Transient Behavior of Two-Machine Geometric Production Lines

Semyon M. Meerkov, *Fellow, IEEE*, Nahum Shimkin, *Senior Member, IEEE* and Liang Zhang, *Member, IEEE*

Abstract

Production system transients characterize the process of reaching the steady state throughput. Reducing transients' duration is important in a number of applications. This paper is intended to analyze transients in serial production lines with machines obeying the geometric reliability model. The Markov chain approach is used, and the second largest eigenvalue of the transition matrices is utilized to characterize the transients. Due to large dimensionality of the transition matrices, only two-machine systems are addressed, and the second largest eigenvalue is investigated as a function of the breakdown and repair probabilities. Conditions under which shorter up- and downtimes lead to faster transients are provided.

Index Terms

Production lines, Geometric reliability model, Production rate, Transient behavior, Effects of up- and downtime.

I. Introduction

Production systems often operate in transient regimes. Examples include paint shops of automotive assembly plants, where some buffers are emptied at the end of each shift due to

S. M. Meerkov is with Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA (e-mail: smm@eecs.umich.edu).

N. Shimkin is with Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: shimkin@ee.technion.ac.il).

L. Zhang is with Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA (e-mail: liangzh@uwm.edu).

The work of S. M. Meerkov and L. Zhang was partially supported by the NSF under Grant CMMI-0758259.

technological requirements; this leads to production losses in the subsequent shift (until the buffer occupancy reaches its steady state). Another example is provided by machining departments operating with so-called floats, where additional work-in-process is built up by slow machines after the end of a shift in order to prevent starvations of fast machines in the subsequent shift, leading to increased production during the transients. Clearly, to quantify the performance of these systems, a method for analysis of their transients is necessary.

Unfortunately, the literature offers very few publications in this regard. Specifically, using the idea of Markov process absorption time, transients in one-machine production systems were studied in [1]. An algorithm for solving the partial differential equation, which describes the evolution of the probability density function of a buffer with Markov-modulated input and output flows, was developed in [2]. The closest to the current study is paper [3], which investigates transients of serial production lines with machines obeying the Bernoulli reliability model. According to this model, each machine, when neither starved nor blocked, produces a part during a cycle time (i.e., the time necessary to process a part) with probability p and fails to do so with probability $1-p$, irrespective of what had happened in the previous cycle time. Thus, Bernoulli machines are memoryless, which simplifies the analysis of the resulting systems. The Bernoulli reliability model is applicable in situations, where the machine downtime is comparable with its cycle time. Such situations typically occur in painting and assembly operations, where the downtime is mostly due to quality problems or conveyor pallet jams. When the downtime is due to breakdowns, the machine typically stays down much longer than the cycle time, making the Bernoulli reliability model not applicable. In such situations, the geometric or exponential reliability models are used (see [4]-[8]), whereby machines stay up and down during geometrically or exponentially distributed periods of time. These models are applicable in many manufacturing systems, such as machining, heat treatment, washing, etc. Thus, an extension of the results reported in [3] is necessary. This is carried out in this Technical Note for machines obeying the geometric reliability model. Due to the complexity of the resulting mathematical description, only the case of two-machine systems is addressed; longer lines are beyond the scope of this Note.

The outline of this paper is as follows: Section II presents the model and the problem formulation. In Section III, transients of individual machines are analyzed. Sections IV and V are devoted to two-machine lines with small and large buffers, respectively. The conclusions and future work are given in Section VI. All proofs and numerical justifications are included in the Appendix.

II. Model and Problem Formulation

A. Model

We consider a two-machine serial production line where parts are processed by the first machine, stored in the buffer, and then processed by the second machine. To facilitate the mathematical study of the system, we introduce the following assumptions:

(i) Both machines, m_1 and m_2, have identical cycle time, τ. The time axis is slotted with slot duration τ. The state of each machine is determined at the beginning of each time slot.

(ii) Both machines obey the geometric reliability model, i.e., if $s(n) \in \{0 = \text{down}, 1 = \text{up}\}$ denotes the state of a machine at time slot n, the transition probabilities are given by

$$P[s(n+1) = 0 | s(n) = 1] = P, \quad P[s(n+1) = 1 | s(n) = 1] = 1 - P,$$

$$P[s(n+1) = 1 | s(n) = 0] = R, \quad P[s(n+1) = 0 | s(n) = 0] = 1 - R,$$

where P and R are referred to as the breakdown and repair probabilities, respectively. Both machines operate independently from each other.

(iii) The buffer, b, which separates the machines, is characterized by its capacity $1 \leq N < \infty$. The state of the buffer is determined at the end of each time slot.

(iv) m_1 is never starved; it is blocked during a time slot if it is up and b is full.

(v) m_2 is never blocked; it is starved during a time slot if it is up and b is empty.

Note that these assumptions imply, in particular, that the failures are *time-dependent* and the *blocked before service* convention is used; that is why $N \geq 1$. Note also that the average up- and downtime of the machines are $T_{up} = 1/P$ and $T_{down} = 1/R$ (in units of cycle times), respectively, and the machine efficiency is $e = R/(R+P) = T_{up}/(T_{up}+T_{down})$. Finally, note that the assumption on identical cycle time is observed in many production systems with automated material handling [8].

B. Problems

Given the above model, the production system at hand is described by an ergodic Markov chain. As it is well known (see [3]), the transients of such a system are characterized by the second largest eigenvalue (SLE) of its transition matrix. With this in mind, the problems addressed in this Technical Note are as follows:

- Analyze the second largest eigenvalue of an individual geometric machine as a function of P and R. In particular, investigate the effect of T_{up} and T_{down} on SLE, under the assumption that the machine efficiency e is fixed. (Note that the effects of T_{up} and T_{down} on SLE could not be analyzed in the framework of the Bernoulli reliability model, since it is defined by one parameter only.)
- Carry out similar analyses for two-machine lines. In addition, investigate explicitly the transients of the production rate, $PR(n)$, $n = 1, 2, \ldots$, i.e., the probability that m_2 is up and the buffer is not empty at time slot n.

The steady state production rate, $PR(\infty) =: PR_{ss}$, of a production line defined by assumptions (i)-(v) can be evaluated using the method developed in [9]. Here, we are interested in how $PR(n)$, $n = 1, 2, \ldots$, approaches the steady state value PR_{ss}.

The interest in the effect of T_{up} and T_{down} on the transients stems from the following: It is well known (see [8]) that for serial lines with unreliable machines and finite buffers,

- for a fixed e, shorter T_{up} and T_{down} lead to a larger PR_{ss} than longer ones;
- decreasing T_{down} by a factor leads to a larger PR_{ss} than increasing T_{up} by the same factor.

Do similar effects exist in the case of transients as well? In other words, do shorter T_{up} and T_{down} lead to faster transients than longer ones? These questions are answered in this paper.

III. Transients of Individual Machines

Let $x_i(n)$, $i \in \{0, 1\}$, be the probability that the machine is in state i during time slot $n = 1, 2, \ldots$. Then, the evolution of the state $x(n) = [x_0(n)\ x_1(n)]^T$ can be described by

$$x(n+1) = Ax(n), \quad x_0(n) + x_1(n) = 1, \tag{1}$$

$$A = \begin{bmatrix} 1-R & P \\ R & 1-P \end{bmatrix}. \tag{2}$$

The eigenvalues of A are

$$\lambda_0 = 1, \quad \lambda_1 = 1 - P - R, \tag{3}$$

and, therefore, the dynamics of the machine states can be expressed as

$$x_0(n) = (1-e) + [x_0(0) - (1-e)](1-P-R)^n = (1-e)\left(1 - \frac{\Delta}{1-e}\lambda_1^n\right), \quad (4)$$

$$x_1(n) = e + [x_1(0) - e]\,(1-P-R)^n = e\left(1 + \frac{\Delta}{e}\lambda_1^n\right), \quad (5)$$

$$\Delta = x_1(0) - e = (1-e) - x_0(0). \quad (6)$$

To investigate the effects of up- and downtime on the transients, consider λ_1 as a function of R for a fixed e, i.e., $\lambda_1(R) = 1 - \frac{R}{e}$. The behavior of $|\lambda_1|$ as a function of R is illustrated in Figure 1. From this figure, we conclude:

- For $0 < R < e$, $|\lambda_1|$ is decreasing in R. Thus, longer up- and downtimes lead to longer transients.
- For $R = e$, $\lambda_1 = 0$. Thus, the machine has no transients. Such a machine can be viewed as a Bernoulli machine.
- For $e < R < 1$, the evolution of the machine states is oscillatory (since $\lambda_1 < 0$) and, more importantly, shorter up- and downtimes lead to longer transients.

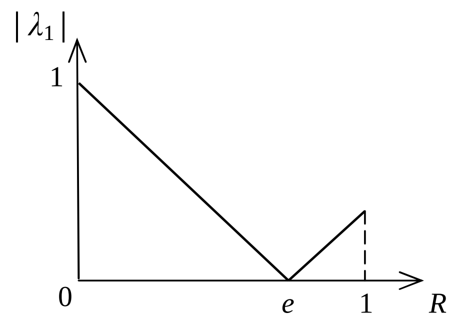


Fig. 1. Behavior of $|\lambda_1|$ as a function of R

Next, we address the issue of separate effects of uptime and of downtime on the transients. Recall that, as mentioned in Section II, increasing the uptime by a factor $(1+\alpha)$, $\alpha > 0$, or decreasing the downtime by the same factor leads to the same steady state performance for an individual machine since the new efficiency, e', in both cases is the same, i.e.,

$$e' = \frac{1}{1 + \frac{T_{down}}{(1+\alpha)T_{up}}}. \quad (7)$$

However, the transient properties resulting from both cases are different. Indeed, consider a geometric machine with breakdown and repair probabilities P and R, respectively. Let λ_1^u denote the SLE of the machine with the uptime increased by $(1+\alpha)$, $\alpha > 0$ and λ_1^d denote the SLE for the same machine with the downtime decreased by the same factor. Then,

Theorem 3.1: For a geometric machine,

$$|\lambda_1^u| > |\lambda_1^d|, \tag{8}$$

if

$$e > 0.5, \quad \frac{T_{down}}{1+\alpha} > 2. \tag{9}$$

This theorem implies that if the machine efficiency is larger than 0.5 and the decreased downtime is larger than two cycle times, decreasing the downtime leads to faster transients than increasing the uptime, while preserving the steady state production rate in both cases the same. We note that condition (9) generically takes place on the factory floor.

To conclude this section, we evaluate the eigenvalues of a system consisting of two geometric machines operating independently. In this case, the state of the system is $x(n) = [x_{00}(n)\ x_{01}(n)\ x_{10}(n)\ x_{11}(n)]^T$, where $x_{ij}(n)$ denotes the probability that m_1 is in state i and m_2 is in state j during time slot n. The transition matrix is:

$$A = \begin{bmatrix} (1-R)^2 & P(1-R) & P(1-R) & P^2 \\ (1-R)R & (1-P)(1-R) & RP & P(1-P) \\ R(1-R) & RP & (1-P)(1-R) & (1-P)P \\ R^2 & (1-P)R & (1-P)R & (1-P)^2 \end{bmatrix}, \tag{10}$$

which implies that the four eigenvalues are:

$$1,\ 1-P-R,\ 1-P-R,\ (1-P-R)^2. \tag{11}$$

These eigenvalues are used in Sections IV and V for the analysis of transients in two-machine geometric serial lines.

IV. Transients of Two-Machine Lines with $N = 1$

For a serial line with two geometric machines, the state of the system is denoted by a triple (h, s_1, s_2), where $h \in \{0, 1, \dots, N\}$ is the state of the buffer and $s_i \in \{0, 1\}$, $i = 1, 2$, are the states of the first and the second machine, respectively. The behavior of the system is described by an ergodic Markov chain. In this section, we study the case of $N = 1$. This case is important, since it corresponds to the case of bufferless systems. Indeed, under the blocked before service assumption (see Section II), the first machine itself serves as a buffer of capacity 1 and, therefore, the case $N = 1$ describes systems with no additional buffering between the two machines.

For $N = 1$, the system state can be defined as $x(n) = [x_{000}(n)\, x_{001}(n)\, x_{010}(n)\, x_{011}(n)\, x_{100}(n)\, x_{101}(n)\, x_{110}(n)\, x_{111}(n)]^T$, where $x_{hij}(n)$ denotes the probability that the buffer has h parts, $s_1 = i$ and $s_2 = j$. The transition matrix is:

$$A = \begin{bmatrix} A_1 & \mathbf{0} & A_2 & \mathbf{0} \\ \mathbf{0} & A_3 & \mathbf{0} & A_4 \end{bmatrix}, \tag{12}$$

where

$$A_1 = \begin{bmatrix} (1-R)^2 & (1-R)\,P \\ (1-R)\,R & (1-R)\,(1-P) \\ R\,(1-R) & R\,P \\ R^2 & R\,(1-P) \end{bmatrix}, \quad A_2 = \begin{bmatrix} (1-R)\,P \\ (1-R)\,(1-P) \\ R\,P \\ R\,(1-P) \end{bmatrix},$$

$$A_3 = \begin{bmatrix} (1-R)\,P & P^2 & (1-R)^2 \\ R\,P & P\,(1-P) & (1-R)\,R \\ (1-P)\,(1-R) & (1-P)\,P & R\,(1-R) \\ (1-P)\,R & (1-P)^2 & R^2 \end{bmatrix}, \quad A_4 = \begin{bmatrix} (1-R)\,P & P^2 \\ R\,P & P\,(1-P) \\ (1-P)\,(1-R) & (1-P)\,P \\ (1-P)\,R & (1-P)^2 \end{bmatrix}$$

and **0**'s are zero-matrices of appropriate dimensionality. The eight eigenvalues of A are

$$[1,\, 1-P-R,\, 1-P-R,\, (1-P-R)^2,\, (1-R)^2,\, 0,\, 0,\, 0], \tag{13}$$

where the algebraic and geometric multiplicities of all repeated eigenvalues are equal.

Clearly, the two eigenvalues $1-P-R$ represent, as it follows from Section III, the dynamics of the individual machines; the eigenvalue $(1-P-R)^2$ represents the transients of a pair of individual machines (note that the states of the machines in model (i)-(v) are determined independently); therefore, the remaining non-zero eigenvalue $(1-R)^2$ can be viewed as describing the transients of the buffer. The last statement is supported by the following two arguments:

First, using the notations $\lambda_m = 1-P-R$, $\lambda_b = (1-R)^2$, the transients of the states can be represented as

$$x_{hij}(n) = x_{hij}\left(1 + B\lambda_b^n + C\lambda_m^n + D(\lambda_m^2)^n\right), \quad h \in \{0,1\},\, i,j \in \{0,1\},\, n = 0,1,2,\ldots, \tag{14}$$

where $x_{hij} = \lim_{n\to\infty} x_{hij}(n)$ and B, C and D are constants defined by initial conditions.

Theorem 4.1: Consider a serial line defined by assumptions (i)-(v) and $N = 1$. Assume that initially the machines are in the steady states, i.e.,

$$P[s_1(0) = 1] = P[s_2(0) = 1] = e. \tag{15}$$

Then, in expression (14), $C = D = 0$, $\forall i, j, h \in \{0, 1\}$.

Thus, if the machines are in the steady states, the eigenvalue $(1 - R)^2$ indeed characterizes the transients of the buffer.

The second argument is as follows: Recall that if $R = e$, the machines can be viewed as obeying the Bernoulli reliability model. In this case, the machines have no transients, and the transients of the system are defined by $\lambda_b = (1-e)^2$, which, as it follows from [3], is equivalent to the Bernoulli case with $p = e$.

From (13), it is not immediately clear which of the eigenvalues is the SLE. Obviously, the SLE can be either $1 - P - R$ or $(1 - R)^2$, i.e., either λ_m or λ_b. Which one, in fact, is the SLE depends on the relationship between P and R. To investigate when λ_m or λ_b is SLE, consider the simplex $0 < P < R < 1$ in the (P, R)-plane (see Figure 2). Each point (P, R) implies $e > 0.5$ and each line, $P = kR$, $k < 1$, represents a set of points (P, R) with identical efficiency $e = \frac{1}{1+k}$. Let λ_1 denote the SLE, i.e., $|\lambda_1| = \max\{|\lambda_m|, \lambda_b\}$. Then, it can be shown that

$$|\lambda_1| = \begin{cases} \lambda_m, & \text{if } 0 < P < R(1-R), \\ \lambda_b, & \text{if } R(1-R) < P < (1-R)(2-R), \\ -\lambda_m, & \text{if } (2-R)(1-R) < P < 1. \end{cases} \tag{16}$$

This leads to the partitioning of the simplex according to SLE as shown in Figure 2. Specifically, in area I, the transients of the system are defined mostly by an individual machine; in area II, the transients are defined mostly by the buffer; in area III, the transients are again defined mostly by the machine, however, since the eigenvalue in this area is negative, the transients in area III are oscillatory.

Next, we characterize the effects of shorter and longer up- and downtimes on the duration of the transients.

Theorem 4.2: Consider a serial line defined by assumptions (i)-(v) and $N = 1$. Then, for any fixed $e > 0.5$, the SLE is a monotonically decreasing function of R for $R \in (0, 0.5)$.

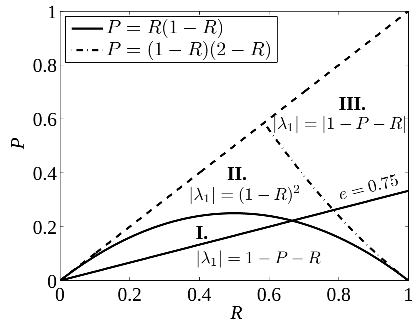
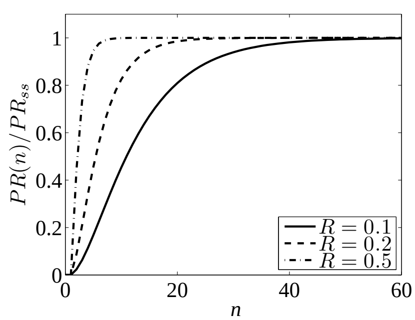
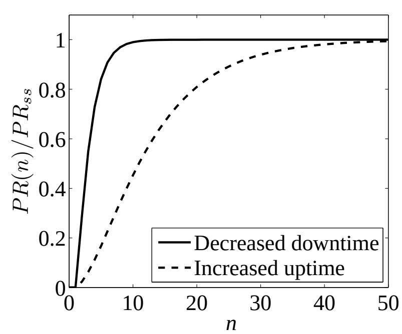


Fig. 2. Partitioning of the simplex $0 < P < R < 1$ according to SLE

Thus, for $T_{down} > 2$, shorter up- and downtimes lead to faster transients than longer ones, even if the efficiency $e > 0.5$ remains the same. This phenomenon is illustrated in Figure 3(a).



(a) Transients of the production rate PR for $e = 0.9$



(b) Transients of PR with increased uptime or decreased downtime for $e = 0.7$, $R = 0.1$ and $e' = 0.9$

Fig. 3. Effects of up- and downtime on the transients of two-machine lines with $N = 1$

In addition, the following can be obtained regarding the effects of increasing uptime or decreasing downtime on system transients:

Theorem 4.3: Consider a serial line defined by assumptions (i)-(v) and $N = 1$. Let $|\lambda_1^u|$ and $|\lambda_1^d|$ denote the SLEs resulting from increasing the uptime by $(1+\alpha)$, $\alpha > 0$, or decreasing its downtime by the same factor, respectively. Then, under assumption (9), $|\lambda_1^u| > |\lambda_1^d|$.

Thus, the qualitative effect of the uptime and the downtime on the transients in two-machine lines with $N = 1$ remains the same as that for individual machines: under condition (9), it is better to reduce the downtime than increase the uptime in order to shorten the transients. This phenomenon is illustrated in Figure 3(b).

V. Transients of Two-Machine Lines with $N \geq 2$

Due to high dimensionality of the resulting Markov transition matrices for two-machine geometric lines with $N \geq 2$, a complete analysis of their eigenvalues is all but impossible. However, some of the eigenvalues, namely those that characterize the dynamic behavior of the machines themselves, can be identified. This is carried out below.

As in the previous section, we denote the state of the system by (h, s), where $h \in H = \{0, 1, \ldots, N\}$ is the buffer state and $s = (s_1, s_2) \in S = \{0, 1\} \times \{0, 1\}$ denotes the states of the two machines. The transition matrix for this serial line is given by $A_{SL} = \{p(h', s'|h, s)\}$, where all variables vary over their ranges. Recall that the transition matrix, A, of two machines operating independently is given in (10). Finally, note that by assumption (ii) of Section II, the machine states are independent of the buffer state, so that

$$p(h', s'|h, s) = p(h'|h, s)p(s'|s)\,. \tag{17}$$

Theorem 5.1: The eigenvalues (11) of A given in (10) constitute a subset of the eigenvalues of A_{SL}.

The eigenvalues of A, as computed in Section III, are 1, λ_m, λ_m, λ_m^2. Thus, these eigenvalues are also eigenvalues of the serial line. Therefore, λ_m provides a *lower bound* on the SLE of the transition matrix A_{SL}.

To estimate the dynamics of PR in systems with $N \geq 2$, we resort to approximations. Clearly, the dynamics of the production rate under the time-dependent failures are given by

$$PR(n) = P[\text{buffer is not empty at } n | m_2 \text{ is up at } n]P[m_2 \text{ is up at } n]. \tag{18}$$

The second factor in the right hand side of (18), as it follows from Section III, is given by $e(1+\frac{\Delta}{e}\lambda_m^n)$, where Δ is defined in (6). We approximate the first factor by reducing the geometric line to a Bernoulli one. This reduction is similar to that developed in [8] and is referred to as geometric-Bernoulli transformation. For this transformation, we assume that the reliability parameter, p^{Ber}, of the Bernoulli machine equals to the efficiency e of the geometric machine, and each unit of buffer capacity in the Bernoulli line is equivalent to the buffer capacity of the geometric line necessary to protect against one downtime. In addition, the cycle time of the Bernoulli machine is selected as the average downtime of the geometric machine. This leads to the following expressions:

$$p^{Ber} = e = \frac{R}{P+R}, \quad N^{Ber} = \left[\frac{N}{T_{down}} + 1\right], \tag{19}$$

where $[x]$ denotes the nearest integer to x and the "1" is added to ensure $N^{Ber} \geq 1$.

For the resulting Bernoulli line, P^{Ber}[the buffer is not empty at slot k], $k = 0, 1, 2, \ldots$, can be calculated using the results of [3]. Since

$$eP^{Ber}[\text{the buffer is not empty at slot } k] = PR^{Ber}(k), \quad k = 0, 1, 2, \ldots, \tag{20}$$

where $PR^{Ber}(k)$ is the production rate of the Bernoulli line at time slot k, we obtain the following estimate of the production rate for the geometric line:

$$\widehat{PR}(kT_{down}) \;=\; PR^{Ber}(k)\left(1 + \frac{\Delta}{e}\lambda_m^{kT_{down}}\right)^2, \quad k = 0, 1, 2, \ldots, \tag{21}$$

where the additional multiplier $(1 + \frac{\Delta}{e}\lambda_m^{kT_{down}})$ is intended to account for the transients of the first machine (as indicated in Section III).

The accuracy of (21) has been investigated numerically using 50,000 lines constructed by selecting the machine and buffer parameters randomly and equiprobably from the sets:

$$e \in (0.5, 0.95], \quad R \in [0.05, 0.5], \quad N \in \{2, 3, \ldots, 40\}. \tag{22}$$

A typical example is shown in Figure 4, where the accuracy $\epsilon(kT_{down})$ is defined by

$$\epsilon(kT_{down}) = \frac{\widehat{PR}(kT_{down})}{\widehat{PR}(\infty)} - \frac{PR(kT_{down})}{PR(\infty)}, \quad k = 0, 1, 2, \ldots \tag{23}$$

and $\widehat{PR}(kT_{down})$ is calculated using (21), while $PR(kT_{down})$ is obtained by simulations. The average of $|\epsilon(kT_{down})|$ is given in Figure 5. As one can see, for small k, the error is within 0.08 and within 0.02 for k large.

Using approximation (21), the effects of up- and downtime on the transients can be evaluated. Since this is carried out numerically, we formulated the results as numerical facts.

Numerical Fact 5.1: Consider a geometric line with two identical machines having $e > 0.5$ and $N \geq 2$. Then, for any $T_{down} > 2$, shorter up- and downtimes practically always lead to faster transients than longer ones.

Numerical Fact 5.2: Under condition (9), reducing downtime practically always leads to shorter transients than increasing uptime.

As it is shown in the justification of these numerical facts, the term "practically always" is quantified as 99% for Numerical Fact 5.1 and 96% for Numerical Fact 5.2.

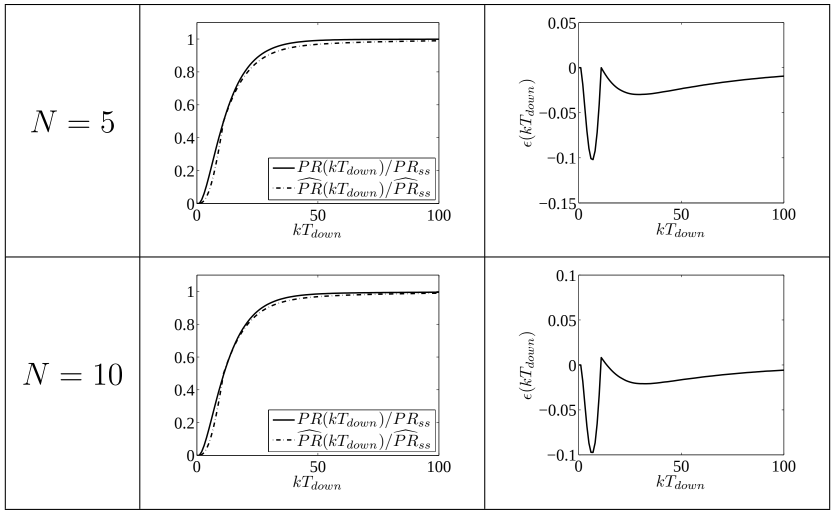


Fig. 4. Illustration of the accuracy of expression (21) for $e = 0.9$ and $R = 0.1$

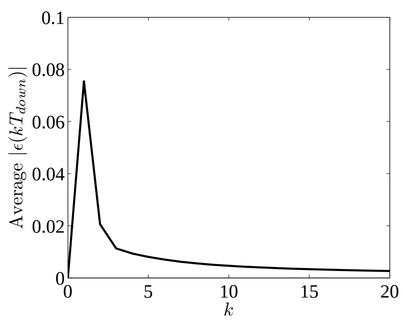


Fig. 5. Average accuracy of expression (21)

VI. Conclusions and Future Work

This Technical Note provides a characterization of transients in two-machine geometric production lines. It shows that the system's transients can be analyzed by separating the transients of the machines and the transients of the buffer. When the buffer is of capacity 1, this separation is exact; for larger buffers the separation is approximate. In either case, it is shown that if the machines' efficiency is greater than 0.5 and the average downtime is larger than two cycle times, shorter up- and downtimes lead to faster transients than longer ones. Under the same condition, it is shown that a reduction in downtime leads to faster transients than a similar increase of the uptime.

Future work will address transients in geometric lines with more than two machines and production lines with other machine reliability models, e.g., exponential, Weibull, log-normal, etc. For non-Markovian machines, the effect of the coefficients of variation of up- and downtime on the duration of transients will be investigated.

APPENDIX

Proof of Theorem 3.1: It follows from (7) that under condition (9),

$$e' = \frac{R}{R + \frac{P}{1+\alpha}} = \frac{(1+\alpha)R}{(1+\alpha)R + P} > e > 0.5, \quad R < (1+\alpha)R < 0.5. \tag{A.1}$$

Therefore,

$$R + \frac{P}{1+\alpha} < 2R < 1, \quad (1+\alpha)R + P < 2(1+\alpha)R < 1. \tag{A.2}$$

Thus, according (3),

$$\lambda_1^u = 1 - \frac{P}{1+\alpha} - R > 0, \quad \lambda_1^d = 1 - P - (1+\alpha)R > 0. \tag{A.3}$$

Hence, under condition (9),

$$|\lambda_1^u| - |\lambda_1^d| > 0 \Longleftrightarrow 1 - \frac{P}{1+\alpha} - R > 1 - P - (1+\alpha)R \Longleftrightarrow \alpha R + \frac{\alpha P}{1+\alpha} > 0, \tag{A.4}$$

which completes the proof.

Proof of Theorem 4.1: Since the algebraic and geometric multiplicities of every eigenvalue of matrix A, defined in (12), are equal to each other, there exists a nonsingular matrix Q such that

$$A = Q^{-1}\tilde{A}Q, \quad \tilde{A} = diag[1 \quad \lambda_b \quad \lambda_m \quad \lambda_m \quad \lambda_m^2 \quad 0 \quad 0 \quad 0].$$

Thus,

$$x(n) = Ax(n-1) = Q^{-1}\tilde{A}Qx(n-1) = Q^{-1}\tilde{A}_nQx(0),$$

where $\tilde{A}_n = diag[1 \quad \lambda_b^n \quad \lambda_m^n \quad \lambda_m^n \quad (\lambda_m^2)^n \quad 0 \quad 0 \quad 0]$. Hence, the evolution of the states can be expressed as

$$\begin{aligned} x_{hij}(n) &= x_{hij}[1 + \tilde{B}\tilde{x}_2(0)\lambda_b^n + (\tilde{C}_1\tilde{x}_3(0) + \tilde{C}_2\tilde{x}_4(0))\lambda_m^n + \tilde{D}\tilde{x}_5(0)(\lambda_m^2)^n], \\ & \quad h \in \{0,1\},\ i,j \in \{0,1\},\ n = 1,2,\ldots, \end{aligned} \tag{A.5}$$

where $\tilde{B}$, $\tilde{C}_1$, $\tilde{C}_2$ and $\tilde{D}$ are constants, $\tilde{x}_i(0) = q_i x(0)$ and q_i is the i-th row of Q.

Then, it follows from (14) that

$$C = \tilde{C}_1\tilde{x}_3(0) + \tilde{C}_2\tilde{x}_4(0), \quad D = \tilde{D}\tilde{x}_5(0). \tag{A.6}$$

For matrix Q, it can be shown that

$$\begin{bmatrix} q_3 \\ q_4 \\ q_5 \end{bmatrix} = \frac{P^2}{(-R+P+R^2)\,(R+P)^2} \begin{bmatrix} R^2 & -RP & R^2 & -RP & R^2 & -RP & R^2 & -RP \\ R & R & -P & -P & R & R & -P & -P \\ -\frac{R^3}{P^2} & \frac{R^2}{P} & \frac{R^2}{P} & -R & -\frac{R^3}{P^2} & \frac{R^2}{P} & \frac{R^2}{P} & -R \end{bmatrix}.$$

Moreover, initial condition (15) implies that

$$\sum_{h,j} x_{h1j}(0) = \sum_{h,i} x_{hi1}(0) = e, \quad \sum_{h,j} x_{h0j}(0) = \sum_{h,i} x_{hi0}(0) = 1-e.$$

In addition, since m_1 and m_2 are independent,

$$\sum_{h,i\neq j} x_{hij}(0) = 2e(1-e), \quad \sum_{h} x_{h00}(0) = (1-e)^2, \quad \sum_{h} x_{h11}(0) = e^2.$$

Thus, under (15),

$$\tilde{x}_3(0) = \frac{P^2[R^2(1-e)-RPe]}{(-R+P+R^2)\,(R+P)^2} = 0, \;\; \tilde{x}_4(0) = \frac{P^2[R(1-e)-Pe]}{(-R+P+R^2)\,(R+P)^2} = 0,$$

$$\tilde{x}_5(0) = \frac{R[2RPe(1-e)-R^2(1-e)^2-P^2e^2]}{(-R+P+R^2)\,(R+P)^2} = 0.$$

Therefore, due to (A.6), $C = D = 0$.

Proof of Theorem 4.2: Since $|1-P-R|$ and $(1-R)^2$ are both monotonically decreasing functions of R on $(0, 0.5)$ for a fixed e, the SLE of the system is a monotonically decreasing function of R on $(0, 0.5)$.

Proof of Theorem 4.3: It follows from Theorem 3.1 that $|\lambda_m^u| > |\lambda_m^d|$. In addition, $\lambda_b^u = (1-R)^2 > [1-(1+\alpha)R]^2 = \lambda_b^d$. Thus, $|\lambda_1^u| = \max(|\lambda_m^u|, \lambda_b^u) > \max(|\lambda_m^d|, \lambda_b^d) = |\lambda_1^d|$.

Proof of Theorem 5.1: Let λ be an eigenvalue of A given in (10), with left eigenvector $v = (v(s))_{s\in S}$. Thus, $(vA)(s) = \sum_{s'\in S} v(s')p(s', s) = \lambda v(s)$, $s \in S$. Consider now the vector $w = (w(h, s))_{h\in H, s\in S}$ with $w(h, s) = v(s)$ for all h and s; that is, w is a row vector indexed by the elements of the set $H \times S$, and the value of its (h, s) component equals $v(s)$, independently of h. Then, using (17), we obtain that

$$\begin{aligned}(wA_{SL})(s,h) &= \sum_{h',s'} w(h',s')p(h',s'|h,s) = \sum_{h',s'} v(s')p(s'|s)p(h'|h,s) \\ &= \lambda v(s)\sum_{h'} p(h'|h,s) = \lambda v(s) \equiv \lambda w(s,h)\end{aligned}$$

for every s and h. Hence w is a left eigenvector of A_{SL} with eigenvalue λ. This establishes that any eigenvalue of the matrix A is an eigenvalue of the matrix A_{SL} as well, as claimed.

Justification of Numerical Fact 5.1: This justification is carried out by evaluating the settling time of production rate, t_{sPR}, which is the time necessary for PR to reach and remain within $\pm 5\%$ of its steady state value, provided that the buffer is initially empty. A total of 10,000 lines were generated with e and N randomly and equiprobably selected from the sets (22), respectively. For each line, thus constructed, t_{sPR} was evaluated using approximation (21) as a function of R. As a result, we obtained that t_{sPR} is a monotonically decreasing function of R on $R \in (0, 0.5)$ in 99% of all cases studied. Thus, we conclude that shorter up- and downtimes lead, practically always, to faster transients.

Justification of Numerical Fact 5.2: To justify this numerical fact, the 50,000 lines generated as mentioned in Section V were used to investigate the effects of increasing uptime or decreasing downtime on t_{sPR}. To accomplish this, we selected α randomly and equiprobably from the set $\alpha \in [0.05, 1]$ and evaluated the settling times t^u_{sPR} and t^d_{sPR}, resulting from increasing uptime by $(1+\alpha)$ and decreasing downtime by $(1+\alpha)$, respectively. It turned out that t^u_{sPR} was longer than t^d_{sPR} in 96.12% of all cases studied. For the remaining 3.88% of cases, t^u_{sPR} was shorter than t^d_{sPR} by at most 1 cycle time. Therefore, we conclude that Numerical Fact 5.2 takes place.

REFERENCES

[1] Y. Narahari and N. Viswanadham, “Transient analysis of manufacturing systems performance,” *IEEE Transactions on Robotics and Automation*, vol. 10, no. 2, pp. 230–244, 1994.

[2] S. Mocanu, “Numerical algorithms for transient analysis of fluid queues,” in *Proceedings of 5th International Conference on the Analysis of Manufacturing Systems*, Zakynthos, Greece, 2005, pp. 15–20.

[3] S. M. Meerkov and L. Zhang, “Transient behavior of serial production lines with Bernoulli machines,” *IIE Transactions*, vol. 40, no. 3, pp. 297–312, 2008.

[4] J. A. Buzacott and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*. Prentice Hall, Englewood Cliff, NJ, 1993.

[5] H. T. Papadopoulos, C. Heavy and J. Browne, *Queueing Theory in Manufacturing Systems Analysis and Design*. Chapman & Hill, London, UK, 1993.

[6] S. B. Gershwin, *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliff, NJ, 1994.

[7] T. Altiok, *Performance Analysis of Manufacturing Systems*. Springer-Verlag, New York, NY, 1997.

[8] J. Li and S. M. Meerkov, *Production Systems Engineering*. Springer, 2009.

[9] ——, “Due-time performance in production systems with markovian machines,” in *Analysis and Modeling of Manufacturing Systems*, S. B. Gershwin, Y. Dallery, C. T. Papadopolous, and J. Smith, Eds. Boston, MA: Kluwer Academic, 2003, ch. 10, pp. 221–253.