Iie Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie20

Analysis of maintenance policies for *M* machines with deteriorating performance

EMRE BERK & KAMRAN MOINZADEH

IIE Transactions (2000) **32**, 433–444

Analysis of maintenance policies for M machines with deteriorating performance

EMRE BERK[1] and KAMRAN MOINZADEH[2,*]

[1]*Faculty of Business Administration, Bilkent University, Ankara, Turkey*
E-mail: eberk@bilkent.edu.tr
[2]*School of Business, University of Washington, Seattle, WA 98195, USA*
E-mail: kamran@u.washington.edu

Received October 1998 and accepted June 1999

In this paper, we consider the maintenance scheduling of a group of M identical machines, the performance of which deteriorates with usage. Examples of such situations are frequently found in the heavy machine tooling, petro-chemical and semi-conductor industries among others. Assuming a limited maintenance resource and that the maintenance times are i.i.d., we propose a dynamic maintenance policy which utilities the information about the number of operating machines and their ages. We analyze the system for the special cases of constant and exponentially distributed maintenance times. We investigate the impact of maintenance time variability on system performance and evaluate the performance of various maintenance policies within the proposed policy class when the expected profit rate is maximized.

1. Introduction

In this paper, we consider the maintenance scheduling of a group of M identical machines the performance of which deteriorates over time. That is, we assume that the performance of each machine measured by the quality of its output, is inversely related to the operation time since the last maintenance. This general setup may be found in a variety of environments such as the heavy machine tooling. (e.g., metal cutting), petro-chemical, semi-conductor industries (e.g., various stages of wafer fabrication as in etching and deposition).

Maintenance models for deteriorating systems have been widely investigated in the past three decades. Comprehensive reviews of the literature can be found in the surveys by McCall [1], Pierskalla and Voelker [2], Valdez-Flores and Feldman [3], and Cho and Parlar [4]. In order to highlight the contribution of this paper, we classify and discuss the previous work in two broad categories: (i) failure models versus strict deterioration models; and (ii) single versus multiple machine environments.

In general, failure models treat the deterioration of a machine as a stochastic progression through a finite set of states from a new or renovated state to an inoperative or failed state. Early studies which have focused on a single machine are due to Barlow and Proschan [5] and Cox [6], Kolesar [7], Derman [8] and Kao [9] who showed that the optimal policy that minimizes the discounted or time average costs is of the control limit type. Under a control limit rule, the machine is maintained or replaced at failure or at a certain age, whichever comes first. For the extensions of these basic results and various constructs of the single machine setting with random failure, we refer the reader to the surveys in McCall [1] and Valdez-Flores and Feldman [3].

When dealing with a group of machines that are subject to random failures, the preventive maintenance problem becomes more complex due to possible economic or stochastic dependencies among the machines. Economic dependence implies that there are economies of scale in repair or replacement. Stochastic dependence may occur in two ways: (i) the failure rate of a machine may be affected by the failure of another machine; and, (ii) in environments with limited maintenance resources, the existence of a queue of machines at the maintenance facility may result in maintenance delays. In the presence of either type of dependence, a control limit rule on the repair or replacement of one machine is not necessarily optimal anymore.

In order to exploit the economic dependence in a multiple machine environment, the so-called opportunistic maintenance policies have been proposed in the literature. The models using such policies focus primarily on the effect of economies of scale and invariably assume instantaneous

*Corresponding author

0740-817X

replacement. Under the opportunistic replacement policy, the maintenance decision is made at the instances of machine failure. In such cases, either a group or all of the machines are replaced. The former replacement policy is called *group replacement* and the latter is known as *block replacement*. Of the two types of stochastic dependence, failure rate dependence among the machines has received very little attention in the literature (see Ozekici [10] for a discussion and review of preventive maintenance in the presence of interdependent machine failure). As for the stochastic dependence due to limited repair resources, the vast existing literature on failure models falls into the category of machine interference and employs queueing methodology to calculate the optimal repair rates (in terms of number of servers, allocation of units to non-identical servers, establishing priorities, etc.) and the optimal number of machines and spares. For an extensive review of these models, we refer the reader to the excellent surveys by Pierskalla and Voelker [2], and Cho and Parlar [4].

An equally important maintenance scheduling problem that requires attention arises in situations where the machines are, with usage, subject to performance deterioration but not to failure. In the single machine setting, it follows from Kolesar [7] that the machine is repaired or replaced when its deterioration (or, its deterioration measured as the age of the machine) exceeds a certain level. Other works which deal with the maintenance of a single machine subject to strict deterioration include Barlow and Preschan [5], Derman [11], Jorgensen *et al.* [12], Gibra [13], Drezner and Wesolowsky [14], Jeang and Yang [15] and Zhou *et al.* [16]. In a multiple machine setting, however, the optimal maintenance policy class is an open question. One may conjecture that it would be a function of the deterioration status of all machines in the system. The only study in which the maintenance decision is based on the status of all the machines in the group is by Sivazlian and Mahoney [17]. Thus, their model is unique in that the state of the system is represented by the respective deterioration of each machine within the group. However, the simultaneous assumptions of block replacement and that replacements are instantaneous render the model inappropriate for production environments where repair resources are limited and repair times are consequently non-zero. Other work on machine groups subject to strict performance deterioration also suffer from the same simplifying assumptions (e.g., Brosh *et al.* [18], Berg and Epstein [19] and Love *et al.* [20]. In fact, the observation in McCall [1] is still valid: "The only maintenance models that have dealt successfully with queueing are those in which the equipment is replaced at failure. [...] Non-trivial maintenance policies have never been considered within a queueing environment. Mathematical analysis of these problems constitutes an important area for further research."

In this paper, we consider a multiple machine environment in which identical machines are subject to performance deterioration but not to random failure and the maintenance is done by a maintenance facility or crew that has limited resources. Assuming that machines are monitored continuously, we develop a model under a general class of "dynamic" maintenance policies which utilizes the information about the number of machines in operation and their ages. A number of studies have appeared on dynamic maintenance policies (e.g., van der duyn Schouten and Vanneste [21,22], Gurler and Kaya [23]. However, these works have assumed instantaneous replacements. Recently, Noh *et al.* [24] have considered a class of age-replacement policies for the setting with two machines subject to failure and one repair person. The proposed control limit is such that a machine is taken out of service for replacement either when it has failed or as soon as it reaches a critical age when both machines are operating; that is, the control limit is operational only when the repair facility is idle. To the best of our knowledge, ours is the first work which studies a general class of *age-based* maintenance policies in the multiple machine environment subject to strict performance deterioration in the presence of *limited* maintenance resources.

To motivate our model, we cite a specific example dealing with the maintenance scheduling of the plasma etching machines in the semi-conductor fabrication process. Plasma etching is used at various points in the microcircuit fabrication process. The term etching is used to describe all the techniques by which the material can be either uniformly removed from the wafer as in surface polishing, or locally removed as in the delineation of a pattern for microcircuit. In plasma etching, the material to be removed undergoes conversion to a volatile state by chemical reaction with one or more energetic species which are produced in a gas discharge (see Ghandi [25] for more details). As the wafers are etched in the etching chamber of the machine, etch byproducts referred to as polymers are gradually deposited in the chamber walls. With the change in the environment (e.g., temperature of the chamber), the polymer will start to flake off creating particle problems which result in the deterioration of the process yield (e.g., the fraction of good dies in a wafer). This problem necessitates the periodic cleaning and maintenance of the etching chamber which is usually performed by a single maintenance crew. The restoration is almost perfect; therefore, one can reasonably assume that the machine is put into service (taken on line) in excellent condition. Round the clock production is a common industry practice and, hence, all online etching machines can be assumed to be in continuous operation. At any point in time, any of the on-line machines can be taken off line for maintenance. The operating cost of a machine associated with performance increases due to deterioration, therefore, it is desirable to employ intelligent maintenance policies which avoid operating machines in a *highly* deteriorated condition.

The remainder of the paper is organized as follows: in Section 2 we present the model, state the underlying assumptions and propose the general control policy class. In Sections 3 and 4, for deterministic and random maintenance times, respectively, we derive the steady state behavior and obtain expressions for the operating characteristics of the system. In Section 5, we provide the optimal policy within the proposed class and discuss a number of 'simple and intuitive' alternative maintenance policies. We investigate the effectiveness of such policies *vis a vis* the optimal policy via a numerical experiment and provide sensitivity results. In Section 6, we develop an extension of the model herein with state-dependent random maintenance times; and, finally, in Section 7, we summarize our model and provide suggestions for future work.

2. The model and assumptions

Consider a group of M identical machines the performance of which, measured by the quality of the output of a machine (e.g., the fraction of good dies), deteriorates with the operating time since the last maintenance. We shall interchangeably refer to the operating time elapsed since the last maintenance of the machine as the age of the machine. The operating machines are run continuously and are subject only to performance deterioration. That is, no random failures occur. Any of the on line machines can be taken off line for maintenance at any point in time. We assume that maintenance is performed by a single maintenance crew, although we shall remove this assumption in Section 6, and, that the maintenance times are identically and independently distributed. In general, the maintenance time of a machine would depend on the wear on the machine and, therefore, be a function of the age of the machine; however, age-dependent maintenance times render the analytical treatment of the problem complex. Therefore, for analytical tractability, we assume that maintenance time of a machine is independent of the age at which it is taken off line and has a mean of $1/\mu$. Upon maintenance, the machine is restored to perfect condition. Each machine produces output at a constant rate ψ and the output yield (e.g., the expected fraction of good dies produced in a wafer per unit time) of a machine with an age of x is $\theta(x)$, where $\partial\theta(x)/\partial x \leq 0$.

Define the state of the system as $\mathbf{x}_n(t) = \{x_1(t), \ldots, x_n(t)\}$, here n is the number of machines in operation at time t and $x_i(t)$ denotes the time elapsed prior to t since the i-th operational machine has undergone maintenance (i.e., the age of the i-th operational machine at time t). Note that $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$ for $0 \leq n \leq M$. Let $p_n(t, x_1, \ldots, x_n)$ and P_n denote, respectively, the probability density of the system being in state $\mathbf{x}_n$ at time t, and the probability distribution of n operational machines at steady state.

The form of the optimal maintenance policy in this general setup is an open question and depends on the operational objective(s) of the firm; however, one may conjecture that it would involve the number of machines that are undergoing maintenance and the ages of the all the machines that are operational. We propose the following general "dynamic" maintenance policy class which captures the essence of the above conjecture by considering both the number and the age of machines in operation:

POLICY. *When there are n machines in operation (M–n in the maintenance queue), a machine is taken off line for maintenance if its age is greater than or equal to a maintenance trigger time, T_n.*

Note that $\mathbf{T} = \{T_1, \ldots, T_M\}$ is the decision vector. It is practical to assume that the maintenance trigger times are a non-increasing function of the number of on-line machines; that is $T_1 \geq T_2 \geq \cdots \geq T_M$. In other words, with more machines in operation, the age dependent maintenance decisions are made at least as early as when there are fewer machines on-line. Clearly, $0 \leq x_1(t) \leq \cdots \leq x_n(t) \leq T_n$ for $1 \leq n \leq M$. Since the system is monitored continuously, the maintenance decisions apply only to the oldest operational machine.

The measures that describe the performance of a system under consideration will depend on the specifics of the firm's operational objectives. For example, in the semiconductor fabrication, process efficiency is a critical factor of performance in wafer production, given the commodity (high volume, low cost) nature of computer chips. The expected throughput (e.g., the *number* of good dies that are produced per time unit), $E[TP]$, and the expected number of operational machines (e.g., the average machine utilization), $E[N]$, have a direct impact on the cost of production and, hence, are commonly used as efficiency measures in practice. The expected yield of the process (e.g., the *fraction* of good dies that are produced per time unit) is also monitored as a benchmark efficiency measure.

Expected cost per unit time has been the traditional measure of the steady state behaviour of a maintenance policy. This is appropriate when the value of the system output does not vary with the age of the system. However, when the output performance of a machine deteriorates over time, one needs to consider an expected profit per unit time since the trade off is now between the incurred costs of maintenance and idleness (the latter in the form of foregone revenue) and those of output deterioration (in the form of scrap).

We consider the following cost/revenue structure: each unit produced by a machine incurs a cost of c (regardless of whether the unit produced is good or defective), and each good unit produced brings in a revenue of r. Without loss of generality, we assume that defective units are scrapped at no additional charge. Furthermore, every time a machine is taken off line for maintenance, a fixed maintenance cost, K is incurred. Here, we are assuming that the maintenance cost is independent of the age of the machine. This assumption is reasonable in situations where the maintenance often involves the replacement of modules and parts. the expected profit per unit time, $\Pi(\mathbf{T})$ is, then given by:

$$\Pi(\mathbf{T}) = rE[TP] - c\psi E[N] - KE[R], \tag{1}$$

where $E[R]$ is the expected number of maintenances undertaken per unit time.

In the next two sections, we analyze two special cases of the general setup described above: one with deterministic (constant) maintenance times and the other with random (exponential) maintenance times.

3. The case with deterministic maintenance times

In this section, we analyze the case when maintenance times are constant. Such deterministic maintenance times may be observed when maintenance procedures are highly routine and involve perhaps replacement of a whole unit or module.

Consider a group of M machines to be operated under a given maintenance policy $\mathbf{T} = \{T_1, \ldots, T_M\}$ of the proposed class. Without loss of generality, assume that the machines are initially all new. For any given policy $\mathbf{T}$, the behavior of the system will depend on the value of T_M. Specifically:

(i) If $T_M < (M-1)/\mu$, since $T_1 \geq T_2 \geq \cdots \geq T_M$, then there exist a $j < M$ such that $(j-1)/\mu \leq T_j < j/\mu$. Let $\hat{j}$ be the *smallest* value of $j (j < M)$ such that $(j-1)/\mu \leq T_j < j/\mu$ for $T_j \in \{T_1, \ldots T_{M-1}\}$. Suppose that $T_{\hat{j}} = \tau$ and note that $\tau < (M-1)/\mu$. One can easily see that the system reaches steady state after exactly one maintenance of all the machines. At steady state, there are either $\hat{j}$ or $(\hat{j}-1)$ machines on line (the remainder being in the queue for maintenance), the ages of the operational machines are exactly $1/\mu$ apart and, the system renews itself every M/μ time units. Within each renewal cycle, defined as the instances at which a tagged machine is taken on line, there are exactly M maintenances and when a machine is taken off line, it has operated for exactly τ time units. Let N denote the number of on-line machines at steady state and $P_n (1 \leq n \leq M)$ be the probability distribution of N. Using renewal theory, we get

$$P_n = \begin{cases} \dfrac{\hat{j}/\mu - \tau}{1/\mu} & n = \hat{j} - 1, \\ \dfrac{\tau - (\hat{j}-1)/\mu}{1/\mu} & n = \hat{j}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(ii) If $T_M \geq (M-1)/\mu$, then let $T_M = \tau$, and note that that $\tau \geq (M-1)/\mu$. In such cases, the system also reaches steady state after exactly one maintenance of all the machines. At steady state, there are either M or $(M-1)$ machines on line, the ages of the operational machines are exactly $1/\mu$ apart and, the system renews itself every $(\tau + 1/\mu)$ time units. Within each renewal cycle, there are exactly M maintenances and when a machine is taken off line, it has operated for exactly τ time units, as before. Therefore, we have

$$P_n = \begin{cases} \dfrac{M/\mu}{\tau + 1/\mu} & n = M-1, \\ \dfrac{\tau - (M-1)/\mu}{\tau + 1/\mu} & n = M, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that the above steady-state probability distribution holds irrespective of the queue discipline (FCFS, random etc.) employed at the maintenance facility except for preemptive service. To see that this is so, first consider $T_M < (M-1)/\mu$. At steady state, there are $(M - \hat{j} + 1)$ or $(M - \hat{j})$ machines at the maintenance facility; that is, the maintenance facility is always busy. (Recall that $\hat{j} < M$.) As long as preemption is not allowed during maintenance, it does not matter how a machine is chosen from the queue for maintenance and there will be a machine joining the running group every $(1/\mu)$ time units. Therefore Equation (2) holds for all service routines except the primitive ones. For $T_M \geq (M-1)/\mu$, there is at most one machine undergoing maintenance and no queue at all times; therefore, the service routine is not an issue in this case.

For a given policy $\mathbf{T} = \{T_1, \ldots, T_M\}$, the expected number of operating machines/time will be:

$$E[N] = \begin{cases} \dfrac{M\tau}{\tau + 1/\mu} & \text{if } \tau \geq (M-1)/\mu, \\ \mu\tau & \text{otherwise.} \end{cases} \tag{4}$$

Furthermore, the expected throughput rate can be derived by dividing the expected total output of the system in a cycle by the expected cycle time which yields:

$$E[TP] = \begin{cases} \mu\psi \int_0^\tau \theta(x)\mathrm{d}x & \text{if } \tau < (M-1)/\mu, \\ \dfrac{M\psi}{\tau + 1/\mu} \int_0^\tau \theta(x)\mathrm{d}x & \text{otherwise.} \end{cases} \tag{5}$$

Similarly, it can shown that the expected number of machines maintained/time can be written as:

$$E[R] = \begin{cases} \dfrac{M}{\tau + 1/\mu} & \text{if } \tau \geq (M-1)/\mu, \\ \mu & \text{otherwise.} \end{cases} \tag{6}$$

Using (4) through (6), the expected profit per unit time is given as:

$$\Pi(\mathbf{T}) = \begin{cases} \dfrac{Mr\psi \int_0^\tau \theta(x)\mathrm{d}x - c\psi M\tau - KM}{\tau + 1/\mu} & \text{if } \tau \geq (M-1)/\mu, \\ r\psi\mu \int_0^\tau \theta(x)\mathrm{d}x - c\psi\mu\tau - K\mu & \text{otherwise,} \end{cases} \tag{7}$$

where τ is as defined before. Note that the steady state behaviour of the system above is determined solely by τ and the policy reduces to a single parameter policy. That is, all policies with $T_M < (M-1)/\mu$ and those with $T_M \geq (M-1)/\mu$ result in the same steady state behavior, respectively. Therefore, it suffices to specify τ to achieve a particular operational objective when maintenance times are deterministic.

4. The case with random maintenance times

In this section, we consider the case when maintenance times are random. Following the customary assumption in machine-interference problems for analytical tractability, we specifically assume that maintenance times are exponentially distributed with a mean of $1/\mu$. Our approach in the sequel is similar to the one employed by Cox [26], Gnedenko and Kovalenko [27], Schmidt and Nahmias [28], and Moinzadeh [29]; we shall derive the steady state distribution of $\mathbf{x}_n(t)$ which will then be employed to obtain the operating characteristics of the system for a given maintenance policy. Let $p_n(t,x_1,\ldots,x_n)$ denote the probability density of the system being in state $\mathbf{x}_n$ at time t. The system of partial differential equations and their boundary conditions which describe the state of the system are found as follows.

Case 1(a). $1 \leq n < M$, $x_1 > 0$ and $x_n < T_{n+1}$. Let $h > 0$ be a small number. Then

$$p_n(t+h,x_1+h,\ldots,x_n+h) = (1-\mu h)p_n(t,x_1,\ldots,x_n) + (1-\mu h)\int_{T_{n+1}-h}^{T_{n+1}} p_{n+1}(t,x_1,\ldots,x_n,\zeta)\mathrm{d}\zeta + o(h). \quad (8)$$

This case corresponds to the instance when there is at least one machine undergoing maintenance at time $t+h$ and no machine has been put into service in the time interval $(t,\ t+h)$. The first order probability of the occurrence of this event is $1-\mu h$. The state $\mathbf{x}_n$ can then be reached at time $t+h$ *either* if there were n machines running at time t, *or* there were $n+1$ machines running at time t and the oldest machine was taken off line for maintenance after having reached the age of T_{n+1} during the interval $(t,t+h)$. All other transitions have probability $o(h)$. Note that $x_1 > 0$ prohibits the possibility of a machine finishing maintenance during the time interval $(t,t+h)$; hence, a transition is not possible from a state with $n-1$ machines running.

By adding and subtracting successive terms and using the Mean Value Theorem, we have

$$\begin{aligned}
&p_n(t+h,x_1+h,\ldots,x_n+h) \\
&= (1-\mu h)p_n(t,x_1,\ldots,x_n) + [p_n(t,x_1+h,\ldots,x_n+h) - p_n(t,x_1+h,\ldots,x_n+h)] \\
&+ \sum_{i=1}^{n}[p_n(t,x_1,\ldots,x_i,x_{i+1}+h,\ldots,x_n+h) - p_n(t,x_1,\ldots,x_i,x_{i+1}+h,\ldots,x_n+h)] \\
&+ hp_{n+1}(t,x_1,\ldots,x_n,T_{n+1}-\varepsilon) + o(h),
\end{aligned} \quad (9)$$

where $0 \leq \varepsilon \leq h$. Rearranging terms and dividing by h, we get

$$\begin{aligned}
&[p_n(t+h,x_1+h,\ldots,x_n+h) - p_n(t,x_1+h,\ldots,x_n+h)]/h \\
&+ \sum_{i=1}^{n}[p_n(t,x_1,\ldots,x_{i-1},x_i+h,\ldots,x_n+h) - p_n(t,x_1,\ldots,x_i,x_{i+1}+h,\ldots,x_n+h)]/h \\
&= -\mu p_n(t,x_1,\ldots,x_n) + p_{n+1}(t,x_1,\ldots,x_n,\{T_{n+1}-\varepsilon\}) + o(h).
\end{aligned} \quad (10)$$

If we assume that a steady state distribution exists, letting $h \to 0$ and then $t \to \infty$, we obtain

$$\sum_{i=1}^{n}\frac{\partial}{\partial x_1}p_n(x_1,\ldots,x_n) = -\mu p_n(x_1,\ldots,x_n) + p_{n+1}(x_1,\ldots,x_n,T_{n+1}). \quad (11)$$

Note that we have dropped the time argument in notation at steady state for the sake of brevity and $p_n(x_1,\ldots,x_n)$ denotes the steady state probability density of $\mathbf{x}_n(t)$.

Case 1(b). $1 \leq n < M, x_1 > 0$ and $T_{n+1} \leq x_n < T_n$. This case is similar to the one above except that, since $T_{n+1} \leq x_n < T_n$, transitions from states with $(n+1)$ operating machines in the interval $(t,t+h)$ are now allowed. Hence,

$$p_n(t+h,x_1+h,\ldots,x_n+h) = (1-\mu h)p_n(t,x_1,\ldots,x_n) + o(h). \quad (12)$$

Using similar operations, at steady state, we obtain

$$\sum_{i=1}^{n}\frac{\partial}{\partial x_i}p_n(x_1,\ldots,x_n) = -\mu p_n(x_1,\ldots,x_n). \quad (13)$$

Case 2. $n = M, x_1 > 0$ and $x_M < T_M$. Then

$$p_M(t+h,x_1+h,\ldots,x_M+h) = p_M(t,x_1,\ldots,x_M). \quad (14)$$

This follows since all machines are operational at t, and $x_1 > 0$ implies that no machine has finished maintenance and $x_M < T_M$ implies that no machine has been taken off line for maintenance during $(t,t+h)$. Hence, machines age uniformly over time. At steady state, we obtain

$$\sum_{i=1}^{M}\frac{\partial}{\partial x_i}p_M(x_1,\ldots,x_M) = 0. \quad (15)$$

Case 3. $n = 0$. In this case, all machines are off-line and at the maintenance queue. This state can be reached only if there are no operational machines at t and no maintenance is completed in $(t, t+h)$ or there is one machine operational at t and its age reaches T_1 during $(t, t+h)$ and is, therefore, taken off line for maintenance. Hence,

$$p_0(t+h,\cdot) = (1-\mu h)p_0(t+h,\cdot) + (1-\mu h)\int_{T_1-h}^{T_1} p_1(t,\zeta)\mathrm{d}\zeta + o(h). \quad (16)$$

Once again, at steady state we have

$$\mu p_0(\cdot) = p_1(T_1). \quad (17)$$

Next we consider the boundary conditions for the above system of partial differential equations. The boundary conditions are found by considering the discontinuities in the motion of the state of the system caused by the completion of a maintenance and the introduction of a new machine (i.e. the instances at which the age of the youngest machine is equal to zero). We derive the boundary conditions at steady state as follows. (For a detailed discussion on the derivation of the boundary conditions, see Schmidt and Nahmias [8])

(i) For $1 \le n < M$ and $x_{n-1} < T_n$, there are $(M-n)$ machines in the maintenance queue. The completion of a maintenance introduces a new machine to join the group of machines in operation. Transitions to state $(0, x_1, \ldots, x_{n-1})$ occur if

(a) there are $(n-1)$ machines in operation and a maintenance is completed, or

(b) there are n machines operating and the age of the oldest machine has exceeded T_{n+1} when a maintenance is completed. In such situations, the arrival of the new machine will bring the state of the system to $(n+1)$ machines, resulting in the oldest machine being taken off line leaving n machines in operation.

These transitions occur in an infinitesimal time. To obtain the boundary condition, we consider the system immediately before and after a maintenance completion. Define:

$$Q_n(t,0,x_2,\ldots,x_n) = \int_0^h\int_{x_2}^{x_2+h}\cdots\int_{x_n}^{x_n+h} p_n(t,\xi_1,\xi_2,\ldots,\xi_n)\mathrm{d}\xi_n\cdots\mathrm{d}\xi_1. \quad (18)$$

We note that $Q_n(t, 0, x_2, \ldots, x_n)$ represents the probability mass in the neighbourhood of the boundary $(t, 0, x_2, \ldots, x_n)$.

By the mean value theorem, for some $0 \le \varepsilon_i \le h$ $(i = 1, \ldots, n)$, we have

$$Q_n(t,0,x_2,\ldots,x_n) = h^n p_n(t,\varepsilon_1,x_2+\varepsilon_2,\ldots,x_n+\varepsilon_n) + o(h^n). \quad (19)$$

Furthermore,

$$Q_n(t+h,0,x_2,\ldots,x_n) = \mu h\int_{x_2-h}^{x_2}\cdots\int_{x_n-h}^{x_n} p_{n-1}(t,\xi_2,\ldots,\xi_n)\mathrm{d}\xi_n\ldots\mathrm{d}\xi_2$$
$$+ \mu h\int_{(T_{n+1}\vee x_{n-1})}^{T_n}\left\{\int_{x_2-h}^{x_2}\cdots\int_{x_n-h}^{x_n}\int_{\zeta-h}^{\zeta} p_n(t,\xi_2,\ldots,\xi_n,\xi_{n+1}) \times \mathrm{d}\xi_{n+1}\mathrm{d}\xi_n\cdots\mathrm{d}\xi_2\right\}\mathrm{d}\zeta + o(h), \quad (20)$$

where $(x \vee y) = \max(x, y)$. The RHS of (20) represents the probability flow over $(t, t+h)$ for the boundary neighbourhood about $(0, x_2, \ldots, x_n)$ as explained above. By using Taylor's Theorem, together with the Mean Value Theorem, $Q_n(t+h, 0, x_2, \ldots, x_n)$ can also be written as

$$Q_n(t+h,0,x_2,\ldots,x_n) = h^n p_n(t,\varepsilon_1,x_2+\varepsilon_2,\ldots,x_n+\varepsilon_n) + o(h^n). \quad (21)$$

Equating (20) and (21), dividing by h^n, taking the limit $h \to 0$ and then $t \to \infty$, we get, at steady state

$$p_n(0,x_1,\ldots,x_{n-1}) = \mu p_{n-1}(x_1,\ldots,x_{n-1}) + \mu\int_{(T_{n+1}\vee x_{n-1})}^{T_n} p_n(x_1,\ldots,x_{n-1},\zeta)\mathrm{d}\zeta. \quad (22)$$

(ii) When $n = M$, the boundary condition is similar to (i) except that transitions from states with M machines to themselves are not allowed, since there are no machines undergoing maintenance in such cases. Therefore, the second term of the RHS of (22) will not be present and we have,

$$p_M(0,x_1,\ldots,x_{M-1}) = \mu\, p_{M-1}(x_1,\ldots,x_{M-1}). \quad (23)$$

It can be verified that a solution to the above system of partial differential equations and their boundary conditions is:

$$p_0(\cdot) = c_0\exp(-\mu T_1), \quad (24a)$$

$$p_n(x_1,\ldots,x_n) = \begin{cases} c_0\mu^n\exp(-\mu T_{n+1}) & x_n \le T_{n+1},\ 1 \le n \le M-1 \\ c_0\mu^n\exp(-\mu x_n) & T_{n+1} < x_n \le T_n, \end{cases} \quad (24b)$$

$$p_M(x_1,\ldots,x_M) = c_0\mu^M\exp(-\mu T_M), \quad (24c)$$

where c_0 is the normalizing constant obtained by integrating the densities to unity and can be written as:

$$c_0 = \frac{1}{1-F(M+1,\mu T_M)},$$

with $f(\cdot, \mu t)$ being the probability mass of the Poisson distribution with mean μt, and $F(y, \mu t) = \sum_{i=y}^{\infty} f(i, \mu t)$.

The probability distribution of the number of on-line machines at steady state, P_n, can be found by integrating the probability density over the appropriate n-dimensional space:

$$P_n = \begin{cases} \dfrac{f(0, \mu T_1)}{1 - F(M+1, \mu T_m)} & n = 0, \\ \dfrac{F(n, \mu T_n) - F(n+1, \mu T_{n+1})}{1 - F(M+1, \mu T_M)} & 1 \le n < M, \\ \dfrac{f(M, \mu T_M)}{1 - F(M+1, \mu T_M)} & n = M. \end{cases} \quad (25)$$

Note that the steady state distribution derived in Equations (8)–(25) holds irrespective of the queue discipline employed at the maintenance facility. In this case, the queue discipline may even be preemptive due to the memoryless property of the exponential assumption. From the above distribution, we obtain

$$E[N] = \frac{\sum_{n=1}^{M} F(n, \mu T_n) - MF(M+1, \mu T_M)}{1 - F(M+1, \mu T_M)}. \quad (26)$$

Furthermore, as shown in the Appendix,

$$E[TP] = \psi \sum_{n=1}^{M} \int_{x_n=0}^{T_n} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \sum_{i=1}^{n} \theta(x_i) p_n(\mathbf{x}) d\mathbf{x}$$

$$= \psi c_0 \mu \left\{ \int_0^{T_1} \theta(x) dx - \left(\int_0^{T_M} \theta(x) \, dx \right) F(M, \mu T_M) - \sum_{n=1}^{M-1} \int_{T_{n+1}}^{T_n} \theta(x) F(n, \mu x) dx \right\}. \quad (27)$$

The expected number of maintenances undertaken per time, $E[R]$, is given by

$$E[R] = \mu(1 - P_M) = \frac{1 - F(M, \mu T_M)}{1 - F(M+1, \mu T_M)}. \quad (28)$$

The expected profit per unit time can be obtained by substituting these expressions in (1). In the next section, we present results on the effect of maintenance time variability and the performance of various maintenance policies within the general policy class when the objective is expected profit maximization.

5. Optimization and numerical results

In this section, we examine the behaviour of the system under the optimal proposed policy with respect to certain system parameters. Furthermore, we examine the performance of a number of simple but intuitively appealing maintenance policies constructed as special cases within the general policy class.

We start by describing the policies considered.

(i) The first policy refers to the optimal policy within the proposed general class which maximizes the expected profit per unit time, and we refer to it as P1. It will constitute our benchmark in the comparison of other policies. In the following proposition, we characterize the optimal policy parameter for P1 for both cases when maintenance times are constant and exponential.

Proposition 1

(a) *When maintenance times are constant,* $\Pi(\mathbf{T})$ *is concave in* τ *and; if* $\theta((M-1)/\mu) < c/r$, *then optimal policy parameter* τ^* *which maximizes* $\Pi(\mathbf{T})$ *will be the solution to* $\theta(\tau) = c/r$; *otherwise,* τ^* *will be the solution to*:

$$\theta(\tau)(\tau + 1/\mu) - \int_0^{\tau} \theta(x) dx = \frac{c}{r\mu} - \frac{K}{r\psi}.$$

(b) *when maintenance times are exponentially distributed,* $\Pi(\mathbf{T})$ *is (separately) concave in* T_i *for* $i = 1, \ldots, M-1$ *and is maximized when* $T_i^* = \tau$ *for* $i = 1, \ldots, M-1$ *where* τ *solves* $\theta(\tau) = c/r$, *and* T_M^* *lies in the region* $[0, \tau]$.

Proof. Follows from standard optimization results and that $\theta(x)$ is decreasing in x. ■

We are unable to show the concavity of $\Pi(\mathbf{T})$ with respect to T_M when the maintenance times are exponential; therefore, to find T_M^* a univariate numerical search was conducted.

(ii) One special class of maintenance policies that we consider is the single parameter policies. Such policies assign a single (identical) age for all machines to trigger preventive maintenance and, thus, ignore the *number* of operational machines when making decisions as to which machine(s) should be maintained. They are attractive because the maintenance decision for a particular machine is made independent of the status of the other machines, and, thus, there is less of a need in practice for coordination and supervision among the operators of the machines. We consider two single parameter policies which we describe below:

(a) In the first single parameter policy, the maintenance age for all machines is determined within the proposed general policy class by setting them equal. That is, under this policy, the profit maximization problem is solved subject to the condition that all $T_i (i = 1, \ldots, M)$ are equal. We shall refer to this policy as P2. Note that P2 is an optimal policy within the proposed class when maintenance times are deterministic since it solves for the same unique τ^* as P1.

(b) In the second single parameter policy considered, the identical maintenance age for all machines is based on the single machine environment that ignores possible queueing delays encountered for maintenance. Under this policy, the maintenance age of the machines is found by maximizing the expected profit rate for a *single* machine which requires an average maintenance time of $(1/\mu)$. We refer to this policy as P3. This policy is important in two respects: first, it is often used in practice and has been the focus of many of the previous work as mentioned above. Secondly, its performance *vis a vis* the proposed policy P1 illustrates the importance of the queueing delays in a multiple machine environment and enables us to identify the conditions under which ignoring the queueing effects in the analysis may be appropriate. When maintenance times are deterministic, employing P3 forces the policy parameter τ to be set equal to $\hat{\tau}$. Clearly, we would expect P3 to be inferior to P2 in general.

(iii) Finally, we consider a simple maintenance policy which takes the oldest machine off line as soon as a maintenance is finished. We refer to this policy as P4. Under P4, there are always $(M-1)$ operational machines; therefore, the maintenance crew is always busy. The advantages of this policy lies in its simplicity since it does not require keeping records of the operational ages of the machines. Note that, P4 is also a special case of the proposed general policy class: when maintenance times are exponentially distributed, $T_i \to \infty$ $(i = 1, \ldots, M-1)$ and $T_M = 0$. And, when maintenance times are constant, $\tau = (M-1)/\mu$.

In our numerical study, we assumed that the expected deterioration in the output of a machine as a function of operation time follows an exponential form; that is, $\theta(x) = a\exp(-bx)$. We set $a = 1$ and considered a group of four machines $(M = 4)$. We also fixed the maintenance rate $(\mu = 1)$ and varied the deterioration rate, b, between 0.1 and 1. It can easily be shown that the cost parameters in (1) can be normalized by ψr. Let $c' = c/r$ and $K' = K/\psi r$ denote the normalized unit production and maintenance costs. This way, one needs to consider only c' and K' as the cost parameters in the problem. In our experiment, we varied K' between 0 and 0.5 and c' between 0 and 1. We examined the behavior of the system under both constant exponentially distributed maintenance times. The optimization problem for the random maintenance case under all policies was solved numerically. We used increments of 0.01 to determine the optimal policy parameters when a search was needed. In all of our test cases, we observed unimodality.

Figure 1 illustrates the impact on expected profit rate under P1 of the deterioration rate, b, and the normalized unit production cost, c' for both deterministic and random maintenance cases. As expected, increasing b and c' results in a lower profit rate. Very low deterioration rates results in larger maintenance trigger times and therefore queuing delays at the maintenance facility are minimal.

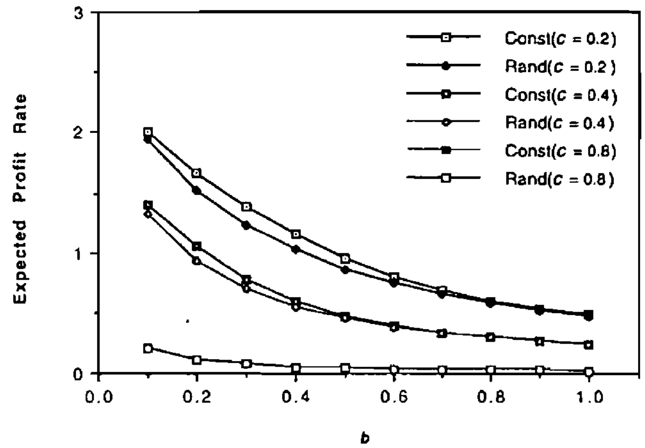

Fig. 1. The behavior of expected profit rate versus b. $(M = 4$, $\mu = 1$, $a = 1$, and $K' = 0)$.

Very high deterioration rates, on the other hand, result in smaller trigger times causing big queuing delays. Therefore, we see that the effect of maintenance time variability becomes negligible for very small and very large values of b, and it remains small for moderate values of deterioration rates, as well. The maximum difference in expected profit rates with constant and exponential maintenance times is observed to be around 20% for the parameter set considered. The effect of maintenance time randomness increases as the fixed maintenance cost, K, gets bigger, and it decreases as the unit production cost, c, increases.

Next, we discuss the performance of the special policies P2, P3 and P4 when compared to P1. Since the observations hold for both deterministic and random maintenance times, we only provide the results for the case of exponentially distributed maintenance times. We present our findings in terms of the percentage deviation in the expected profit per unit time under each policy (P2, P3 and P4) as compared to the profit rate obtained by the optimal policy within the general class, P1, defined as:

$$\%\text{ deviation} = \left[\frac{\text{Expected profit rate under P1}}{\text{Expected profit rate under P1}} - \frac{\text{Expected profit rate under } P_i (i = 2,3,4)}{\text{Expected profit rate under P1}}\right] \times 100.$$

In Fig. 2, the performance of each policy is depicted as c' varies. This illustrates the trade off between the incurred maintenance costs due to maintenance and production costs due to the deteriorated performance of the machines under each policy.

First, we note that P2 dominates P3 in all cases, as expected. We also observe that, in general, there exist three distinct regions in which single parameter policies (P2 and P3) and P4 alternately dominate one another. When the unit production cost is small relative to the unit revenue, P2 and P3 perform better than P4. For

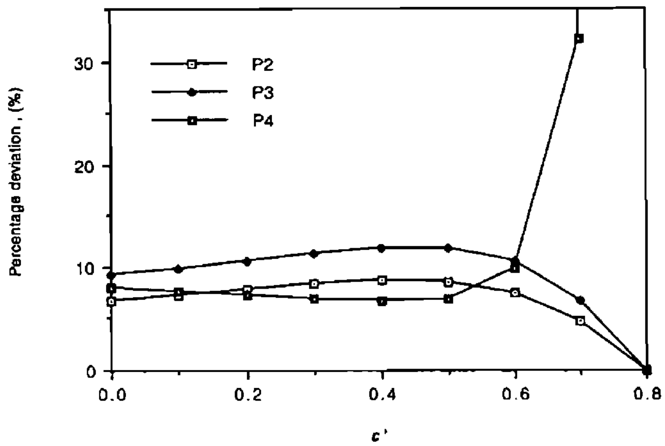

Fig. 2. The behavior of percentage deviation in average profit rate versus c'. ($M = 4$, $\mu = 1$, $a = 1$, $b = 0.1$, and $K' = 0.25$).

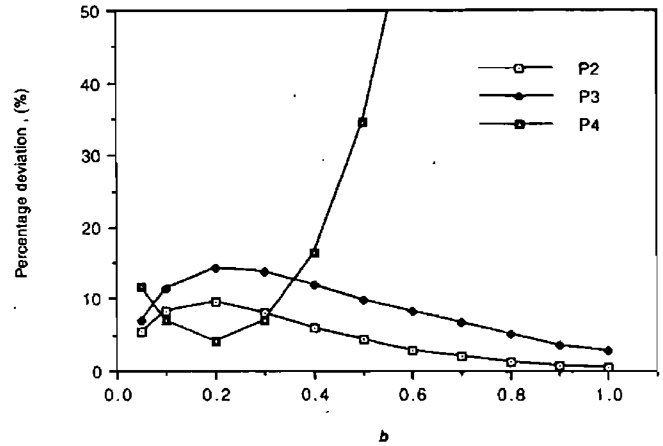

Fig. 3. The behavior of percentage deviation in average profit rate versus deterioration rate, b. ($M = 4$, $\mu = 1$, $a = 1, c' = 0.3$, and $K' = 0.25$).

small unit production costs, P4 takes machines off line sooner (faster) than necessary and thereby incurs unnecessary maintenance costs. Policies P2 and P3, on the other hand, allow for some possible wait if the oldest machine has not deteriorated much when the maintenance on the off line machine is over. As the unit production cost increases, it becomes more attractive to take the machines off line sooner and incur some additional maintenance costs which compensate for the lost revenues due to deteriorated performance. Hence, in this region P4 dominates P2 and P3 since the former is more reactive. However, when the unit production cost exceeds a certain level, single parameter policies (P2 and P3) become more responsive than P4.

When the unit maintenance cost gets smaller, all else being equal, the region in which single parameter policies dominate P4 disappears. That is, for small unit maintenance costs, taking the oldest machine off line as soon as a maintenance is finished is the better policy except for high unit production costs. For large unit maintenance costs, we notice that the middle region also disappears; that is, when the unit maintenance cost increases, P2 and P3 are superior to P4 regardless of the unit production cost.

Next, we consider the effect of the deterioration rate, b, on the performance of the maintenance policies. The typical behavior of the policies as a function of the deterioration rate is shown in Fig. 3.

In Fig. 3, we also notice the existence of three regions in which policies alternately dominate one another. For reliable systems (small deterioration rates), P2 and P3 perform better than P4. This is again due to the fact that P4 is unnecessarily reactive taking machines off line too soon. As the system becomes less reliable (as the deterioration gets larger), P4 becomes more profitable since its reactiveness matches the deterioration in the performance of the machines. That is, the additional costs of maintenance compensate for the increasing production costs due to low output performance. However, beyond a certain value of the deterioration rate, P4 loses its attractiveness as it becomes under-reactive. In this case, single parameter policies (P2 and P3) start performing better since they do not allow for highly deteriorated machines to operate even when there is a queue for maintenance.

6. The case with state-dependent random maintenance times

An interesting generalization of our model is the case when the maintenance time is exponential, but the mean restoration rate depends on the number of operational machines (the state of the system). Such a case occurs when, for instance, the maintenance is performed by multiple teams (crews) with identical mean maintenance rates or the crew(s) adjusts the mean maintenance rate according to the number of off-line machines waiting for maintenance. The dynamics of the system with such state dependent service is similar to that obtained in Section 4 with the exception that, now, the restoration rate is a function of the number of operational machines.

Let μ_n be the machine restoration rate when there are n machines in operation. For example, if there are s identical maintenance crews present and the mean maintenance rate is μ, then $\mu_n = \min[s, (M - n)]\mu$. (Note that $\mu_M = 0$.) Rewriting the integro-differential equations with the appropriate restoration rate, one can show that the steady state density is now given by

$$p_0(\cdot) = c_0 \exp(-\mu_1 T_1), \tag{29a}$$

$$p_n(x_1, \ldots, x_n) = \begin{cases} c_n \exp(-\mu_n T_{n+1}) & x_n \le T_{n+1} \\ & 1 \le n \le M - 1 \\ c_n \exp(-\mu_n x_n) & T_{n+1} < x_n \le T_n, \end{cases} \tag{29b}$$

$$p_M(x_1, \ldots, x_M) = c_M, \tag{29c}$$

where

$$c_n = c_0 \prod_{i=0}^{n-1} \mu_1 \exp\{(\mu_{i+1} - \mu_i)T_{i+1}\}. \quad (30)$$

As before, c_0 is the normalizing factor and is found by integrating the densities to unity. The operating characteristics of the system with state dependent service can then be computed in a similar fashion as in Section 4.

7. Summary

In this paper, we considered a group of M identical machines, the performance of which deteriorates with usage. We proposed a "dynamic" maintenance policy class which utilizes the information about the number of operating machines and the performance status of the machines. We developed expressions for the operating characteristics of the system and optimization results for the cases when maintenance times are deterministic and exponentially distributed. We investigated the effect of maintenance time randomness and compared the performance of maintenance policies within the proposed policy class on the average profit rate.

Possible extensions to this research include the case where the maintenance times follow a general distribution and, possibly, are a function of the age of the machine taken off line, and the case where there are random failures of the machines due to part breakdowns in addition to the gradual performance deterioration considered in our model. Also maintenance facility design decisions such as having one faster crew or more crews with slow maintenance rates under various cost structures constitute an interesting research area (e.g. van der duyn Schouten and Wartenhorst [21,22]). The authors intend to pursue the analysis of the model with state-dependent maintenance times in this problem setting in the future.

Acknowledgements

The authors wish to thank the DE, Izak Dueynas and the anonymous referee for their valuable comments and Mr. A. Harriri of Lam Research Co., Fremont, CA for suggesting the problem. Funding for the second author was provided through a research grant from the Program for Engineering and Manufacturing Management (PEMM), University of Washington, Seattle, WA and the Burlington Northern Foundation.

References

[1] McCall, J.J. (1965) Maintenance policies for stochastically failing equipment: a survey. *Management Science*, **11**, 493–524.
[2] Pierskalla, W.P. and Voelker, J.A. (1976) A survey of maintenance models: the control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, **23**, 353–388.
[3] Valdez-Flores, C. and Feldman, R.M. (1989) A survey of preventive maintenance models for stochastically deteriorating single unit systems. *Naval Research Logistics*, **36**, 419–446.
[4] Cho, D.I. and Parlar, M. (1991) A survey of maintenance models for multi-unit systems. *European Journals of Operational Research*, **51**, 1–23.
[5] Barlow, R.E. and Proschan, F. (1963) *Mathematical Theory of Reliability*, Wiley, New York, NY.
[6] Cox, D.R. (1962) *Renewal Theory*, Methuen, London.
[7] Kolesar, P. (1966) Minimum cost replacement under Markovian deterioration. *Management Science*, **12**, 694–706.
[8] Derman, C. (1970) *Finite State Markovian Decision Processes*, Academic Press, New York, NY.
[9] Kao, E. (1973) Optimal replacement rules. *Operations Research*, **21**, 1231–1249.
[10] Ozekici, S. (1989) Optimal periodic replacement of multi-component reliability systems. *Operations Research*, **36**, 542–552.
[11] Derman, C. (1963) On optimal replacement rules when changes of state are Markovian, in *Mathematical Optimization Techniques* Bellman, R. (ed), University of California Press, Berkeley, CA.
[12] Jorgensen, D., McCall, J. and Radner, R. (1967) *Optimal Replacement Policies*, Rand McNally, Chicago, IL.
[13] Gibra, T.N. (1967) Optimal control processes subject to a linear trend. *The Journal of Industrial Engineering*, **18**, 35–41.
[14] Drezner, Z. and Wesolowsky, G.O. (1989) Optimal control of a linear trend process with quadratic loss. *IIE Transactions*, **21**, 66–72.
[15] Jeang, A. and Yang, K. (1992) Optimal tool replacement with non-decreasing tool wear. *International Journal of Production Research*, **30**, 299–314.
[16] Zhou, C., Chandra, J. and Wysk, R. (1990) Optimal cutting tool replacement based on tool wear status. *International Journal of Production Research*, **28**, 1357–1367.
[17] Sivazlian, B.D. and Mahoney, J.F. (1978) Group replacement of a multi-component system which is subject to deterioration only. *Advances in Applied Probability*, **10**, 867–885.
[18] Brosh, I., Shlifer, E. and Zeira, Y. (1975) Optimal maintenance policy for a fleet of vehicles. *Management Science*, **22**, 401–410.
[19] Berg, M. and Epstein, B. (1979) A note on a modified block replacement policy for units with increasing marginal running costs. *Naval Research Logistics Quarterly*, **23**, 15–24.
[20] Love, C.E., Rodger, A. and Blazenko, G. (1982) Repair limit policies for vehicle replacement. *INFOR*, **20**, 226–236.
[21] van der duyn Schouten, F.A. and Vanneste, S.G. (1993) Two simple control policies for a multicomponent maintenance system. *Operations Research*, **41**, 1125–1136.
[22] van der duyn Schouten, F.A. and Wartenhorst, P. (1993) A two-machine repair model with variable repair rate. *Naval Research Logistics*, **40**, 495–523.
[23] Gurler, U. and Kaya, A. (1998) A policy for the maintenance of a multistate multicomponent series system, Technical Report IEOR-9812, Department of Industrial Engineering, Bilkent University.
[24] Noh, J., Kang, S. and Feldman, R.M. (1992) A control policy for replacement involving two machines and one repairman. *European Journal of Operational Research*, **56**, 413–422.
[25] Ghandi, S. (1983) *VLSI Fabrication Principles*, Wiley, New York, NY.
[26] Cox, D.R. (1955) The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proceedings of the Cambridge Society*, **51**, 433–441.
[27] Gnedenko, B. and Kovalenko, I.N. (1968) *Introduction to Queueing Theory*, Israel Program for Scientific Translations, Jerusalem.

[28] Schmidt, C.P. and Nahmias, S. (1985) (S-1,S) policies for perishable inventory. *Management Science*, **31**, 719–728.
[29] Moinzadeh, K. (1989) Operating characteristics of the (S-1,S) inventory system with partial backorders and constant resupply times. *Management Science*, **35**, 472–477.

Appendix

Derivation of Equation (27)

First, we note that the expected throughput of the system can be written as:

$$E[TP] = \psi \sum_{n=1}^{M} \int_{x_n=0}^{T_n} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \sum_{i=1}^{n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

Now for $1 \leq n \leq M-1$ we have:

$$\psi \int_{x_n=0}^{T_n} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \sum_{i=1}^{n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$= \psi \sum_{i=1}^{n} \int_{x_n=0}^{T_{n+1}} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$+ \psi \sum_{i=1}^{n} \int_{x_n=T_{n+1}}^{T_n} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}. \quad \text{(A1)}$$

Substituting (25) in the first term in the RHS of (A1), we get:

$$\psi \sum_{i=1}^{n} \int_{x_n=0}^{T_{n+1}} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \sum_{i=1}^{n} \int_{x_n=0}^{T_{n+1}} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \theta(x_i) \mu^n \exp(-\mu T_{n+1}) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \sum_{i=1}^{n} \int_{x=0}^{T_{n+1}} \theta(x) \mu \frac{(\mu x)^{i-1}}{(i-1)!} \frac{[\mu(T_{n+1}-x)]^{n-i}}{(n-i)!} \exp(-\mu T_{n+1}) \mathrm{d}x$$

$$= \psi c_0 \mu \exp(-\mu T_{n+1}) \int_{x=0}^{T_{n+1}} \theta(x) \frac{(\mu T_{n+1})^{n-1}}{(n-1)!} \mathrm{d}x$$

$$= \psi c_0 \mu \int_{x=0}^{T_{n+1}} \theta(x) \mathrm{d}x\, f(n-1, \mu T_{n+1}). \quad \text{(A2)}$$

Also, in a similar fashion we can write the second term in the RHS of (A1) as:

$$\psi c_0 \sum_{i=1}^{n} \int_{x_n=0}^{T_{n+1}} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=x_{n-2}}^{x_n} \theta(x_i) \mu^n \exp(-\mu x_n) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \sum_{i=1}^{n-1} \int_{x_n=T_{n+1}}^{T_n} \int_{x_i=0}^{x_n} \theta(x_i) \mu^2 \frac{(\mu x_i)^{i-1}}{(i-1)!} \frac{[\mu(x_n - x_1)]^{n-i-1}}{(n-i-1)!}$$

$$\times \exp(-\mu x_n) \mathrm{d}x_i \mathrm{d}x_n + \psi c_0 \int_{x=T_{n+1}}^{T_n} \theta(x) \mu \frac{(\mu x)^{n-1}}{(n-1)!} \exp(-\mu x) \mathrm{d}x$$

$$= \psi c_0 \int_{y=T_{n+1}}^{T_n} \int_{x=0}^{y} \theta(x) \mu^2 f(n-2, \mu y) \mathrm{d}x\, \mathrm{d}y$$

$$+ \psi c_0 \mu \int_{y=T_{n+1}}^{T_n} \theta(x) f(n-1, \mu x) \mathrm{d}x. \quad \text{(A3)}$$

Changing the order of integration of the first term in the RHS of the above expression and noting that:

$$\int_0^t f(i, \mu x) \mathrm{d}x = \frac{1}{\mu} F(i-1, \mu t),$$

we get:

$$\psi \sum_{i=1}^{n} \int_{x_n=T_{n+1}}^{T_n} \int_{x_1=0}^{x_n} \cdots \int_{x_{n-1}=0}^{x_n} \theta(x_i) p_n(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \mu \left\{ \int_0^{T_n} \theta(x) \mathrm{d}x F(n-1, \mu T_n) - \int_0^{T_{n+1}} \theta(x) \mathrm{d}x F(n, \mu T_{n+1}) \right.$$

$$\left. - \int_{T_{n+1}}^{T_n} \theta(x) F(n-1, \mu x) \mathrm{d}x \right\}. \quad \text{(A4)}$$

For $n = M$, using the same approach, we get:

$$\psi \sum_{i=1}^{M} \int_{x_n=0}^{T_M} \int_{x_1=0}^{x_M} \cdots \int_{x_{M-1}=0}^{x_M} \theta(x_i) p_M(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \sum_{i=1}^{M} \int_{x_M=0}^{T_M} \int_{x_1=0}^{x_M} \cdots \int_{x_{M-1}=x_{M-2}}^{x_M} \theta(x_i) \mu^M \exp(-\mu T_M) \mathrm{d}\mathbf{x}$$

$$= \psi c_0 \mu \int_{x=0}^{T_M} \theta(x) \mathrm{d}x f(M-1, \mu T_M). \quad \text{(A5)}$$

Combining (A2) and (A4) and summing for $n = 1, \ldots, M-1$ together with (A5), after some algebraic manipulation, (27) is derived.

Biographies

Emre Berk is an Assistant Professor at the School of Business, Bilkent University, Turkey. Dr. Berk received his B.Sc. in Mechanical Engineering at Bogazici University, Turkey and his M.Sc. in Mechanical Engineering at Washington State University. He received his M.B.A. and Ph.D. in Operations Management at the University of Washington. His research interests include supply chain management and stochastic models in health care applications. He is a member of INFORMS.

Kamran Moinzadeh is the Burligton Northern/Burlington Resources Professor of Manufacturing Management, Professor of Management Science and the co-director of the Program in Engineering and Manufacturing Management (PEMM) at the University of Washington. He received his B.A. in Computer Science from the University of California, San Diego, his M.S. in Operations Research and Ph.D. in Industrial Engineering from Stanford University. Dr. Moinzadeh's publications have appeared in *Management Science*, *Operations Research*, *Naval Research Logistics*, *EJOR* and *IIE Transactions*. His research interests include Production/Operations management, Inventory management, Multi-echelon distribution systems, Supply Chain management and Quality management. Dr. Moinzadeh is a member of INFORMS and IIE. He is currently the department editor of *IIE Transactions*, and is the Associate Editor for *Operations Research* and *Management Science*. Dr. Moinzadeh has been a consultant to Microsoft Corp., Starbucks Coffee Inc. and Boeing Computer Services.