Optimal Batch Sizing and Repair Strategies for Operations with Repairable Jobs

Kut C. So • Christopher S. Tang
Graduate School of Management, University of California, Irvine, California 92717
Anderson Graduate School of Management, University of California, Los Angeles, California 90024

This paper presents a model of a bottleneck facility that performs two distinct types of operations: "regular" and "repair." Both switch-over time and cost are incurred when the facility switches from performing one type of operation to a different type. Upon the completion of a batch of jobs in the regular mode, each batch is subjected to a test, where the entire batch (of jobs) will be classified accordingly as either nondefective, repairable, or nonrepairable. A nondefective batch continues its process downstream, a nonrepairable batch is scrapped, and a repairable batch can be cycled back to the bottleneck facility for repair. The objective of this paper is to determine the optimal repair policy for the bottleneck facility so that the long run average operating profit is maximized. We first characterize the optimal repair policy by showing that the optimal repair policy must take one of the two forms: a "repair-none" policy under which all repairable batches are scrapped, or a "repair-all" policy under which all repairable batches are repaired. We then develop optimality conditions for the repair-none policy and the repair-all policy. When the repair-all policy is optimal, we further show that there exists an optimal "threshold" operating policy that can be described as follows: upon completion of a regular batch, switch over to the repair mode only if the number of available repairable batches exceeds a certain threshold value. We also evaluate the impact of batch sizes, yield, and switch-over cost on the optimal operating policy.
(*Bottleneck with Yield Loss; Bottleneck with Repairable Jobs; Optimal Repair Policy; Optimal Processing Policy; Dynamic Policy; Static Policy; Optimal Batch Size*)

1. Introduction

Since the mid-80s, most manufacturers have launched various total quality management programs to improve the quality of the manufacturing processes and products. Consequently, quality is no longer a competitive advantage; instead, it is now the prerequisite for many manufacturers to stay in business in the 1990s. At the same time, intense global competition continues to force many high-tech industries to reduce the profit margin. For instance, semiconductor and computer manufacturers are under continuing pressure to lower price while improving product performance. Consequently, while it is important to achieve certain quality standards, manufacturers need to continuously improve productivity so as to lower their manufacturing costs.

Motivated by our experience at one integrated circuit manufacturing facility, we study the productivity issue of a manufacturing system that faces variable processing time and uncertain yield. These characteristics are common in almost all manufacturing environments, especially in high-tech industries such as the semiconductor industry. As an initial step, we shall focus our analysis on the bottleneck stage of a manufacturing system. This bottleneck stage has three distinctive features.

First, each completed batch (of jobs) at the bottleneck stage is classified as nondefective, repairable, or nonrepairable. A nondefective batch continues its process downstream, a nonrepairable batch is scrapped, and a repairable batch can be cycled back to the bottleneck

0025-1909/95/4105/0894$01.25

facility for repair. As an example, consider the chemical tank that performs the etching process in a wafer fabrication facility. In order to etch away the portion of the deposited layer that is not protected by the photoresist during the photolithography operation, a batch (of wafers) is submerged (simultaneously) into a tank of chemical solution during the etching process. The process is very sensitive to several parameters including the temperature and the composition of the chemicals. A uncontrollable change in any of these parameters can disrupt the stability of the process. As a result, each completed batch (of wafers) can belong to one of the these three categories: nondefective, repairable, or nonrepairable.

Second, each repaired batch is considered to be "substandard" and has a lower value than a nondefective batch. Here we offer two examples. The first example arises from the integrated-circuit industry in which most integrated circuit chips have built-in redundant circuits for extra reliability (cf., Gise and Blanchard 1986, and Tang and Demeester 1992). In this case, the repairable batches correspond to those chips whose redundant circuits are defective. These repairable chips have lower reliability and are sold at a lower price (cf., Bitran and Dasu 1992, and Bitran and Gilbert 1992). Another example arises from a clothing or linen (such as sheets and pillow cases) manufacturer in which repairable batches correspond to those "irregular" fabrics that can only be sold at a lower price.

The third feature of the bottleneck facility is that the repair operation is different from the regular operation due to the nature of the process. In our model we consider the general situation where both switch-over time and cost are incurred when the facility switches from performing one type of operation to a different type. The readers are referred to So and Tang (1995) for details regarding two examples that have significant switch-over time and cost, where the first example arises from the chemical tanks that perform the etching process in a wafer fabrication facility, and the second example occurs at the testers that perform test and repair operations in a circuit board assembly system.

These three distinctive features raise the following important operating issues. Since the repaired batch has a lower value and since switch-over time and cost are incurred for the bottleneck facility to perform the repair operation, the productivity of the system is affected by both the *repair policy* that can be specified by the *proportion* of the repairable batches to be repaired as well as the *operating policy* that specifies *when* the bottleneck facility should be switched to process the repairable jobs.

When uncertainty (such as random processing time and yield) is inherent to the bottleneck facility, it is critical to develop a good policy to manage the facility so as to maximize (or minimize) the operating profit (or cost). Gerchak and Parlar (1990), Lee (1992), and Porteus (1986) have developed various models that are based on the classical Economic Order Quantity (EOQ) model for determining the optimal lot size for different systems; see Erlenkotter (1990) for some interesting information regarding this classical model. Yano and Lee (1995) have provided an extensive review on other models that deal with the operating issues in a multistage facility.

Our model is different from previous work in the literature. First, most models in the literature are based on a given repair policy in which all repairable jobs are either scrapped or repaired. In our model we allow a proportion of repairable jobs to be scrapped, but show that the optimal repair policy must take one of the two forms: a "repair-none" policy under which all repairable batches are scrapped, or a "repair-all" policy under which all repairable batches are repaired. Our model also allows the value of a repaired job to be lower than that of a nondefective job. Furthermore, the production environment considered in our model is different from the environment commonly considered in the literature, where some jobs in a batch can be nondefective while the remaining jobs within the same batch can be defective. For example, in So and Tang (1995), the nondefective jobs are "separable" from the defective ones within the same batch, so that the nondefective jobs continue their process downstream while the defective jobs are cycled back for repair. Our model here considers the environment where, due to the nature of the process, *all* jobs within the same batch are either nondefective, repairable, or nonrepairable (e.g., the etching process discussed earlier), and/or that the multiple jobs within the same batch are nonseparable (e.g., multiple chips (jobs) on a single wafer (batch) in the fabrication process). In either case, each batch (of jobs) is considered as a single entity in our model.

This paper is organized as follows. In §2, we describe our model and show that the optimal repair policy must take one of the following two types: a "repair-none" policy under which all repairable batches are scrapped, or a "repair-all" policy under which all repairable batches are repaired. In §3, we first develop some optimality conditions for the repair-none policy and the repair-all policy. Then we evaluate the impact of yield, switch-over cost, and other system parameters on the optimal repair and operating policy. In §4, we deal with the important issue of batch size and show how to determine the optimal batch size. We further compare the optimal operating profit generated by the optimal dynamic operating policy to that of a simple static operating policy. Our analysis shows that the simple static operating policy performs well when the setup cost is small. Finally, in §5 we conclude the paper with a discussion on possible extensions of the model and a few remarks on some remaining issues.

2. Model Formulation

Consider a bottleneck facility that performs two distinct types of operations: "regular" and "repair." We assume that there are always regular batches (of jobs) available for processing. This assumption is reasonable for a well managed bottleneck facility that attempts to maximize the productivity of the system, and our model should closely approximate a bottleneck facility that may starve occasionally. Let p_1, p_2, and p_3 denote the respective probabilities that a finished batch is nondefective, repairable, and nonrepairable, with $p_1 + p_2 + p_3 = 1$. Each nondefective batch continues its process downstream, each non-repairable batch is scrapped, and each repairable batch can be cycled back to the bottleneck facility for repair. We assume that repair is perfect, i.e., no defective items remain after repair, and only a degradation of quality is resulted. Therefore, the yield loss of the operation is equal to p_3 if all repairable batches are repaired, and is equal to $p_2 + p_3$ if all repairable batches are scrapped.

The processing times for regular batches are i.i.d. random variables with arbitrary distribution function $F(\cdot)$ and mean u. The times for repairing those repairable batches are i.i.d. random variables with arbitrary distribution function $G(\cdot)$ with mean v. (When the facility may need several trials before a repairable batch is eventually repaired, $G(\cdot)$ represents the total amount of time for repair.) A switch-over time is incurred when the facility switches from performing one type of operation to a different type. The switch-over times from regular mode to repair mode, and vice versa, are independent random variables with finite means t_1 and t_2, respectively.

The reward structure of each finished batch is given as follows. Each nondefective batch has a value r_1. Each nonrepairable batch is scrapped with a salvage value of r_3. The repairable batches can be cycled back to the facility for repair. However, the performance and/or quality of a repaired batch could be inferior to that of a nondefective batch, so a repaired batch has a value of r_2 that is lower than that of a nondefective batch (i.e., $r_2 \leq r_1$). We shall assume that $r_1 \geq r_2 > r_3 \geq 0$. Without loss of generality, the average cost for repairing a batch is assumed to be zero, as we can simply subtract the repair cost from r_2. Moreover, since $r_2 \leq r_1$ and since switch-over time and cost are incurred for the facility to perform the repair operation, the manufacturer might decide not to repair a repairable batch, but simply scrap a repairable batch. In that case, a salvage value r_3, which is the same as that for a non-repairable batch, is obtained. In addition, there is a (combined) fixed charge s for switching the facility from regular mode to repair mode (and back), a holding cost rate h for each repairable batch waiting to be repaired, and a processing cost (material plus operating cost) c_1 for processing a regular batch.

Since $r_2 \leq r_1$ and since switch-over time and cost are incurred for the facility to perform the repair operation, it may not be cost-effective to repair all of the repairable batches. Our objective is to determine the optimal *repair policy* that maximizes the long run average operating profit. Essentially, the repair policy specifies how many repairable batches should be repaired or scrapped. The general case is very difficult to analyze since the repair/scrap decision can depend on the number of existing repairable batches currently in the facility. The decision is further complicated by the fact that one can keep only a portion of repairable jobs generated in each batch and that one can scrap some or all existing repairable batches in the facility at any instance. A general model incorporating such "dynamic" scrap and repair deci-

sions is apparently intractable. Therefore, as an initial step to analyze the impact of different repair policies and operating policies, we restrict our attention to repair policies that can be specified by the *proportion* of the repairable batches to be repaired. This proportion is specified prior to the actual operation of the bottleneck facility, and is independent of the number of existing repairable batches currently in the facility. Thus, we refer to this class of policies as "static" repair policies.

In our model each static repair policy can be specified by the parameter α, $0 \le \alpha \le 1$. Essentially, α specifies the proportion of the repairable batches to be repaired and represents the probability that a repairable batch will be repaired. When $\alpha = 0$, we scrap all of the repairable batches and call this repair policy the *repair-none* policy (or scrap-all policy). When $\alpha = 1$, we repair all the repairable batches, and we call this repair policy the *repair-all* policy. We refer to the class of policies with $0 < \alpha < 1$ as the *partially-repair* policies. For a static repair policy α, the "effective" probability of a finished batch to be repaired is equal to αp_2, and the effective probability of a finished batch to be scrapped is equal to

$$(1 - \alpha)p_2 + p_3 = 1 - p_1 - \alpha p_2.$$

Let g^α be the average operating profit under the repair policy α. The optimal operating profit, denoted by g^*, can be expressed as *Problem* (P):

$$g^* = \sup_{0 \le \alpha \le 1} \{g^\alpha\}. \tag{P}$$

In addition, let α^* be the optimal solution to problem (P) that corresponds to the optimal (static) repair policy.

2.1. Average Operating Profit

For any given static repair policy α (specified prior to the actual operation), the average operating profit generally depends on the actual *operating policy* that selects the operation mode (regular mode or repair mode) of the bottleneck facility according to the state of the system, i.e., the number of repairable batches waiting to be repaired. Essentially, the operating policy selects the next operation mode for the facility when the bottleneck facility completes a regular batch or a repairable batch. In order to characterize the optimal repair policy α^* and the optimal operating profit g^*, we need to evaluate the average operating profit for any given repair policy α.

First, let us evaluate the average operating profit under the repair-none policy (or the scrap-all policy). In this case, $\alpha = 0$. Under the scrap-all policy, the probability of a finished batch to be nondefective is equal to p_1. The effective probability of a finished batch to be scrapped is equal to $p_2 + p_3$. Therefore, the expected net reward for processing each finished batch is equal to $(p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1)$, and the average operating profit under the scrap-all policy, denoted by g^0, can be written as:

$$g^0 = \frac{p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1}{u}. \tag{1}$$

Next consider $0 < \alpha \le 1$. When a regular batch is completed, the facility can: (1) continue to process another regular batch, or (2) switch over to repair the repairable batches. When a repairable batch is completed, the facility can: (1) continue to repair another repairable batch, or (2) switch over to process a new regular batch. As it turns out, we were able to extend the analysis developed by So and Tang (1995) to show that there exists an optimal operating policy (for any given static repair policy α) that is of a threshold type: once the bottleneck is set up to repair, repair all repairable batches before switching over to process regular batches; and upon completion of a regular batch, switch over to repair the repairable batches only if the number of repairable batches exceeds a certain threshold value. Therefore, we can restrict our attention to this class of operating policies in which each threshold policy is specified by the threshold value n, where $n \ge 1$. Furthermore, we can establish the following result:

PROPOSITION 1. *For any fixed repair policy α ($0 < \alpha \le 1$), there exists an increasing sequence $\{R_n^\alpha\}_{n=0}^\infty$ with $R_0^\alpha = -\infty$, $R_1^\alpha > -\infty$, $R_n^\alpha < \infty$, and $\lim_{n\to\infty} R_n^\alpha = \infty$ such that the threshold policy n is the optimal operating policy for $s \in [R_{n-1}^\alpha, R_n^\alpha]$. Moreover, the 'break-point' R_n^α is given by*

$$R_n^\alpha = (t_1 + t_2)\left\{hn + \frac{h\alpha p_2(v + t_1)}{u + \alpha p_2 v} - \frac{p_1 r_1 + \alpha p_2 r_2 + (1 - p_1 - \alpha p_2) r_3 - c_1}{u + \alpha p_2 v}\right\} + \frac{h(u + \alpha p_2 v)n(n+1)}{2\alpha p_2}, \tag{2}$$

and the average operating profit under the threshold policy n is given by

$$g_n^\alpha = \frac{1}{t_1 + t_2 + n(u/(\alpha p_2) + v)} \left\{ \frac{n(p_1 r_1 + \alpha p_2 r_2 + (1 - p_1 - \alpha p_2) r_3 - c_1)}{\alpha p_2} - s - h\left(\frac{un(n-1)}{2\alpha p_2} + nt_1 + \frac{vn(n+1)}{2} \right) \right\}. \quad (3)$$

PROOF. See the Appendix.

The average operating profit, g_n^α, can be interpreted as follows. To simplify our exposition, we shall focus on the repair-all policy when $\alpha = 1$. In this case, g_n^1 can be rewritten as:

$$g_n^1 = \frac{1}{n(u/p_2 + v) + t_1 + t_2} \left\{ \frac{n(p_1 r_1 + p_2 r_2 + p_3 r_3 - c_1)}{p_2} - s - h\left(\frac{un(n-1)}{2p_2} + nt_1 + \frac{vn(n+1)}{2} \right) \right\}. \quad (4)$$

The denominator of the first term denotes the expected cycle time between successive epochs when the facility is switched from repair mode back to regular mode, where the expression corresponds to the time it takes to process the expected number of regular batches in one cycle (n/p_2), the time to repair n repairable batches in one cycle, and the total switch-over time within one cycle. The terms inside the brace represent the total expected net reward in one cycle, where $(p_1 r_1 + p_2 r_2 + p_3 r_3 - c_1)$ denotes the expected net reward for processing one regular batch, s denotes the switch-over cost, and $hun(n-1)/2p_2$, hnt_1, and $hvn(n+1)/2$ denote the expected total holding costs during the time periods when the facility is, respectively, processing (n/p_2) regular batches until n repairable batches are cumulated, switched to repair mode, and repairing the n repairable batches.

With the average operating profits g^0 and g_n^α given in (1) and (3), problem (P) can be expressed as:

$$g^* = \max\left(g^0, \max_{0<\alpha\le 1} \max_{n\ge 1} \{g_n^\alpha\} \right).$$

For compactness, let us rewrite (3) as:

$$g_n^\alpha = \frac{a(n) + b(n)\alpha}{c(n) + d(n)\alpha}, \quad \text{where} \quad (5)$$

$$a(n) = n(p_1 r_1 + (1 - p_1) r_3 - c_1 - hu(n-1)/2),$$

$$b(n) = p_2[nr_2 - nr_3 - s - hnt_1 - hvn(n+1)/2],$$

$$c(n) = nu, \quad \text{and} \quad d(n) = p_2(t_1 + t_2 + nv).$$

Consider $\alpha = 0$ in (5) and define

$$g_n^0 = \frac{a(n)}{c(n)} = \frac{p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1 - hu(n-1)/2}{u}.$$

Observe that g_n^0 is maximized when $n = 1$ and that $g_1^0 = g^0$. Hence, we can express g^0 as $g^0 = \max_{n\ge 1}\{g_n^0\}$ and rewrite problem (P) compactly as *Problem* (Q):

$$g^* = \max_{0\le\alpha\le 1}\left\{ \max_{n\ge 1}\{g_n^\alpha\} \right\} = \max_{n\ge 1}\left\{ \max_{0\le\alpha\le 1}\{g_n^\alpha\} \right\}, \quad (Q)$$

where g_n^α is given in (5). The optimal solution to Problem (Q) corresponds to the (joint) optimal repair and operating policy, which will be denoted by (α^*, n^*).

2.2. Characteristics of the Optimal Repair Policy

For any given n, problem (Q) can be reduced to the following subproblem: $\max_{0\le\alpha\le 1}\{g_n^\alpha\}$. It follows from (5) that g_n^α is decreasing in α if $a(n)d(n) > b(n)c(n)$ and g_n^α is increasing in α if $a(n)d(n) < b(n)c(n)$ for all $0 \le \alpha \le 1$. Therefore, for any given n, g_n^α is monotone in α and the maximum of g_n^α occurs at either $\alpha = 0$ or $\alpha = 1$, depending on the sign of $(a(n)d(n) - b(n)c(n))$. Thus, we have the following result.

PROPOSITION 2. *There exists an optimal repair policy α^* with $\alpha^* = 0$ or $\alpha^* = 1$. In other words, either the repair-none or the repair-all policy is optimal.*

Proposition 2 implies that we can restrict our attention to two cases: $\alpha = 0$ and $\alpha = 1$. Thus, g^* is given by

$$g^* = \max_{n\ge 1}\{\max(g_n^0, g_n^1)\} = \max\left(g_1^0, \max_{n\ge 1}\{g_n^1\} \right),$$

where the second equality follows from the fact that g_n^0 is maximized when $n = 1$.

For notational convenience, we shall use the following notation for the remainder of this paper. The scrap-all policy is denoted by the (0, 0)-policy, and the repair-all policy with operating threshold policy n is denoted by the $(1, n)$-policy. Let $g^0 = g_1^0$ represent the average operating profit under the (0, 0)-policy, and let g_n^1 be the average operating profit under the $(1, n)$-policy. Also, let $g^1 = g_{n*}^1 = \max_{n\geq 1}\{g_n^1\}$, where n^* corresponds to the optimal threshold policy among all repair-all policies. The optimal average operating profit g^* can now be rewritten as:

$$g^* = \max(g^0, g^1). \tag{6}$$

Hence, in order to determine the optimal repair policy, we only need to compare g^0 and g^1.

3. Characteristics of the Optimal Repair and Operating Policy

In the previous section, we show that the optimal repair policy can be determined by simply comparing g^0 and $g^1 = \max_{n\geq 1}\{g_n^1\}$. Substituting $\alpha = 1$ into (2), we obtain

$$R_n^1 = (t_1 + t_2)\left\{hn + \frac{hp_2(v + t_1)}{u + p_2v} - \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1}{u + p_2v}\right\} + \frac{h(u + p_2v)n(n + 1)}{2p_2}. \tag{7}$$

Using Proposition 1, we can use the following simple algorithm to find the optimal repair and operating policy (α^*, n^*):

Algorithm

1. Find the smallest positive n such that $s \leq R_n^1$. Denote this n by n^*.
2. Then, $g^1 = g_{n*}^1$, where g_{n*}^1 can be evaluated by (4).
3. Compute g^0 from (1) and compare with g^1. If $g^0 \geq g^1$, then the scrap-all policy (or the (0, 0)-policy) is optimal. Otherwise, the $(1, n^*)$-policy is optimal.

We next provide some sufficient conditions for the optimality of the (0, 0)-policy and the (1, 1)-policy (i.e., the repair-all policy under which all repairable batches are repaired immediately). To this end, we use the notation $g_n^1(s)$ to express g_n^1 as a function of s. It is easy to see from (4) that $g_n^1(s)$ is linear and strictly decreasing in s for all n. Also, observe from Proposition 1 that $g^1(s) = \max_{n\geq 1}\{g_n^1(s)\}$ is continuous, piecewise linear, strictly decreasing and convex in s, with breakpoints R_n^1 given in (7). Furthermore, Proposition 1 implies that both $(1, n)$-policy and $(1, n + 1)$-policy are optimal operating policies at $s = R_n^1$. Hence, $g^1 = g_n^1(R_n^1) = g_{n+1}^1(R_n^1)$. On the other hand, observe from (1) that g^0 is independent of s. These observations and (6) imply that the optimal average operating profit g^* is also continuous, piecewise linear, decreasing and convex in s. Figure 1 depicts the structure of g^0, g^1, and g^* as a function of s.

It is intuitive from Figure 1 that either the (0, 0)-policy or the (1, 1)-policy is optimal for all s if $g^0 \geq g_1^1(R_1^1) = g_2^1(R_1^1)$, where

$$R_1^1 = (t_1 + t_2)\left(h + \frac{hp_2(v + t_1)}{u + p_2v} - \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1}{u + p_2v}\right) + \frac{h(u + p_2v)}{p_2}. \tag{8}$$

Substituting $s = R_1^1$ into (4) and after simplification, we obtain

$$g_1^1(R_1^1) = \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - h.$$

Thus, $g^0 \geq g_1^1(R_1^1)$ is equivalent to

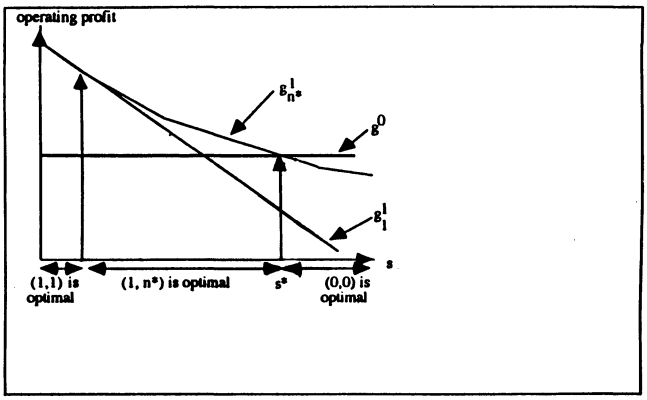
$$\frac{p_1r_1 + p_2r_3 + p_3r_3 - c_1}{u} \geq \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - h,$$

which, after simplification, is equivalent to

$$\frac{p_1r_1 + p_2r_3 + p_3r_3 - c_1}{u} \geq \frac{r_2 - r_3}{v} - h\left(2 + \frac{u}{p_2v} + \frac{t_1}{v}\right). \tag{9}$$

The next result provides sufficient conditions for the scrap-all policy and the repair-all policy to be optimal.

Figure 1 Average Operating Profits



PROPOSITION 3.

(*a*) *Consider the case when condition* (9) *holds. Then, either the* (0, 0)*-policy or the* (1, 1)*-policy is optimal. Specifically, the* (1, 1)*-policy is optimal if* $s \le s^1$, *where*

$$s^1 = (r_2 - r_3) - (p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1) \times \left(\frac{t_1 + t_2 + v}{u}\right) - h(t_1 + v).$$

Otherwise, the (0, 0)*-policy is optimal.*

(b) *Consider the case when condition* (9) *does not hold. Then:*

(i) *The* (1, 1)*-policy is optimal when* $s \le R_1^1$.

(ii) *There exists some constant* $s^* > R_1^1$ *such that the* (0, 0)*-policy is optimal when* $s \ge s^*$.

(iii) *For any fixed* s *with* $R_1^1 < s \le s^*$, *there exists some* $n^* > 1$ *such that the* (1, n^*)*-policy is optimal. Moreover,* n^* *is increasing in* s.

(iv) *An upper bound of* s^* *is given by*

$$s^u = \frac{h}{2(u/p_2 + v)}\left\{\left(\frac{r_2 - r_3}{h} + t_2 - v - \frac{v(p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1)}{uh}\right)^2 - (t_1 + t_2)^2\right\} + \frac{t_1 + t_2}{u/p_2 + v}\left\{h(t_1 + v) - \frac{p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1}{p_2}\right\} - \frac{(t_1 + t_2)h}{2}.$$

Therefore, if $s \ge s^u$, *the* (0, 0)*-policy is optimal.*

PROOF. See the Appendix.

Proposition 3 states that the (1, 1)-policy is optimal only when the switch-over cost s is less than or equal to a certain cutoff value. This value is given by either s^1 or R_1^1, depending on whether condition (9) holds or not, i.e., part (a) or (b), respectively. Similarly, the (0, 0)-policy is optimal only when the switch-over cost s is larger than or equal to a certain cutoff. The cutoff value is given by either s^1 or s^* (where s^* is bounded above by s^u), depending on whether condition (9) holds or not. This implies that when the switch-over cost is sufficiently small, e.g., $s \le \min(s^1, R_1^1)$, it is optimal to always repair all repairable batches immediately. On the other hand, when the switch-over cost is too large, e.g., $s \ge \max(s^1, s^u)$, it is optimal to always scrap all repairable batches. We summarize this observation in the next corollary, which follows directly from Proposition 3.

COROLLARY 1. *The* (1, 1)*-policy is optimal if* $s \le \min(s^1, R_1^1)$. *On the other hand, the* (0, 0)*-policy is optimal if* $s \ge \max(s^1, s^u)$.

In situations where the required switch-over operations when the facility is switched between the process mode and repair mode can be mostly performed off-line, i.e., $t_1 = t_2 = 0$, the optimality conditions displayed in Proposition 3 and Corollary 1 can be further simplified. The next result follows immediately from Proposition 3 and Corollary 1 for this special case.

COROLLARY 2.

(*a*) *Suppose that* $t_1 = t_2 = 0$. *Then, the* (1, 1)*-policy is optimal if*

$$s \le \min\left\{\frac{h(u + p_2 v)}{p_2}, (r_2 - r_3) - vh - \frac{(p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1)v}{u}\right\};$$

and the (0, 0)*-policy is optimal if*

$$s \ge \max\left\{h\left(\frac{u}{p_2} + v\right), \frac{h}{2(u/p_2 + v)}\left(\frac{r_2 - r_3}{h} - v - \frac{v(p_1 r_1 + p_2 r_3 + p_3 r_3 - c_1)}{uh}\right)^2\right\}.$$

(*b*) *Furthermore, suppose that* $t_1 = t_2 = s = 0$. *Then, the* (0, 0)*-policy* (*i.e., the scrap-all policy*) *is optimal if*

$$\frac{p_1r_1 + p_2r_3 + p_3r_3 - c_1}{u} \geq \frac{(r_2 - r_3) - hv}{v}.$$

Otherwise, the (1, 1)*-policy is optimal.*

When $t_1 = t_2 = 0$, one can further use Proposition 1 to characterize the behavior of the optimal n^* and g^1_{n*} with respect to the various cost parameters. For instance, it can be easily shown that n^* is increasing in r_1, r_2, r_3, and s, but is decreasing in c_1 and h. Similarly, it can be shown that both g^1_{n*} and g^* are increasing in r_1, r_2 and r_3, but are decreasing in c_1, h and s. Furthermore, since g^0 is independent of h and s while g^1_{n*} decreases in h and s, we can obtain the intuitive result that we should be more inclined to scrap all repairable batches when the holding cost and switch-over cost are high.

When the switch-over cost s is also zero, Corollary 2(b) gives the intuitive result that the optimal decision will depend on the difference between $(r_2 - r_3)$, the extra reward earned by repairing a repairable batch rather than scrapping it, and the cost of holding the repairable batch. Specifically, it is optimal to scrap all repairable batches only when the average operating reward under the scrap-all policy is higher than the net reward rate of gaining the extra reward $(r_2 - r_3)$ and paying a holding cost of hv during the repair period for each repairable batch.

4. The Issue of Batch Size

As discussed earlier, our model applies to the production environment in which multiple jobs within the same batch will be either all nondefective, repairable, or nonrepairable. Alternatively, our model also applies to the situation where all jobs in each batch are nonseparable throughout the process. In either case, we treat each batch (of jobs) as one individual unit in our model. In this section we address the important issue of selecting the optimal batch size.

To simplify the mathematical analysis, we shall restrict our attention to the case in which no switch-over time is incurred when the facility switches from one operating mode to a different operating mode, i.e., $t_1 = t_2 = 0$. Consider that there are K jobs in each batch, and the rewards for each nondefective, repaired, and scrapped job are given by ρ_1, ρ_2, and ρ_3, respectively. To process each batch (of K jobs), the processing cost consists of a set-up cost τ and a unit cost χ. The holding cost for each repairable job is denoted by ν. Using the notation in our model, the reward and cost structure for each batch that consists of K jobs is therefore given by $r_1 = \rho_1 K$, $r_2 = \rho_2 K$, $r_3 = \rho_3 K$, $h = \nu K$, $c_1 = \tau + \chi K$. Also assume that the expected processing time and repair times for each job are given by μ and ϑ, respectively, so that the corresponding expected processing and repair times for each batch (of K jobs) are given by $u = \mu K$ and $v = \vartheta K$.

Let $\beta = p_1\rho_1 + p_2\rho_3 + p_3\rho_3 - \chi$, and let $g^0(K)$ be the average operating profit under the scrap-all policy; i.e., the (0, 0) policy. Substituting r_1, r_2, r_3, c_1, and u into (1), we obtain

$$g^0(K) = \frac{\beta}{\mu} - \frac{\tau}{\mu K}. \tag{10}$$

Notice that $g^0(K)$ is increasing in the batch size K and that as the batch size K increases, the set-up cost per unit time, $\tau/\mu K$, decreases. Similarly, let $g^1_n(K)$ be the average operating profit under the (1, n) policy. Substituting the parameters into (4), we obtain

$$g^1_n(K) = z - aK - b/K - cKn - d/(Kn), \tag{11}$$

where

$$z = \frac{\beta}{\mu + p_2\vartheta}, \tag{12}$$

$$a = \frac{\nu(-\mu + p_2\vartheta)}{2(\mu + p_2\vartheta)}, \tag{13}$$

$$b = \frac{\tau}{(\mu + p_2\vartheta)}, \tag{14}$$

$$c = \frac{\nu}{2}, \tag{15}$$

$$d = \frac{sp_2}{(\mu + p_2\vartheta)}. \tag{16}$$

Notice that the terms b, c, d, and z are positive while the term a can be positive or negative, depending on the sign of $(-\mu + p_2\vartheta)$.

For any given batch size K, one can substitute the corresponding parameters into our model and use the results in § 3 to determine the optimal repair and operating policy associated with K. In particular, the scrap-all policy is optimal if

$$g^0(K) \geq \max_{n\geq 1}\{g_n^1(K)\},$$

where $g^0(K)$ is given in (10) and $g_n^1(K)$ in (11). We next provide further analysis to gain more managerial insights regarding the impact of batch size on the optimal repair and operating policy.

4.1. Characteristics of Optimal Batch Size and Optimal Threshold Policy

Our goal is to provide useful managerial insight via the analysis of our model so that we can better understand the impact of batch size. Specifically, we shall characterize the optimal batch size and establish the relationship between the optimal batch size K^* and the optimal threshold policy n^*. Since $g^0(K)$ is given by the simple equation (10), we restrict our attention to the repair-all policy.

We consider the batch size K as a decision variable and analyze the impact of K on the repair-all policy. To simplify the analysis, we shall treat the variables K and n as positive continuous variables for the remainder of this section. We also extend the definition of the average operating profit function $g_n^1(K)$ given by (11) for continuous K and n. Of course, the corresponding optimal K^* and n^* will somehow need to be "rounded" to obtain an integer solution. However, our empirical study suggests that the results obtained in our analysis remain fairly accurate even after the rounding. Hence, the insights provided in our analysis remain valid after the rounding.

PROPOSITION 4. *(a) Suppose that $a \leq 0$. Then,*

$$n^* = 1,$$

$$K^* = \sqrt{\frac{b+d}{a+c}} = \sqrt{\frac{\tau + sp_2}{\nu p_2 \vartheta}}, \quad \text{and}$$

$$g_{n*}(K^*) = z - 2\sqrt{(b+d)(a+c)}$$

$$= \frac{\beta - 2\sqrt{(\tau + sp_2)\nu p_2 \vartheta}}{\mu + p_2\vartheta}.$$

(b) Suppose that $a > 0$. Then,

$$n^* = \sqrt{\frac{ad}{bc}} = \sqrt{\frac{sp_2(p_2\vartheta - \mu)}{\tau(\mu + p_2\vartheta)}},$$

$$K^* = \sqrt{\frac{b}{a}} = \sqrt{\frac{2\tau}{\nu(p_2\vartheta - \mu)}}, \quad \text{and}$$

$$g_{n*}(K^*) = z - 2\sqrt{ab} - 2\sqrt{cd}$$

$$= \frac{\beta - \sqrt{2\nu\tau(p_2\vartheta - \mu)} - \sqrt{2\nu s p_2(\mu + p_2\vartheta)}}{\mu + p_2\vartheta}.$$

PROOF. See the Appendix.

When $a \leq 0$, (13) implies that the repair time for each job is shorter than the time it takes to process $(1/p_2)$ jobs. In this case, Proposition 4(a) states that it is optimal to switch-over whenever there are repairable batches. Furthermore, the optimal batch size K^* has a similar structure to the EOQ formula and has the intuitive properties that K^* is increasing in τ (the set-up cost) and s (the switch-over cost), but is decreasing in ν (the unit holding cost rate for each repairable job).

When $a > 0$, (13) implies that the repair time for each job is longer than the time it takes to process $(1/p_2)$ jobs. In this case, Proposition 4(b) shows that the optimal batch size K^* is increasing in τ (the set-up cost), is decreasing in ν (the unit holding cost rate), but is independent of s (the switch-over cost), while the optimal threshold n^* is increasing in s (the switch-over cost), is decreasing in τ (the set-up cost), but is independent of ν (the unit holding cost rate). Therefore, this result suggests that the set-up cost τ has major impact on the optimal batch size as well as the threshold policy, while the switch-over cost s affects the threshold policy only when $a > 0$. Furthermore, observe that

$$n^*K^* = \sqrt{2sp_2/\nu(\mu + p_2\vartheta)},$$

which is independent of τ. Therefore, as the set-up cost τ increases, the optimal batch size K^* increases, but the optimal switch-over threshold n^* decreases accordingly.

4.2. Two Simple Classes of Policies

Here we provide some useful insights about the effectiveness of some simple policies. Specifically, we consider two special classes of policies. The first class of policies restricts $K = 1$ (each batch has one job), and we analyze the corresponding optimal threshold $n^*(1)$. The second class of policies restricts $n = 1$ (always repair

batches immediately), and we analyze the corresponding optimal batch size $K^*(1)$. For each class, we compare the average operating profits associated with $(1, n^*(1))$ and $(K^*(1), 1)$ to the optimal operating profit associated with (K^*, n^*).

We shall establish some simple conditions under which the best policy among these two simple classes, $K = 1$ (each batch has one job) and $n = 1$ (always repair batches immediately) would yield average operating profit that is closed to the optimal average operating profit. Parallel to Proposition 4, these conditions depend on whether $a \le 0$ or $a > 0$.

First, consider the case when $a \le 0$. In this case, the optimal K^*, n^*, and $g_{n*}(K^*)$ are given by Proposition 4(a). In particular, $n^* = 1$. Suppose that we restrict $K = 1$. From (12),

$$g_n(1) = z - a - b - cn - d/n.$$

It is easy to show that the corresponding optimal threshold policy

$$n^*(1) = \sqrt{d/c} \quad \text{with } g_{n*(1)}(1) = z - a - b - 2\sqrt{cd}.$$

Therefore, the best policy restricting to $K = 1$ will be effective when

$$g_{n*}(K^*) - g_{n*(1)}(1) = a + b + 2\sqrt{cd} - 2\sqrt{(b+d)(a+c)}$$
$$= \frac{\nu(p_2\vartheta - \mu)/2 + \tau + \sqrt{2\nu s p_2(\mu + p_2\vartheta)} - 2\sqrt{(\tau + sp_2)\nu p_2\vartheta}}{\mu + p_2\vartheta} \approx 0,$$

or alternatively, when

$$\frac{\nu(p_2\vartheta - \mu)}{2} + \tau + \sqrt{2\nu s p_2(\mu + p_2\vartheta)} \approx 2\sqrt{(\tau + sp_2)\nu p_2\vartheta}. \tag{17}$$

Next, consider the case when $a > 0$. In this case, the optimal K^*, n^*, and $g_{n*}(K^*)$ are given by Proposition 4(b). Let us first restrict $K = 1$. Then, $n^*(1) = \sqrt{d/c}$ and $g_{n*(1)}(1) = z - a - b - 2\sqrt{cd}$. Therefore, the best policy restricting $K = 1$ will be effective when

$$g_{n*}(K^*) - g_{n*(1)}(1) = a + b - 2\sqrt{ab} = (\sqrt{a} - \sqrt{b})^2$$
$$= \frac{(\sqrt{\nu(p_2\vartheta - \mu)/2} - \sqrt{\tau})^2}{\mu + p_2\vartheta} \approx 0.$$

In other words, one can restrict $K = 1$ when

$$\tau \approx \frac{\nu(p_2\vartheta - \mu)}{2}, \tag{18}$$

which is equivalent to $K^* \approx 1$ from Proposition 4(b). Hence, rounding K to one should be very close to the optimum when K^* is close to one.

Let us now restrict $n = 1$. Then,

$$g_1(K) = z - aK - b/K - cK - d/K.$$

In this case, it is straightforward to show that

$$K^*(1) = \sqrt{(b+d)/(a+c)} \quad \text{and}$$
$$g_1(K^*(1)) = z - 2\sqrt{(b+d)(a+c)}.$$

Therefore, the best policy restricting $n = 1$ will be effective when

$$g_{n*}(K^*) - g_1(K^*(1)) = 2\sqrt{(b+d)(a+c)} - 2\sqrt{ab} - 2\sqrt{cd}$$
$$= \frac{2\sqrt{(\tau + sp_2)\nu p_2\vartheta} - \sqrt{2\tau\nu(p_2\vartheta - \mu)} - \sqrt{2sp_2\nu(\mu + p_2\vartheta)}}{\mu + p_2\vartheta} \approx 0.$$

In other words, one can restrict $n = 1$ when

$$\sqrt{2(\tau + sp_2)p_2\vartheta} \approx \sqrt{\tau(p_2\vartheta - \mu)} + \sqrt{sp_2(\mu + p_2\vartheta)}$$
$$\Leftrightarrow 2(\tau + sp_2)p_2\vartheta \approx \tau(p_2\vartheta - \mu) + sp_2(\mu + p_2\vartheta) + 2\sqrt{\tau(p_2\vartheta - \mu)sp_2(\mu + p_2\vartheta)}$$
$$\Leftrightarrow \tau(\mu + p_2\vartheta) + sp_2(p_2\vartheta - \mu) \approx 2\sqrt{\tau(p_2\vartheta - \mu)sp_2(\mu + p_2\vartheta)}$$
$$\Leftrightarrow (\sqrt{\tau(\mu + p_2\vartheta)} - \sqrt{sp_2(p_2\vartheta - \mu)})^2 \approx 0$$
$$\Leftrightarrow \tau(\mu + p_2\vartheta) \approx sp_2(p_2\vartheta - \mu)$$
$$\Leftrightarrow \tau \approx \frac{sp_2(p_2\vartheta - \mu)}{\mu + p_2\vartheta}, \tag{19}$$

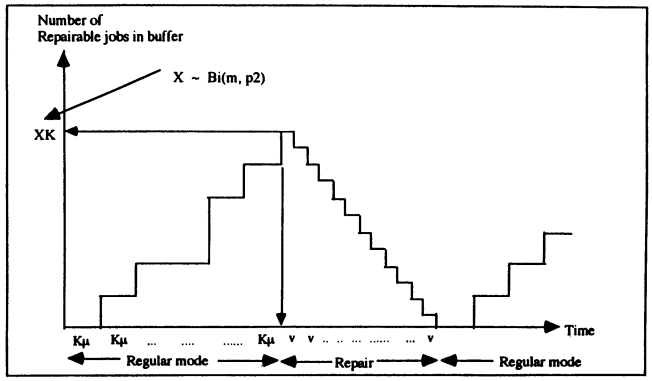
which is equivalent to $n^* \approx 1$ from Proposition 4(b). Again, this result implies that rounding n to one should be close to the optimum when n^* is close to one.

We illustrate some of the above results with the following example: $\rho_1 = 10$, $\rho_2 = 9$, $\rho_3 = 1$, $p_1 = 0.1$, $p_2 = 0.8$, $p_3 = 0.1$, $\mu = 1$, $\vartheta = 2.5$, $\chi = 0.05$, $\nu = 0.02$, $s = 10$, and $\tau = 1$. In this example, $a > 0$. From Proposition 4(b), $n^* = 1.63$ and $K^* = 10.0$. Restricting n^* and K^* to positive integers gives the optimal solution $(n^*, K^*) = (2, 8)$ with optimal average operating profit $g_2(8) = 2.355$. Since n^* is close to one, our previous analysis suggests that the policy that restricts $n = 1$ should be effective, with the corresponding $K^*(1) = 15$ and $g_1(15) = 2.350$. On the other hand, since K^* is much larger than one, we expect that the policy that restricts $K = 1$ would not be effective. Indeed, the corresponding $n^*(1) = 16$ with $g_{16}(1) = 2.087$.

4.3. Static Operating Policy

The optimal operating policy as specified by n^* corresponds to a *dynamic* operating policy in the sense that the switch-over decision depends on the number of repairable batches at the decision epochs. Implementing this dynamic operating policy may sometimes generate extra burden on the operator, as the operator needs to continuously monitor the state of the system to control the bottleneck facility accordingly. Hence, it is desirable to study the performance of some simple static operating policies and evaluate their effectiveness as compared to the optimal dynamic operating policy (K^*, n^*).

Figure 2 A Schematic View of a Static Operating Policy



Consider the following simple *static* operating policy that is much easier to implement: The system is reviewed *only* after processing every m batches (each batch has K jobs), and the facility is switched to repair all repairable batches if there exist repairable batches at that time; otherwise, the facility continues to process m new batches of jobs. Figure 2 depicts one complete cycle (processing m regular batches and repairing all available repairable batches) under this static operating policy. In each cycle, the facility processes m regular batches, and generates and repairs, on the average, mp_2 repairable batches. Therefore, the expected cycle time is equal to $m(K\mu) + mp_2(K\vartheta)$. Furthermore, the expect net profit generated in one cycle can be written as:

$$mK\beta - m\tau - s(1 - (1 - p_2)^m) - (\nu K)\left\{\frac{m(m-1)p_2}{2}(K\mu) + \frac{mp_2(mp_2 + 2 - p_2)}{2}(K\vartheta)\right\}.$$

The first term corresponds to expected revenue minus processing cost for m batches of jobs. The second term corresponds to the expected set-up cost. The third term represents the expected switch-over cost, where the probability of having at least one repairable batch after processing m regular batches is equal to $(1 - (1 - p_2)^m)$. The two expressions in the last term correspond to the expected holding cost incurred during the process mode and during the repair mode in one cycle, respectively, where the last expression is derived from the fact that

$$E(X(X+1)) = np(np + 2 - p)$$

for a binomial random variable X with parameters (n, p). Therefore, the operating profit per unit time under this static operating policy, denoted by $g_{\text{stat}}(K, m)$, is given by

$$g_{\text{stat}}(K, m) = z - a'K - b/K - c'Km - d'/(Km), \tag{20}$$

where z is given in (12), b is given in (14), and

$$a' = \frac{\nu p_2(-\mu + (2 - p_2)\vartheta)}{2(\mu + p_2\vartheta)},$$

$$c' = \frac{\nu p_2}{2},$$

$$d' = \frac{s(1 - (1 - p_2)^m)}{(\mu + p_2\vartheta)}. \tag{21}$$

Let

$$g^*_{\text{stat}} = \max_{K\geq 1} \max_{m\geq 1} \{g_{\text{stat}}(K, m)\}$$

denote the optimal average operating profit among this class of static operating policies.

For $a \leq 0$, Proposition 4(a) shows that $n^* = 1$. Therefore, the optimal dynamic operating policy is equivalent to the static operating policy with $m = 1$. (In this case, $d = d'$ and $a + c = a' + c'$.) Hence, we shall only consider the case when $a > 0$, or equivalently, $\vartheta > \mu/p_2$. In this case, it is easy to show that $a' > 0$. To simplify our analysis so that we can compare the performance of the static operating policy with the optimal dynamic operating policy, let us ignore the term $(1 - p_2)^m$ in (21) so that d' is reduced to d'', where

$$d'' = \frac{s}{(\mu + p_2\vartheta)}.$$

Alternatively, $d' = d''$ when the switch-over cost is always charged in each cycle even when no repairable batches are generated in one cycle. Such an assumption applies, for example, to a situation in which the switch-over operation is scheduled regularly (say, during the last hour in each shift), and thus the switch-over cost can be incurred due to the preparation even when it turns out that no repair is required.

Let g'_{stat} and g'^*_{stat} denote the corresponding g_{stat} and g^*_{stat} with d' replaced by d''. Clearly, $g'_{\text{stat}}(K, m) \leq g_{\text{stat}}(K, m)$ for any K and m. Therefore, $g'^*_{\text{stat}} \leq g^*_{\text{stat}}$. Applying the same analysis as in the proof of Proposition 4(b) (since $a' > 0$), we can show that

$$g'^*_{\text{stat}} = z - 2\sqrt{a'b} - 2\sqrt{c'd''}.$$

Observe that $c'd'' = cd$. Hence, using Proposition 4(b) again,

$$g_{n*}(K^*) - g'^*_{\text{stat}} = 2\sqrt{a'b} - 2\sqrt{ab}$$
$$= \frac{\sqrt{2\nu\tau}(\sqrt{p_2(2\vartheta - p_2\vartheta - \mu)} - \sqrt{p_2\vartheta - \mu})}{\mu + p_2\vartheta}. \quad (22)$$

Notice from the above expression that $g'^*_{\text{stat}} \approx g_{n*}(K^*)$ when $\tau \approx 0$, i.e., when the setup cost for each batch is negligible. Since

$$p_2(2\vartheta - p_2\vartheta - \mu) - (p_2\vartheta - \mu)$$
$$= p_2\vartheta + \mu + p_2(-p_2\vartheta - \mu) = (1 - p_2)(p_2\vartheta + \mu),$$

it follows that $g'^*_{\text{stat}} \approx g_{n*}(K^*)$ when p_2 is close to 1. Since $(1 - p_2)^m$ is small when p_2 is close to 1, $g^*_{\text{stat}} \approx g'^*_{\text{stat}}$ when p_2 is close to 1. Therefore, we have the following result.

PROPOSITION 6. *When τ is small or when p_2 is large, the static policy is effective. Specifically,*

$$g_{n*}(K^*) - g^*_{\text{stat}}$$
$$\leq \frac{\sqrt{2\nu\tau}(\sqrt{p_2(2\vartheta - p_2\vartheta - \mu)} - \sqrt{p_2\vartheta - \mu})}{\mu + p_2\vartheta}.$$

Consider the same numerical example given at the end of § 4.2. Recall that the optimal average operating profit is equal to 2.355 with the corresponding $(n^*, K^*) = (2, 8)$. Using the static operating policy (with K, m being positive integers), it turns out that $g'^*_{\text{stat}} = 2.338$ (with $K = 7$ and $m = 3$) and $g^*_{\text{stat}} = 2.350$ (with $K = 15$ and $m = 1$). Hence, $g'^*_{\text{stat}} < g^*_{\text{stat}} < g_{n*}(K^*)$. Furthermore, $g_{n*}(K^*) - g'^*_{\text{stat}} = 0.017$, whereas the right-hand side of (22) is equal to 0.018 in this example.

5. Concluding Remarks

In this paper we have developed a model for a situation in which the bottleneck facility processes batches of jobs. Upon the completion of each batch, the entire batch is either nondefective, repairable, or nonrepairable. This situation occurs in certain manufacturing processes or in a situation when the jobs in a batch are not detachable (or separable). For instance, the chips on each wafer are not detachable during the wafer fabrication process. This distinctive feature of the process has motivated us to develop a model that focused on the repair issue as well as the operating issue. By extending the analysis presented in So and Tang (1995), we were able to determine the operating profit for a given repair policy. In this paper we have shown that it is optimal to repair all repairable jobs or to scrap all repairable jobs. We have also developed optimality conditions for each of the repair policies to be optimal. Moreover, we have extended our analysis to deal with the issue of batch size and the issue of static operating policy versus dynamic operating policy.

For the case in which a certain proportion of the jobs in a batch is nondefective and a certain proportion is either repairable or nonrepairable, one can extend the

analysis presented in So and Tang (1995) to analyze the operating profit associated with different repair policies and operating policies. However, the analysis is rather complex, and it is difficult to characterize the optimal repair policy and operating policy analytically. In this case, one may need to consider numerical analysis for evaluating different repair and operating policies.[1]

[1] This research was supported in part by a research fellowship from the University of California at Irvine and by the Center for Technology Management at UCLA and the UCLA Committee of Research Grant No. 92.

Appendix

PROOF OF PROPOSITION 1. We shall prove Proposition 1 by extending the results in So and Tang (1995) to our model here. Since the approach is exactly the same as So and Tang (1995) with some minor adaptations, we only briefly illustrate the required modifications here. Interested readers are referred to So and Tang (1995) for details.

As in So and Tang (1995), we formulate our problem as a semi-Markov decision process. To follow their analysis, consider for now our objective as being minimizing the long run average operating costs, rather than maximizing average operating profit. Following the notation developed there, the one-step expected costs for our model here are given by

$$c(i;1) = \sum_{j=0}^{B-1} \binom{B-1}{j} p_2^j (1-p_2)^{B-1-j} \times \sum_{k=0}^{j} \frac{Bu}{j+1} h_{i+k} - B(p_1r_1 + p_3r_3 - c_1), \quad i \geq 0,$$

$$c(i;2) = h_i t_1 + \sum_{j=1}^{i} h_j v + s - ir_2, \quad i \geq 1,$$

while $p_{i,j}(a)$ and $\tau(i;a)$ are the same as before with q replaced by $p_2\alpha$. We can repeat the same argument there to establish the results in Theorem 4.1 and Propositions 4.2–4.4 to our problem here. In particular, Proposition 4.4 states that all repairable jobs will eventually be repaired. Under the average cost criterion we can then assume, without loss of generality, that the reward for each repairable job, r_2, is received when the repairable job is generated instead of when it is repaired. Then, the one-step expected costs $c(i;a)$ can be rewritten as

$$c(i;1) = \sum_{j=0}^{B-1} \binom{B-1}{j} p_2^j (1-p_2)^{B-1-j} \times \sum_{k=0}^{j} \frac{Bu}{j+1} h_{i+k} - B(p_1r_1 + p_2r_2 + p_3r_3 - c_1), \quad i \geq 0,$$

$$c(i;2) = h_i t_1 + \sum_{j=1}^{i} h_j v + s, \quad i \geq 1.$$

Therefore, our model here is equivalent to the previous model with $B = 1$, q replaced by $p_2\alpha$, and ru by $(p_1r_1 + p_2r_2 + p_3r_3 - c_1)$. Thus, all previous analysis and results remain valid. Equations (2) and (3) follows directly from the analysis in So and Tang (1995) with linear holding cost and batch size equal to one. □

We remark that, following the discussions after Theorem 4.5 in So and Tang (1995), we can also show that the average operating reward under an n-policy, denoted by g_n, is unimodal in n. Furthermore, the simple procedure given there can be directly adapted to compute the optimal threshold policy.

PROOF OF PROPOSITION 3. (a) If $s \leq R_1^1$, it follows from Proposition 1 and the definition of R_1^1 that $g^1 = g_1^1$. If $s > R_1^1$,

$$g^1(s) < g^1(R_1^1) = g_1^1(R_1^1) \leq g^0,$$

where the first inequality follows from the fact that $g^1(s)$ is strictly decreasing in s and the second inequality follows from (9). Therefore, $g^* = \max(g^0, g^1) = \max(g^0, g_1^1)$ for all s. From (1) and (4), $g_1^1 \geq g^0$ is, after simplification, equivalent to $s \leq s^1$.

(b) If $s \leq R_1^1$, $g^1(s) = g_1^1(s)$. Thus, $g^1(s) \geq g^1(R_1^1) > g^0$, where the first inequality follows $g^1(s)$ decreasing in s and the second inequality follows from the assumption that condition (9) does not hold. Therefore, $g^*(s) = g_1^1(s)$ for $s \leq R_1^1$. This proves (i). Since $g^1(s)$ is strictly decreasing to $-\infty$ as $s \to \infty$, and g^0 is independent of s. There exists some finite constant s^* such that the (0, 0)-policy is optimal for $s \geq s^*$. Again, the assumption that condition (9) does not hold implies that $s^* > R_1^1$. This proves (ii). Part (iii) follows directly from the discussions for (i) and (ii).

To derive an upper bound for s^*, consider n as a continuous variable in (4) and differentiate g_n^1 with respect to n. To simplify notation, let

$$w = (p_1r_1 + p_2r_2 + p_3r_3 - c_1)/p_2.$$

Then,

$$\begin{aligned} f(n) \equiv \frac{\partial}{\partial n} g_n^1 &= \frac{w - hun/p_2 + hu/2p_2 - ht_1 - hvn - hv/2}{t_1 + t_2 + n(u/p_2 + v)} - \frac{(u/p_2 + v)(nw - s - hun(n-1)/2p_2 - hnt_1 - hvn(n+1)/2)}{(t_1 + t_2 + n(u/p_2 + v))^2} \\ &= \frac{1}{(t_1 + t_2 + n(u/p_2 + v))^2} \left\{ \left(w - \frac{hun}{p_2} + \frac{hu}{2p_2} - ht_1 - hvn - \frac{hv}{2} \right) \left(t_1 + t_2 + \frac{nu}{p_2} + nv \right) \right. \\ &\qquad \left. - \left(\frac{u}{p_2} + v \right) \left(nw - s - \frac{hun(n-1)}{2p_2} - hnt_1 - \frac{hvn(n+1)}{2} \right) \right\} \\ &= \frac{1}{(t_1 + t_2 + n(u/p_2 + v))^2} \left\{ -\frac{n^2h}{2} \left(\frac{u}{p_2} + v \right)^2 - nh(t_1 + t_2) \left(\frac{u}{p_2} + v \right) + \theta \right\}, \end{aligned}$$

where θ is given by

$$\theta = s\left(\frac{u}{p_2} + v\right) - (t_1 + t_2)\left(h(t_1 + v) - \frac{p_1r_2 + p_2r_2 + p_3r_3 - c_1}{p_2}\right) + \frac{(t_1 + t_2)h}{2}\left(\frac{u}{p_2} + v\right).$$

Since $s^* > R_1^1$, we can consider only $s \geq R_1^1$. When $s \geq R_1^1$, (8) implies that $\theta > 0$. Therefore,

$$y^* = \frac{h(t_1 + t_2)(u/p_2 + v) - \sqrt{h^2(t_1 + t_2)^2(u/p_2 + v)^2 + 2h(u/p_2 + v)^2\theta}}{-h(u/p_2 + v)^2} = \frac{\sqrt{(t_1 + t_2)^2 + 2\theta/h} - (t_1 + t_2)}{(u/p_2 + v)}$$

is the only positive root for $f(y) = 0$. Since $g_n^1(s) \to -\infty$ as $n \to \infty$, y^* is a maximum point, and $g^1 = g_{n^*}^1$ with $n^* = \lfloor y^* \rfloor$ or $\lceil y^* \rceil$, i.e., $g^1 = \max(g_{\lfloor y^* \rfloor}^1, g_{\lceil y^* \rceil}^1)$. Furthermore, it follows from Proposition 1 and the definition of R_n^1 that $s \geq R_{\lfloor y^* \rfloor}^1$. Since $g_n^1(s)$ is decreasing in s,

$$g^1(s) = \max(g_{\lfloor y^* \rfloor}^1(s), g_{\lceil y^* \rceil}^1(s))$$
$$\leq \max(g_{\lfloor y^* \rfloor}^1(R_{\lfloor y^* \rfloor}^1), g_{\lceil y^* \rceil}^1(R_{\lfloor y^* \rfloor}^1)) = g_{\lfloor y^* \rfloor}^1(R_{\lfloor y^* \rfloor}^1),$$

where the last equality follows from the fact that the threshold policy $\lfloor y^* \rfloor$ is optimal at $s = R_{\lfloor y^* \rfloor}^1$ among repair-all policies; see Proposition 1. On the other hand, substituting $s = R_{\lfloor y^* \rfloor}^1$ from (7) into (4) for $n = \lfloor y^* \rfloor$ and after simplification, we obtain

$$g_{\lfloor y^* \rfloor}^1(R_{\lfloor y^* \rfloor}^1) = \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - h\lfloor y^* \rfloor.$$

Now, consider that

$$g^0 \geq \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - hy^*.$$

From (1) and the definition of y^*, the above inequality is equivalent to

$$\frac{p_1r_1 + p_2r_3 + p_3r_3 - c_1}{u} \geq \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - h\frac{\sqrt{(t_1 + t_2)^2 + 2\theta/h} - (t_1 + t_2)}{(u/p_2 + v)},$$

which, after rearranging terms and simplification, reduces to

$$\sqrt{(t_1 + t_2)^2 + 2\theta/h} \geq \frac{r_2 - r_3}{h} + t_2 - v - \frac{v(p_1r_1 + p_2r_3 + p_3r_3 - c_1)}{uh}. \quad (23)$$

Observe that the right-hand side of (23) is positive from the assumption that condition (9) does not hold. Finally, using (23) and the definition of s^u and θ, we obtain, for $s \geq s^u$,

$$g^0 \geq \frac{p_1r_1 + p_2r_2 + p_3r_3 - c_1 - hp_2(v + t_1)}{u + p_2v} - hy^*$$
$$\geq g_{\lfloor y^* \rfloor}^1(R_{\lfloor y^* \rfloor}^1) \geq g^1(s).$$

Therefore, the (0, 0)-policy is optimal for $s \geq s^u$ and s^u is an upper bound for s^*. □

PROOF OF PROPOSITION 4. (a) For any fixed n, it follows from (11) that $g_n^1(K)$ is maximized at

$$K^*(n) = \sqrt{(b + d/n)/(a + cn)}.$$

Substituting $K^*(n)$ into (12) to get

$$g_n^1(K^*(n)) = z - 2\sqrt{(b + d/n)(a + cn)}.$$

Since $a < 0$, one can take the derivative with respect to n to show that the term $2\sqrt{(b + d/n)(a + cn)}$ is increasing in n. Therefore, $g_n^1(K^*(n))$ is decreasing in n. Hence, $n^* = 1$,

$$K^* = \sqrt{(b + d)/(a + c)} \quad \text{and} \quad g_{n^*}(K^*) = z - 2\sqrt{(b + d)(a + c)}.$$

(b) Differentiate $g_n^1(K)$ given in (12) with respect to K and n to obtain the following first-order conditions:

$$-a + \frac{b}{K^2} - cn + \frac{d}{K^2n} = 0,$$

$$-cK + \frac{d}{Kn^2} = 0,$$

which yield $n^* = \sqrt{ad/bc}$ and $K^* = \sqrt{b/a}$. Differentiate twice to obtain the Hessian matrix H of $g_n^1(K)$,

$$H = \begin{pmatrix} -2b/K^3 - 2d/K^3n & -c - d/K^2n^2 \\ -c - d/K^2n^2 & -2d/Kn^3 \end{pmatrix}.$$

It is easy to check that the determinant of $H(K^*, n^*)$ is equal to $4\sqrt{abc^3/d} > 0$. Clearly, all elements of $H(K^*, n^*)$ are negative. Therefore, $H(K^*, n^*)$ is negative definite and the point (K^*, n^*) is a local maximum (e.g., see Avriel 1976, p. 15 and Theorem 2.2). Observe also that $g_n^1(K) \to -\infty$ as $K \to 0$, $n \to 0$, $K \to \infty$, and $n \to \infty$. Since (K^*, n^*) is the unique nonnegative solution to the first-order conditions, we can conclude that the point (K^*, n^*) is global maximum. Finally, substitute n^* and K^* into (12) to obtain $g_{n^*}(K^*)$. □

References

Avriel, M., *Nonlinear Programs: Analysis and Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1976.

Bitran, G. and S. Dasu, "Ordering Policies in an Environment of Stochastic Yields and Substitutable Demands," *Oper. Res.*, 40 (1992), 999–1017.

—— and S. Gilbert, "Co-Production Processes with Random Yields

in the Semi-Conductor Industry," Working Paper, Sloan School of Management, MIT, Cambridge, MA, 1992.

Erlenkotter, D., "Ford Whitman Ford and the Economic Order Quantity Model," *Oper. Res.*, 38 (1990), 937–946.

Gerchak, Y. and M. Parlar, "Yield Randomness, Cost Tradeoffs, and Diversification in the EOQ Model," *Naval Res. Logistics Quarterly*, 37 (1990), 341–354.

Gise, P. and R. Blanchard, *Modern Semiconductor Fabrication Technology*, Prentice-Hall, Englewood Cliffs, NJ, 1986.

Lee, H., "Lot Sizing in a Production Process with Defective Items, Process Corrections, and Rework," *Management Sci.*, 38 (1992), 1314–1328.

Porteus, E., "Optimal Lot Sizing, Process Quality Improvement, and Setup Cost Reduction," *Oper. Res.*, 34 (1986), 137–144.

So, K. C. and C. Tang, "Optimal Operating Policy for a Bottleneck with Random Rework," *Management Sci.*, 41 (1995), 620–636.

Tang, C. and L. Demeester, "Mathematical Models in IC Manufacturing," in R. Sarin (Ed.), *Perspectives in Operations Management*, Kluwer Publishers, 1992.

Yano, C. and H. Lee, "Lot Sizing with Random Yields: A Review," *Oper. Res.*, 43 (1995), 311–334.

Accepted by Hau Lee; received July 1, 1993. *This paper has been with the authors* 1 *month for* 1 *revision.*