MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 9, No. 1, Winter 2007, pp. 62–83
ISSN 1523-4614 | EISSN 1526-5498 | 07 | 0901 | 0062
DOI 10.1287/msom.1060.0118

Performance Evaluation of an Automatic Transfer Line with WIP Scrapping During Long Failures

George Liberopoulos, George Kozanidis
Department of Mechanical and Industrial Engineering, University of Thessaly, 38334 Volos, Greece
{glib@uth.gr, gkoz@uth.gr}

Panagiotis Tsarouhas
Department of Electronics, Technological Education Institute of Lamia,
35100 Lamia, Greece, ptsarouh@teilam.gr

We develop a model of a failure-prone, bufferless, paced, automatic transfer line in which material flows through a number of workstations in series, receiving continuous processing along each workstation. When a workstation fails, it stops operating, and so do all the other workstations upstream of it. The quality of the material trapped in the stopped workstations deteriorates with time. If this material remains immobilized beyond a certain critical time, its quality becomes unacceptable and it must be scrapped. We develop analytical expressions for important system performance measures for two cases. In the first case, the in-process material has no memory of the quality deterioration that it experienced during previous stoppages, whereas in the second case it has. In both cases, we assume that the workstation uptimes and downtimes follow memoryless distributions. We use the analytical expressions to numerically study the effect of system parameters on system performance. To evaluate the memoryless assumption, we compare the performance of the original model to that of a modified model in which the workstation downtimes do not follow memoryless distributions. The performance of the modified model is obtained via simulation.

Key words: transfer line; material scrapping; performance evaluation
History: Received: December 20, 2005; accepted: June 2, 2006.

1. Introduction

A traditional, widespread way of organizing high-volume, low-variety production is the manufacturing *flow line*, *production line*, or *transfer line*. Transfer lines require all material to visit workstations in the same sequence, thus simplifying material handling. Transfer lines are common in both discrete-parts and continuous-processing manufacturing. Discrete-parts manufacturing is characterized by individual parts that are clearly distinguishable and is often encountered in the industries of computer and electronic products, electrical equipment and appliances, transport equipment, machinery, fabricated metal, wood, furniture products, etc. Process industries, on the other hand, operate on material that is continually flowing, as is the case with petroleum and coal products, metallurgical products, nonmetallic mineral products (e.g., ceramics, glass, and cement), basic chemicals, food and beverage products, paper products, etc.

Mass production of discrete parts shares many of the characteristics of the process industries. Conversely, sometimes in process industries, fluids are processed in distinct batches that can be viewed as discrete parts. Generally, process industries are capital intensive and are concerned with capacity. With increasing production volume, it becomes economically attractive to automate individual workstations, integrate them into one system by a common automated transfer mechanism and a common control system, and link them with synchronized material movement so they can begin their tasks simultaneously. Transfer lines with these characteristics are often referred to as *synchronized* or *paced automatic transfer lines*.

The literature on transfer lines is vast. The earliest papers on transfer lines appeared in the 1950s and '60s. Notable examples of such research are Vladzievskii (1950–1951), Koenigsberg (1959), Zimmern (1956), Sevastyanov (1962), Freeman (1967), and Buzacott (1967). Much of the research on transfer lines that was done until the mid-90s can be found in the review papers by Dallery and Gershwin (1992) and Papadopoulos and Heavey (1996) and the books by Askin and Standridge (1993), Buzacott and Shanthikumar (1993), Gershwin (1994), and Altiok (1996).

Most of this work and the work that followed it dealt with the influence of buffer stock between workstations and the development of exact and approximate analytical methods for evaluating system performance, particularly *throughput* or *production capacity*. Inserting buffers in a low- to medium-speed transfer line can be very effective for increasing the actual production capacity of the line, because the buffers absorb some of the downtime of the failed workstations. An important issue that arises in designing such a line is where to insert the buffers and how big to make them to maximize the actual production capacity of the line. This issue is often referred to in the literature as the *buffer allocation problem*.

The research around this problem is very active even today. Some recent references on this topic are Helber (2000), Spinellis and Papadopoulos (2000), Papadopoulos and Vidalis (2001), Sorensen and Janssens (2001), Tolio et al. (2002), Shi and Men (2003), Tempelmeier (2003), Daskalaki and Smith (2004), Diamantidis and Papadopoulos (2004), Colledani et al. (2005), Enginarlar et al. (2005), Helber (2005), and Van Vuuren et al. (2005). Although inserting buffers in a low- to medium-speed transfer line can be beneficial, placing buffers in a high-speed line is impractical, because buffers can hold only a relatively small amount of material. Hence, they get full quickly, causing the entire line upstream of a failed workstation to stop within a short period of time, as if there were no buffers between the workstations. Moreover, in many high-speed lines, especially in the process industries, it may not be possible to store in-process material, because buffers may hurt the quality of the material by allowing it to deteriorate over time, as we will see below. For these reasons, high-speed lines generally do not have buffers between workstations, as is mentioned in Dogan and Altiok (1998); there are exceptions, however, such as the one reported in Liberopoulos and Tsarouhas (2002).

When an unexpected failure occurs in a bufferless line, the failed workstation stops and forces the part of the line upstream of the failure to go on standby mode, i.e., operate at a minimum level without transferring material. This causes a gap in production downstream. Because a workstation on standby mode is for the most part operating, except that material movement has stopped, it may fail itself.

To better see how a stopped workstation can fail, think of a failure that has occurred downstream of a furnace. The material movement in the furnace has stopped because of the failure, but the furnace itself has not been shut down, so it can fail itself. Under these conditions, it is reasonable to assume that the times between failures of the workstations are *time dependent* and not *operation dependent*.

Moreover, in many industries, when a failure takes place, the quality of the material that is trapped in the stopped section of the line deteriorates with time. If this material remains immobilized beyond a certain critical time, its quality becomes unacceptable and it must be scrapped. Besides the havoc that this will bring about, it will also create a significant additional gap in production upstream of the interruption. As a result, the actual output rate of the line can be substantially lower than its nominal production rate. The costs associated with the drop in output and the scrapping of material can be significant, especially if the scrapped material is not reusable.

We first encountered the problem of scrapping work in process (WIP) during long failures in automatic transfer lines producing bread and bakery products. The most important type of quality deterioration in such products is the rise of the dough in the stages before baking. In a study of an automated pizza transfer line, we found that the *effective input rate* of the line was equal to 95.45% of its nominal production rate, because of the unavailability of the system during failures (Liberopoulos and Tsarouhas 2005). Yet, because of the scrapping of material during long failures, the *effective output rate* of the line dropped to only 90.43% of its nominal production rate.

In other words, approximately half of the 10% drop in efficiency of the line was due to the gap in production caused by failures, while the other half was due to the gap caused by scrapping of material during long failures. In another study of a croissant transfer line, our analysis showed that the effect of scrapping was less severe, because the dough used in croissants can remain immobilized for twice as long as the dough used in pizzas without having to be scrapped, because different types of yeast are utilized in the two products (Liberopoulos and Tsarouhas 2002). We subsequently became aware that the problem of scrapping during long failures is encountered in the production of many other products in the food industry, where certain processes (e.g., pasteurization, fermentation, proofing, carbonation, etc.) must be performed in a timely and carefully controlled manner, and disruptions in the production process may cause severe quality deterioration.

In fact, there is a plethora of manufacturing processes where material is scrapped because its physical and/or chemical characteristics fall out of specifications during a stoppage. A typical type of quality deterioration is material solidification. A plant manager in a large metallurgical products industry reported a severe scrapping problem in the rolling process of his plant, where hot ingots are rolled into intermediate shapes, such as blooms, billets, and slabs, which are further rolled into plates, sheets, bar stock, structural shapes, or foils. If there is a failure somewhere in the plastic deformation process (e.g., a hot billet is blocked in a rolling mill), the material that is trapped upstream of the failure stops moving. If the failure lasts long enough, the trapped material solidifies and has to be scrapped.

A similar problem occurs in the thermosetting plastics industry. There, when a failure takes place in the molding phase (e.g., the mold starts to leak), the process upstream of the failure freezes, and the material in the die starts being polymerized or cured, which means that it solidifies. If the failure lasts long enough, the trapped material must be scraped because the polymerization process is irreversible. This phenomenon, where interruptions in one station cause state changes and quality problems in the WIP upstream of the stopped station, is also addressed by Lee and Allwood (2003) in the context of temperature-dependent processes, such as aluminum extrusion.

Another common cause of quality deterioration is the extensive exposure of material to certain environments (e.g., heat, humidity, acidity, etc.). For instance, in the process described above of rolling ingots, the ingots trapped in the furnace upstream of the plastic deformation process may have to be scrapped because of overheating.

Katok et al. (1999) report a similar situation in aluminum can transfer lines. In the washing stage of such lines, the cans are cleaned with acids in a washer to remove all the cooling fluids and lubricants that were used during the preceding cup- and body-pressing stages. If the cans remain trapped in the washer for too long because of a failure further downstream, they are overexposed to the acids in the washer. This affects their surface and consequently the quality of the graphics applied by the printer. If the damage is severe, the cans must be "canned."

In push-pull strip pickling lines in the iron, steel, and aluminum industries, the material trapped in the rinsing section may have to be scrapped, because strip staining may occur during long standstill times. In the production of machine-made glass tableware (e.g., beverage glasses), a mixture of raw material, which is referred to as glass batch, is fed into a storage bin and then melted in a furnace. Molten glass is drawn from the furnace and delivered to molding and forming equipment, where the glass is progressively formed into finished tableware, heat treated, and cooled. The mixed-batch storage bin may show dramatic signs of quality deterioration, called segregation (e.g., visual blemishes or residual stresses, seeds and blisters, and composition shifts), which increase in magnitude when the batching process is suspended for a period. In picture tube manufacturing, the aluminizing rail process is extremely precise and should not be interrupted. When an aluminizing conductor system goes down for a certain amount of time, the product on the loop during the aluminizing process becomes damaged and must be scrapped.

There are countless other examples of manufacturing processes where WIP may be damaged and may have to be scrapped because of long stoppages, so the problem is crucial from a manufacturing systems practitioner's point of view. Yet, to the best of our knowledge, the problem has not been studied from a manufacturing systems engineering perspective, perhaps because most of the research effort in this area has been directed toward discrete-parts manufacturing, where the problem of WIP scrapping during long failures is less severe than in automated continuous and semicontinuous processes.

There are a few works that explicitly consider scrapping of parts when a workstation on a transfer line fails. Okamura and Yamashina (1977) consider a two-workstation model in which when a workstation fails, the part in it is scrapped. Shanthikumar and Tien (1983) analyze a synchronous two-workstation transfer line with geometrically distributed uptimes and downtimes in which, when a workstation fails, the part in it is scrapped with a certain fixed probability. Jafari and Shanthikumar (1987) extend this analysis to a synchronous transfer line with more than two workstations and present an approximation technique to determine production rates and buffer levels. This model is also briefly discussed in Buzacott and Shanthikumar (1993, Chapter 6.5.2). Altiok (1996, Chapter 5.2) considers a two-workstation line with no intermediate buffer. Each workstation can accommodate one part. When a workstation fails, the part in it is scrapped as soon as the workstation becomes operative. Dogan and Altiok (1998) consider a pharmaceutical transfer line with several workstations in series connected with conveyor segments and circular buffers with limited bank capacity in between workstations. Each workstation can accommodate one part. When a workstation fails, the part in it is scrapped.

Apart from the above works on the relationship between scrapping and productivity, some work has been recently reported in the remotely related area of quality and productivity (Kim and Gershwin 2005) and in the more closely related area of rework/scrap and productivity. More specifically, Pourbabai (1990) describes a model with more than two workstations and nonzero buffers but assumes that if a blockage occurs, the trapped parts are scrapped. Yu and Bricker (1993) consider scrapping and/or rework for multistage systems with unlimited buffer capacity. Gopalan and Kannan (1994) present a two-workstation, zero-buffer model in which bad parts can be reworked or scrapped. Rework starts immediately at the workstation where the bad part is produced. Helber (2000) considers a transfer line in which a fixed percentage of the parts processed at some stations in the line may be scrapped or reworked at dedicated machines to meet product quality requirements. Reworked parts are not sent back into the main line. Li (2004) presents an overlapping decomposition method to approximate the throughput of production systems with rework loops but no scrapping.

In all the above works, wherever scrapping is involved, it is assumed that when a failure occurs at a workstation, a single part—the one that is on the workstation—is either always scrapped or is scrapped with a given probability, independently of the failure time.

In this paper, we address the issue of scrapping WIP material during long failures in bufferless, paced, automatic transfer lines. We consider a model of a transfer line in which material flows through different workstations in series where it receives continuous processing. No storage buffers are allowed in between the workstations. When a workstation fails, it stops operating, and so do all its upstream workstations. The quality of the material that is trapped in the stopped workstations deteriorates with time. If this material remains still in the same place for a long enough time, its quality becomes unacceptable and it must be scrapped.

The maximum-allowable standstill time of material in any workstation depends on the workstation. If this time is zero, then the material must be immediately scrapped. If it is infinite, then there is no quality deterioration and no material is ever scrapped. If the maximum-allowable standstill time is neither zero nor infinite, there are two cases to consider. In the first case, the quality deterioration during a stoppage caused by a failure is fully recoverable, so that when the workstation starts operating after the failure is repaired, the material in it—assuming that it has not been scrapped—is as good as new. In other words, the material has no memory of the damage done to it during previous stoppages. In the second case, the deterioration is cumulative, so that when the workstation starts operating again, the remaining maximum-allowable standstill time of the material in it—assuming that it has not been scrapped—is less than before the stoppage. In other words, the material has memory of the damage done to it during previous stoppages.

In practice, the second case is encountered much more often than the first. All the examples of manufacturing processes with material damaged during stoppages discussed earlier in this section belong to the second case.

To see how the first case would come about, think of a deformable material that starts stretching when a stoppage occurs. If the stoppage is not too long, the material will be elastically deformed during the stoppage, so that after the stoppage it will return to its original shape and have no memory of the deformation that it went through during the stoppage. If the stoppage is long enough, however, the material will stretch so much during the stoppage that it will be plastically deformed, so that after the stoppage it will not return to its original shape; therefore, it will no longer be acceptable and will have to be scrapped.

It is worth noting that in some cases it may be possible to install a "strip" or evacuation segment immediately following a potentially damaging workstation. Such a segment is dedicated to emptying the workstation at the end of the shift or if a downstream workstation fails. This evacuation segment is not a throughput buffer, but a quality buffer reserved for emptying the potentially destructive station. For example, oven stations in the paint shops of automotive assembly lines virtually always have evacuation segments immediately downstream. Because these evacuation segments can accommodate the maximum number of jobs in the potentially destructive station, they virtually eliminate the risk of damaging WIP due to blocking.

Of course, if the product deteriorates with time (instead of the actions or environment of the workstation), as in the food processing examples, then evacuation segments have little benefit. However, for the example of washing aluminum cans, evacuation segments could dramatically reduce the risk of mass scrapping due to machine failures elsewhere on the line. In this paper, we do not consider evacuation buffers.

The rest of this paper is organized as follows. In §2, we develop analytical expressions for key performance measures for the two cases discussed above, i.e., for the cases where material either has or does not have memory of the damage during previous stoppages, under the assumption that the uptimes and downtimes of the workstations are geometrically distributed. In §3, we use these expressions to study the effects of system parameters on system performance. To evaluate the geometric distribution assumptions, in §4 we obtain simulation-based results for the same transfer line, under the assumption that the uptimes of the workstations are not geometrically distributed. We then compare these results with the exact results that are obtained under the geometric distribution assumptions. Finally, we conclude in §5.

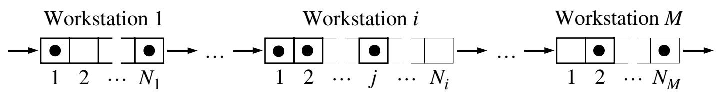
2. Transfer Line Analysis

Consider an automatic transfer line consisting of M workstations in series with no buffers between the workstations. Material flows from outside the system to workstation 1, to workstation 2 , ... , to workstation M, and then out of the system. Once material enters a workstation, it flows at a constant speed through the workstation, where it receives continuous processing, and then exits the workstation.

Time and space are discretized; specifically, the length of each workstation is broken into a number of discrete positions. Let N_i denote the number of discrete positions of workstation i. Material is also broken into discrete parts so that each position in each workstation can accommodate one part. Thus, at each workstation i, a part moves from outside the workstation to position 1, to position 2, ... , to position N_i, and then out of the workstation. The transfer line is paced so that all the workstation positions have equal and constant processing times. Time is scaled so that the processing time of every part in every position of every workstation is one time period. This period is sometimes referred to as a *cycle* in the literature. The *nominal production rate* of the line is the inverse of the cycle, which, in our case, is equal to one part per time period. An inexhaustible supply of raw parts is available upstream of the first workstation, and an unlimited storage area is present downstream of the last workstation. A schematic representation of the line is shown in Figure 1.

Workstations are subject to random failures; therefore, at any time, a workstation is either *up* or *down*. A workstation is *blocked* if one or more of its downstream workstations are down. This means that a workstation is *stopped* if it is either down or blocked, and it is *operating* if it is up and not blocked. Failures are time dependent. This means that a workstation may fail even when it is blocked. At each workstation, the times between failures, or *uptimes*, and the

Figure 1 **Schematic Representation of a Transfer Line with *M* Workstations**



times to repair, or *downtimes*, are i.i.d. random variables with geometric distributions.

Let p_i denote the probability that workstation i will be down in period t, given that it is up in period $t-1$. Let r_i denote the probability that workstation i will be up in period t, given that it is down in period $t-1$. The *availability* or *efficiency* "in isolation" of a workstation is the percentage of time that the workstation is up. Let e_i denote the efficiency in isolation of workstation i. It is easy to see that

$$e_i = \frac{r_i}{r_i + p_i}, \quad \text{for } i = 1, \ldots, M. \tag{1}$$

The "system" availability or efficiency of a workstation is the percentage of time that the workstation is operating (up and not blocked). Let E_i^d denote the system efficiency of workstation i. It is easy to see that

$$E_i^d = \prod_{k=i}^{M} e_k, \quad \text{for } i = 1, \ldots, M. \tag{2}$$

Let p_i^d be the probability that workstation i will be stopped in period t, given that it is operating in period $t-1$. It is easy to see that p_i^d is equal to one minus the probability that none of the workstations from i to M will fail in period t, given that they are all up in period $t-1$; that is,

$$p_i^d = 1 - \prod_{k=i}^{M} (1 - p_k), \quad \text{for } i = 1, \ldots, M. \tag{3}$$

Let r_i^d be the probability that workstation i will be operating in period t, given that it is stopped in period $t-1$. Then, E_i^d can be expressed in terms of p_i^d and r_i^d as

$$E_i^d = \frac{r_i^d}{r_i^d + p_i^d}, \quad \text{for } i = 1, \ldots, M,$$

from which it follows that

$$r_i^d = p_i^d \frac{E_i^d}{1 - E_i^d}, \quad \text{for } i = 1, \ldots, M, \tag{4}$$

where E_i^d and p_i^d can be computed from (2) and (3), respectively.

When a workstation is stopped, the parts in it remain still. When a part is still, its quality deteriorates with time. If a part remains still in the same place for a long enough time, its quality becomes unacceptable and it must be scrapped.

In what follows, we will consider two cases of quality deterioration. In the first case, we assume that parts have no memory of the damage done to them during previous stoppages. In the second case, we assume that parts have memory of the damage done to them during previous stoppages of the workstation that they are in. Next, we analyze these two cases separately.

2.1. Case 1: Material with No Memory of Damage During Previous Stoppages

In case 1, we assume that if a part remains still in the same position of a workstation for more than a fixed maximum-allowable standstill time associated with this workstation, it must be scrapped. If a part "survives" a stoppage at a particular position of a workstation, however, its quality condition after the stoppage is "as good as new." This means that it can remain still in the next position without having to be scrapped for up to the same fixed maximum-allowable standstill time. In other words, the material has no memory of the damage done to it during previous stoppages.

For this case, let n_i denote the fixed maximum-allowable standstill time of a part in any position of workstation i. Let R_i denote the time that it takes for workstation i to start operating once it is stopped; i.e., R_i is the stoppage time of workstation i. It is easy to see that R_i is geometrically distributed with mean $1/r_i^d$. The probability that any part in any position of workstation i will not be scrapped, given that the workstation is stopped, is given by

$$P\{R_i \le n_i\} = \sum_{t=1}^{n_i} (1 - r_i^d)^{t-1} r_i^d = 1 - (1 - r_i^d)^{n_i}, \quad \text{for } i = 1, \ldots, M. \tag{5}$$

Note that the extreme values of n_i are zero and infinity. If $n_i = 0$, any stoppage of workstation i immediately causes severe quality deterioration to all the parts in it; therefore, all parts in workstation i are immediately scrapped on stoppage of the workstation. If $n_i = \infty$, on the other hand, no stoppage of workstation i ever causes any damage to any of the parts in it; therefore, no part in workstation i is ever scrapped from the workstation. Our model includes both these situations as special cases.

Now let us follow a part from the moment it moves to a particular position of a workstation, say workstation i, at the beginning of period t, which implies that workstation i is operating at the beginning of period t. The time that the part spends in this position is one time period of processing plus an extra waiting time. This extra time is zero if the workstation keeps operating after one time period, i.e., at the beginning of period $t+1$. If the workstation is stopped at the beginning of period $t+1$, however, then the extra waiting time of the part depends on whether the stoppage time of workstation is greater than n_i. More specifically, if the time that workstation i is stopped is strictly greater than n_i, the part waits for n_i time periods and then is scrapped, because its quality has become unacceptable. If the stoppage time of workstation i is smaller than or equal to n_i, on the other hand, then the part waits on average for $E[R_i \mid R_i \le n_i]$ time periods and then moves on to the next position. The conditional expected time $E[R_i \mid R_i \le n_i]$ is given by

$$E[R_i \mid R_i \le n_i] = \frac{\sum_{t=1}^{n_i} tP\{R_i = t\}}{P\{R_i \le n_i\}} = \frac{\sum_{j=1}^{n_i} t(1-r_i^d)^{t-1} r_i^d}{1-(1-r_i^d)^{n_i}} = \frac{1}{r_i^d} - n_i \frac{(1-r_i^d)^{n_i}}{1-(1-r_i^d)^{n_i}}, \quad \text{for } i = 1, \ldots, M, \quad (6)$$

where, in the above expression, we have substituted $P\{R_i \le n_i\}$ from (5).

Let l_i denote the conditional expected time that a part spends in any position of workstation i, given that it has moved into this position. Then, in view of our discussion above, l_i is given by

$$\begin{aligned} l_i &= 1 + p_i^d (E[R_i \mid R_i \le n_i] P\{R_i \le n_i\} + n_i P\{R_i > n_i\}) \\ &= 1 + p_i^d \left(\left(\frac{1}{r_i^d} - n_i \frac{(1-r_i^d)^{n_i}}{1-(1-r_i^d)^{n_i}} \right) \cdot (1-(1-r_i^d)^{n_i}) + n_i((1-r_i^d)^{n_i}) \right) \\ &= 1 + \frac{p_i^d}{r_i^d}(1-(1-r_i^d)^{n_i}), \quad \text{for } i = 1, \ldots, M, \end{aligned} \quad (7)$$

where, in the above expression, we have substituted $P\{R_i \le n_i\}$ from (5) and $E[R_i \mid R_i \le n_i]$ from (6).

Let q_i denote the conditional probability that a part will not be scrapped from a particular position of workstation i, given that it has entered this position. This probability is given by

$$\begin{aligned} q_i &= 1 - p_i^d + p_i^d P\{R_i \le n_i\} = 1 - p_i^d + p_i^d (1-(1-r_i^d)^{n_i}) \\ &= 1 - p_i^d (1-r_i^d)^{n_i}, \quad \text{for } i = 1, \ldots, M, \end{aligned} \quad (8)$$

where, in the above expression, we have substituted $P\{R_i \le n_i\}$ from (5).

Let $q_{i,j}$ denote the conditional probability that a part will enter position j of workstation i, given that it has entered workstation i. This probability is given by

$$q_{i,j} = q_i^{j-1}, \quad \text{for } i = 1, \ldots, M, \; j = 1, \ldots, N_i. \quad (9)$$

Let $l_{i,j}$ denote the conditional expected time that a part spends in position j of workstation i, given that it has entered workstation i. Then, $l_{i,j}$ is given by

$$l_{i,j} = l_i q_{i,j} = l_i q_i^{j-1}, \quad \text{for } i = 1, \ldots, M, \; j = 1, \ldots, N_i, \quad (10)$$

where, in the above expression, we have substituted $q_{i,j}$ from (9).

Let L_i denote the conditional expected flow time of a part at workstation i, given that it has entered this workstation. It is easy to see that

$$L_i = \sum_{j=1}^{N_i} l_{i,j} = \sum_{j=1}^{N_i} l_i q_i^{j-1} = l_i \frac{1-q_i^{N_i}}{1-q_i}, \quad \text{for } i = 1, \ldots, M, \quad (11)$$

where, in the above expression, we have substituted $l_{i,j}$ from (10).

Let Q_i denote the conditional probability that a part will move from workstation i (or from outside the system, if $i = 0$) to workstation $i+1$ (or out of the system, if $i = M$), given that it has entered workstation i. In other words, Q_i is the *yield* of workstation i. It is easy to see that

$$Q_0 = 1 \quad \text{and} \quad Q_i = q_i^{N_i}, \quad \text{for } i = 1, \ldots, M.$$

Let $\widehat{Q}_i$ denote the unconditional probability that a part will move from workstation i (or from outside the system, if $i = 0$) to workstation $i+1$ (or out of the system, if $i = M$), given that it has entered the system. In other words, $\widehat{Q}_i$ is the yield of the section of the line that includes workstations 1 through i. It is easy to see that

$$\widehat{Q}_i = \prod_{k=0}^{i} Q_k, \quad \text{for } i = 0, \ldots, M. \tag{12}$$

Let $\hat{L}_i$ denote the unconditional expected flow time of a part at workstation i. It is easy to see that

$$\hat{L}_i = L_i \widehat{Q}_{i-1}, \quad \text{for } i = 1, \ldots, M. \tag{13}$$

Finally, let $\hat{L}_{TOT}$ denote the total unconditional expected flow time of a part in the line. Then,

$$\hat{L}_{TOT} = \sum_{i=1}^{M} \hat{L}_i. \tag{14}$$

The *input rate* of a workstation is the average number of parts that enter the workstation per time period. The *output rate* of a workstation is the average number of good parts that exit the workstation per time period. The *scrap rate* of a workstation is the average number of bad parts that are scrapped from the workstation per time period. Let I_i denote the input rate, let O_i denote the output rate, and let S_i denote the scrap rate of workstation i. It is easy to see that the input rate of workstation 1 is equal to the nominal production rate of the line multiplied by the efficiency of workstation 1. Because the nominal production rate is assumed to be equal to one, the input rate of workstation 1 is simply equal to its efficiency, that is,

$$I_1 = E_1^d. \tag{15}$$

It is also easy to see that

$$I_i = O_{i-1}, \quad \text{for } i = 2, \ldots, M, \tag{16}$$

$$O_i = I_i Q_i, \quad \text{for } i = 1, \ldots, M, \tag{17}$$

$$S_i = I_i(1 - Q_i), \quad \text{for } i = 1, \ldots, M. \tag{18}$$

From (12) and (15)–(18), it follows that

$$I_i = E_1^d \widehat{Q}_{i-1}, \quad \text{for } i = 1, \ldots, M, \tag{19}$$

$$O_i = E_1^d \widehat{Q}_i, \quad \text{for } i = 1, \ldots, M, \tag{20}$$

$$S_i = E_1^d \widehat{Q}_{i-1}(1 - Q_i) = E_1^d(\widehat{Q}_{i-1} - \widehat{Q}_i), \quad \text{for } i = 1, \ldots, M.$$

Note that the effective throughput or production capacity of the entire line is the output rate of the last workstation, that is,

$$O_M = E_1^d \widehat{Q}_M, \tag{21}$$

where E_i^d is the efficiency of workstation 1 and $\widehat{Q}_M$ is the yield of the entire line.

Let $\hat{S}_i$ denote the scrap rate of the line segment that includes workstations 1 to i, that is,

$$\hat{S}_i = \sum_{k=1}^{i} S_k = E_1^d \sum_{k=1}^{i} (\widehat{Q}_{k-1} - \widehat{Q}_k) = E_1^d(1 - \widehat{Q}_i). \tag{22}$$

Then that scrap rate of the entire line is given by

$$\hat{S}_M = E_1^d(1 - \widehat{Q}_M). \tag{23}$$

Let B_i denote the average number of parts in workstation i. Then, from Little's law, B_i is given by

$$B_i = I_i L_i = E_1^d L_i \widehat{Q}_{i-1} = E_1^d \hat{L}_i, \quad \text{for } i = 1, \ldots, M. \tag{24}$$

Finally, let B_{TOT} denote the average number of parts in the entire line. Then,

$$B_{TOT} = \sum_{i=1}^{M} B_i. \tag{25}$$

In our analysis above and in the rest of the paper, we have assumed that whenever a part is blocked during n_i time units, it is damaged and immediately leaves the transfer line. Other plausible assumptions are that damaged parts leave the system when exiting a workstation or when exiting the line. Such assumptions would lead to different expressions for the total unconditional expected flow time of a part in the line, $\hat{L}_{TOT}$, and the average number of parts in the entire line, B_{TOT}; however, they would not change the expression for the most important performance measure, which is the yield at the end of the line, $\widehat{Q}_M$.

2.2. Case 2: Material with Memory of Damage During Previous Stoppages Within Each Workstation

In case 2, we assume that a part has memory of the damage done to it during all the previous stoppages of the workstation that hosts it. More specifically, the maximum-allowable standstill time of any part in any position of any workstation is equal to the initial maximum-allowable standstill time associated with this workstation minus the cumulative standstill time of this part in all the preceding positions of the workstation, where the initial maximum-allowable standstill time is the maximum-allowable standstill time of the part when it enters the workstation. If the part "survives" all the stoppages at a workstation, however, its quality condition after exiting this workstation is "as good as new." This means that it can remain still in the next workstation for up to the initial maximum-allowable standstill time associated with that workstation without having to be scrapped.

The justification for this assumption is that in many real-life manufacturing processes, the causes of quality deterioration during a stoppage differ from one workstation to another. For instance, the cause of quality deterioration may be overheating in one workstation, excessive exposure to a solvent in another, too much humidity in a third, and so on. This means that if a part has survived the threat to its quality in one workstation, when it enters the next workstation, it is as good as new as far as the threat to its quality in the latter workstation is concerned. However, we can easily extend our analysis to the situation where material has memory of the damage it experiences throughout more than one workstation.

As in case 1, let n_i denote the initial maximum-allowable standstill time of a part in workstation i, and let R_i denote the time it takes for workstation i to start operating once it is stopped. Note that, as in case 1, the extreme values of n_i are zero and infinity, and that R_i is geometrically distributed with mean $1/r_i^d$.

Initially, let us assume that the issue of quality deterioration during a stoppage does not exist; therefore, a part may never be scrapped from workstation i. Let $W_{i,j}$ denote the time that workstation i is stopped, from the moment that a part enters position j of workstation i until it exits this position. Clearly,

$$W_{i,j} = \begin{cases} 0, & \text{with probability } (1-p_i^d), \\ R_i, & \text{with probability } p_i^d. \end{cases} \tag{26}$$

Let $S_{i,j}$ denote the cumulative time that workstation i is stopped from the moment that a part enters position 1 of workstation i until it exits position j of workstation i, that is,

$$S_{i,j} = \sum_{k=1}^{j} W_{i,k}. \tag{27}$$

We are interested in finding the probability distribution of $S_{i,j}$, i.e., $P\{S_{i,j} = k\}$, $k = 0, 1, \ldots$. First of all, it is easy to see that

$$P\{S_{i,j} = 0\} = (1-p_i^d)^j, \quad \text{for } i = 1, \ldots, M,\ j = 1, \ldots, N_i. \tag{28}$$

To find $P\{S_{i,j} = k\}$, for $k > 0$, let us follow a part from the moment it enters position 1 of workstation i until it exits position j of the same workstation. Consider the event that, in its trajectory from position 1 to position j, the part does not stop in m out of the j positions and stops in the remaining $j-m$ positions, and the total time that the part is stopped is k, where $k > 0$. The probability of this event is

$$\frac{j!}{m!(j-m)!}(1-p_i^d)^m (p_i^d)^{j-m} \frac{(k-1)!}{(j-m-1)!(k-j+m)!} \cdot (1-r_i^d)^{k-(j-m)} (r_i^d)^{j-m}. \tag{29}$$

In the above expression, the first term, $[j!/(m! \cdot (j-m)!)](1-p_i^d)^m (p_i^d)^{j-m}$, is the probability that the part does not stop in m positions and stops in the remaining $j-m$ positions, which is given by the probability function of the binomial distribution. The second term, $[(k-1)!/((j-m-1)!(k-j+m)!)] \cdot (1-r_i^d)^{k-(j-m)} (r_i^d)^{j-m}$, is the probability that the time until the $(j-m)$th resumption of operation of workstation i following a stoppage is equal to k, which is given by the probability function of the Pascal or negative binomial distribution. For a definition of the binomial and the negative binomial distributions, see any introductory book on probability, such as Hines and Montgomery (1990). Note that if $k \ge j$, then m can take any value from 0 to $j-1$, whereas if $k < j$, then m

can take any value from $j-k$ to $j-1$. In other words, m can take any value from $\max(0, j-k)$ to $j-1$. With this in mind, the desired probability, $P\{S_{i,j}=k\}$ for $k>0$, is the sum of the probabilities given by expression (29) over all possible values of m; that is,

$$P\{S_{i,j}=k\} = \sum_{m=\max(0,j-k)}^{j-1} \frac{j!}{m!(j-m)!}(1-p_i^d)^m (p_i^d)^{j-m} \cdot \frac{(k-1)!}{(j-m-1)!(k-j+m)!}(1-r_i^d)^{k-(j-m)}(r_i^d)^{j-m}, \quad \text{for } i=1,\ldots,M,\ j=1,\ldots,N_i,\ k=1,2,\ldots. \tag{30}$$

Now let us bring back into the foreground our original assumption that a part *is* scrapped from workstation i if its cumulative standstill time in this workstation is greater than n_i. The analysis that is needed to compute the conditional expected flow time of a part at workstation i, L_i, is similar to the analysis carried out in the no memory case in §2.1. The only difference is that each time a part enters a new position j of workstation i, where $j=2,\ldots,N_i$, its remaining maximum-allowable standstill time is $n_i - S_{i,j-1}$ instead of n_i, which was the situation in the no memory case, except when $j=1$, where the two cases are identical. With this in mind, let $l_{i,j}$ be defined as in the no memory case. For $j=1$, the remaining maximum-allowable standstill time of the part is n_i, as in the no memory case; therefore, $l_{i,1}$ is given by the same expression as that derived for l_i in the no memory case, i.e., expression (7). Hence,

$$l_{i,1} = 1 + \frac{p_i^d}{r_i^d}(1-(1-r_i^d)^{n_i}), \quad \text{for } i=1,\ldots,M. \tag{31}$$

For $j=2,\ldots,N_i$, let $l_{i,j,k}$ denote the conditional expected time that a part spends in position j of workstation i, given that it has entered workstation i and that the cumulative stoppage time until it enters position j is k, i.e., $S_{i,j-1}=k$. Then, $l_{i,j,k}$ is given by

$$l_{i,j,k} = \begin{cases} 1 + \dfrac{p_i^d}{r_i^d}(1-(1-r_i^d)^{n_i-k}), \\ \qquad \text{for } i=1,\ldots,M,\ j=2,\ldots,N_i,\ k \le n_i, \\ 0, \quad \text{for } k > n_i, \end{cases} \tag{32}$$

where the above expression is the same as (7) except that we have replaced n_i by $n_i - k$. Then, $l_{i,j}$, for $j=2,\ldots,N_i$, is given by

$$l_{i,j} = \sum_{k=0}^{n_i} l_{i,j,k} P\{S_{i,j-1}=k\} = \sum_{k=0}^{n_i}\left(1+\frac{p_i^d}{r_i^d}(1-(1-r_i^d)^{n_i-k})\right)P\{S_{i,j-1}=k\}, \quad \text{for } i=1,\ldots,M,\ j=2,\ldots,N_i, \tag{33}$$

where, in the above expression, we have substituted $l_{i,j,k}$ from (32), and $P\{S_{i,j}=k\}$ can be computed from (30).

Let L_i be defined as in the no memory case. It is easy to see that

$$L_i = \sum_{j=1}^{N_i} l_{i,j}, \quad \text{for } i=1,\ldots,M, \tag{34}$$

where $l_{i,j}$ is given by (33).

Finally, let Q_i be defined as in the no memory case. It is easy to see that

$$Q_0 = 1 \quad \text{and} \quad Q_i = P\{S_{i,N_i} \le n_i\} = \sum_{k=0}^{n_i} P\{S_{i,N_i}=k\}, \quad \text{for } i=1,\ldots,M.$$

The rest of the analysis is identical to the analysis of the no memory case. In other words, Equations (12)–(25), which were derived for the no memory case, also hold for the memory case.

3. Effect of Parameters

In this section, we use the expressions that we developed in §2 to investigate how the system parameters, p_i, r_i, and n_i, $i=1,\ldots,M$, affect the system's performance. The system performance measures that are of particular interest are (a) the system efficiency of workstation 1, E_1^d, which from (15) is equal to the input rate capacity of the line, I_1; (b) the yield of the line, $\hat{Q}_M$, which, when multiplied with E_1^d, gives the output rate capacity of the line, O_M (see Equation (21)); and (c) the scrap rate of the line, $\hat{S}_M$. A performance measure of less importance is the total unconditional expected flow time, $\hat{L}_{TOT}$. Finally, the average number of parts in the entire line, B_{TOT}, is perhaps the least significant performance measure.

To explore the effect of system parameters on the above system performance measures, we carried out a numerical study of an instance of a bufferless, paced, automatic transfer line with $M = 6$ identical workstations with parameters: $N_i = 10$, $p_i = 0.1$, $r_i = 0.8$, and $n_i = 10$, $i = 1, \ldots, M$. We refer to this system as the "nominal" system. For the nominal system, we varied parameter p_i from 0.04 to 0.16 in steps of 0.02 and parameter r_i from 0.3 to 0.9, in steps of 0.1, for one workstation i at a time, for $i = 1, \ldots, M$. We also varied parameter n_i from 1 to 25 in steps of one, for all workstations at the same time. In all our tests, we observed the system's performance measures, for cases of both material with and without memory of the damage during a stoppage. The results are plotted in Figures 2–18.

From Figures 2–18, we can observe how the system parameters p_i, r_i, and n_i affect the system's performance. First, let us look at p_i. As p_i increases, workstation i fails more frequently; therefore, it and the entire section of the line upstream stop more frequently. This causes a drop in the efficiency of workstations $1, \ldots, i$ and hence in the input rate capacity of the system, E_1^d (Figure 2). The increase in the frequency of stoppages also raises the frequency of repairs, because every stoppage is followed by a repair of the failed workstation. As the frequency of repairs increases, statistically the number of downtimes in any time interval increases. This raises the probability that at least one of the downtimes will exceed the maximum-allowable standstill time of one of the stopped workstations. Naturally, this causes a decrease in the yield of the stopped section of the line.

Figure 2 Input Rate Capacity of the Line, E_1^d, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for Both the No Memory and Memory Cases

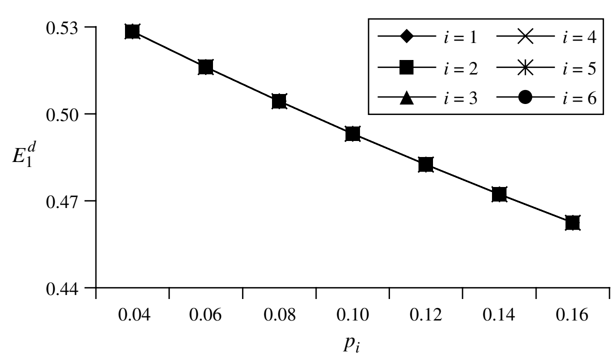
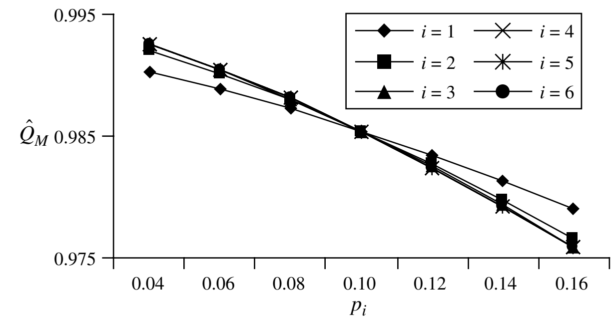


Figure 3 Yield of the Line, $\hat{Q}_M$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the No Memory Case



Eventually, it also decreases the yield of the entire line, $\hat{Q}_M$ (Figures 3 and 4), as well as the average number of parts in the entire line, B_{TOT}, although this is not shown here for space considerations.

The effect of p_i on the scrap rate of the entire line, $\hat{S}_M$, and on the total unconditional expected flow time, $\hat{L}_{TOT}$, is somewhat less straightforward. First, let us look at $\hat{S}_M$. From Equation (23), $\hat{S}_M$ is equal to the product $E_1^d(1 - \hat{Q}_M)$. As was mentioned above, as p_i increases, the input rate capacity of the system (effective availability of workstation 1), E_1^d, and the yield (proportion of good parts) of the entire line, $\hat{Q}_M$, decrease, which means that the proportion of bad parts, $1 - \hat{Q}_M$, increases. Therefore, as p_i increases, $\hat{S}_M$ tends to both decrease (as E_1^d decreases and fewer parts enter the system) and increase (because $1 - \hat{Q}_M$ increases and a larger proportion of the parts that enter the system are scrapped) at the same time. Which of the two tendencies predominates depends on whether the "decrease-of-input-rate-capacity" effect or the "decrease-of-yield" effect

Figure 4 Yield of the Line, $\hat{Q}_M$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the Memory Case

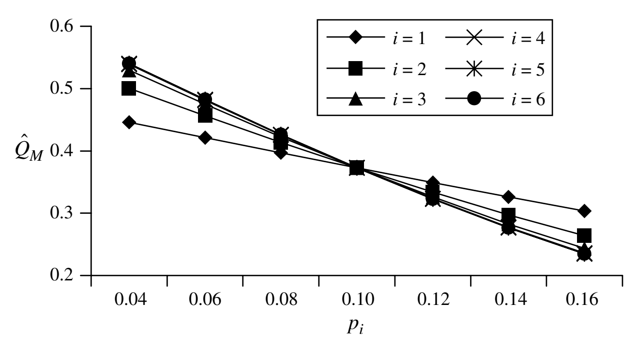


Figure 5 Scrap Rate of the Line, $\hat{S}_M$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the No Memory Case

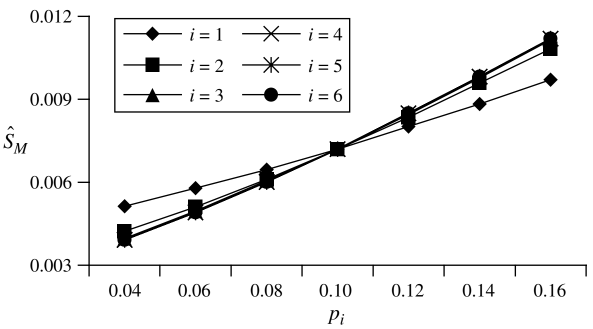
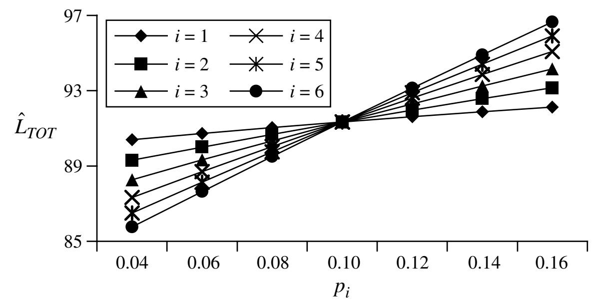


Figure 7 Total Unconditional Expected Flow Time, $\hat{L}_{TOT}$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the No Memory Case



is stronger. In the transfer line instance that we studied, the decrease-of-yield effect seems to predominate in both the no memory case and the memory case; hence, as p_i increases, the scrap rate of the entire line, $\hat{S}_M$, increases (Figure 5 and Figure 6). It is worth mentioning, however, that we have constructed counter-instances where the decrease-of-input-rate-capacity effect predominates, so as p_i increases, the scrap rate of the entire line, $\hat{S}_M$, decreases.

Now, let us look at $\hat{L}_{TOT}$. As was mentioned above, an increase in p_i raises the frequency of stoppages and causes a decrease in the input rate capacity of the system. This raise causes the parts that are trapped in the stopped section of the line to spend more time in the line, and this increases $\hat{L}_{TOT}$. On the other hand, as was also mentioned above, an increase in p_i reduces the probability that a part will exit from the stopped section of the line, and this decreases $\hat{L}_{TOT}$. In other words, as p_i increases, $\hat{L}_{TOT}$ tends to both increase and decrease at the same time, as was the case with $\hat{S}_M$. Which of the two tendencies predominates depends on whether the decrease-of-input-rate-capacity effect or the decrease-of-yield effect is stronger. In the transfer line instance that we studied, the decrease-of-input-rate-capacity effect seems to predominate in the no memory case, because, as p_i increases, $\hat{L}_{TOT}$ increases (Figure 7).

In the memory case, on the other hand, the decrease-of-yield effect predominates, because, as p_i increases, $\hat{L}_{TOT}$ decreases (Figure 8). This is natural, because in the memory case, scrapping is more intense than in the no memory case. In fact, the difference in scrapping rates between the no memory and memory cases becomes evident when one compares the range of values of $\hat{Q}_M$ for the two cases. Namely, $\hat{Q}_M$ ranges approximately between 96% and 99% in the no memory case (Figure 3) and between 16% and 54% in the memory case (Figure 4). Of course, this example is rather extreme, because a 16% line yield would not be acceptable in most real manufacturing systems. It was used on purpose to emphasize the difference between the two cases. An example with more

Figure 6 Scrap Rate of the Line, $\hat{S}_M$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the Memory Case

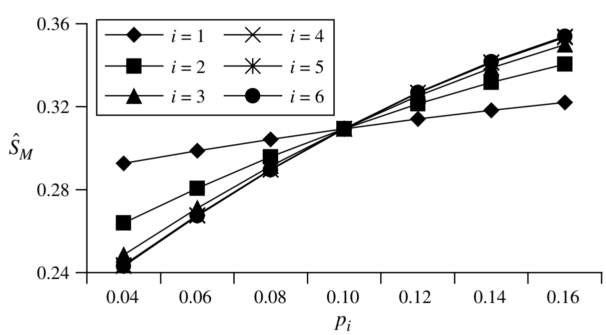


Figure 8 Total Unconditional Expected Flow Time, $\hat{L}_{TOT}$, vs. Failure Probability of Workstation i, p_i, for $i = 1, \ldots, M$, for the Memory Case

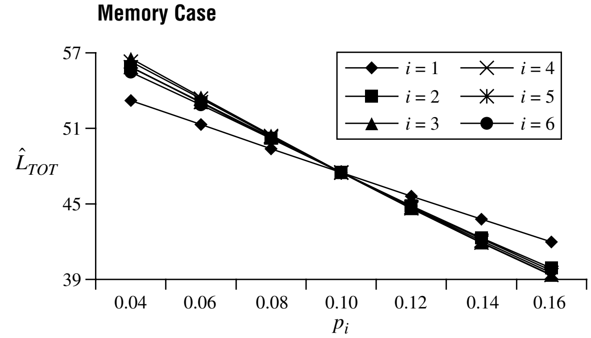


Figure 9 Input Rate Capacity of the Line, E_1^d, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for Both the No Memory and Memory Cases

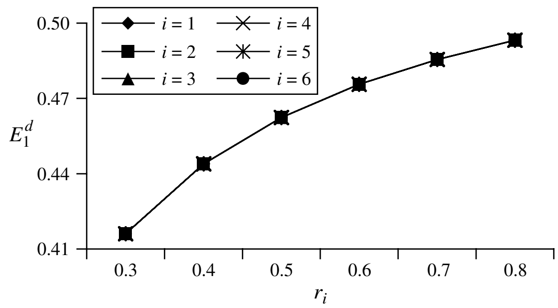
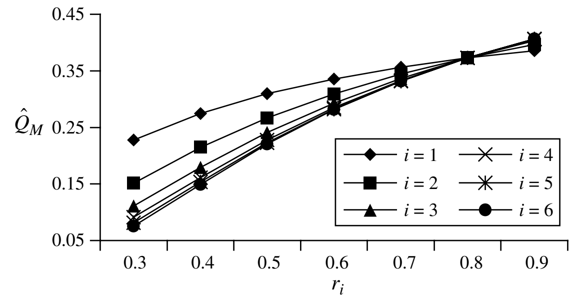


Figure 11 Yield of the Line, $\hat{Q}_M$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the Memory Case



realistic parameter values is discussed in the following section, where the difference in scrapping rates between the two cases is actually very small.

The effect of the repair probability r_i on system performance is the opposite from the effect of p_i. Namely, as r_i increases, E_1^d and $\widehat{Q}_M$ increase (Figures 9–11) and so does B_{TOT}, although this is not shown here for space considerations. Also, as r_i increases, $\widehat{S}_M$ and $\hat{L}_{TOT}$ tend to both increase and decrease at the same time depending on whether the increase-in-input-rate-capacity effect or the increase-of-yield effect predominates. In the transfer line instance that we studied, as far as $\widehat{S}_M$ is concerned, the increase-of-yield effect seems to predominate in both the no memory and the memory cases, because in both cases, as r_i increases, the scrap rate of the entire line, $\widehat{S}_M$, decreases (Figures 12 and 13). As far as $\hat{L}_{TOT}$ is concerned, in the no memory case, sometimes one effect predominates and sometimes the other, depending on the workstation i and its repair rate r_i (Figure 14). In the memory case, however, the increase-of-yield effect is predominant, because as r_i increases, $\hat{L}_{TOT}$ always increases (Figure 15).

A noteworthy observation is that the effects of p_i and r_i on system performance described above are more intense for downstream workstations than for upstream workstations. This is natural, because the further downstream a workstation is, the greater the impact of its failures on the entire transfer line, as, when a workstation fails, all its upstream workstations are stopped. This means that efforts to improve the uptime and downtime distribution parameters should start first at the last workstation and then move upstream toward the first workstation. The only exceptions are the effects of p_i and r_i on E_1^d, which seems to be independent of i (Figures 2 and 9). This, however, is true only because in the particular transfer line instance that we examined all workstations i have the same parameters p_i and r_i. If these parameters were different, the curves in Figures 2 and 9 would be different for each workstation i.

Another noteworthy observation is that the effects of p_i and r_i on system performance are more intense in the memory case than they are in the no memory

Figure 10 Yield of the Line, $\widehat{Q}_M$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the No Memory Case

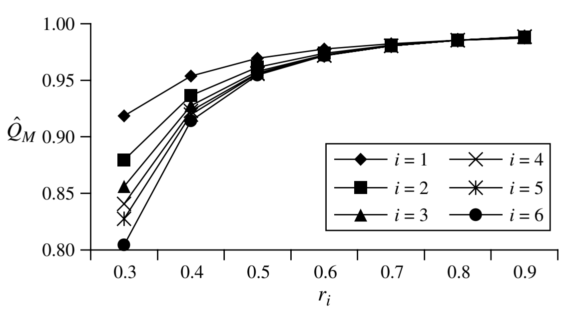


Figure 12 Scrap Rate of the Line, $\widehat{S}_M$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the No Memory Case

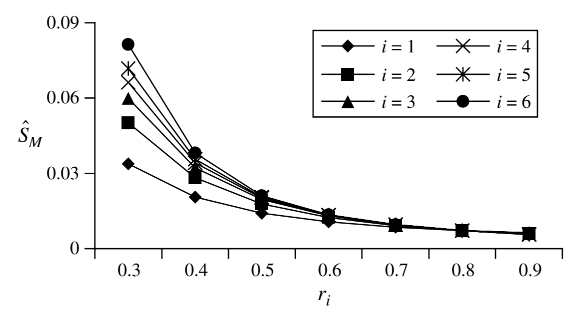
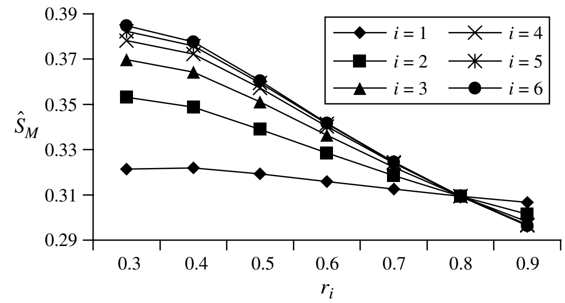


Figure 13 Scrap Rate of the Line, $\hat{S}_M$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the Memory Case



case, because scrapping is more intense in the former case than in the latter. This means that improvements in the uptime and downtime distribution parameters can bring in a bigger performance gain in the memory case than in the no memory case.

A qualitative difference between the effects of p_i and r_i on the input rate capacity of the line, E_1^d, and the yield of the line, $\widehat{Q}_M$, is that, although all the curves of these two performance measures versus p_i and r_i are concave, in the case of p_i they appear to be almost linear. This implies that the marginal gain in performance (capacity and yield) that can be achieved by decreasing the probability of failure p_i (e.g., by improving maintenance operations) is close to constant in p_i. On the other hand, the marginal gain in performance that can be achieved by increasing the probability of repair r_i (e.g., by speeding up repair operations) is clearly decreasing in r_i.

A question that arises is whether it is more beneficial to use the repair/maintenance crew for preventive maintenance to reduce p_i or for speedy repair to increase r_i. To answer this question, we must compare

Figure 14 Total Unconditional Expected Flow Time, $\hat{L}_{TOT}$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the No Memory Case

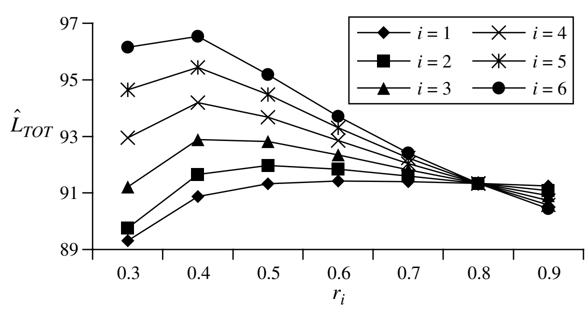
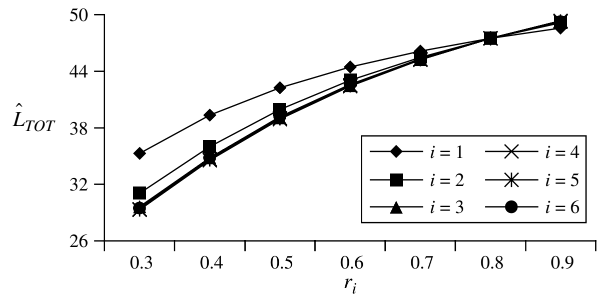


Figure 15 Total Unconditional Expected Flow Time, $\hat{L}_{TOT}$, vs. Repair Probability of Workstation i, r_i, for $i = 1, \ldots, M$, for the Memory Case



the partial derivatives of the system performance measure of interest with respect to p_i and r_i.

To illustrate such a comparison, consider a system that is identical to the nominal system except that we allow the failure and repair probabilities of workstation 1, p_1 and r_1, to vary. For this system, suppose that the performance measure of interest is the effective production capacity of the entire line, O_M. Table 1 shows the numerically computed partial derivatives of O_M with respect to p_1 and r_1, for different combinations of p_1 and r_1 values, for the no memory case. The first line in each entry shows $\partial O_M/\partial p_1$, and the second line shows $\partial O_M/\partial r_1$. From this table, we can see that for all combinations of p_1 and r_1 values, $\partial O_M/\partial p_1 < 0$ and $\partial O_M/\partial r_1 > 0$. This is natural, because any decrease in p_1 or increase in r_1 improves the reliability characteristics of workstation 1 and results in an increase in the effective production capacity of the entire line. Moreover, we can see that for all combinations of p_1 and r_1 values, $|\partial O_M/\partial p_1| > |\partial O_M/\partial r_1|$. This means that the increase in O_M that can be achieved if we decrease p_1 by an infinitesimal amount Δ is greater than the increase in O_M that can be achieved if we increase r_1 by the same *absolute* amount Δ.

Let us illustrate this for $p_1 = 0.04$ and $r_1 = 0.3$. If we decrease p_1 from 0.04 to $0.04 - \Delta$, O_M will increase by 1.7211Δ. If, on the other hand, we increase r_1 from 0.3 to $0.3 + \Delta$, O_M will increase only by 0.26964Δ.

The above observation is reversed, however, if we consider how O_M is affected by a *relative* decrease in p_1 and the same relative increase in r_1. Let us illustrate this for $p_1 = 0.04$ and $r_1 = 0.3$. If we decrease p_1 by an infinitesimal percentage δ from 0.04 to $0.04 - 0.04\delta$, O_M will increase by $(1.7211)(0.04\delta) = 0.0688\delta$. However, if we increase r_1 by the same infinitesimal per-

Table 1 **Partial Derivatives of the Effective Production Capacity of the Entire Line, O_M, with Respect to p_1 and r_1, for Different Combinations of p_1 and r_1 Values, for the No Memory Case**

	r_1						
p_1	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.04	−1.7211	−1.2994	−1.0452	−0.8747	−0.7521	−0.6595	−0.5871
	0.26964	0.14483	0.09109	0.06281	0.04601	0.0352	0.02782
0.06	−1.6349	−1.2375	−0.9973	−0.8365	−0.721	−0.6338	−0.5655
	0.39044	0.21049	0.13233	0.09125	0.06689	0.05122	0.04051
0.08	−1.5507	−1.1807	−0.9536	−0.8013	−0.6921	−0.6097	−0.5452
	0.49906	0.27229	0.17134	0.11812	0.08659	0.06632	0.05249
0.1	−1.4647	−1.1269	−0.9129	−0.7685	−0.6651	−0.5871	−0.526
	0.59264	0.3298	0.20822	0.14357	0.10523	0.08061	0.06382
0.12	−1.3765	−1.0746	−0.8745	−0.7379	−0.6397	−0.5657	−0.5078
	0.66989	0.38253	0.24297	0.1677	0.12291	0.09414	0.07455
0.14	−1.2875	−1.0235	−0.838	−0.709	−0.6158	−0.5455	−0.4906
	0.73099	0.43013	0.27554	0.19055	0.13969	0.107	0.08473
0.16	−1.1992	−0.9735	−0.803	−0.6816	−0.5931	−0.5264	−0.4741
	0.77711	0.47241	0.30586	0.21215	0.15562	0.1192	0.09439

centage δ from 0.3 to $0.3+0.3\delta$, O_M will increase by $(0.26964)(0.3\delta)=0.080889\delta$; i.e., the increase of O_M is greater in the second case.

Qualitatively similar results were obtained when we varied the failure and repair probabilities of other workstations. The only difference is that the further downstream the workstation is, the higher the magnitude of the partial derivatives of O_M with respect to the failure and repair probabilities of the workstation. This is natural, because, as we mentioned above, the further downstream a workstation is, the greater the impact of its failures on the entire transfer line. To illustrate this, consider a system that is identical to the nominal system except that we allow the failure and repair probabilities of workstation 6, p_6 and r_6, to vary. For this system, Table 2 shows the numerically computed partial derivatives of O_M with respect to p_6 and r_6 values, for the no memory case. For space consideration, we show the results for fewer combinations of p_6 and r_6 values than the combinations of p_1 and r_1 values shown in Table 1. By comparing Table 2 with Table 1, we can see that the partial derivatives in Table 2 are greater in magnitude than the corresponding partial derivatives in Table 1.

Finally, the effect of the maximum-allowable standstill time n_i on system performance is straightforward. Namely, as n_i increases, $\hat{Q}_M$ and $\hat{L}_{TOT}$ increase (Figures 16 and 18), whereas $\hat{S}_M$ decreases (Figure 17), converging to certain limiting values as n_i goes to infinity. B_{TOT} also increases, but this is not shown here for space considerations. The limiting values that the performance measures approach as n_i goes to infinity correspond to the performance measures of the

Table 2 **Partial Derivatives of the Effective Production Capacity of the Entire Line, O_M, with Respect to p_6 and r_6, for Different Combinations of p_6 and r_6 Values, for the No Memory Case**

	r_6	
p_6	0.3	0.9
0.04	−2.1786	−0.5987
	0.3803	0.02885
0.16	−1.6552	−0.4897
	1.32875	0.1022

Figure 16 **Yield of the Line, $\hat{Q}_M$, vs. Maximum-Allowable Standstill Time, n_i, for Both the No Memory and Memory Cases**

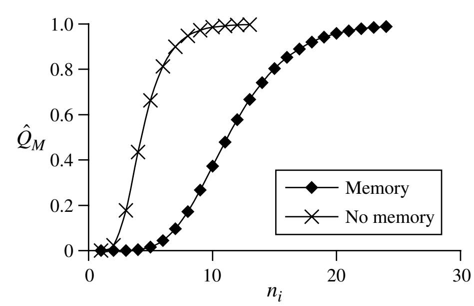
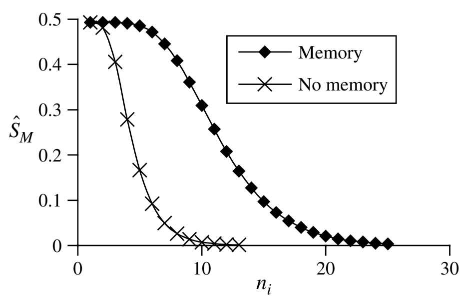


Figure 17 Scrap Rate of the Line, $\hat{S}_M$, vs. Maximum-Allowable Standstill Time, n_i, for Both the No Memory and Memory Cases

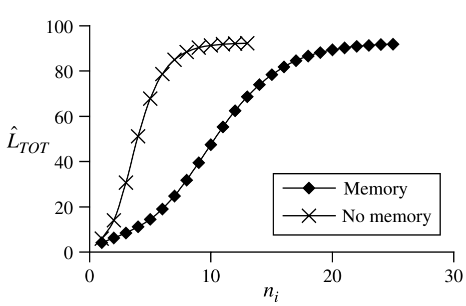


same system with no scrapping. Note that the convergence rate is much faster for the no memory case than for the memory case. This means that the marginal gain in performance that can be achieved by increasing the maximum-allowable standstill time n_i (e.g., by improving the material and/or process characteristics) is more significant in the no memory case than in the memory case.

4. Simulation Study

Thus far in this paper, we have assumed that the uptimes and downtimes of the workstations are i.i.d. random variables with geometric distributions. This means that the probability that a workstation will fail in the next time period, given that it is presently up, does not depend on the number of periods that it has been up, and that the probability that a workstation will be repaired in the next time period, given that it is presently down, does not depend on the number of periods that it has been down. This implies that the vector of operational states of the workstations is a discrete-time Markov process. Without the Markovian property, the mathematical analysis becomes much more difficult. An important question is how realistic the Markovian assumption is and how it affects the performance of the system.

Figure 18 Total Unconditional Expected Flow Time, $\hat{L}_{TOT}$, vs. Maximum-Allowable Standstill Time, n_i, for Both the No Memory and Memory Cases



In a study of failure data of a transfer line producing transmission cases at Chrysler Corporation, reported by Hanifin (1975) and Buzacott and Hanifin (1978), it was found that the observed uptimes of machines could be reasonably well approximated by exponential distributions, where the exponential distribution is the equivalent of the geometric distribution in continuous time. That same study also found that the observed downtimes were due to a mixture of different repairs, where the time to make each repair could be reasonably well approximated by exponential distributions. In another study of failure data of two automotive body welding lines (Inman 1999), it was concluded that the exponential distribution assumption appeared reasonable for the uptimes but not for the downtimes. The above studies concerned asynchronous transfer lines.

In a study of failure data of an automatic transfer line producing pizzas (Liberopoulos and Tsarouhas 2005), it was found that the coefficient of variation of the uptimes of all workstations was between 1 and 1.7, with an average of 1.33; however, the uptimes were not "pure" because they included the times between successive proactive maintenance operations on the line. The same analysis showed that the coefficient of variation of the downtimes of all workstations was between 0.5 and 1, with an average of 0.7. Recall that the coefficient of variation of an exponentially distributed random variable is one. In both Inman (1999) and Liberopoulos and Tsarouhas (2005), it was concluded that the downtimes and uptimes of the workstations are independent. In another study that analyzed actual downtime data from a paper-manufacturing facility over a period of three years (Boyd and Radson 1998), the gamma distribution was identified as a good fit for downtimes, and the best fit shape parameters for the gamma distribution ranged from 1.299 to 3.069.

Based on the above studies, it appears that although the Markovian assumption may not be too unreasonable for the uptimes, it is certainly less suitable for the downtimes. To evaluate the Markovian assumption that was used to develop the analytical expressions of the original transfer line model in §2, we compared the performance of that model to the performance of a modified model in which the downtimes of each workstation are i.i.d. random variables that are distributed as the sum of two independent, identical, geometrically distributed random variables.

To obtain the performance of the modified model we used simulation. We chose to model the downtime distributions of the workstations as the sum of two geometric distributions for two reasons—first, because the sum of two geometric distributions is the discrete counterpart of the sum of two exponential distributions, which is equivalent to a gamma distribution with shape parameter equal to 2. As mentioned above, the gamma distribution with a shape parameter of 2 was reported as a realistic fit in Boyd and Radson (1998). Second, if one wanted to extend the analytical results developed in §2 to systems with non-Markovian downtime distributions, a good bet for obtaining analytically tractable exact or approximate results would be to assume that the downtimes are distributed as the sum of two geometrically distributed random variables. Our simulation-based results could then be used to validate such potential analytical results.

With this in mind, let $1/s_i$ denote the mean of each of the two independent, geometrically distributed random variables whose sum has the same distribution as the downtime of workstation i in the modified model. The variance of each random variable is $(1-s_i)/s_i^2$, because each variable is geometrically distributed. Given that the random variables are independent, the mean and variance of their sum are $2/s_i$ and $2(1-s_i)/s_i^2$, respectively. In order for the comparison between the original and the modified models to be fair, we require that the mean downtime of each workstation be the same for both models. Recall that the downtimes of each workstation i in the original transfer line model are geometrically distributed with mean $1/r_i$ and therefore variance $(1-r_i)/r_i^2$. Because the mean downtime is the same for both models, i.e., $2/s_i = 1/r_i$, it follows that $s_i = 2r_i$. With this in mind, we can write the following expression for the ratio of the variances of the downtimes of the modified and the original model:

$$\frac{\text{Var of downtimes in modified model}}{\text{Var of downtimes in original model}} = \frac{2(1-s_i)/s_i^2}{(1-r_i)/r_i^2} = \frac{2(1-2r_i)/4r_i^2}{(1-r_i)/r_i^2} = \frac{1-2r_i}{2-2r_i}. \quad (35)$$

The above ratio is zero, when $r_i = 0.5$ and approaches 0.5 as r_i goes to zero. This means that the variance of the downtimes of the modified model is at most half the variance of the downtimes of the original model.

To compare the performance of the original and modified models, we performed a numerical study of an instance of a bufferless, paced, automatic transfer line with $M = 6$ identical workstations. Each workstation i, $i = 1, \ldots, M$, had $N_i = 30$ positions, a mean uptime of 1,600 time units, a mean downtime of 30 time units, and a maximum-allowable standstill time n_i. We considered four different values for n_i, namely, 10, 20, 40, and 50. The first two values correspond to the case where the maximum-allowable standstill time is smaller than the mean downtime. The last two values correspond to the case where the maximum-allowable standstill time is greater than the mean downtime. For each value of n_i, we considered both cases of material with and without memory of damage during a stoppage; i.e., we considered a total of eight sets of parameters. Our choice for the parameter values used in the numerical study was inspired by the failure data statistics of a real automated transfer line reported in Liberopoulos and Tsarouhas (2005).

For each set of parameters, we computed the system performance measures, E_1^d, $\widehat{Q}_M$, $\widehat{S}_M$, $\hat{L}_{TOT}$, and B_{TOT}, for both the original and the modified models. For the modified model we used simulation. For the original model we used the analytical expressions developed in §2, as well as simulation, to verify our simulation algorithm. More specifically, for each set of parameters, we ran 60 simulation replications of both the original and the modified models and computed 95% confidence intervals for the system performance measures. Each replication had a time horizon of 100 million time units and took approximately 18 minutes on a Pentium IV PC @ 2.66 GHz. The results are reported in Table 3. The third column of Table 3 indicates the model and method used to evaluate the performance measures; "Org anl" denotes the analytical

Table 3 Performance Measures of the Original and Modified Models for Different Values of n_i and Both Cases of Material With and Without Memory of Damage During a Stoppage

n_i	Memory of damage	Model & method	E_1^d	$\hat{Q}_M$	$\hat{S}_M$	$\hat{L}_{TOT}$	B_{TOT}
10	No	Org anl	0.894529	0.753069	0.220887	153.750000	137.534000
		Org sim	0.894427	0.754470	0.219609	153.949547	137.696622
			±0.000067	±0.000138	±0.000112	±0.015072	±0.022615
		Mod sim	0.894409	0.712180	0.257429	148.982286	133.251008
			±0.000052	±0.000126	±0.000102	±0.013308	±0.018718
	Yes	Org anl	0.894529	0.752624	0.221285	153.674000	137.466000
		Org sim	0.894475	0.754173	0.219886	153.885355	137.646487
			±0.000051	±0.000115	±0.000097	±0.011732	±0.016345
		Mod sim	0.894387	0.711827	0.257738	148.928328	133.199440
			±0.000049	±0.000119	±0.000098	±0.012446	±0.017453
20	No	Org anl	0.894529	0.815279	0.165238	163.595000	146.340000
		Org sim	0.894462	0.816809	0.163858	163.785563	146.499965
			±0.000053	±0.000125	±0.000104	±0.011719	±0.018038
		Mod sim	0.894425	0.783496	0.193646	160.460665	143.519916
			±0.000043	±0.000125	±0.000105	±0.011666	±0.016116
	Yes	Org anl	0.894529	0.814089	0.166303	163.395000	146.161000
		Org sim	0.894388	0.815486	0.165027	163.575056	146.299482
			±0.000057	±0.000117	±0.000097	±0.010921	±0.017912
		Mod sim	0.894419	0.782194	0.194810	160.233243	143.315585
			±0.000043	±0.000112	±0.000094	±0.010193	±0.014779
40	No	Org anl	0.894529	0.899484	0.089914	176.721000	158.082000
		Org sim	0.894437	0.899814	0.089610	176.723203	158.067620
			±0.000061	±0.000090	±0.000077	±0.008160	±0.016412
		Mod sim	0.894451	0.902673	0.087055	178.220523	159.409557
			±0.000055	±0.000096	±0.000082	±0.008514	±0.015025
	Yes	Org anl	0.894529	0.897274	0.091891	176.355000	157.754000
		Org sim	0.894411	0.897571	0.091614	176.355364	157.734225
			±0.000064	±0.000082	±0.000068	±0.007382	±0.016228
		Mod sim	0.894456	0.898835	0.090488	177.605143	158.859928
			±0.000049	±0.000096	±0.000082	±0.008083	±0.014192
50	No	Org anl	0.894529	0.926535	0.065717	180.897000	161.817000
		Org sim	0.894437	0.926349	0.065877	180.818517	161.730827
			±0.000049	±0.000063	±0.000054	±0.005439	±0.011344
		Mod sim	0.894435	0.938894	0.054656	183.437015	164.072415
			±0.000045	±0.000076	±0.000066	±0.007368	±0.011564
	Yes	Org anl	0.894529	0.924187	0.067817	180.510000	161.471000
		Org sim	0.894452	0.924115	0.067876	180.450259	161.404019
			±0.000049	±0.000080	±0.000070	±0.005949	±0.012211
		Mod sim	0.894458	0.934722	0.058389	182.783067	163.491779
			±0.000046	±0.000070	±0.000061	±0.005976	±0.010751

evaluation of the original model, "Org sim" denotes the simulation-based evaluation of the original model, and "Mod sim" denotes the simulation-based evaluation of the modified model. From Table 3, we can make the following observations.

The first observation is that the analytically obtained performance measures of the original model slightly miss the fairly tight confidence intervals of the simulation-obtained performance measures of the same model. The reason for this discrepancy is that the sample means of the uptimes and downtimes of the different workstations generated for the simulation were slightly different than the values assumed in the analytical model, which were 1,600 and 30 time units, respectively, even though a sophisticated random number generator was employed in the simulation. Specifically, the sample mean of the uptimes were 1,601.6514, 1,600.7875, 1,599.5739, 1,598.5693, 1,599.4704, and 1,601.0981 for workstations 1–6, respectively, instead of 1,600.00. Similarly, the sample means of the downtimes were 30.0186, 30.0400, 30.0180, 30.0068, 30.0344, and 30.0183 for workstations 1–6, respectively, instead of 30.00. This slight apparent bias of the randomly generated uptimes and downtimes is due to the fact that the probabilities of failure and repair of the workstations are quite small, particularly the former. When we used the above observed values of the sample mean uptimes and downtimes in our analytical expressions, then the analytically obtained performance measures of the original model fell at the heart of the confidence intervals of the simulation-obtained performance measures of the same model.

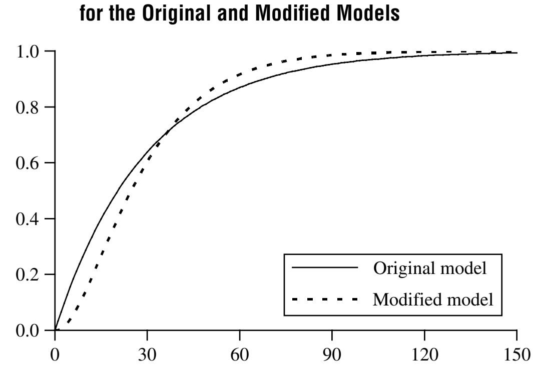
The second observation is that for all sets of parameters, the no memory case has higher values of $\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT} and lower values of $\widehat{S}_M$ than the memory case, as is expected (see also the discussion of results in §3). The difference in performance between the two cases, however, is very small. This is because the mean uptime of each workstation—i.e., 1,600 time units—is very large compared to the number of positions of the workstation—i.e., 30. Thus, it is very unlikely that a part will experience more than one stoppage in any particular workstation. Note that if every part experiences at most one stoppage in any particular workstation, then the two cases (with and without memory) behave identically.

Another observation is that as n_i increases, $\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT} increase, whereas $\widehat{S}_M$ decreases, as is expected (see also our discussion of the results in §3). When n_i is equal to 40 or 50—i.e., greater than the mean downtime—$\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT} are higher in the modified model than in the original model, whereas $\widehat{S}_M$ is higher in the original model than in the modified model. When n_i is equal to 10 or 20, i.e., smaller than the mean downtime, however, the opposite is true; i.e., $\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT} are lower in the modified model than in the original model, whereas $\widehat{S}_M$ is higher in the modified model than in the original model. At first this may seem a little surprising, because one would expect that the values of $\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT} should always be higher in the modified model than in the original model and the values of $\widehat{S}_M$ should always be higher in the original model than in the modified model, as the downtimes in the modified model have less than half the variance of the downtimes in the original model. In fact, in the transfer line instance that we studied, the ratio of the two variances, computed from (35), is 0.4828.

However, if we look at this issue more carefully, we will note that when n_i is equal to 10 or 20, the probability that a part will not be scrapped, which is equal to the cumulative distribution function of the downtimes evaluated at n_i, is higher in the original model than in the modified model, as is shown in Figure 19, precisely because of the higher variance of the downtime in the original model. This means that when n_i is equal to 10 or 20, fewer parts are scrapped in the original model than in the modified model. Those parts that are not scrapped in the original model but that are scrapped in the modified model increase the values of $\widehat{Q}_M$, $\hat{L}_{TOT}$, and B_{TOT}, making them higher than their respective values in the original model.

To better see how a lower variance of the downtimes can lead to more scrapping, think of the extreme case where the variance is zero, i.e., where all downtimes are exactly equal to 30, while the maximum-allowable standstill time is either 10 or 20. In this case, every time there is a stoppage, clearly all parts in the stopped section of the line are scrapped.

Figure 19 **Cumulative Distribution Function of Workstation Downtimes for the Original and Modified Models**



Perhaps the most important observation, from the point of view of the usefulness of analytical results of §2, is that the difference in the performance measures between the original and the modified models is quite small, especially when n_i is equal to 40 or 50, i.e., greater than the mean downtime. When n_i is equal to 10 and 20—i.e., smaller than the mean downtime—the difference in performance between the two models is bigger but still less than 6%. This suggests that the original, analytically tractable model can provide fairly accurate performance estimates for transfer lines with nongeometrically distributed downtimes.

5. Conclusions

We developed analytical expressions for two variants of a failure-prone, bufferless, paced, automatic transfer line model. The model assumes memoryless workstation uptimes and downtimes and material quality deterioration during failures that may lead to WIP scrapping if the failures last long.

We used these expressions to study the effect of system parameters on system performance. We found that as the probability of failure of a workstation increases or as the probability of its repair decreases, the input rate capacity of the system, the yield of the line, and the average number of parts in the entire line decrease. The scrap rate of the line and the total unconditional expected flow time may increase or decrease, depending on whether the decrease-of-input-rate-capacity effect or the decrease-of-yield effect predominates. We found that the input rate capacity and the yield of the line are close to linear in the probability of failure of a workstation, but clearly concave in the probability of repair. This implies close to constant performance gains as the failure probabilities decrease, but diminishing gains as the repair probabilities increase.

We also found that all these effects are more intense for downstream workstations than for upstream workstations. This means that efforts to improve the uptime and downtime distribution parameters should start first at the last workstation and then move upstream toward the first workstation. Finally, we found that transfer lines processing material that has memory of the damage during stoppages perform worse than lines processing material having no memory of the damage. Also, the marginal gains in system performance that are obtained by increasing the maximum-allowable standstill time of material are more important for material with no memory of the damage than for material with memory.

To evaluate the memoryless assumption of workstation uptimes and downtimes, we compared the performance of the original model with that of a modified model in which the workstation downtimes do not follow a memoryless distribution. The performance of the modified model was obtained via simulation. We found that when the maximum-allowable standstill time is greater than the mean downtime, more parts are scrapped in the original model than in the modified model, whereas when the maximum-allowable standstill time is smaller than the mean downtime, the opposite is true. More important, we found that the differences in the performance measures between the original and the modified models were fairly small. This suggests that the original model, which is analytically tractable, can provide fairly accurate performance estimates for transfer lines with downtimes that do not follow memoryless distributions, and this makes it a powerful analytical tool for designing such lines.

A promising direction for future research would be to extend to analysis the cases where in-process storage buffers or evacuation buffers are allowed in between workstations. Another direction for future research would be to try to develop exact or approximate analytical results for systems with non-Markovian downtime distributions. Finally, it would be worthwhile to use the analytical results developed in §2 to optimally design automatic transfer lines.

To this end, one would need to develop an average profit (reward minus cost) function of the parameters of the system, such as the nominal production rate of the line, the length and the maximum-allowable standstill time of each workstation, and the failure and repair probabilities of the workstations, and then try to optimize it. The average reward function could be related to the revenue of selling good items, which is in turn related to effective production capacity of the entire line, O_M, given by (21). The average operational cost may have several components. One component could be related to the material cost of scrapped items, which is in turn related to the scrap rate of the entire line, given by (23). This cost could be reduced if the scrapped material had some value, for example, if a part of it could be retransformed into raw material or if it could be sold to a less demanding customer for a lower price. Another component could be related to the preventive maintenance and repair service levels of the repair/maintenance crew, which determine the failure and repair probabilities of the workstations.

A third component of the average cost could be related to the resources that are needed to increase the maximum-allowable standstill time of each workstation. For example, refrigeration or heating or some other intervention on the environmental conditions might help delay the deterioration of the parts that are trapped in the stopped workstations. Finally, the cost of accrued investment and equipment depreciation could also be accounted for in the model.

References

Altiok, T. 1996. *Performance Analysis of Manufacturing Systems*. Springer Series in Operations Research, Springer, New York.

Askin, R. G., C. R. Standridge. 1993. *Modeling and Analysis of Manufacturing Systems*. Wiley, New York.

Boyd, A. H., D. Radson. 1998. Statistical analysis of downtime severity rates. *Internat. J. Production Res.* **36**(3) 603–615.

Buzacott, J. A. 1967. Automatic transfer lines with buffer stocks. *Internat. J. Production Res.* **5** 183–200.

Buzacott, J. A., L. E. Hanifin. 1978. Models of automatic transfer lines with inventory banks: A review and comparison. *AIIE Trans.* **10**(2) 197–207.

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

Colledani, M., A. Matta, T. Tolio. 2005. Performance evaluation of production lines with finite buffer capacity producing two different products. *OR Spectrum* **27**(2–3) 243–263.

Dallery, Y., S. B. Gershwin. 1992. Manufacturing flow line systems: A review of models and analytical results. *Queueing Systems* **12** 3–94.

Daskalaki, S., J. M. Smith. 2004. Combining routing and buffer allocation problems in series-parallel queueing networks. *Ann. Oper. Res.* **125** 47–68.

Diamantidis, A. C., C. T. Papadopoulos. 2004. A dynamic programming algorithm for the buffer allocation problem in homogeneous asymptotically reliable serial production lines. *Math. Problems Engrg.* (3) 209–223.

Dogan, E., T. Altiok. 1998. Blocking policies in pharmaceutical transfer lines. *Ann. Oper. Res.* **79** 323–347.

Enginarlar, E., J. Li, S. M. Meerkov. 2005. Lean buffering in serial production lines with non-exponential machines. *OR Spectrum* **27**(2–3) 195–219.

Freeman, M. C. 1967. The effect of breakdowns and interstage storage on production line capacity. *J. Indust. Engrg.* **15** 194–200.

Gershwin, S. B. 1994. *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliffs, NJ.

Gopalan, M. N., S. Kannan. 1994. Expected duration analysis of a two-stage transfer-line production system subject to inspection and rework. *J. Oper. Res. Soc.* **45**(7) 797–805.

Hanifin, L. E. 1975. Increased transfer line productivity using systems simulation. Unpublished doctoral dissertation, University of Detroit, Detroit, MI.

Helber, S. 2000. Approximate analysis of unreliable transfer lines with splits in the flow of material. *Ann. Oper. Res.* **93** 217–243.

Helber, S. 2005. Analysis of flow lines with Cox-2-distributed processing times and limited buffer capacity. *OR Spectrum* **27**(2–3) 221–242.

Hines, W. W., D. C. Montgomery. 1990. *Probability and Statistics in Engineering and Management Science*. Wiley, New York.

Inman, R. R. 1999. Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. *Production Oper. Management* **8**(4) 409–432.

Jafari, M. A., J. G. Shanthikumar. 1987. An approximate model of multi-stage automatic transfer lines with possible scrapping of work pieces. *IIE Trans.* **19**(3) 252–265.

Katok, E., T. Serrander, M. Wennstrom. 1999. Throughput improvement and scrap reduction in aluminum can manufacturing. *Production Inventory Management J.* **40**(1) 36–42.

Kim, J., S. B. Gershwin. 2005. Integrated quality and quantity modeling of a production line. *OR Spectrum* **27**(2–3) 287–314.

Koenigsberg, E. 1959. Production lines and internal storage—A review. *Management Sci.* **5** 410–433.

Lee, W. L., J. M. Allwood. 2003. Lean manufacturing in temperature dependent processes with interruptions. *Internat. J. Oper. Production Management* **23**(11) 1377–1400.

Li, J. S. 2004. Performance analysis of production systems with rework loops. *IIE Trans.* **36**(8) 755–765.

Liberopoulos, G., P. Tsarouhas. 2002. System analysis speeds up Chipita's food-processing line. *Interfaces* **32**(3) 62–76.

Liberopoulos, G., P. Tsarouhas. 2005. Reliability analysis of an automated pizza processing line. *J. Food Engrg.* **69**(1) 79–96.

Okamura, K., H. Yamashina. 1977. Analysis of the effect of buffer storage capacity in transfer line systems. *AIEE Trans.* **9** 127–135.

Papadopoulos, H. T., C. Heavey. 1996. Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines. *Eur. J. Oper. Res.* **92** 1–27.

Papadopoulos, H. T., M. I. Vidalis. 2001. A heuristic algorithm for the buffer allocation in unreliable unbalanced production lines. *Comput. Indust. Engrg.* **41**(3) 261–277.

Pourbabai, B. 1990. Optimal utilization of a finite capacity integrated assembly system. *Internat. J. Production Res.* **28**(2) 337–352.

Sevastyanov, B. A. 1962. Influence of storage bin capacity on the average standstill time of a production line. *Theory Probab. Appl.* **7** 429–438.

Shanthikumar, J. C., C. C. Tien. 1983. An algorithmic solution to two-stage transfer lines with possible scrapping of units. *Management Sci.* **29** 1069–1086.

Shi, L. Y., S. Men. 2003. Optimal buffer allocation in production lines. *IIE Trans.* **35**(1) 1–10.

Sorensen, K., G. K. Janssens. 2001. Buffer allocation and required availability in a transfer line with unreliable machines. *Internat. J. Production Econom.* **74**(1–3) 163–173.

Spinellis, D. D., C. T. Papadopoulos. 2000. A simulated annealing approach for buffer allocation in reliable production lines. *Ann. Oper. Res.* **93**(1–4) 373–384.

Tempelmeier, H. 2003. Practical considerations in the optimization of flow production systems. *Internat. J. Production Res.* **41**(1) 149–170.

Tolio, T., A. Matta, S. B. Gershwin. 2002. Analysis of two-machine lines with multiple failure modes. *IIE Trans.* **34**(1) 51–62.

Van Vuuren, M., I. J. B. F. Adan, S. A. E. Resing-Sassen. 2005. Performance analysis of multi-server tandem queues with finite buffers and blocking. *OR Spectrum* **27**(2–3) 315–338.

Vladzievskii, A. P. 1950–1951. The theory of internal stocks and their influence on the output of automatic lines. *Stanki i Instrumenty* **21**(12) 4–7, **22**(1) 16–17.

Yu, K.-Y. C., D. L. Bricker. 1993. Analysis of a Markov chain model of a multistage manufacturing system with inspection, rejection, and rework. *IIE Trans.* **25**(1) 109–112.

Zimmern, B. 1956. Etudes de la propagation des arrêts aléatoires dans les chaînes de production. *Revue de Statistique Appliquée* **4** 85–104.