IIE Transactions (2006) **38**, 837–846

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170600733228

Quantitative analysis of a transfer production line with Andon

JINGSHAN LI* and DENNIS E. BLUMENFELD

Manufacturing Systems Research Laboratory, GM Research & Development Center, Mail Code 480-106-359, 30500 Mound Road, Warren, MI 48090-9055, USA
E-mail: jingshan.Li@gm.com

Received December 2004 and accepted February 2006

In this paper, analytical models are developed to study the performance of a transfer production line featuring Andon. In addition to providing analytical expressions to evaluate effective production rates, we investigate conditions under which Andon should be introduced and implemented. Some practical rules are derived to guide operations management on the factory floor. We show that when average repair times are short, introducing Andon to stop the line and repair all defects on the line is an effective way to achieve a high throughput of non-defective jobs.

1. Introduction

Andon, derived from the Japanese word for paper lantern, is a term for a visual control system using an electric light board (or other signal device) hung in a factory, so that a worker can call for help and stop the line (Monden, 1997; Liker, 2004). It originates from the Toyota Production System and has been used in many Japanese and American manufacturing plants as an effective approach to improve product quality (Mayne *et al.*, 2001; Strozniak, 2001; Inman *et al.*, 2003; Tierney, 2004). The idea of Andon is that the worker can pull the so-called Andon cord, triggering the light and/or music as a call for help and stopping the line when a defect is discovered. It has been claimed that, although productivity is lost due to line stoppages, overall system performance improves. By implementing Andon, problems are not hidden anymore, but are detected and fixed so that good quality can be achieved the first time. Such principles have been used throughout the production systems at Toyota (Liker, 2004).

The current literature on Andon contains many popular articles that are descriptive or provide qualitative studies. However, to the best of our knowledge, there is no quantitative analysis of how Andon improves product quality or what tradeoffs exist between quality and productivity. Therefore, there is a need to develop quantitative models to analyze the performance of a production system with Andon, to identify under what conditions introducing Andon can improve system performance, and to determine how Andon can be used successfully. The goal of this paper is to contribute to this objective.

In many transfer production lines with Andon, time is slotted into cycles and all machines have identical cycle times. When a defect is discovered (or a problem arises) within a cycle, the Andon cord is pulled and the team leader is the first to be called for help. If the problem is solved before the end of the cycle, all jobs move to the downstream machines at the end of the cycle smoothly. If the defect is not fixed by the end of the cycle, all machines linked to this Andon cord may stop. Extra time is then taken to repair the defect. We refer to this case as "pulling the Andon cord to stop the line" throughout this paper. Specifically, we consider three types of transfer lines: (i) *no Andon*; (ii) *full Andon*; and (iii) *partial Andon* systems.

A line with no Andon is also known as a paced line, where a job is passed to the next machine at the end of the cycle no matter whether it is complete (with good quality) or incomplete (with defects). In a full Andon system, for every defect that cannot be fixed within the cycle, the line is stopped to allow additional time to fix the problem.

In some manufacturing plants, Andon is used only for signalling the problem without stopping the line every time, or the workers are encouraged to reduce the number of line stoppages. Although the throughput of jobs in total increases due to fewer stoppages in these cases, the result might be to encourage workers to overlook glitches, and ultimately, to decrease the throughput of non-defective jobs. In such systems, the line stops for repair only when a severe defect is found. Jobs with minor defects (which typically are the majority of cases) are passed on to the next machine. We refer to such a line as a partial Andon system.

*Corresponding author

The major contribution of this paper is the development of analytical models and closed-form expressions to analyze and compare the performance of these three types of transfer production lines (no Andon, full Andon and partial Andon) in terms of the effective production rate (i.e., the average number of non-defective jobs produced per unit time). In addition, we investigate the conditions for successful Andon use, i.e., the conditions under which Andon should be introduced into the system and how it should be implemented. Some simple practical rules are presented for use in operations management on the factory floor. It is shown that introducing Andon can improve the throughput of non-defective jobs when the average repair times are short, and that solving all the problems has more benefits than simply fixing severe defects.

The remainder of the paper is structured as follows. Section 2 reviews the literature related to Andon and paced/unpaced lines. The problem addressed in this paper is formulated in Section 3. Section 4 discusses the multiple-machine line with one Andon cord. A system with multiple Andon cords is described in Section 5. Finally, Section 6 formulates the conclusions. The proofs, unless provided in the main text, are presented in the Appendix.

2. Literature review

It has been reported in Mayne *et al.* (2001), Strozniak (2001), Liker (2004) and Tierney (2004) that every assembly line worker at Toyota is empowered to stop the line by pulling the Andon cord when they see defects or problems, in order to correct them. A production line at Toyota may stop hundreds of times during each shift. As addressed by Liker (2004), implementing Andon is one of the approaches used to "build a culture of stopping to fix problems, to get quality right the first time", and to "use visual control so no problems are hidden". Inman *et al.* (2003) have recognized the study of Andon systems as an important research opportunity to address the tradeoffs between quality and throughput.

Paced and unpaced production lines have been studied for many years, see for example Buffa (1961), Sury (1964), Franks and Sury (1966), Buxey and Sadjadi (1976), and Buzacott and Shanthikumar (1993). In paced lines, a job that cannot be finished within the designed cycle time will be incomplete and defective, whereas in unpaced lines, it is assumed that there is always sufficient time to complete each job with good quality. In particular, Franks and Sury (1966) study the expected cost of incomplete jobs for a paced line with stochastic processing times.

In Buzacott and Shanthikumar (1993), both paced and unpaced lines are discussed, with variability in task time and interactions among the machines included in the analysis. The proportion of non-defective jobs required at each machine to meet the overall quality target of a product is analyzed. Their study may be considered as the first attempt to analyze transfer lines operating under Andon-type or similar strategies. The present paper extends from Buzacott and Shanthikumar (1993, pp. 166–171) and provides quantitative models and insights regarding when and how to use an Andon system.

3. Problem formulation

A transfer production line with Andon is shown in Fig. 1, where the circles represent machines and the rectangles represent buffers. The following assumptions describe the details of three types of production systems (no Andon, full Andon, and partial Andon) studied in this paper:

1. The line consists of M segments, each containing k machines that are linked to one Andon cord. There is a buffer separating every pair of adjacent segments. Thus, there are M Andon cords, $M-1$ buffers, and Mk machines in total.
2. All machines have the same cycle time, c units of time, to finish their job functions. Machines linked to each Andon cord are synchronized, i.e., jobs start at the same time. At the end of each cycle, there exists a probability $\lambda_{i,j}$, $i = 1, \ldots, M$, $j = 1, \ldots, k$, that the job cannot be finished with good quality, i.e., has a defect. We refer to the parameter $\lambda_{i,j}$ as the *quality failure rate*.
3. For the no Andon system, the job is transferred to the next machine at the end of the cycle, no matter whether or not it is finished with good quality. For the full or partial Andon systems, if a job is not finished with good quality at the end of the cycle, extra repair time may be needed. The maximum repair time is $t_{\rm m}$, which is identical for all machines. When $t_{\rm m}$ is reached, a job must be sent to the next machine even if it is not fully repaired.
4. When the Andon cord is pulled, all machines linked to the same cord stop working. For full and partial Andon systems, the extra time needed to repair the defect is described by an exponential distribution with *quality*

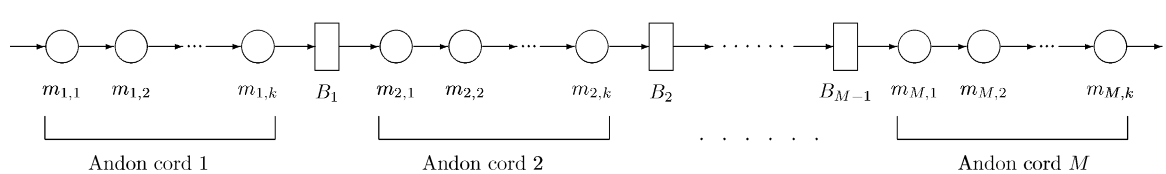

Fig. 1. Transfer production line featuring Andon.

repair rates $\mu_{i,j}$ and $\nu_{i,j}$, respectively, $i = 1, \ldots, M,\ j = 1, \ldots, k$, but constrained by maximum time t_m. In addition, we assume $\nu_{i,j} \leq \mu_{i,j}$ (i.e., the repair rate for severe defects is lower than that for all defects).
5. For the partial Andon system, a fraction $\alpha_{i,j},\ i = 1, \ldots, M, j = 1, \ldots, k$, of all the defective jobs have severe defects and are repaired during the extra time. All other jobs are transferred to the next machine at the end of the cycle without repair.
6. At the end of each cycle, there is at most one machine that cannot finish the job with good quality among all the machines linked to the same Andon cord.

   Remark 1. This assumption is introduced to simplify the analysis. It is consistent with most practical situations, where the probability that workers at two or more machines pull the cord in the same cycle is typically small. For example, consider a production line with 100 machines and ten Andon cords that has 1000 Andon calls in total per shift. This implies that each section (one cord) on average only has about 12 calls per hour. Typically, the cycle time is roughly 1 minute. Therefore, the probability that there are two or more Andon calls per cycle on one Andon cord is negligible (less than 2%).

7. Each buffer has a finite capacity N_i, $0 < N_i < \infty$, $i = 1, \ldots, M-1$.
8. Machine $m_{1,1}$ is never starved and $m_{M,k}$ is never blocked. For each segment i, $i = 2, \ldots, M$, if $m_{i,1}$ is starved at the beginning of the cycle, then all machines linked to the same Andon cord, $m_{i,j},\ j = 1, \ldots, k$, are idle during the cycle. Analogously, if $m_{i,k}$ is blocked at the beginning of the cycle, then all machines, $m_{i,j}$, $j = 1, \ldots, k$, stop working during the cycle.
9. Quality defects are independent as well as all operations. In other words, each operator only works on his/her own specified task, and no correlation exists between sequential job quality failures.

The problem addressed in this paper is as follows: Given a production system defined by assumptions 1–9, develop a method to evaluate the production rate of non-defective jobs as a function of the system parameters.

Solutions to the problem are given in Section 4 for the k-machine line with a single Andon cord. Approximate solutions for the case of multiple Andon cords are presented in Section 5.

4. System with one Andon cord

Suppose there are k machines linked to one Andon cord. This implies that if one machine has introduced a quality failure at the end of a cycle, and the worker pulls the Andon cord to stop the line, then all k machines stop and wait until either the problem is fixed or the limit of the extra time, t_m, is reached.

4.1. *Analytical expressions*

We begin the analysis with partial Andon systems, and then consider full Andon and no Andon systems as special cases.

For the partial Andon system, the Andon cord is pulled and the line is stopped only when a severe defect is discovered, i.e., extra repair time is only permitted for some defective jobs in order to reduce the number of line stoppages. For such a system, let $f_{\text{extra}}(t)$ denote the probability density function of extra time t, described by an exponential distribution with parameter ν and maximum extra time t_m. Then $f_{\text{extra}}(t)$ is given by:

$$f_{\text{extra}}(t) = \nu e^{-\nu t} + e^{-\nu t_m}\delta(t - t_m), \quad 0 < t \leq t_m, \tag{1}$$

where parameter ν defines the quality repair rate for severe defects, and $\delta(\cdot)$ is the delta function, i.e.,

$$\delta(x) = 0 \text{ for } x \neq 0, \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1.$$

Similarly, for a full Andon system, the extra time $f_{\text{extra}}(t)$ can be described as:

$$f_{\text{extra}}(t) = \mu e^{-\mu t} + e^{-\mu t_m}\delta(t - t_m), \quad 0 < t \leq t_m, \tag{2}$$

where parameter μ is the quality repair rate for all defects.

Let $G_{p,k}$ denote the effective production rate in the k-machine partial Andon case. The subscript "p" denotes the partial Andon case (similarly, "f" and "n" are used to denote full and no Andon cases, respectively) and "k" implies the number of machines in the system ("kM" will be used to denote the case of M-Andon cords, each with k machines). For the single Andon cord case, $M = 1$. For simplicity, the first subscript for the machine parameters is omitted in this section. From assumptions 5 and 6, at each cycle, either all k machines can finish the jobs with good quality, or machine m_i, $i = 1, \ldots, k$, has a quality failure with probability α_i that the line will be stopped for repair. Therefore, we obtain the following results:

Proposition 1. *Under assumptions 1–9 with $M = 1$, the effective production rate of a k-machine partial Andon system can be calculated as*:

$$G_{p,k} = \frac{\prod_{i=1}^{k}(1-\lambda_i) + \sum_{i=1}^{k}\lambda_i\alpha_i(1-e^{-\nu_i t_m})\prod_{j=1,j\neq i}^{k}(1-\lambda_j)}{c\big[\prod_{i=1}^{k}(1-\lambda_i) + \sum_{i=1}^{k}\lambda_i\prod_{j=1,j\neq i}^{k}(1-\lambda_j)\big] + \sum_{i=1}^{k}\lambda_i\alpha_i\prod_{j=1,j\neq i}(1-\lambda_j)(1-e^{-\nu_i t_m}/\nu_i)}. \tag{3}$$

In the case of identical machines, $\lambda_i = \lambda$, $\nu_i = \nu$, $\alpha_i = \alpha$, $i = 1, \ldots, k$, we have that:

$$G_{\mathrm{p},k} = \frac{1-\lambda+k\lambda\alpha(1-e^{-\nu t_\mathrm{m}})}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_\mathrm{m}})/\nu}. \tag{4}$$

Proof. Considering that only two events can happen (all machines produce non-defective jobs, or only one machine has a quality problem), we have that:

$$\begin{aligned} G_{\mathrm{p},k} = & \left[\frac{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i\alpha_i(1-e^{-\nu_i t_\mathrm{m}})\prod_{j=1,j\neq i}^k(1-\lambda_j)}{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}^k(1-\lambda_j)}\right] \\ & \Big/ \left[c\frac{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i(1-\alpha_i)\prod_{j=1,j\neq i}^k(1-\lambda_j)}{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}^k(1-\lambda_j)}\right. \\ & + \frac{\sum_{i=1}^k \lambda_i\alpha_i\prod_{j=1,j\neq i}^k(1-\lambda_j)}{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}^k(1-\lambda_j)} \\ & \left.\times\left(\int_0^{t_\mathrm{m}}(c+t)\nu_i e^{-\nu_i t}\mathrm{d}t+(c+t_\mathrm{m})e^{-\nu_i t_\mathrm{m}}\right)\right]. \end{aligned}$$

The general result is obtained after some algebraic manipulation, and the case of identical machines follows immediately. ■

Note that the two terms in the numerator of Equation (3) denote the probabilities that a good job is produced when all operations have good quality or when one operation produces defective jobs that are corrected through extra time. The denominator includes the cycle time and extra time for line stoppage.

When $\alpha_i = 1$ and $\nu_i = \mu_i$, $i = 1, \ldots, k$, we obtain the effective production rate for the full Andon system, i.e.,

$$G_{\mathrm{f},k} = \frac{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i(1-e^{-\mu_i t_\mathrm{m}})\prod_{j=1,j\neq i}^k(1-\lambda_j)}{c\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}(1-\lambda_j)(c+(1-e^{-\mu_i t_\mathrm{m}})/\mu_i)}. \tag{5}$$

In the identical machine case, $\mu_i = \mu$, $i = 1, \ldots, k$, and we have that:

$$\begin{aligned} G_{\mathrm{f},k} &= \frac{(1-\lambda)^k+k\lambda(1-e^{-\mu t_\mathrm{m}})(1-\lambda)^{k-1}}{c(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}(c+(1-e^{-\mu t_\mathrm{m}})/\mu)} \\ &= \frac{1-\lambda+k\lambda(1-e^{-\mu t_\mathrm{m}})}{c-c\lambda+ck\lambda+k\lambda(1-e^{-\mu t_\mathrm{m}})/\mu}. \end{aligned} \tag{6}$$

Moreover, if $t_\mathrm{m} = 0$, Equation (5) corresponds to a no Andon system:

$$G_{\mathrm{n},k} = \frac{\prod_{i=1}^k(1-\lambda_i)}{c\left[\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}^k(1-\lambda_j)\right]}. \tag{7}$$

Again, when all machines are identical, we have that:

$$G_{\mathrm{n},k} = \frac{(1-\lambda)^k}{c[(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}]} = \frac{1-\lambda}{c(1-\lambda+k\lambda)}. \tag{8}$$

4.2. *Structural properties*

4.2.1. *Asymptotic property*

Clearly, when $t_\mathrm{m} \to \infty$, we obtain the asymptotic behavior of the effective production rates for the partial and full Andon cases:

$$\lim_{t_\mathrm{m}\to\infty} G_{\mathrm{p},k} = \frac{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i\alpha_i\prod_{j=1,j\neq i}^k(1-\lambda_j)}{c\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}^k(1-\lambda_j)(c+\alpha_i/\nu_i)}, \tag{9}$$

$$\lim_{t_\mathrm{m}\to\infty} G_{\mathrm{f},k} = \frac{\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i\prod_{j=1,j\neq i}^k(1-\lambda_j)}{c\prod_{i=1}^k(1-\lambda_i)+\sum_{i=1}^k \lambda_i \prod_{j=1,j\neq i}(1-\lambda_j)(c+1/\mu_i)}. \tag{10}$$

When all machines are identical, we have that:

$$\lim_{t_\mathrm{m}\to\infty} G_{\mathrm{p},k} = \frac{1-\lambda+k\lambda\alpha}{c-c\lambda+ck\lambda+k\lambda\alpha 1/\nu}, \tag{11}$$

$$\lim_{t_\mathrm{m}\to\infty} G_{\mathrm{f},k} = \frac{1-\lambda+k\lambda}{c-c\lambda+ck\lambda+k\lambda 1/\mu}. \tag{12}$$

Below we will show that these limits define the maximum or minimum effective production rates, depending on the conditions on how λ, ν (or μ), c and k are related. To investigate these conditions, we first study the monotonic properties of $G_{\mathrm{p},k}$.

4.2.2. *Monotonicity*

From Equations (3), (5) and (7), it is possible to show that the effective production rate is a monotonically decreasing or increasing function of λ_i, μ_i, etc. In order to avoid messy notation and expressions, we only illustrate here the monotonic properties for the identical machine case.

Theorem 1. *Under assumptions 1–9 with $M = 1$, assume $\lambda_i = \lambda$, $\nu_i = \nu$, $\alpha_i = \alpha$, $i = 1, \ldots, k$. Then, $G_{\mathrm{p},k}$ is monotonically decreasing with respect to k, c and λ, and is monotonically increasing with respect to ν. It is a monotonic decreasing function of α if $(k-1)c\lambda\nu+\lambda+c\nu > 1$. In addition, $G_{\mathrm{p},k}$ is monotonically decreasing or increasing with respect to t_m if $(k-1)c\lambda\nu+\lambda+c\nu < 1$, or $(k-1)c\lambda\nu+\lambda+c\nu > 1$, respectively.*

Proof. See the Appendix. ■

Analogously, for a full Andon system, by letting $\alpha = 1$, $\nu = \mu$, we can show that $G_{\mathrm{f},k}$ is monotonically decreasing with respect to k, c, λ, and increasing with respect to μ. Again, $G_{\mathrm{f},k}$ is a monotonic increasing or decreasing function of t_m if $(k-1)c\lambda\mu+\lambda+c\mu > 1$, or $(k-1)c\lambda\mu+\lambda+c\mu < 1$, respectively.

Examples of $G_{\mathrm{p},k}$ and $G_{\mathrm{f},k}$ as functions of t_m are shown in Figs. 2 and 3, respectively. The straight solid lines in Figs. 2 and 3 represent the special cases where $(k-1)c\lambda\nu+\lambda+c\nu$

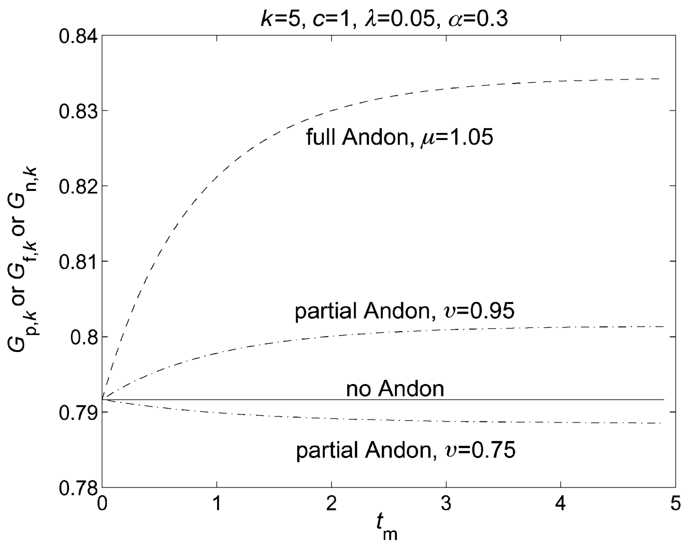

Fig. 2. Monotonicity of $G_{p,k}$ with $M = 1$.

and $(k-1)c\lambda\mu + \lambda + c\mu$ are equal to one, respectively. The expressions for $G_{p,k}$ and $G_{f,k}$ in these cases reduce to the effective production rates for the no Andon system.

It is clear from Theorem 1 and Figs. 2 and 3 that increasing repair rates can improve the effective production rate. Counter-intuitively, we see that increasing t_m can lead to either an increase or decrease in effective production rate. If the repair rate is high, which implies that the average repair time is short, then increasing t_m can improve the effective production rate, since fewer jobs need a long time to repair. However, if the repair rate is low, then the average repair time is long, and increasing t_m results in more jobs that need longer repair times. In this case, increasing t_m can be harmful to the effective production rate. Moreover, it can be shown that $\partial^2 G_{p,k}/\partial t_m^2 < 0$ if $(k-1)c\lambda\nu + \lambda + c\nu > 1$, and $\partial^2 G_{f,k}/\partial t_m^2 < 0$ if $(k-1)c\lambda\mu + \lambda + c\mu > 1$. In other words, the growth rate of $G_{p,k}$ and $G_{f,k}$ with respect to t_m is decreasing, as we can see from Figs. 2 and 3. Therefore, it is not necessary to set t_m very high, since this would not result in much additional improvement.

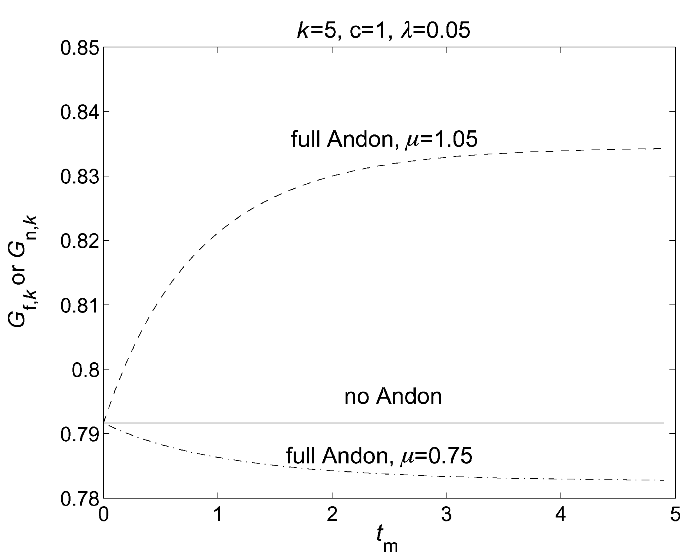

Fig. 3. Monotonicity of $G_{f,k}$ with $M = 1$.

4.2.3. *Comparisons*

Using the monotonicity results, we compare all three systems: no Andon, full Andon, and partial Andon. We obtain the conditions that show when Andon is effective and which Andon system should be used. Since $\nu < \mu$, we have:

Corollary 1. *Under assumptions 1–9 with identical machines and $M = 1$:*
If $(k-1)c\lambda\nu + \lambda + c\nu > 1$, *then*:

$$G_{f,k} > G_{p,k} > G_{n,k}.$$

If $(k-1)c\lambda\nu + \lambda + c\nu < 1 < (k-1)c\lambda\mu + \lambda + c\mu$, *then*:

$$G_{p,k} < G_{n,k} < G_{f,k}.$$

Moreover, if $(k-1)c\lambda\mu + \lambda + c\mu < 1$, *then*:

$$G_{p,k} < G_{n,k} \quad \text{and} \quad G_{f,k} < G_{n,k}.$$

Proof. The result follows directly from Theorem 1 by monotonicity with respect to t_m and α. ∎

The above results show that the effective production rates are higher for the full Andon or no Andon systems, depending on the relative values of λ, μ, ν, c and k. When the repair rates μ and ν are high, the full Andon system is the best system. When the repair rates are low, the no Andon system is shown to be better than either the full or partial Andon systems. In neither case is the partial Andon system the best system.

Intuitively, with a reasonably small quality failure rate, one may expect that, if the repair rate is high and defects can be fixed quickly, introducing the Andon system improves the system throughput of non-defective jobs. If the repair rate is low and it takes a much longer time to repair defects, then the line without Andon performs better. In other words, the Andon system works well only when problems can be solved within a short time period. Although introducing the Andon system decreases productivity, the results show that, when repair times are short, more non-defective jobs are produced due to the improvement in product quality, and it is better to repair all defects. Thus, partial Andon is *never* the best policy. A more effective way to use Andon is to solve all the problems on the line.

In automotive assembly lines, the majority of the Andon calls are due to over cycle, push button stop, reset, carrier jam, etc. Usually the time needed to correct such problems is much less than the cycle time (roughly 1 minute). Therefore, Andon can be implemented successfully. The workers should be trained to be capable of solving the problems quickly by being "aware of the importance of bringing the problems to surface", and "a problem-solving process should be already in place" (Liker, 2004).

Remark 2. It is reported by Tierney (2004) that, although maybe up to thousands of Andons are called per shift in Toyota plants, the total line stoppage time per shift is about 10 to 15 minutes. This implies a very short stoppage time for each Andon call, which is consistent with the results obtained in this paper.

Corollary 1 provides exact conditions for when Andon can improve a system's effective production rate. To make it easier to apply on the factory floor, we summarize them below in terms of some practical rules. To do that, we define $\beta_\nu = e^{-\nu t_m}$ as the *unrepairable rate for severe defects* (i.e., the fraction of severe defects that cannot be repaired by the end of maximum extra time t_m). Then we can show the following: If ν is large so that $\beta_\nu < \lambda$, and if the average extra time $(1-\beta_\nu)/\nu < c + (k-1)c\lambda = c(1-\lambda) + kc\lambda$, then:

$$(k-1)c\lambda\nu + \lambda + c\nu > 1 - \beta_\nu + \lambda > 1.$$

Similarly, if μ is large and *unrepairable rate* $\beta_\mu = e^{-\mu t_m} < \lambda$, and if $(1-\beta_\mu)/\mu < c + (k-1)c\lambda = c(1-\lambda) + kc\lambda$, then:

$$(k-1)c\lambda\mu + \lambda + c\mu > 1 - \beta_\mu + \lambda > 1.$$

Here $c\lambda$ and $c(1-\lambda)$ represent the average defective and non-defective working time, i.e., the average time spent working on defective jobs (excluding the repair times) and non-defective jobs, for one machine, respectively. Thus, if we define $kc\lambda$ as the system's average defective working time on all machines, we obtain the following practical rules which can be used on the factory floor to guide operation management:

- If the unrepairable rate is low (smaller than the quality failure rate) and the average time to repair a defect is short (less than the average non-defective working time on one machine plus average defective working time on all machines), then introducing (full) Andon will improve the effective production rate.
- If the unrepairable rate for severe defects is low (smaller than the quality failure rate) and the average time to repair a severe defect is short (less than the average non-defective working time on one machine plus the average defective working time on all machines), then introducing any type of Andon (full or partial) will improve the effective production rate,
- When Andon is applicable, full Andon gives a greater improvement than partial Andon. In other words, it is more beneficial to repair all defects rather than only severe ones.

4.3. *Special case: one-machine system*

A simple but important special case is the one-machine system. It is of interest because, first, the results are easy to understand, and second, assumption 6 is not necessary. Since only one machine is considered here, we omit all subscripts in the machine parameters. Setting $k = 1$, we obtain:

$$G_{p,1} = \frac{1 - \lambda + \lambda\alpha(1 - e^{-\nu t_m})}{c + (\lambda\alpha/\nu)(1 - e^{-\nu t_m})}, \tag{13}$$

$$G_{f,1} = \frac{1 - \lambda e^{-\mu t_m}}{c + (\lambda/\mu)(1 - e^{-\mu t_m})}, \tag{14}$$

$$G_{n,1} = \frac{1-\lambda}{c}. \tag{15}$$

When $t_m \to \infty$, we obtain:

$$G_{p,1} = \frac{1 - \lambda + \lambda\alpha}{c + \lambda\alpha/\nu},$$

$$G_{f,1} = \frac{1}{c + \lambda/\mu}.$$

In particular, when $c = 1$, $G_{f,1} = \mu/(\lambda + \mu)$, i.e., the effective production rate of an individual machine is equal to its "quality efficiency", analogous to the definition of machine efficiency in throughput analysis.

The monotonic properties obtained above still hold for the one-machine system. As shown in Figs. 4 and 5, increasing t_m can either benefit or harm the effective production rate, depending on the repair rate.

In addition, the practical rules obtained in the previous subsection can be simplified as follows:

- If the unrepairable rate is low (smaller than the quality failure rate) and the average time to repair a defect is short (less than the cycle time), then introducing (full) Andon will improve the effective production rate.

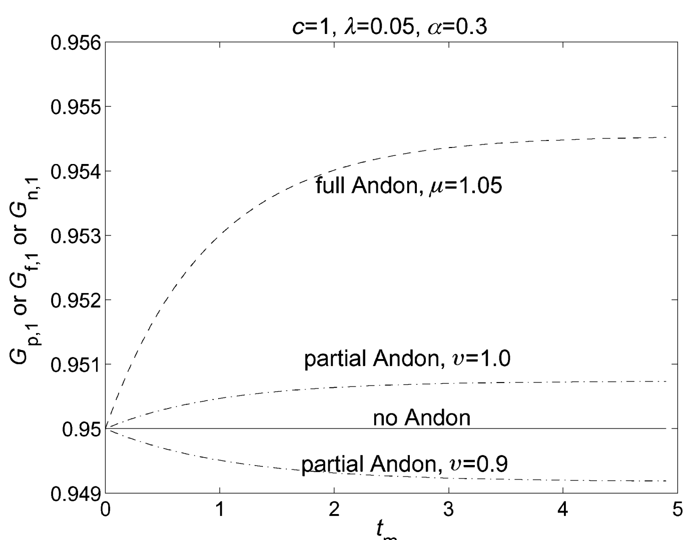

Fig. 4. Monotonicity of $G_{p,1}$ with $M = k = 1$.

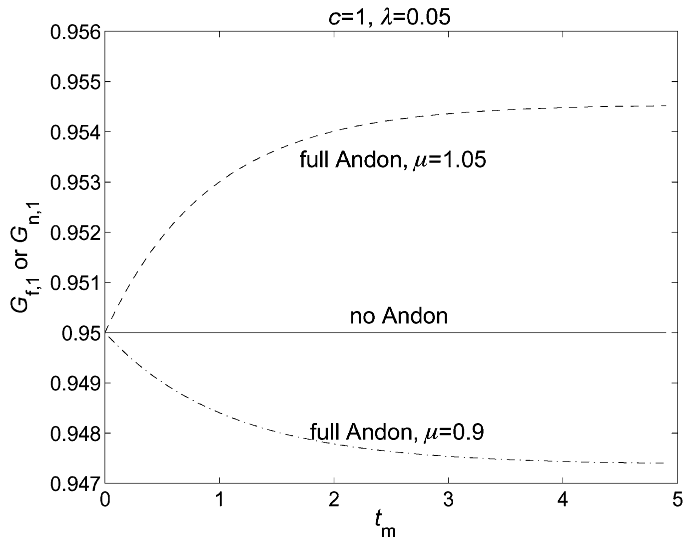

Fig. 5. Monotonicity of $G_{\mathrm{f},1}$ with $M = k = 1$.

- If the unrepairable rate for severe defects is low (smaller than the quality failure rate) and the average time to repair a severe defect is short (less than the cycle time), then introducing any type of Andon will improve the effective production rate.
- When Andon is applicable, full Andon gives greater improvement than partial Andon. Therefore, it is worth repairing all defects rather than only severe ones.

5. Systems with multiple Andon cords

In this section, we extend our study to a system with multiple Andon cords (shown in Fig. 1). The line consists of M segments, separated by $M-1$ buffers. Each segment has k machines linked to one Andon cord. We consider the special case where all machines are identical, and all buffers are identical. In addition, for simplicity and without loss of generality, we assume cycle time $c = 1$ for all segments.

To analyze this system, consider each segment as an aggregated single machine (Fig. 6). For the case of no Andon, the system can be viewed as a transfer line of "reliable" machines having identical processing times (here "reliable" implies the aggregated machines have no downtimes). Then the effective production rate, $G_{\mathrm{n},kM}$, is only dependent on the fraction of non-defective jobs at each machine. Therefore, we have that:

$$G_{\mathrm{n},kM} = \left(\frac{1-\lambda}{1-\lambda+k\lambda}\right)^M. \tag{16}$$

The full and partial Andon systems can be viewed as transfer lines with reliable machines and random processing times. It is shown in Blumenfeld (1990) that for a serial production line with reliable machines and random processing times, in which all machines and buffers are identical, the system throughput TP can be approximated by the following equation:

$$TP \approx 1\Bigg/\left(\bar{T}\left[1+\frac{1.67(M-1)CV}{1+M+0.31CV+1.67MN/(2CV)}\right]\right), \tag{17}$$

where $\bar{T}$ and CV are the mean and coefficient of variation of processing times, respectively, and N is the buffer size.

In order to use this result, we consider each segment (with k machines) as an aggregated machine (Fig. 6). The mean and coefficient of variation of the processing times of this aggregated machine (full or partial Andon line) are obtained first:

Proposition 2. *Under assumptions 1–9 with $M = 1$, and assuming $\lambda_i = \lambda$, $\nu_i = \nu$, $\alpha_i = \alpha$, $i = 1, \ldots, k$, the mean, $\bar{T}_{\mathrm{p}}$, and coefficient of variation, CV_{p}, of processing time T_{p} for a k-machine partial Andon system are given by*:

$$\bar{T}_{\mathrm{p}} = \frac{1}{1-\lambda+k\lambda}\left(c - c\lambda + kc\lambda + k\lambda\alpha\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}\right), \tag{18}$$

$$CV_{\mathrm{p}} = \frac{\sqrt{k\lambda\alpha[2(1-e^{-\nu t_{\mathrm{m}}} - t_{\mathrm{m}}\nu e^{-\nu t_{\mathrm{m}}})(1-\lambda+k\lambda) - k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})^2]}}{c\nu - c\lambda\nu + kc\lambda\nu + k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})}. \tag{19}$$

Proof. See the Appendix. ■

When $\alpha = 1$ and $\nu = \mu$, these results reduce to the full Andon system case, and we obtain:

$$\bar{T}_{\mathrm{f}} = \frac{1}{1-\lambda+k\lambda}\left(c - c\lambda + kc\lambda + k\lambda\frac{1-e^{-\mu t_{\mathrm{m}}}}{\mu}\right), \tag{20}$$

$$CV_{\mathrm{f}} = \frac{\sqrt{k\lambda[2(1-\lambda+k\lambda)(1-e^{-\mu t_{\mathrm{m}}} - t_{\mathrm{m}}\mu e^{-\mu t_{\mathrm{m}}}) - k\lambda(1-e^{-\mu t_{\mathrm{m}}})^2]}}{c\mu - c\lambda\mu + kc\lambda\mu + k\lambda(1-e^{-\mu t_{\mathrm{m}}})}, \tag{21}$$

where $\bar{T}_{\mathrm{f}}$ and CV_{f} denote the mean and coefficient of variation of processing times, respectively, for the full Andon system.

Then, for kM-machine partial and full Andon production lines, we obtain

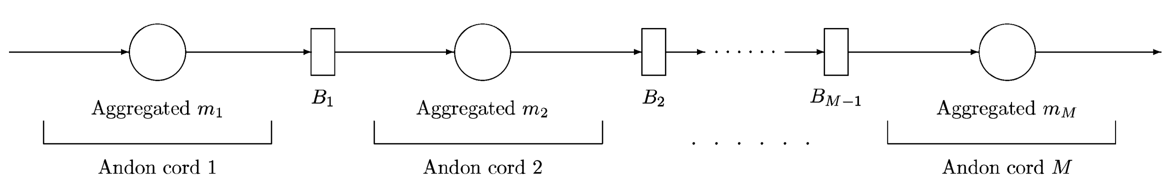

Fig. 6. Transfer line with multiple Andon cords: aggregating machines linked to one Andon cord into a single machine.

$$G_{\mathrm{p},kM} \approx \left(\frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_{\mathrm{m}}}}{1-\lambda+k\lambda}\right)^{M} \Bigg/ \left(\bar{T}_{\mathrm{p}}\left[1+\frac{1.67(M-1)CV_{\mathrm{p}}}{1+M+0.31CV_{\mathrm{p}}+1.67MN/(2CV_{\mathrm{p}})}\right]\right), \tag{22}$$

$$G_{\mathrm{f},kM} \approx \left(\frac{1-\lambda+k\lambda-k\lambda e^{-\mu t_{\mathrm{m}}}}{1-\lambda+k\lambda}\right)^{M} \Bigg/ \left(\bar{T}_{\mathrm{f}}\left[1+\frac{1.67(M-1)CV_{\mathrm{f}}}{1+M+0.31CV_{\mathrm{f}}+1.67MN/(2CV_{\mathrm{f}})}\right]\right), \tag{23}$$

where $\bar{T}_{\mathrm{p}}$ and CV_{p} are given by Equations (18) and (19), and $\bar{T}_{\mathrm{f}}$ and CV_{f} by Equations (20) and (21), respectively.

If $t_{\mathrm{m}} \to \infty$, we have that:

$$\lim_{t_{\mathrm{m}}\to\infty} G_{\mathrm{p},kM} \approx \left(\frac{1-\lambda+k\lambda\alpha}{1-\lambda+k\lambda}\right)^{M} \Bigg/ \left(\bar{T}_{\mathrm{p}}\left[1+\frac{1.67(M-1)CV_{\mathrm{p}}}{1+M+0.31CV_{\mathrm{p}}+1.67MN/(2CV_{\mathrm{p}})}\right]\right), \tag{24}$$

$$\lim_{t_{\mathrm{m}}\to\infty} G_{\mathrm{f},kM} \approx 1 \Bigg/ \bar{T}_{\mathrm{f}}\left[1+\frac{1.67(M-1)CV_{\mathrm{f}}}{1+M+0.31CV_{\mathrm{f}}+1.67MN/(2CV_{\mathrm{f}})}\right], \tag{25}$$

where $\bar{T}_{\mathrm{p}}$, CV_{p} and $\bar{T}_{\mathrm{f}}$, CV_{f} are given by:

$$\bar{T}_{\mathrm{p}} = \frac{1}{1-\lambda+k\lambda}(c-c\lambda+kc\lambda+k\lambda\alpha/\nu), \tag{26}$$

$$CV_{\mathrm{p}} = \frac{\sqrt{k\lambda\alpha(2-2\lambda+2k\lambda-k\lambda\alpha)}}{c\nu-c\lambda\nu+kc\lambda\nu+k\lambda\alpha}, \tag{27}$$

$$\bar{T}_{\mathrm{f}} = \frac{1}{1-\lambda+k\lambda}(c-c\lambda+kc\lambda+k\lambda/\mu), \tag{28}$$

$$CV_{\mathrm{f}} = \frac{\sqrt{k\lambda(2-2\lambda+k\lambda)}}{c\mu-c\lambda\mu+kc\lambda\mu+k\lambda}. \tag{29}$$

Note that Equation (25) results in the same expression as in Blumenfeld (1990).

Using the above equations, we can show that monotonicity properties still hold for $G_{\mathrm{p},kM}$ and $G_{\mathrm{f},kM}$ as a function of the failure and repair rates (see Li and Blumenfeld (2004) for details). In addition, increasing the buffer size N can improve the effective production rate, as expected. However, larger buffers can result in more inventory. Therefore, investigation of tradeoffs between the effective production rate and work in process would be an important topic for future work.

6. Conclusions

In this paper, analytical models to evaluate the performance (effective production rate) of a transfer production line with Andon are presented. Three types of transfer lines, no Andon, full Andon and partial Andon systems, are analyzed. Closed-form expressions are derived and system performances are compared for the three types of line. The results show that introducing Andon can improve product quality in terms of effective production rate when repair times are short. In addition, compared with only solving severe problems (i.e., stopping the line only for severe defects), solving all problems on the line (i.e., stopping the line for all defects) typically can achieve a higher throughput of non-defective jobs.

A topic for future work is to extend the study to systems with simultaneous Andon calls, i.e., systems where multiple workers can trigger a line stoppage in the same cycle. In addition, lines with arbitrary configurations, such as non-identical machines and buffers, different numbers of machines linked to each Andon cord, and non-homogenous systems, etc., may also be studied. Moreover, an important topic would be the investigation of tradeoffs among productivity, quality and cost. For example, in many assembly plants, a clinic or repair center is typically used to fix problems that have been skipped or unsolved in the production line. For such plants, an important issue is the capacity of the repair center. Should the plant keep a smaller repair center and fix more problems on the production line or a large one and skip some problems on the line in order to achieve higher line throughput? A cost model delineating the tradeoffs between line productivity and additional labor cost could be set up to answer this question. An integrated model to include team size, quality information feedback, etc., would be of additional interest.

The results of these studies could provide rules or principles for implementation and management of successful Andon systems.

Acknowledgements

The authors thank Dr. Samuel P. Marin of General Motors, IIE Transactions Department Editor Prof. Saif Benjaafar and the anonymous reviewers for their valuable comments and suggestions.

References

Blumenfeld, D.E. (1990). A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity. *International Journal of Production Research,* **28**, 1163–1182.

Buffa, E.S. (1961) Pacing effects in production lines. *Journal of Industrial Engineering,* **7**, 383–386.

Buxey, G.M. and Sadjadi, D. (1976) Simulation studies of conveyor-paced assembly lines with buffer capacity. *International Journal of Production Research*, **14**, 607–624.

Buzacott, J.A. and Shanthikumar, J.G. (1993) *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Franks, I.T. and Sury, R.J. (1966) The performance of operators in conveyor-paced work. *International Journal of Production Research*, **5**, 97–112.

Inman, R.R., Blumenfeld, D.E., Huang, N. and Li, J. (2003) Designing production systems for quality: research opportunities from an automotive industry perspective. *International Journal of Production Research*, **41**, 1953–1971.

Li, J. and Blumenfeld, D.E. (2004) Quantitative analysis of a transfer production line with Andon. Technical Report R&D-10031, General Motors Research & Development Center, Warren, MI.
Liker, J.K. (2004) *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, McGraw-Hill, New York, NY.
Mayne, E., Murphy, T. and Winter, D. (2001) Quality crunch. *WARD'S AutoWorld*, **7**, 32–37.
Monden, Y. (1997) *Toyota Production System: An Integrated Approach to Just-In-Time*, 3rd edn., Engineering & Management Press, Norcross, GA.
Strozniak, P. (2001) Toyota alters faces of production. *Industrial Week*, Aug. 13, pp. 46–48.
Sury, R.J. (1964) An industrial study of paced and unpaced operator performance in a single stage work task. *International Journal of Production Research*, **3**, 91–101.
Tierney, C. (2004) Big three play catch-up to Toyota plant prowess. *Detroit News Special Report: Auto Plants: Battle to Be the Best*, Feb. 22.

Appendix

For brevity, we omit most of the algebraic manipulation and only provide sketches of the proofs. The complete proofs can be found in Li and Blumenfeld (2004).

Proof of Theorem 1. First we consider monotonicity of $G_{p,k}$ with respect to c. From:

$$\frac{\partial G_{p,k}}{\partial c} = -\frac{1-\lambda+k\lambda\alpha(1-e^{-\nu t_m})}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2}(1-\lambda+k\lambda) < 0,$$

we obtain that $G_{p,k}$ is a monotonically decreasing function of c.

Next we address monotonicity with respect to k.

$$\begin{aligned}\frac{\partial G_{p,k}}{\partial k} &= \frac{\lambda\alpha-\lambda\alpha e^{-\nu t_m}}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu} \\ &\quad - \frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2} \\ &\quad \times\left(c\lambda+\lambda\alpha\frac{1-e^{-\nu t_m}}{\nu}\right) \\ &= \frac{-\lambda(1-\lambda)}{[c-c\lambda+ck\lambda+k\lambda\alpha((1-e^{-\nu t_m})/\nu)]^2} \\ &\quad \times\left(c-c\alpha+c\alpha e^{-\nu t_m}+\alpha\frac{1-e^{-\nu t_m}}{\nu}\right) \\ &< 0.\end{aligned}$$

Therefore, $G_{p,k}$ is a monotonically decreasing function of k.

In addition, we discuss monotonicity with respect to λ.

$$\begin{aligned}\frac{\partial G_{p,k}}{\partial \lambda} &= \frac{-1+k\alpha-k\alpha e^{-\mu t_m}}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu} \\ &\quad - \frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2} \\ &\quad \times\left(-c+ck+k\alpha\frac{1-e^{-\nu t_m}}{\nu}\right) \\ &= \frac{-k}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2} \\ &\quad \times\left(c-c\alpha+c\alpha e^{-\nu t_m}+\alpha\frac{1-e^{-\nu t_m}}{\nu}\right) \\ &< 0.\end{aligned}$$

Again, $G_{p,k}$ is a monotonically decreasing function of λ.

Next, we consider monotonicity with respect to t_m.

$$\begin{aligned}\frac{\partial G_{p,k}}{\partial t_m} &= \frac{k\lambda\alpha\nu e^{-\nu t_m}}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu} \\ &\quad - \frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2}k\alpha\lambda e^{-\nu t_m}, \\ &= \frac{k\lambda\alpha e^{-\nu t_m}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2} \\ &\quad \times[(k-1)c\lambda\nu+\lambda+c\nu-1].\end{aligned}$$

Therefore, $\partial G_{p,k}/\partial t_m$ is positive (in other words, monotonically increasing with respect to t_m) if $(k-1)c\lambda\nu+\lambda+c\nu > 1$, and is negative (i.e., monotonically decreasing with respect to t_m) if $(k-1)c\lambda\nu+\lambda+c\nu < 1$.

Concerning monotonicity with respect to ν, we have:

$$\begin{aligned}\frac{\partial G_{p,k}}{\partial \nu} &= \frac{k\lambda\alpha t_m e^{-\nu t_m}}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu} \\ &\quad - \frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_m})/\nu]^2} \\ &\quad \times\left(\frac{k\lambda}{\nu}t_m e^{-\nu t_m}-\frac{1-e^{-\nu t_m}}{\nu^2}k\lambda\right) \\ &= \frac{k\lambda}{\nu^2[c-c\lambda+ck\lambda+k\lambda\alpha((1-e^{-\nu t_m})/\nu)]^2} \\ &\quad \times[t_m e^{-\nu t_m}\nu[(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha-1] \\ &\quad +(1-e^{-\nu t_m})(1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m})].\end{aligned}$$

Clearly, if $(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha > 1$, we have $\partial G_p/\partial\nu > 0$. For the case $(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha < 1$, we let:

$$\begin{aligned}A &= t_m e^{-\nu t_m}\nu[(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha-1] \\ &\quad +(1-e^{-\nu t_m})(1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m}),\end{aligned}$$

and obtain

$$\begin{aligned}\frac{\partial A}{\partial \nu} &= \nu e^{-\nu t_m}[(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha-1] \\ &\quad - t_m\nu[(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha-1]\nu e^{-\nu t_m} \\ &\quad +(1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_m})\nu e^{-\nu t_m} \\ &\quad +(1-e^{-\nu t_m})k\lambda\alpha\nu e^{-\nu t_m}, \\ &= \nu e^{-\nu t_m}((k-1)\lambda\nu\alpha+2k\lambda\alpha(1-e^{-\nu t_m}) \\ &\quad - t_m\nu[(k-1)c\lambda\nu\alpha+\lambda+c\nu\alpha-1]), \\ &> 0.\end{aligned}$$

Thus, A is a monotonically increasing function of ν. When $t_m = 0$, we obtain $A = (k-1)c\lambda\nu\alpha \geq 0$. Therefore, $A > 0$ if $t_m > 0$. Finally, we have $\partial G_{p,k}/\partial t_m > 0$ for $t_m \in [0, \infty)$, i.e., it is a monotonically increasing function of ν.

Finally, we consider monotonicity with respect to α. Let $\nu^\star = \partial\nu/\partial\alpha$. Since ν is monotonically increasing with respect to α, we have $\nu^\star > 0$ and

$$\begin{aligned}\frac{\partial G_{\mathrm{p},k}}{\partial\alpha} &= -\frac{1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_{\mathrm{m}}}}{[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu]^2}\\ &\quad\times\left(k\lambda\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}+\frac{k\lambda\alpha}{\nu}\nu^\star t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}-\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu^2}k\lambda\alpha\nu^\star\right)\\ &\quad+\frac{k\lambda-k\lambda e^{-\nu t_{\mathrm{m}}}+k\lambda\alpha t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}\nu^\star}{c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu}\\ &=\frac{k\lambda}{\nu^2[c-c\lambda+ck\lambda+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu]^2}\\ &\quad\times[\nu(\nu^\star t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}\alpha+1-e^{-\nu t_{\mathrm{m}}})[(k-1)\lambda\nu+\lambda+\nu-1]\\ &\quad+\nu^\star(1-e^{-\nu t_{\mathrm{m}}})(1-\lambda+k\lambda\alpha-k\lambda\alpha e^{-\nu t_{\mathrm{m}}})].\end{aligned}$$

Therefore, if $(k-1)c\lambda\nu+\lambda+c\nu>1$, we have $\partial G_{\mathrm{p}}/\partial\alpha>0$.

Proof of Proposition 2. The average processing time, $\bar{T}_{\mathrm{p}}=E(T_{\mathrm{p}})$, is obtained by considering the following three cases: (i) all machines finish the jobs with a good quality; (ii) one machine finishes the job with a defect but does not stop the line; and (iii) one machine finishes the job with a defect and stops the line. Summing the resulting terms, we obtain:

$$\begin{aligned}\bar{T}_{\mathrm{p}} &= \frac{c(1-\lambda)^k+ck\lambda(1-\alpha)(1-\lambda)^{k-1}}{(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}}\\ &\quad+\frac{k\lambda\alpha(1-\lambda)^{k-1}\left(\int_0^{t_{\mathrm{m}}}(c+t)\nu e^{-\nu t}\mathrm{d}t+e^{-\nu t_{\mathrm{m}}}(c+t_{\mathrm{m}})\right)}{(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}},\\ &=\frac{1}{1-\lambda+k\lambda}\left(c-c\lambda+kc\lambda+k\lambda\alpha\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}\right).\end{aligned}$$

Similarly, for the second moment of processing times, we obtain:

$$\begin{aligned}E\left(T_{\mathrm{p}}^2\right) &= \frac{c^2(1-\lambda)^k+c^2k\lambda(1-\alpha)(1-\lambda)^{k-1}}{(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}}+\frac{k\lambda\alpha(1-\lambda)^{k-1}\left(\int_0^{t_{\mathrm{m}}}(c+t)^2\nu e^{-\nu t}\mathrm{d}t+\lambda e^{-\nu t_{\mathrm{m}}}(c+t_{\mathrm{m}})^2\right)}{(1-\lambda)^k+k\lambda(1-\lambda)^{k-1}}\\ &=\frac{c^2-c^2\lambda+k\lambda c^2+2k\lambda\alpha c(1-e^{-\nu t_{\mathrm{m}}})/\nu-2k\lambda\alpha t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}/\nu+2k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu^2}{1-\lambda+k\lambda}.\end{aligned}$$

Thus, the variance is

$$\begin{aligned}\mathrm{Var}(T_{\mathrm{p}}) &= \frac{c^2-c^2\lambda+k\lambda c^2+2k\lambda\alpha c(1-e^{-\nu t_{\mathrm{m}}})/\nu-2k\lambda\alpha t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}/\nu+2k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu^2}{1-\lambda+k\lambda}\\ &\quad-\left(\frac{c-c\lambda+kc\lambda+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu}{1-\lambda+k\lambda}\right)^2\\ &=k\lambda\alpha\left(2\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}-2t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}+2\lambda t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}-2\lambda\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}\right.\\ &\quad\left.-2k\lambda t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}+2k\lambda\frac{1-e^{-\nu t_{\mathrm{m}}}}{\nu}-k\lambda\alpha\frac{(1-e^{-\nu t_{\mathrm{m}}})^2}{\nu}\right)\Big/[\nu(1-\lambda+k\lambda)^2],\end{aligned}$$

and the coefficient of variation, is:

$$\begin{aligned}CV(T_{\mathrm{p}}) &= \sqrt{\frac{k\lambda\alpha2((1-e^{-\nu t_{\mathrm{m}}})/\nu-2t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}+2\lambda t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}-2\lambda(1-e^{-\nu t_{\mathrm{m}}})/\nu-2k\lambda t_{\mathrm{m}}e^{-\nu t_{\mathrm{m}}}+2k\lambda(1-e^{-\nu t_{\mathrm{m}}})/\nu-k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})^2/\nu}{\nu(1-\lambda+k\lambda)^2}}\\ &\quad\Big/(1/(1-\lambda+k\lambda))(c-c\lambda+kc\lambda+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})/\nu)\\ &=\frac{\sqrt{k\lambda\alpha[2(1-e^{-\mu t_{\mathrm{m}}}-t_{\mathrm{m}}\nu e^{-\nu t_{\mathrm{m}}})(1-\lambda+k\lambda)-k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})^2]}}{c\nu-c\lambda\nu+kc\lambda\nu+k\lambda\alpha(1-e^{-\nu t_{\mathrm{m}}})}.\end{aligned}$$

Biographies

Jingshan Li is a Staff Researcher in Manufacturing Systems Research Lab, General Motors Research & Development Center, Warren, Michigan. He received a BS from Tsinghua University, Beijing, China, an MS from Chinese Academy of Sciences, Beijing, China, and a PhD from University of Michigan, Ann Arbor, Michigan in 1989, 1992, and 2000, respectively. He has published over 20 referred journal articles in *IIE Transactions*, *IEEE Transactions*, *Annals of Operations Research*, *International Journal of Production Research*, etc. He is the associate editor of *Mathematical Problems in Engineering*, and *Encyclopedia of Manufacturing Systems and Supply Chain Management*. He was in the finalist of Best Automation Paper Award in 2005, IEEE International Conference on Robotics and Automation. His primary research interests are in modeling, analysis and control of complex manufacturing systems. He is a member of IIE, IEEE and INFORMS.

Dennis Blumenfeld is a staff research scientist at the General Motors R&D Center. He previously held faculty positions at Princeton University and University College London. He has a B.Sc. in mathematics and M.Sc. in statistics and operations research from Imperial College, London, and a Ph.D. in civil engineering from University College London. He is a member of the Institute for Operations Research and the Management Sciences, and a fellow of the Royal Statistical Society. He is author of a recent reference book on OR results, *Operations Research Calculations Handbook*, and has published articles on transportation modeling, traffic flow and queueing, inventory control, and production systems. He was an Edelman Award finalist for implementation of logistics network modeling, and has served on the editorial advisory board of *Transportation Research*.