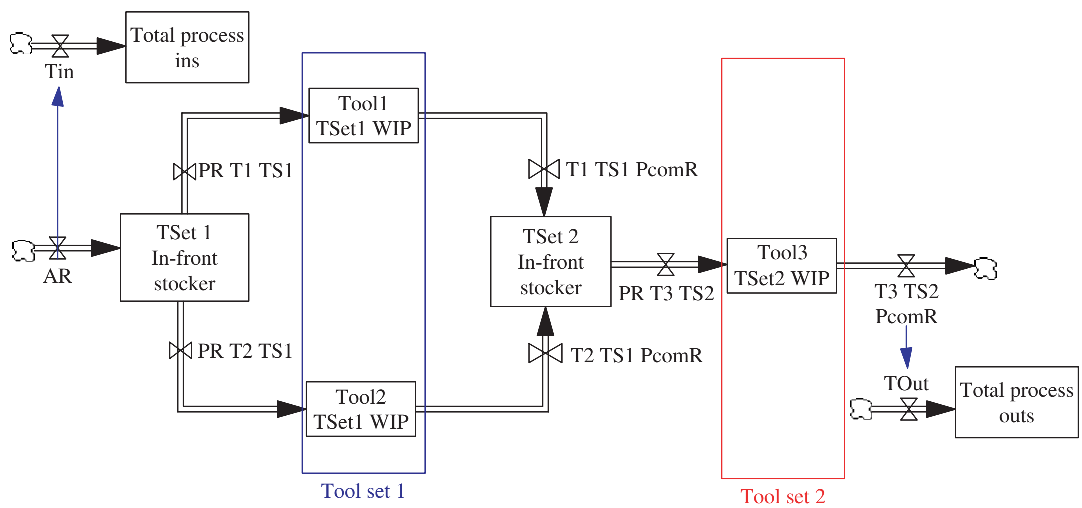


Figure 5. Production flow configuration for scenario 2.

Table 1. Tool processing time for the scenarios (in minutes).

	Scenario 1	Scenario 2
$Tool_1Tset_1$	2	4
$Tool_2Tset_1$	–	4
$Tool_2Tset_2$	2	–
$Tool_3Tset_2$	–	2

there are clear areas of higher performance of the wafer production process (more than 3% output improvement for the optimal policy). At the same time, and for a certain area, simulation results for Class Max1, Max2 policies are better than those obtained for age based maintenance policy (this is 14 885 wafers), assuming age for PM to be 1440.

Initially, results are a bit counter-intuitive. For the cases where the tool failure rate does not increase, we would expect to find policies delaying maintenance as much as possible (Class $0, 0$) to produce better performance (less down time due to less number of PM along the simulation horizon). However, this is not happening. Class $77, 171$ was found to be the best of the high performance classes. Although this is a sample result for a certain set of seeds values, we could simulate other cases (seed sets) and obtain similar results.

Figure 7 presents the results corresponding to another 400 simulations considering the same set of

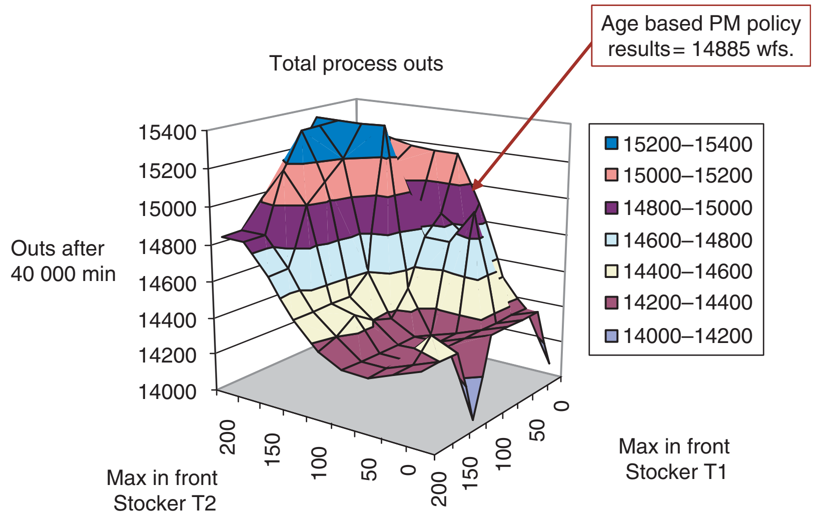


Figure 6. Comparison of age based versus age and in-front buffer based scheduling policy results for different levels of maximum inventory (Max in figure 3) in both stockers, scenario 1.

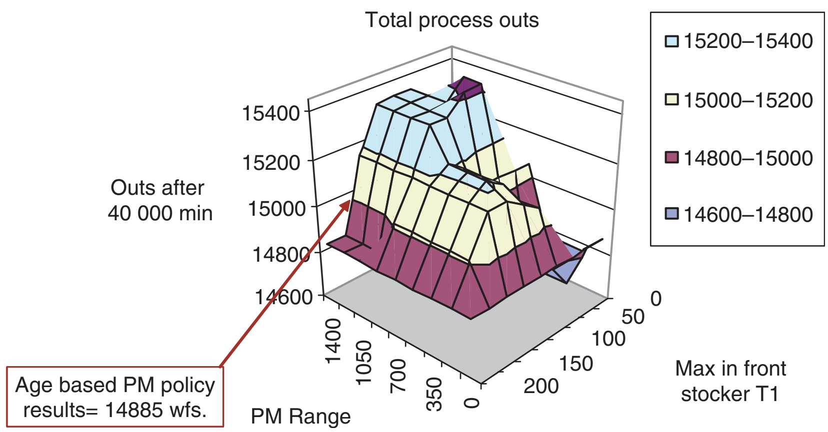


Figure 7. Comparison of age based versus age and in-front buffer based scheduling policy results for different levels of maximum inventory of stocker 1 and range for the PM $(N-n)$ in equation (17), scenario 1.

seeds for the pseudo-random number as in figure 6. We fixed Max $2=170$ wafers, since this value offered good performance in previous graph for a wide range of Max 1 values. We now wish to explore whether by shortening PM range values, we could maintain the system performance. Surprisingly, in this example, selecting values of Max 1 within the interval [75, 150] results for total output are reasonably good for values of PM range $(N-n)$, even closer to 300 minutes. This result supports the idea that too much relaxation of the PM interval does not yield good payback.

3.2 *Results for scenario 2*

For scenario 2, figure 8 shows results for Class Max1, Max2 maintenance policies. We have maintained the same seed of random numbers generation for simulations as in previous figures. Now, policy Class 124, 138 produced the best results, although age and in-front buffer policy offers some lower relative improvements than in previous scenario (around 2% potential improvement for this example). For this scenario, comments regarding the optimal policy resulting in this example are similar to those for scenario 1 in figure 6.

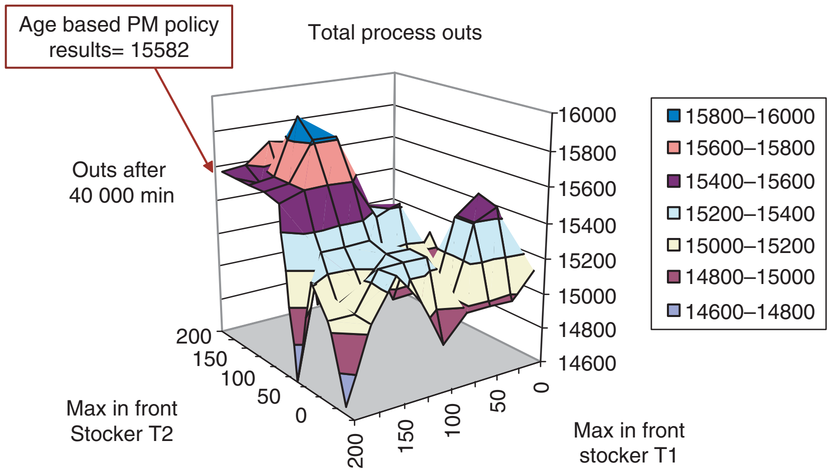


Figure 8. Comparison of age based versus age and in-front buffer based scheduling policy results for different levels of maximum inventory (Max in figure 3) in both stockers, scenario 2.

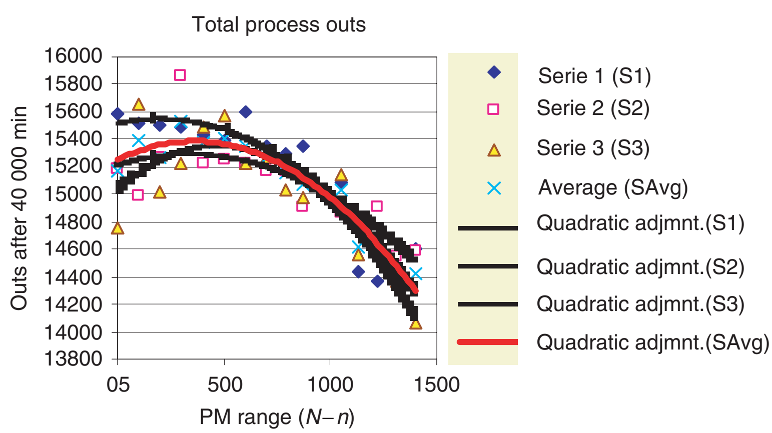


Figure 9. Comparison of age and tool availability scheduling policy results for different values of the range for the PM $(N-n)$ in equation (16), scenario 2.

For this second scenario, let us now consider the age and availability based scheduling maintenance policy. As we mentioned in section 3.4.2, this policy depends on the range $(N-n)$ that allows for maintenance to be delayed in case one tool is down when releasing the scheduled maintenance of a parallel tool. We refer of course to tools within the same tool-set. In the these simulations, we explore the importance of the PM delay range for improving total process output when using this class of policy (Class $N-n$, with $n=1440$ fixed).

In figure 9, we present results for three seed sets plus the average of the three. Notice that for $N-n=0$, we would be talking about the simple age based policy with $n=1440$. These results show that increasing the range of maintenance delay improves performance (output) until the range is around 1000 minutes. For this example, the average performance would be around 1.3% for the optimal policy (Class 500).

3.3 *Confidence in simulation results*

So far we have presented results for a given set of random numbers for the stochastic mechanism of arrival times and failure rates. Let us now try to build confidence in the results obtained by using some statistical inference techniques applied to results in areas showing important potential improvements. To obtain acceptable insights from these tests, we will constrain this exercise to the number of simulations required. Notice that the stochastic mechanism of the mentioned variables is functioning every single minute of the 40 000 minutes of the planning horizon. This means that we initially do not expect big changes by selecting different seeds and adding replications. Of course, results depend on many other factors and we want to be sure about subsequent conclusions we may arrive in this paper (see comments about this aspect in Andijani and Duffuaa 2002).

Table 2. Results obtained for the selected scenario class policies and for 10 replications of the experiments with different seeds (in minutes).

Seed	Process outs		Process outs		Process outs	
	ABP	AIFSBP Class 77, 170	ABP	AIFSBP Class 124, 138	ABP	AIFSBP Class $N-n=500$
	Scenario 1		Scenario 2		Scenario 2	
1	14802	14796	15169	14435	15169	15356
2	15189	15182	15018	14737	14919	15019
3	14885	15365	15582	15718	15582	15466
4	15170	15564	15043	14564	14040	15555
5	14531	14125	14898	15379	14749	15565
6	14567	14452	15370	14782	15372	15423
7	14988	14867	14810	14852	14810	15090
8	14687	14471	14968	14614	15180	15314
9	15471	14264	15239	15115	14760	14907
10	14657	14865	15239	15115	15241	15186

ABP, age based policy; AIFSBP, age and in-front Stocker based policy.

Table 3. Confidence intervals for policy results in table 2 (in minutes).

	ABP		AIFSBP		95% confidence interval			
	μ	s	μ	s	Min	Max	Min	Max
Scenario 1 Class 77, 170	14894.7	307.5	14795.1	474.9	14704.1	15085.3	14500.8	15089.4
Scenario 2 Class 124, 138	15133.6	233.2	14931.1	399.3	14989.0	15278.2	14683.6	15178.6
Scenario 2 Class $N-n=500$	14982.2	431.8	15288.1	228.6	14714.6	15249.8	15146.4	15429.8

ABP, age based policy; AIFSBP, age and in-front Stocker based policy.

Results in table 2 for the policies selected are not what we thought they would be. The age and in-front Stocker based policy (AIFSBP) does not perform as expected in both scenarios (1 and 2). We can confirm this impression by looking at table 3, where we can appreciate that age based policy (ABP) performs slightly better than AIFSBP in terms of the mean value for output and expected variability of the output.

Results for the age and availability based policy are positive. We can appreciate an average 2% increase in throughput plus closer bounds for the 95% confidence interval, which also may imply indirectly lower variability wafer output when applying the selected class of this policy. This seems to be a promising result, and prompts us to further explore this type of maintenance scheduling policy.

4. Conclusions

In this paper, we explored the opportunity to use Monte Carlo continuous time simulation modelling to improve preventive maintenance scheduling in semiconductor fabs. Using this technique, we showed that we can produce a reasonable assessment of alternative scheduling policies that could be implemented dynamically on the shop floor. Policies considered were based on the status of current manufacturing tools, tools in-front stocker status and tools availability status. We compared and discussed the relative benefits of different scheduling policies in terms of the number of process outputs produced. In order to do, so we treated simulation results as a series of realistic experiments and used statistical inference to reach reasonable confidence intervals of performance parameters.

Based on the simulation results obtained in this paper, we conclude that the use of age and availability based maintenance scheduling policy (AABP) maximises process throughput and provides better results than those obtained by simple age based maintenance scheduling policy (ABP) for the scenario 2. Results for both scenarios in the paper also show that scheduling policies based on age and in-front stocker level (AIFSBP) may result in higher variability of the flow while producing no real performance improvement in these particular environments. Nevertheless, setting up a maintenance scheduling policy based on age and in-front buffer levels may not be easy and requires extensive computational efforts that may render it less practical for day to day operations. Future research needs to investigate the implications of the use of AABP for more specific tool set configurations in specific scenarios. Also future research to ascertain the process and steps to follow in articulating a final policy to be given to managers on the shop floor would be beneficial and interesting.

Acknowledgements

We would like to thank Jonathan Mathews from Intel for his interest and data gathering work for this project and to two anonymous reviewers whose critical and constructive comments improved the presentation in this paper. This research has been funded by the Spanish Ministry of Science and Education, Project DPI 2004-01843, in addition to FEDER funds.

References

Abboud, N.E., A discrete-time Markov production-inventory model with machine breakdowns. *Comp. Indust. Eng.*, 2001, **39**, 95–107.

Andijani, A. and Duffuaa, S., Critical evaluation of simulation studies in maintenance systems. *Prod. Plan. & Cont.*, 2002, **13**(4), 336–341.

Becker, G., Camarinopoulos, L. and Zioutas, G., A semi-Markovian model allowing for inhomogeneities with respect to process time. *Reliab. Eng. Syst. Safety*, 1999, **70**(1), 41–48.

Bellman, R., *Dynamic Programming*, 1957 (Princeton University Press: Princeton, NJ).

Campbell, J.D. and Jardine, A.K.S., *Maintenance Excellence: Optimising Equipment Life-cycle Decisions*, 2001 (Marcel Dekker: New York, NY).

Charles, A.S., Floru, I.R., Azzaro-Pantel, C., Pibouleau, L. and Doménech, S., Optimisation of preventive maintenance strategies in a multi purpose batch plant: Application to semiconductor manufacturing. *Comp. Chem. Eng.*, 2003, **27**, 449–467.

Crespo Márquez, A. and Sánchez Herguedas, A., Models for maintenance optimisation: A study for repairable systems and finite time periods. *Reliab. Eng. Syst. Safety*, 2002, **75**(3), 367–377.

Crespo Marquez, A., Gupta, J.N.D. and Sánchez Herguedas, A., Maintenance policies for a production system with constrained production rate and buffer capacity. *Int. J. Prod. Res.*, 2003, **41**(9), 1909–1926.

Dekker, R. and Groenendijk, W., Availability assessment methods and their application in practice. *Microelectron. Reliab.*, 1995, **35**(9–10), 1257–1274.

EN 13306:2001, *European Standard on Maintenance Terminology*, 2001 (CEN: Brussels).

Gerstbakh, I.B., Sufficient optimality conditions for control-limit policy in a semi-Markov model. *J. Appl. Prob.*, 1976, **13**, 400–406.

Gerstbakh, I.B., *Models of Preventive Maintenance*, 1977 (North-Holland: New York).

Gerstbakh, I.B., *Reliability Theory. With Applications to Preventive Maintenance*, 2000 (Springer-Verlag: Berlin).

Hopp, W.J., Pati, N. and Jones, P.C., Optimal inventory control in a production flow system with failures. *Int. J. Prod. Res.*, 1989, **27**, 1367–1384.

Hopp, W.J. and Spearman, M.L., *Factory physics. Foundations of Manufacturing Management*, 1996 (Irwin: Chicago, IL).

Howard, R.A., *Markov Processes and Dynamic Programming*, 1960 (Technology Press and Wiley Press: New York, NY).

Hsu, L.F., Simultaneous determination of preventive maintenance and replacement policies in a queue-like production system with minimal repair. *Reliab. Eng. Syst. Safety*, 1999, **63**(2), 161–167.

Hoyland, A. and Rausand, M., *System Reliability Theory. Models, Statistical Methods and Applications* (Wiley Series in Probability and Statistics), 2004 (John Wiley & Sons Inc.: Hoboken, NJ).

Ignizio, J.P., Optimal maintenance headcount allocation: An application of Chebyshev goal programming. *Int. J. Prod. Res.*, 2004, **42**(1), 201–210.

Liu, B. and Cao, J., Analysis of a production-inventory system with machine breakdowns and shutdowns. *Comp. Oper. Res.*, 1999, **26**, 73–91.

Marseguerra, M. and Zio, E., Optimising maintenance and repair policies via combination of genetic algorithms and Monte Carlo Simulation. *Reliab. Eng. Syst. Safety*, 2000, **68**, 69–83.

Marseguerra, M. and Zio, E., *Basics of the Monte Carlo Method with Applications to System reliability*, 2002 (LiLoLe-Verlag GmbH: Hagen, Germany).

Papazoglou, I.A., Semi-Markovian reliability models for systems with testable components and general test/outage times. *Reliab. Eng. Syst. Safety*, 2000, **47**, 175–185.

Pidd, M., *Tools for Thinking. Modelling in Management Science*, 2nd ed., 2003 (Wiley: ChichesterEngland).

Scarf, P.A., On the application of mathematical models in maintenance. *Euro. J. Oper. Res.*, 1997, **99**, 493–506.

Simon, J.T. and Hopp, W.J., Throughput and average inventory in discrete balanced assembly systems. *IIE Trans.*, 1995, **267**, 368–373.

Van Der Duyn Schouten, F.A. and Vanneste, S.G., Maintenance optimisation of a production system with buffer capacity. *Euro. J. Oper. Res.*, 1995, **82**, 323–338.

VENSIM, Version 5.4, 2004 (Ventana Systems Inc.: Harvard, MA).

Yao, X., Fernandez-Gaucherand, E., Fu, M.C. and Marcus, S.I., Optimal preventive maintenance scheduling in semiconductor manufacturing. *IEEE Trans. on Semicon. Manuf.*, 2004, **17**(3), 345–356.

Adolfo Crespo Marquez is currently Associate Professor at the School of Engineering of the University of Seville, in the Department of Industrial Management. He holds a PhD in Industrial Engineering from this same University. His works have been published, or are about to be published, in journals such as the *International Journal of Production Research*, *International Journal of Production Economics*, *European Journal of Operations Research*, *Journal of Purchasing and Supply Management*, *International Journal of Agile Manufacturing*, *Omega*, *Journal of Quality in Maintenance Engineering*, *Decision Support Systems*, *Computers in Industry and Reliability Engineering and System Safety*, among others. Within the area of maintenance he has extensively participated in many engineering and consulting projects for different companies and for the Spanish Ministries of Defense and Science and Education. He has been the President of INGEMAN (a National Association for the Development of Maintenance Engineering in Spain) for 3 years.

Dr James Ignizio received his PhD in Industrial Engineering and Operations Research from Virginia Tech in 1971. He is a Fellow of the Institute of Industrial Engineering, the British Operational Research Society and the World Academy of Productivity Science. Dr Ignizio is the author of eight books and more than 350 publications, including over 150 refereed papers in international professional journals. In 1980 he was awarded The First Hartford Prize by the US National Safety Council for his contributions to USA's manned moon-landing program. In 2002 Dr Ignizio was inducted into the Academy of Distinguished Alumni of Virginia Tech. He has also served as a consultant to numerous industrial and governmental organisations in both the United States and abroad, including the US Army, US Air Force, US Navy, NASA, SDI ('Star Wars'), Litton, Exxon, Texaco, Boeing, SAI, GRC, SRI, Bell Labs, Finland's Ministry of Economics, Quantas Airways, KAIST, Virginia Manufacturing Institute, Chase Bank, and the Commercial Bank of Greece.

Jatinder N.D. Gupta is currently Eminent Scholar of Management Information of Technology, Professor of Management Information Systems, Industrial and Systems Engineering and Engineering Management, and former Chairperson of the Department of Accounting and Information Systems at the University of Alabama in Huntsville. Most recently, he was Professor of Management, Information and Communication Sciences, and Industry and Technology at Ball State University, Muncie, Indiana. He holds a PhD in Industrial Engineering (with specialisation in Production Management and Information Systems) from Texas Tech University. Dr Gupta is a co-author of a textbook on operations research and serves on the editorial boards of several national and international journals. He is the recipient of the Outstanding Faculty and Outstanding Research awards from Ball State University, he has published numerous papers in such journals as *Journal of Management Information Systems*, *International Journal of Information Management*, *Operations Research*, *INFORMS Journal of Computing*, *Annals of Operations Research*, and *Mathematics of Operations Research*. More recently, he served as a co-editor of several special issues including the Neural Networks in Business of Computers and Operations Research and books that included *Decision Making Support Systems: Achievements and Challenges for the New Decade* published by Idea Group Publishing. His current research interests include e-commerce, supply-chain management, information and decision technologies, scheduling, planning and control, organisational learning and effectiveness, systems education, knowledge management, and enterprise integration. Dr Gupta has held elected and appointed positions in several academic and professional societies including the Association for Information Systems, Production and Operations Management Society (POMS), the Decision Sciences Institute (DSI), and the Information Resources Management Association (IRMA).