Production Lead Time in Serial Lines: Evaluation, Analysis, and Control

Semyon M. Meerkov and Chao-Bo Yan

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122

Abstract

Production Lead Time (LT) is the average time a part spends in the system, being processed or waiting for processing. In systems with unlimited buffers, LT may be orders of magnitude larger than the total processing time, leading to serious economic and quality problems. At present, no systematic analytical methods for evaluation, analysis, and control of LT in systems with machines having up- and downtime characterized by continuous random variables are available. This paper is intended to develop such methods. Specifically, we address synchronous serial lines with exponential machines and derive formulas for LT as a function of machine parameters and raw material release rate. Using these formulas, we develop methods for open- and closed-loop raw material release, which result in the desired LT. For asynchronous exponential lines, we provide an upper bound on LT. For non-exponential lines (e.g., Weibull, gamma, and log-normal), we offer an empirical formula for LT as an affine function of the coefficient of variation. The results reported in this paper enable a new paradigm for production systems management, namely: manage a production system so that the desired LT is ensured, while the throughput is maximized.

Keywords: Production systems, Lead time, Exponential and non-exponential reliability models, Open- and closed-loop raw material release control.

1 Introduction

1.1 Issues addressed and results obtained

Production Lead Time (*LT*) is the average time a part spends in the system, being processed or waiting for processing. In systems with unlimited buffers, *LT* may be orders of magnitude larger than the total processing time, leading to serious economic and quality problems. Unlike other performance measures, such as throughput (*TP*), work-in-process (*WIP*), blockages (*BL*), and starvations (*ST*), evaluation and analysis/control of *LT* received relatively little attention in the current literature. The goal of this paper is to develop systematic analytical methods for evaluation and analysis of *LT* in serial lines with machines having their up- and downtime described by continuous random variables and, using these methods, provide recommendations for *LT* control and management.

To accomplish this, first we consider the simplest case – the serial lines with identical exponential machines. This case offers a possibility of detailed analytical investigation of *LT* and, more importantly, provides insights into the behavior of *LT* as a function of machine parameters and raw material release rate. In addition, it permits the development and investigation of simple open- and closed-loop raw material release policies, which ensure the desired *LT*'s, while maximizing *TP*. Then, we extend these results to more practical cases. Specifically, we consider synchronous exponential lines with non-identical machines and carry out an analytical investigation, similar to that of the simplest case. Finally, we address asynchronous exponential and non-exponential lines; for the former, we derive an upper bound on *LT*, expressed through auxiliary synchronous exponential lines; for the latter, we provide an empirical formula for their *LT*'s as a function of up- and downtime coefficient of variation.

From the mathematical perspective, this work is based on evaluating *LT* for synchronous serial lines with exponential machines (using the expressions for *TP* and *WIP* and Little's law) and investigating analytical and structural properties of the resulting expressions. In addition, these expressions are used to design open- and closed-loop raw material release policies, which ensure the desired *LT*.

From the operational perspective, this work is based on associating a raw material release control mechanism with production lines under consideration. Initially, this mechanism is modeled as an auxiliary exponential machine, leading to random, once-per-cycle, raw material release. Then, the results are extended to deterministic, e.g., once-per-hour or once-per-shift, release. Finally, for the case when only nominal values of machine parameters are known and the real ones may differ from the nominal, we offer a relay-type closed-loop release policy, which is based on the real-time *WIP*.

From the managerial perspective, the results of this paper enable a new paradigm of production management: rather than maximizing *TP* (which may lead to unlimited *LT*), manage the production systems so that the desired *LT* is ensured, while *TP* is maximized.

1.2 Related literature

The current literature on production lead time can be classified into three groups. The first one considers the lead time as a function of the dispatch, rather than release, rule [1–7]. Dispatch rules indicate which job must be selected for processing at a given machine or workcenter. The main result here is that, under a wide range of conditions, jobs with the shortest processing time should be selected first in order to minimize the lead time. The second, and the most prolific, group addresses the problem of feedback control of raw material release. The main techniques considered here are kanban [8–20], CONWIP [21–27], and their comparisons [28–32]. However, with a few exceptions, this literature does not provide rigorous methods for selecting parameters of these control strategies (e.g., the numbers of kanbans and CONWIP), which ensure the desired lead time. A notable exception is a recent article [33], where a method for calculating the smallest number of kanbans that ensures the desired level of customer service in exponential lines with finite buffers are provided. The third, and the least copious, group consists of recent publications on lead time analysis using analytical tools. It includes papers [34–37], where distributions of lead time are derived for production systems with exponential processing time and Poisson order arrival. In [38], these results are extended to multi-product systems. An approach to lead time reduction based on manufacturing intelligence is developed in [39], where neural networks have been employed to evaluate *WIP* in a semiconductor manufacturing facility and used subsequently for lead time analysis and control. Finally, [40] considers serial lines with finite buffers and derives lower bounds on the production completion time, which is closely related to the lead time.

Our initial results on LT evaluation, analysis, and control have been reported in [41] and [42] for serial and cellular lines, respectively. Both of these papers assume that the machines obey the Bernoulli reliability model, according to which a machine produces a part during a cycle time with probability p and fails to do so with probability $1-p$ (see [43] for details). While this reliability model is applicable to assembly operations with downtimes close to the machine cycle time, it is not applicable to machining and other operations, where the downtime is typically much longer than the cycle time. Also, the Bernoulli model, being defined by a single parameter, does not offer a possibility of investigating various qualitative properties of LT, e.g., the effects of up- and downtime on LT. These shortcomings necessitate the development of methods for evaluation, analysis, and control of LT in production systems with more practical reliability models, e.g., exponential, Weibull, gamma, log-normal, etc. This is carried out in the current paper.

1.3 Paper outline

The outline of this paper is as follows: In Section 2, a model of production systems with a raw material release mechanism is introduced, and the problems addressed are stated. Sections 3-5 provide a solution of these problems for the simplest case – serial lines with identical exponential producing machines. As mentioned in Subsection 1.1, this case is considered in details because it provides clear insights into the lead time behavior and control. In Sections 6 and 7, extensions to systems with a wider practical applicability are provided. Namely, in Section 6, we report analytical results on the evaluation, analysis, and control of LT in synchronous exponential lines with non-identical producing machines. In Section 7, we offer an upper bound on LT in asynchronous exponential lines and an empirical formula for LT in lines with non-exponential machines. The conclusions and topics for future research are given in Section 8. The proofs are included in the Appendix.

2 Modeling and Problems Addressed

Consider a serial line shown in Figure 2.1, where the circles represent the machines and the open rectangles are the buffers. While $m_1, m_2, \dots, m_M$ and $b_1, b_2, \dots, b_{M-1}$ are the usual producing machines and unlimited work-in-process buffers, respectively, m_0 represents the raw material release machine and b_0 unlimited raw material buffer (to indicate this, m_0 and b_0 are shown in gray). Controlling the efficiency of the release machine, m_0, one can control the availability of raw material in the system and, thus, the lead time.

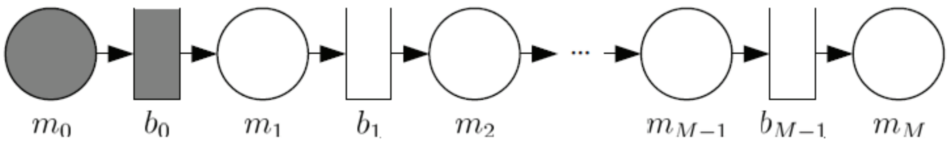


Figure 2.1: Serial production line with a release machine

To formalize this model, we introduce the following assumptions:

(i) The system consists of M producing machines, $m_1, m_2, \dots, m_M$, a release machine, m_0, $M-1$ work-in-process buffers, $b_1, b_2, \dots, b_{M-1}$, and a raw material buffer, b_0.

(ii) Each machine is characterized by its cycle time, τ_i (in min), $i = 0, 1, \dots, M$. If cycle times of all machines (including the release machine) are identical, the system is called *synchronous*; otherwise, it is *asynchronous*. While in the asynchronous case, τ_i, $i = 1, 2, \dots, M$, are fixed, τ_0 is free and can be selected at will.

(iii) In addition, each machine is characterized by its *reliability model*, i.e., continuous random variables that define its up- and downtime. If these distributions are exponential, i.e., defined by the breakdown rate λ_i and repair rate μ_i, $i = 0, 1, \dots, M$, (both in 1/min), the line is called *exponential*; otherwise, it is *non-exponential*. While for the producing machines, λ_i and μ_i, $i = 1, 2, \dots, M$, are fixed, for the release machine, λ_0 and μ_0 are design parameters that can be selected at will.

(iv) Each buffer is of infinite capacity.

(v) The flow model [43] is assumed, i.e., infinitesimal quantity of parts, produced during an infinitesimal time interval, are transferred to and from the buffers. A machine is starved, if the buffer in front of it is empty; m_0 is never starved. Machine failures are time-dependent [43], i.e., a machine can be down even if it is starved.

Assumption (iv) is introduced to reflect the fact that the lead time control problem is of particular importance for systems with practically unlimited storage, e.g., with no hardware-constrained buffers. Assumption (v) is introduced for technical reasons: it permits a precise formulation of the equations describing the systems at hand (a justification of this assumption is given in [43]).

Let $T_{up,i}$ and $T_{down,i}$ denote the average up- and downtime of the machines, $i = 0, 1, \ldots, M$. Then the machine *efficiency* for any continuous reliability model is:

$$e_i := \frac{T_{up,i}}{T_{up,i} + T_{down,i}}, \; i = 0, 1, \ldots, M, \tag{2.1}$$

and its *throughput in isolation* (i.e., when the machine is not starved) is

$$TP_{isol,i} := \frac{T_{up,i}}{\tau_i(T_{up,i} + T_{down,i})}, \; i = 0, 1, \ldots, M. \tag{2.2}$$

Since for exponential machines, $T_{up,i} = \frac{1}{\lambda_i}$ and $T_{down,i} = \frac{1}{\mu_i}$,

$$e_i = \frac{\mu_i}{\lambda_i + \mu_i} \text{ and } TP_{isol,i} = \frac{\mu_i}{\tau_i(\lambda_i + \mu_i)}, \; i = 0, 1, \ldots, M. \tag{2.3}$$

Clearly, to obtain meaningful results, it should be assumed that $e_0 < e_i$, $i = 1, 2, \ldots, M$, for synchronous lines or $TP_{isol,0} < TP_{isol,i}$, $i = 1, 2, \ldots, M$, for asynchronous ones (otherwise, LT is unbounded).

For the case of finite buffer capacity, an analytical method for evaluating the throughput and work-in-process in each buffer (WIP_i, $i = 0, 1, \ldots, M$) of exponential serial lines is given in [43]. In the current paper, we extend this method to the case of infinite buffers and address the following problems:

- Derive an analytical expression for LT in synchronous exponential lines as a function of λ_i

and μ_i, $i = 0, 1, \dots, M$.

- Analyze structural properties of LT as a function of e_0 and (λ_i, μ_i), $i = 1, 2, \dots, M$.
- Derive a formula for e_0, as a function of λ_i and μ_i, $i = 1, 2, \dots, M$, which ensures the desired LT, while maximizing TP (i.e., solve the open-loop LT control problem).
- Extend the above open-loop control method to deterministic (e.g., once-per-hour or once-per-shift) raw material release.
- Develop a closed-loop deterministic raw material release policy applicable in systems where λ_i and μ_i, $i = 1, 2, \dots, M$, are not know precisely.
- Generalize the above results for asynchronous exponential lines and non-exponential lines (with Weibull, gamma, and log-normal machine reliability models) by providing an upper bound for the former and an empirical formula for the latter.

Solutions of these problems are described in the subsequent sections.

3 Analysis of *LT* in Exponential Lines with Identical Producing Machines

3.1 *LT* evaluation

Proposition 3.1 *Consider an exponential line defined by assumptions (i)-(v) with $\lambda_i = \lambda$, $\mu_i = \mu$, $i = 1, 2, \dots, M$, $\tau_i = \tau$, $i = 0, 1, \dots, M$, and $e_0 < e$. Then, an estimate of LT (in min) is given by*

$$\widehat{LT} = M\tau + \left[\frac{e_0}{\mu_0} + (2M-1)\frac{e}{\mu}\right]\left(\frac{1-e}{e-e_0}\right). \tag{3.1}$$

As mentioned in Subsection 1.3, the derivation of this proposition, as well as all others cited throughout the paper, is given in the Appendix.

The accuracy of this estimate has been evaluated by simulating exponential lines with parameters M, e, e_0, T_{down}, and $T_{down,0}$ selected randomly and equiprobably from the following sets:

$$M \in [3, 10],\ e \in [0.7, 0.99],\ e_0 \in [0.7e, 0.99e],\ T_{down} \in [10\text{min}, 100\text{min}],\ T_{down,0} = T_{down}. \quad (3.2)$$

For each line, thus formed, the simulations were carried out for two τ's: $\tau = 0.5$min and $\tau = 5$min. The total of 1000 lines have been simulated using the following procedure: For each line, in addition to a warm-up period of $2,000,000$min, the simulation was carried out for $22,000,000$min; 20 repetitions of this procedure were carried out to evaluate LT. This simulation procedure results in a 95% confidence interval of $\pm 0.87\%$ of LT for both $\tau = 0.5$min and $\tau = 5$min. The accuracy of (3.1) was quantified by $\epsilon_{LT} = \frac{|\widehat{LT} - LT|}{LT} \times 100\%$. As a result, we obtained: For $\tau = 0.5$min, the smallest and the largest errors were 0.0025% and 8.97%, respectively, and the average error was 2.17%; for $\tau = 5$min, the smallest and the largest errors were 0.0007% and 7.21%, respectively, and the average error was 1.99%. Based on these results and recognizing that machine parameters on the factory floor are rarely known with accuracy better than $\pm 5\%$, we conclude that estimate (3.1) is precise enough for the lead time analysis and control.

Expression (3.1) leads to the following:

Observation 3.1 *For fixed e and e_0, shorter up- and downtimes lead to smaller $\widehat{LT}$. As $e \to 1$, $\widehat{LT}$ tends to its minimum value, $M\tau$. $\widehat{LT}$ is monotonically increasing in M, hyperbolically increasing as $e_0 \to e$, and is an affine function of τ with the slope M.*

To further analyze the behavior of $\widehat{LT}$, introduce the following parametrization:

$$\rho := \frac{e_0}{e}, \qquad \widehat{lt} := \frac{\widehat{LT}}{M\tau}. \quad (3.3)$$

We refer to $0 < \rho < 1$ as the *relative workload* imposed on the system and to $\widehat{lt} > 1$ as the *relative lead time* (dimensionless), i.e., the lead time in units of the smallest possible lead time. In terms of these parameters, (3.1) becomes

$$\widehat{lt} = 1 + \frac{1}{\tau}\Big(\frac{\rho}{M\mu_0} + \frac{2M-1}{M\mu}\Big)\Big(\frac{1-e}{1-\rho}\Big). \quad (3.4)$$

Clearly, in addition to M, e, and τ, the relative lead time, $\widehat{lt}$, depends on the release machine efficiency, e_0 (through ρ) and on its downtime (through μ_0). However, in the limit as M tends to infinity, the dependence on μ_0 disappears:

$$\widehat{lt}_\infty := \lim_{M\to\infty} \widehat{lt} = 1 + \frac{2}{\mu\tau}\Big(\frac{1-e}{1-\rho}\Big). \tag{3.5}$$

This is important for the lead time control problem, since for sufficiently long lines, only e_0 would need to be selected, rather than μ_0 as well. It can be shown that if $\mu_0 \geqslant \mu$, the accuracy of this approximation in terms of $\Delta = \frac{\widehat{lt}_\infty - \widehat{lt}}{\widehat{lt}_\infty}$ is quantified as $0 < \Delta < \frac{1}{2M}$. Thus, $\widehat{lt}_\infty \geqslant \widehat{lt}$, and the error, Δ, decreases hyperbolically in M. Therefore, (3.5) can be viewed as a relatively tight bound of (3.4).

From (3.5) follows another fact: If $\mu\tau = 2e$, then (3.5) becomes

$$\widehat{lt}_\infty = \frac{e^{-1} - \rho}{1-\rho}, \tag{3.6}$$

which is exactly the same as the expression for $\widehat{lt}$ in Bernoulli serial lines (see [41]), with the Bernoulli machine efficiency p being substituted by the exponential machine efficiency e. Hence, from (3.5) and (3.6), follows:

Observation 3.2 *If $\mu\tau = 2e$, then $\widehat{lt}_\infty$ in exponential lines equals $\widehat{lt}$ in Bernoulli lines with $p = e$. If $\mu\tau < 2e$, then $\widehat{lt}_\infty$ in exponential lines is larger than $\widehat{lt}$ in Bernoulli lines with $p = e$. Since $\mu\tau < 2e$ implies that $T_{down} > \frac{\tau}{2}$* (which is practically always the case)*, we conclude that $\widehat{lt}$ in exponential lines is generically larger than $\widehat{lt}$ in Bernoulli lines with $p = e$.*

3.2 Structural properties of $\widehat{lt}(\rho)$

Figure 3.1 illustrates the behavior of $\widehat{lt}$ given by (3.4) as a function of ρ for $M = 10$, $\tau = 1$min and several values of e, μ, and μ_0; the Bernoulli case, i.e., when $\mu\tau = 2e$, is also shown for comparison. All curves in this figure have a "knee" beyond which $\widehat{lt}$ grows extremely fast. In practice, this knee is refered to as the "sweet point", since *TP* is maximized, while *LT* remains relatively small. Therefore, it is important to quantify the position of the knee.

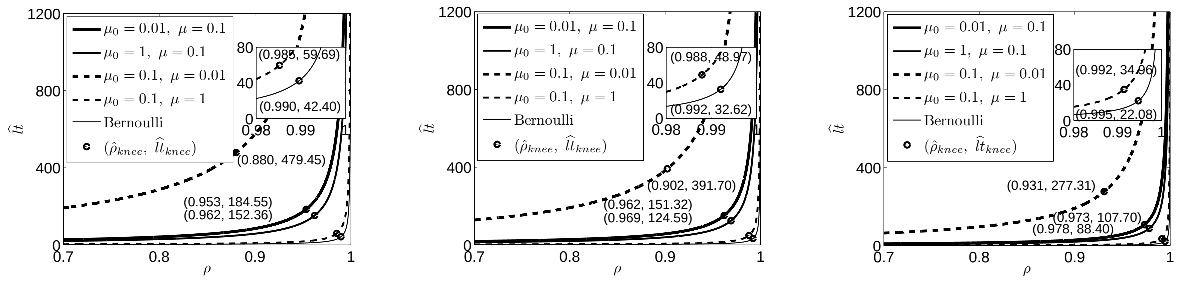
To accomplish this, consider the $(\rho, \widehat{lt})$-plane, where a unit interval of ρ-axis corresponds to

$A > 1$ units of $\widehat{lt}$-axis (in Figure 3.1, $A = 4000$). Introduce the *scaling ratio*, α, defined by

$$\alpha := \frac{1}{A} \tag{3.7}$$

and recall that the *curvature*, κ, of a twice differentiable function, $f(x)$, is given by (see [44])

$$\kappa(f(x)) = \frac{\left|f''_{xx}\right|}{(1 + f'^2_x)^{\frac{3}{2}}}. \tag{3.8}$$



(a) $e = 0.7$ (b) $e = 0.8$ (c) $e = 0.9$

Figure 3.1: Relative lead time, $\widehat{lt}$, as a function of relative workload, ρ, and machine parameters (for $M = 10$, $\tau = 1$min)

Definition 3.1 *The knee, ρ_{knee}, of $\widehat{lt}$ on the $(\rho, \widehat{lt})$-plane with the scaling ratio α is the point on* $[0, 1)$ *at which the curvature of $\alpha\widehat{lt}(\rho)$ reaches its maximum.*

Proposition 3.2 *Under the assumptions of Proposition 3.1,*

$$\rho_{knee} = 1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)(1-e)} \tag{3.9}$$

and

$$\lim_{M\to\infty} \rho_{knee} = 1 - \sqrt{\frac{2\alpha}{\mu\tau}(1-e)}. \tag{3.10}$$

The pairs $(\rho_{knee}, \widehat{lt}(\rho_{knee}))$ are indicated in Figure 3.1 by black dots. Thus, releasing raw material with the rate

$$e_0 < e\Big(1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)(1-e)}\Big), \tag{3.11}$$

or, as $M \to \infty$,

$$e_0 < e\left(1 - \sqrt{\frac{2\alpha}{\mu\tau}(1 - e)}\right), \tag{3.12}$$

results in $\widehat{lt}$ below the knee. From (3.9) and (3.10), we derive:

Observation 3.3 *The position of the knee shifts to the right* (i.e., larger release rates become safe) *if the producing machine efficiency is increased or the up- and downtime of all machines are decreased.*

The results of this section are used below for open- and closed-loop control of $\mathcal{LT}$ by random or deterministic raw material release.

4 Open-loop Control of $\mathcal{LT}$ in Exponential Lines with Identical Producing Machines

In this section, first we quantify the set of attainable lead times (i.e., feasible set) and then derive formulas for the random raw material release rates that ensure the desired feasible lead time, while maximizing the throughput. Next, we extend this result to the deterministic raw material release.

4.1 Random raw material release

Proposition 4.1 *Under the assumptions of Proposition 3.1, the sets of feasible lead times, $\mathscr{F}_{\widehat{lt}}$ and $\mathscr{F}_{\widehat{lt}_\infty}$, are given, respectively, by*

$$\begin{aligned} \widehat{lt} &> 1 + (1 - e)\frac{2M - 1}{M}\frac{T_{down}}{\tau}, \\ \widehat{lt}_\infty &> 1 + 2(1 - e)\frac{T_{down}}{\tau}. \end{aligned} \tag{4.1}$$

Thus, $\mathcal{LT}$ cannot be made arbitrarily small under any raw material release policy. For instance, if $e = 0.8$ and $\frac{T_{down}}{\tau} = 10$, then $\widehat{lt} > 4.8$ (for $M = 10$) and $\widehat{lt}_\infty > 5$, no matter how low the release rate is.

Proposition 4.2 *Under the assumptions of Proposition 3.1, for any feasible desired lead time, $lt_d \in \mathscr{F}_{\widehat{lt}}$, the release rate is given by*

$$\hat{e}_0^* = e\Big[1 - \frac{\mu + (2M-1)\mu_0}{M\mu\mu_0\tau(lt_d - 1) + \mu(1-e)}(1-e)\Big]. \tag{4.2}$$

For this release rate,

$$\widehat{TP}^* = \frac{\hat{e}_0^*}{\tau},\ \widehat{WIP}_0^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{\hat{e}_0^*}{\mu_0} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-\hat{e}_0^*}\Big),\ \widehat{WIP}_i^* = \frac{2\hat{e}_0^* e}{\mu\tau}\Big(\frac{1-e}{e-\hat{e}_0^*}\Big),\ i = 1, 2, \dots, M-1. \tag{4.3}$$

This proposition leads to a solution of the open-loop lead time control problem based on the following arguments:

- Since, as it is possible to show, $\frac{d\hat{e}_0^*}{d\mu_0} > 0$, the throughput $\widehat{TP}^*$ is maximized as $\mu_0 \to \infty$. In this case, the release rate that results in lt_d, becomes:

$$\hat{e}_0^*(\mu_0 = \infty) := \lim_{\mu_0 \to \infty} \hat{e}_0^* = e\Big[1 - \frac{(2M-1)(1-e)}{M(lt_d - 1)} \frac{T_{down}}{\tau}\Big]. \tag{4.4}$$

- Having $\mu_0 \to \infty$, while $\hat{e}_0^*$ being fixed, implies that $\lambda_0 \to \infty$ in such a manner that

$$\lim_{\substack{\lambda_0 \to \infty \\ \mu_0 \to \infty}} \frac{\lambda_0}{\mu_0} = \frac{1 - \hat{e}_0^*(\mu_0 = \infty)}{\hat{e}_0^*(\mu_0 = \infty)}. \tag{4.5}$$

  In other words, both $T_{up,0}$ and $T_{down,0}$ tend to 0 and, thus, raw material is released continuously with the rate (4.4). In practice, this can be accomplished by releasing a part at the beginning of each cycle time with probability

$$p = \hat{e}_0^*(\mu_0 = \infty). \tag{4.6}$$

  This implies that the release machine can be viewed as obeying the Bernoulli reliability model with the probability of success given by (4.6). We refer to this type of release as *once-per-cycle*.

- In the limit as $M \to \infty$, (4.2) becomes

$$\hat{e}_0^*(M = \infty) := \lim_{M\to\infty} \hat{e}_0^* = e\left[1 - \frac{2(1-e)}{lt_d - 1}\frac{T_{down}}{\tau}\right], \tag{4.7}$$

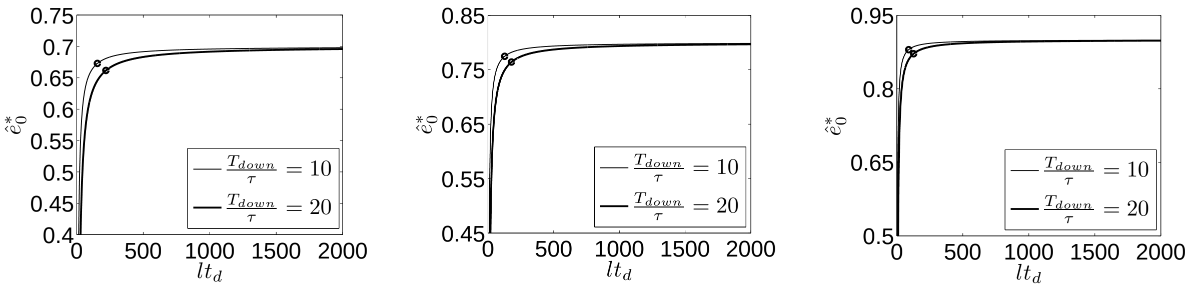
  which is independent of μ_0. Thus, for sufficiently large M, once-per-cycle release also can be implemented with

$$p = \hat{e}_0^*(M = \infty). \tag{4.8}$$

Summarizing the above arguments, we conclude that *a solution of the open-loop lead time control problem is provided by releasing a part into the raw material buffer once-per-cycle with probability (4.6) if M is relatively small (say, $M < 10$) and with probability (4.8) if $M \geqslant 10$.*

The behavior of $\hat{e}_0^*(M = \infty)$ as a function of lt_d is illustrated in Figure 4.1 for various values of e and $\frac{T_{down}}{\tau}$, with black dots indicating $(\widehat{lt}_{knee}, \hat{e}_0^*(\widehat{lt}_{knee}))$. From this figure follows:

Observation 4.1 *For $lt_d < \widehat{lt}_{knee}$, the optimal release rate $\hat{e}_0^*$* (and, therefore, $\widehat{TP}^*$) *is a rapidly increasing function of lt_d. For $lt_d > \widehat{lt}_{knee}$, $\hat{e}_0^*$ is practically constant. Thus, releasing raw material with the rate beyond the knee is not only unnecessary* (since $\widehat{TP}^*$ is practically a constant)*, but detrimental as well* (since $\widehat{WIP}^*$ grows almost linearly in accordance with $\widehat{WIP}^* = \widehat{TP}^*(LT_d - M\tau)$).



(a) $e = 0.7$ (b) $e = 0.8$ (c) $e = 0.9$

Figure 4.1: Optimal release rate, $\hat{e}_0^*$, as a function of the desired relative lead time, lt_d, and machine parameters (for $M = 10$)

4.2 Deterministic raw material release

The random, once-per-cycle, raw material release may be inconvenient for practical implementation. Therefore, below we use the results of Subsection 4.1 to derive deterministic, e.g., once-per-hour or once-per-shift, release policies with guaranteed LT and insignificant losses of the throughput.

Let $\hat{e}_0^*(lt_d)$ be the once-per-cycle release rate calculated using either (4.4) or (4.7). Then, the deterministic hourly release, $\hat{E}_H^*$ (parts/hour), is defined as:

$$\hat{E}_H^* = \lfloor H\hat{e}_0^*(lt_d)\rfloor, \tag{4.9}$$

where $\lfloor x \rfloor$ is the "floor" operator, which denotes the largest integer not greater than x, and H is the number of cycles in an hour, i.e., $H = \frac{60}{\tau}$.

While releasing raw material according to (4.9) results in the obvious inequality

$$\widehat{LT}(\hat{E}_H^*) < \widehat{LT}(\hat{e}_0^*) + 60, \tag{4.10}$$

where $\widehat{LT}(\hat{E}_H^*)$ and $\widehat{LT}(\hat{e}_0^*)$ are the lead times under (4.9) and (4.4) (or (4.7)), respectively, the losses of the throughput under hourly release (4.9) are not obvious and must be evaluated. We carry out this evaluation by simulating three ten-machine lines defined by

$$L_1: \ e = 0.9, \ T_{down} = 70; \quad L_2: \ e = 0.9, \ T_{down} = 7; \quad L_3: \ e = 0.9, \ T_{down} = 0.7, \tag{4.11}$$

with $\tau = 0.5$min and $\tau = 5$min. The lt_d for these simulations has been selected so that, on one hand, it is in the admissible domain (defined by (4.1)) and, on the other hand, the system parameters are in the sets (3.2). Based on lt_d, thus selected, $\hat{e}_0^*$ and $\hat{E}_H^*$ have been evaluated using (4.4) and (4.9), respectively. For each of the systems considered, we ran the simulations with once-per-cycle and once-per-hour release and evaluated the resulting throughputs, TP_C and TP_H (both in parts/min), where the subscripts "C" and "H" denote cycle and hour, respectively. Based on these measurements, we quantified losses in TP by

$$TP_{loss} = \frac{TP_C - TP_H}{TP_C} \times 100\%. \tag{4.12}$$

The results are shown in Tables 4.1 and 4.2 for $\tau = 0.5$min and $\tau = 5$min, respectively. As one can see, when $\tau = 0.5$min, TP_{loss} is about 1%; when $\tau = 5$min, TP_{loss} may be close to 10%. The reason for the latter is that, for large τ, the material released per-hour amounts to just a few parts, even if lt_d is large. To combat this problem, a release for a longer interval of time, e.g., once-per-shift, may be used. In the case of an eight-hour shift, the release becomes

$$\hat{E}_S^* = \left\lfloor \frac{480}{\tau} \hat{e}_0^*(lt_d) \right\rfloor, \tag{4.13}$$

where, as before, $\hat{e}_0^*(lt_d)$ is the release rate per-cycle that ensures lt_d. The simulation results for this release are shown in Table 4.3. Obviously, these data are quite similar to those of Table 4.1. Based on this, we formulate:

Observation 4.2 *The release interval (RI), which leads to practically no losses in the throughput, can be defined as $RI \geqslant 50\tau$; in this case,*

$$\hat{E}_{RI}^* = \left\lfloor \frac{RI}{\tau} \hat{e}_0^*(lt_d) \right\rfloor, \tag{4.14}$$

resulting in LT and TP quantified by

$$\widehat{LT}(\hat{E}_{RI}^*) < \widehat{LT}(\hat{e}_0^*) + RI, \qquad TP_{RI} \approx TP_C. \tag{4.15}$$

5 Closed-loop Control of LT in Exponential Lines with Identical Producing Machines

5.1 Scenario

The previous section provides methods for calculating raw material release rates that ensure the desired lead time, if the parameters of the machines are known precisely. In practice, however, this is seldom the case – the machine parameters (e.g., their efficiencies or up- and downtimes) are known only nominally, and their real values may vary. In this situation, the above methods may result in lead times dramatically different from the expected ones. For instance, if the real machine

Table 4.1: Throughput loss under once-per-hour release for serial lines with identical exponential producing machines (τ = 0.5min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
90	0.6212	74	1.2426	1.2333	0.74
120	0.6907	82	1.3809	1.3667	1.03
300	0.8161	97	1.6318	1.6167	0.93
1500	0.8832	105	1.7665	1.7500	0.93

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
12	0.6738	80	1.3479	1.3333	1.08
16	0.7336	88	1.4671	1.4667	0.03
40	0.8356	100	1.6713	1.6666	0.28
200	0.8873	106	1.7745	1.7667	0.44

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
4.5	0.8283	97	1.656	1.6500	0.38
6	0.8497	101	1.6993	1.6833	0.94
15	0.8820	105	1.7638	1.7500	0.78
75	0.8966	107	1.7932	1.7833	0.55

efficiency, e_{real}, is lower than the nominal one, e_{nom}, and the desired lead time, lt_d, is sufficiently large, it may happen that

$$\hat{e}_0^*(lt_d) > \min_{1\leqslant i\leqslant M} e_{real,i}, \tag{5.1}$$

resulting in an arbitrarily large lead time.

To prevent this situation, feedback control may be used to throttle the raw material release if the work-in-process in the systems exceeds a certain limit. A number of such control strategies can be proposed. Here, we propose and investigate the one which is simple enough for factory floor implementations. Specifically, we consider a relay-type release policy based on the real-time total work-in-process, WIP_{total}: if at the end of the release interval, RI, the WIP_{total} is below $WIP_{nominal}$, the raw material is released; otherwise it is not. In Subsection 5.2 below we formally introduce this control law and in Subsection 5.3 investigate its performance using simulations.

Table 4.2: Throughput loss under once-per-hour release for serial lines with identical exponential producing machines (τ = 5min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
12	0.6738	8	0.1348	0.1333	1.10
16	0.7336	8	0.1466	0.1333	9.06
40	0.8356	10	0.1671	0.1667	0.27
200	0.8873	10	0.1774	0.1667	6.08

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
4.5	0.8283	9	0.1657	0.1500	9.47
6	0.8497	10	0.1699	0.1667	1.90
15	0.8820	10	0.1764	0.1667	5.54
75	0.8966	10	0.1793	0.1667	7.06

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	TP_C	TP_H	TP_{loss} (%)
1.5	0.8497	10	0.1699	0.1667	1.92
2	0.8748	10	0.1749	0.1667	4.73
5	0.8937	10	0.1787	0.1667	6.73
25	0.8990	10	0.1798	0.1667	7.30

5.2 Control law

Consider an exponential serial line with identical producing machines defined by the nominal breakdown and repair rates λ and μ. Let LT_d be the desired lead time. Based on this information, calculate $\hat{e}_0^*$ and $\hat{E}_{RI}^*$ using (4.6) and (4.14), respectively. Also, calculate the nominal total work-in-process using Little's law: since $\widehat{TP}^*$ is given by the first formula in (4.3), and the total waiting time in all buffers is, as it follows from (3.1), $LT_d - M\tau$, we obtain:

$$\widehat{WIP}_{nominal} = \frac{\hat{e}_0^*}{\tau}(LT_d - M\tau). \tag{5.2}$$

Using these data, introduce the following control law:

$$E(s+1) = \begin{cases} \hat{E}_{RI}^*, & \text{if } WIP_{total}(s) \leqslant \widehat{WIP}_{nominal}, \\ 0, & \text{otherwise}, \end{cases} \tag{5.3}$$

Table 4.3: Throughput loss under once-per-shift release for serial lines with identical exponential producing machines (τ = 5min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	TP_C	TP_S	TP_{loss} (%)
12	0.6738	64	0.1348	0.1333	1.10
16	0.7336	70	0.1466	0.1458	0.53
40	0.8356	80	0.1671	0.1667	0.27
200	0.8873	85	0.1774	0.1771	0.21

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	TP_C	TP_S	TP_{loss} (%)
4.5	0.8283	79	0.1657	0.1646	0.67
6	0.8497	81	0.1699	0.1687	0.68
15	0.8820	84	0.1764	0.1750	0.82
75	0.8966	86	0.1793	0.1792	0.09

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	TP_C	TP_S	TP_{loss} (%)
1.5	0.8497	81	0.1699	0.1687	0.69
2	0.8748	83	0.1749	0.1729	1.16
5	0.8937	85	0.1787	0.1771	0.90
25	0.8990	86	0.1798	0.1791	0.38

where $s = 0, 1, \ldots,$ is the index of the release interval; $E(s+1)$ is the raw material release at the beginning of release interval $s+1$; and $WIP_{total}(s)$ is the real-time total work-in-process in the system at the end of release interval s.

Clearly, the "sensor measurement" in this control law is $WIP_{total}(s)$, $s = 0, 1, \ldots$. In some production systems this information is readily available, while in others it is not. In the latter case, the following simple calculation can be used to evaluate $WIP_{total}(s)$:

$$WIP_{total}(s+1) = WIP_{total}(s) + E(s+1) - N(s+1), \;\; s = 0, 1, \ldots, \tag{5.4}$$

where $N(s+1)$ is the number of parts produced during the release interval $s+1$. Thus, the only input to control law (5.3), (5.4) is $WIP_{total}(0)$, i.e., WIP at the start of the system operation.

5.3 Performance evaluation

To evaluate the performance of feedback law (5.3), we use the three exponential lines (4.11) as the nominal ones and form a real one for each of them. The real lines are formed by increasing or decreasing machine up- and downtimes randomly and equiprobably within ±50% of their nominal values. The resulting lines are as follows:

$$
\begin{aligned}
L_1: \ \boldsymbol{e} &= [0.93, 0.89, 0.94, 0.91, 0.86, 0.92, 0.84, 0.93, 0.93, 0.83],\\
\boldsymbol{T_{down}} &= [45.46, 83.00, 51.47, 35.40, 97.05, 81.68, 98.71, 61.90, 55.10, 79.16],\\
L_2: \ \boldsymbol{e} &= [0.83, 0.94, 0.91, 0.90, 0.88, 0.91, 0.90, 0.95, 0.90, 0.84],\\
\boldsymbol{T_{down}} &= [7.66, 4.24, 8.62, 9.72, 10.12, 5.06, 6.43, 4.27, 6.55, 7.45],\\
L_3: \ \boldsymbol{e} &= [0.94, 0.89, 0.89, 0.91, 0.92, 0.95, 0.91, 0.91, 0.93, 0.79],\\
\boldsymbol{T_{down}} &= [0.51, 0.84, 0.44, 0.74, 0.51, 0.36, 0.68, 0.83, 0.38, 1.00].
\end{aligned}
\tag{5.5}
$$

We simulated these lines with and without feedback control (5.3) for $\tau = 0.5$min with hourly release and for $\tau = 5$min with release per shift. The simulations have been carried out using the procedure described in Subsection 3.1. Based on these simulations, the lead times in open- and closed-loop cases (denoted as lt_{OL} and lt_{CL}) have been evaluated. The results are shown in Tables 5.1 and 5.2. From these results follows:

Observation 5.1 *Closed-loop raw material release according to (5.3) maintains the lead time close to the desired, whereas the open-loop release results, in most cases, in an unbounded lt.*

6 Analysis and Control of *LT* in Synchronous Exponential Lines with Non-identical Producing Machines

This section provides a generalization of Sections 3-5 to synchronous exponential lines with non-identical producing machines. While the formulas derived are extensions of those for identical machine case, the interpretations and insights are more obscure and some of the calculations are more complex.

Table 5.1: Lead time under control law (5.3) ($\tau = 0.5$min, once-per-hour release)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
150	0.7324	87	154.63	110.50
300	0.8161	97	521.06	255.05
600	0.8580	102	∞	581.96
1500	0.8832	105	∞	1575.48
3000	0.8916	106	∞	3215.71

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
20	0.7683	92	27.47	19.67
40	0.8356	100	∞	38.45
80	0.8682	104	∞	81.51
200	0.8873	106	∞	213.25
400	0.8937	107	∞	431.84

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
7	0.8581	102	∞	11.04
14	0.8806	105	∞	16.24
28	0.8907	106	∞	32.23
70	0.8963	107	∞	80.31
140	0.8982	107	∞	160.29

6.1 Analysis

6.1.1 LT evaluation

Proposition 6.1 *Consider a synchronous exponential line defined by assumptions (i)-(v) with* $e_0 < \min_{1 \leqslant i \leqslant M} e_i$. *Then, an estimate of* LT *(in min) is given by*

$$\widehat{LT} = M\tau + \sum_{i=0}^{M-1} \Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1 - e_{i+1}}{e_{i+1} - e_0}\Big). \tag{6.1}$$

This expression reduces to (3.1), if all producing machines are identical. Also, the qualitative properties of (3.1) hold for (6.1) as well. For instance, for fixed e_i, $i = 0, 1, \dots, M$, shorter up- and downtimes lead to shorter $\widehat{LT}$, and $\widehat{LT}$ tends to its minimum (i.e., $M\tau$) as $e_i \to 1$, $i = 1, 2, \dots, M$.

Table 5.2: Lead time under control law (5.3) (τ = 5min, once-per-shift release)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
20	0.7683	73	25.98	18.41
40	0.8356	80	∞	38.26
80	0.8682	83	∞	81.70
200	0.8873	85	∞	213.34
400	0.8937	85	∞	431.34

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
7	0.8581	82	∞	9.99
14	0.8806	84	∞	15.23
28	0.8907	85	∞	30.42
70	0.8963	86	∞	76.29
140	0.8982	86	∞	152.60

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
4	0.8916	85	∞	7.87
8	0.8964	86	∞	10.19
16	0.8983	86	∞	18.76
40	0.8994	86	∞	46.16
80	0.8997	86	∞	91.04

The accuracy of this estimate has been evaluated by simulating serial lines with non-identical producing machines and with M, e_i, $T_{down,i}$, $i = 1, 2, \ldots, M$, e_0, and $T_{down,0}$ selected randomly and equiprobably from the sets

$$
\begin{aligned}
&M \in [3, 10],\ e_i \in [0.8, 0.99],\ i = 1, 2, \ldots, M,\ e_0 \in [0.8 \min_{1\leqslant i\leqslant M} e_i,\ 0.99 \min_{1\leqslant i\leqslant M} e_i], \\
&T_{down,0} \in [10\text{min}, 100\text{min}],\ T_{down,i} \in [T_{down,0}, 1.1T_{down,0}],\ i = 1, 2, \ldots, M.
\end{aligned}
\tag{6.2}
$$

For each line, the analysis was carried out with τ = 0.5min and τ = 5min. The total of 1000 lines have been investigated using the simulation procedure outlined in Subsection 3.1 with $\epsilon_{LT} = \frac{|\widehat{LT}-LT|}{LT}\times 100\%$ as the measure of accuracy. The results turned out to be less precise than in the identical producing machine case. Namely, for τ = 0.5min, the smallest and largest errors were 0.0138% and 19.65%, respectively, and the average error was 3.94%; for τ = 5min, the smallest and largest errors were 0.0037% and 19.00%, respectively, and the average error was 3.71%. However, when e_i's were selected from sets $e_i \in [0.9, 0.99]$, $i = 1, 2, \dots, M$, and $e_0 \in [0.9 \min_{1\leqslant i\leqslant M} e_i,\ 0.99 \min_{1\leqslant i\leqslant M} e_i]$, the accuracy was similar to that of the identical producing machine case: for $\tau = 0.5$min, the smallest and largest errors were 0.0002% and 9.71%, respectively, and the average error was 1.42%; for $\tau = 5$min, the smallest and largest errors were 0.0002% and 9.54%, respectively, and the average error was 1.32%.

To further investigate $\widehat{LT}$ defined by (6.1), introduce a modified relative load factor

$$\rho_{max} := \frac{e_0}{e_{min}}, \tag{6.3}$$

where $e_{min} = \min_{1\leqslant i\leqslant M} e_i$, while keeping the relative lead time, $\widehat{lt}$, as in (3.3). Although reducing (6.1) to an expression for $\widehat{lt}(\rho_{max})$ seems to be formidable, the following upper bounds can be derived:

Proposition 6.2 *Under the assumptions of Proposition 6.1,*

$$\widehat{lt} \leqslant \overline{lt} := 1 + \frac{1}{\tau}\left(\frac{\rho_{max}}{M\mu_0} + \frac{2M-1}{M\mu_{min}}\frac{e_{max}}{e_{min}}\right)\left(\frac{1-e_{min}}{1-\rho_{max}}\right), \tag{6.4}$$

where $e_{min} = \min_{1\leqslant i\leqslant M} e_i$, $e_{max} = \max_{1\leqslant i\leqslant M} e_i$, *and* $\mu_{min} = \min_{1\leqslant i\leqslant M} \mu_i$. *Also, in the limit as* $M \to \infty$,

$$\overline{lt}_\infty := \lim_{M\to\infty} \overline{lt} = 1 + \left(\frac{2}{\tau\mu_{min}}\frac{e_{max}}{e_{min}}\right)\left(\frac{1-e_{min}}{1-\rho_{max}}\right). \tag{6.5}$$

These expressions again reduce to (3.4) and (3.5), respectively, if the producing machines are identical. Also, all qualitative properties of (3.4) and (3.5) hold for (6.4) and (6.5) as well. For instance, (6.5) does not depend on μ_0, while (6.4) does. Finally, the rate of convergence of (6.4) to (6.5) for $\mu_0 \geqslant \mu_{min}$, is also $\frac{1}{2M}$ (as in Section 3). So, the following chain of inequalities takes place:

$$\widehat{lt}(\rho_{max}) \leqslant \overline{lt}(\rho_{max}) \leqslant \overline{lt}_\infty(\rho_{max}). \tag{6.6}$$

This implies that if the release rate e_0 is selected so that the bound (6.5) satisfies the desired lead time, LT_d, the system performance will be at least as good as LT_d.

6.1.2 Structural properties of $\widehat{lt}(\rho_{max})$

Similar to the identical machine case, function $\widehat{lt}(\rho_{max})$ exhibits a knee-type behavior. This is illustrated in Figure 6.1 (solid curves) for the following three ten-machine lines:

$$\begin{aligned}
L_1 :\ \boldsymbol{e} &= [0.93, 0.95, 0.92, 0.82, 0.93, 0.94, 0.91, 0.94, 0.85, 0.89],\\
\boldsymbol{T_{down}} &= [13.87, 23.64, 16.06, 20.83, 13.02, 23.96, 17.57, 27.20, 27.07, 21.87],\\
L_2 :\ \boldsymbol{e} &= [0.96, 0.89, 0.96, 0.81, 0.97, 0.92, 0.82, 0.83, 0.89, 0.81],\\
\boldsymbol{T_{down}} &= [10.31, 24.94, 18.90, 28.64, 19.32, 18.37, 26.92, 20.50, 14.05, 23.44],\\
L_3 :\ \boldsymbol{e} &= [0.85, 0.96, 0.90, 0.86, 0.86, 0.81, 0.98, 0.96, 0.88, 0.97],\\
\boldsymbol{T_{down}} &= [24.54, 16.19, 26.77, 21.36, 17.41, 24.06, 20.93, 18.90, 23.89, 22.43],
\end{aligned} \tag{6.7}$$

where $\boldsymbol{e}$ and $\boldsymbol{T_{down}}$ are the vectors of producing machine efficiency and downtime, respectively. The parameters of the producing machines of these lines have been selected randomly and equiprobably from the following sets: $T_{down,i} \in [10\text{min}, 30\text{min}]$ and $e_i \in [0.8, 0.99]$, $i = 1, 2, \ldots, M$. For all three lines, the cycle time τ was selected as 1min and the release machine downtime as 10min.

While it seems impossible to quantify the position of the knee of $\widehat{lt}(\rho_{max})$, it can be lower-bounded by considering the knee of $\overline{lt}(\rho_{max})$ or $\overline{lt}_\infty(\rho_{max})$. The behavior of these functions is also shown in Figure 6.1 (by dashed and dash-dot curves, which practically overlay each other). The knees of $\overline{lt}(\rho_{max})$ and $\overline{lt}_\infty(\rho_{max})$ are quantified as follows:

Proposition 6.3 *Under the assumptions of Proposition 6.1, the knees of $\overline{lt}$ and $\overline{lt}_\infty$ are given, respectively, by*

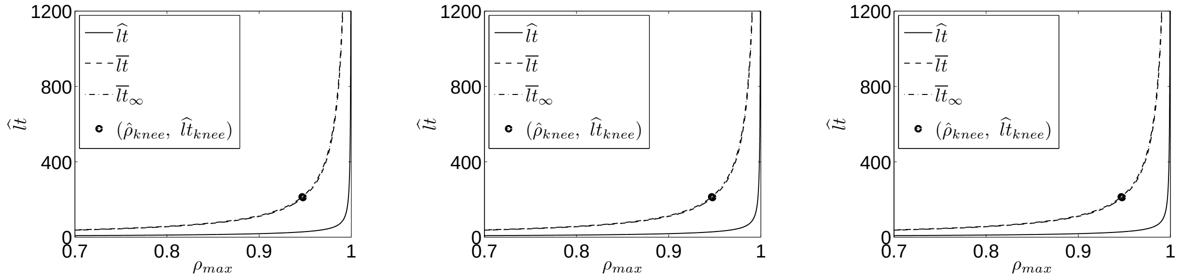
$$\bar{\rho}_{knee}(\overline{lt}) = 1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)(1 - e_{min})}, \tag{6.8}$$

$$\bar{\rho}_{\infty,knee}(\overline{lt}_\infty) = \lim_{M\to\infty} \bar{\rho}_{knee}(\overline{lt}) = 1 - \sqrt{\frac{2\alpha e_{max}}{\tau\mu_{min}e_{min}}(1 - e_{min})}, \tag{6.9}$$

and, if $\mu_0 \geqslant \mu_{min}$,

$$\rho_{knee}(\widehat{lt}) \geqslant \bar{\rho}_{\infty,knee}(\overline{lt}_\infty). \tag{6.10}$$

The knees of $\overline{lt}_\infty(\rho_{max})$ are shown in Figure 6.1 by black dots. Thus, releasing raw material with the load factor $\rho_{max} \leqslant \bar{\rho}_{\infty,knee}$ ensures a safe system operation from the point of view of lead time.



(a) L_1 (b) L_2 (c) L_3

Figure 6.1: Relative lead time, $\widehat{lt}$, as a function of relative workload, ρ_{max} (for $M = 10$, $\tau = 1$min)

6.2 Open-loop control

Proposition 6.4 *Under the assumptions of Proposition 6.1, the sets of feasible lead times, $\mathscr{F}_{\widehat{lt}}$, $\mathscr{F}_{\overline{lt}}$, and $\mathscr{F}_{\overline{lt}_\infty}$, are given, respectively, by*

$$\begin{aligned} \widehat{lt} &> 1 + \frac{1}{M\tau}\Big(\sum_{i=1}^{M}\frac{1-e_i}{\mu_i} + \sum_{i=1}^{M-1}\frac{e_i(1-e_{i+1})}{\mu_i e_{i+1}}\Big), \\ \overline{lt} &> 1 + (1-e_{min})\frac{2M-1}{M\tau\mu_{min}}\frac{e_{max}}{e_{min}}, \\ \overline{lt}_\infty &> 1 + \frac{2(1-e_{min})}{\tau\mu_{min}}\frac{e_{max}}{e_{min}}, \end{aligned} \tag{6.11}$$

where $e_{min} = \min\limits_{1\leqslant i\leqslant M} e_i$, $e_{max} = \max\limits_{1\leqslant i\leqslant M} e_i$, and $\mu_{min} = \min\limits_{1\leqslant i\leqslant M} \mu_i$.

Proposition 6.5 *Under the assumptions of Proposition 6.1, the release rate, $\hat{e}_0^*$, which ensures the desired lead time, $lt_d \in \mathscr{F}_{\widehat{lt}}$, is the unique real root less than $\min\limits_{1\leqslant i\leqslant M} e_i$ of the following M-th order polynomial equation:*

$$(LT_d - M\tau)\prod_{i=0}^{M-1}(e_{i+1}-e_0) - (1-e_1)\Big(\frac{e_0}{\mu_0}+\frac{e_1}{\mu_1}\Big)\prod_{i=1}^{M-1}(e_{i+1}-e_0) - \sum_{i=1}^{M-1}\Big((1-e_{i+1})\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\prod_{j=0,j\neq i}^{M-1}(e_{j+1}-e_0)\Big) = 0. \tag{6.12}$$

For this release rate,

$$\widehat{TP}^* = \frac{\hat{e}_0^*}{\tau},\ \widehat{WIP}_0^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{\hat{e}_0^*}{\mu_0}+\frac{e_1}{\mu_1}\Big)\Big(\frac{1-e_1}{e_1-\hat{e}_0^*}\Big),\ \widehat{WIP}_i^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-\hat{e}_0^*}\Big),\ i = 1,2,\ldots,M-1. \tag{6.13}$$

Solving equation (6.12) might be too complex for practical applications. Therefore, using the upper bounds $\overline{lt}$ and $\overline{lt}_\infty$, we provide below more convenient lower bounds on $\hat{e}_0^*$.

Proposition 6.6 *Let $\bar{e}_0^*$ and $\bar{e}_{0,\infty}^*$ be the release rates that solve the open-loop lead time control problem for $\overline{lt}$ and $\overline{lt}_\infty$ with $lt_d \in \{\mathscr{F}_{\widehat{lt}} \cap \mathscr{F}_{\overline{lt}} \cap \mathscr{F}_{\overline{lt}_\infty}\}$. Then,*

$$\bar{e}_0^* = e_{min}\left[1 - \frac{\mu_{min} + (2M-1)\mu_0 \frac{e_{max}}{e_{min}}}{M\mu_{min}\mu_0\tau(lt_d - 1) + \mu_{min}(1 - e_{min})}(1 - e_{min})\right], \tag{6.14}$$

$$\bar{e}_{0,\infty}^* = e_{min}\left[1 - \frac{2(1 - e_{min})\frac{e_{max}}{e_{min}}}{\tau\mu_{min}(lt_d - 1)}\right], \tag{6.15}$$

and, if $\mu_0 \geqslant \mu_{min}$,

$$\hat{e}_0^* \geqslant \bar{e}_0^* \geqslant \bar{e}_{0,\infty}^*. \tag{6.16}$$

Similar to the identical machine case, the solution of the open-loop lead time control problem for non-identical machines can be implemented by releasing raw material once-per-cycle with probability $\bar{e}_0^*$ given by (6.14) with $\mu_0 = \infty$, i.e.,

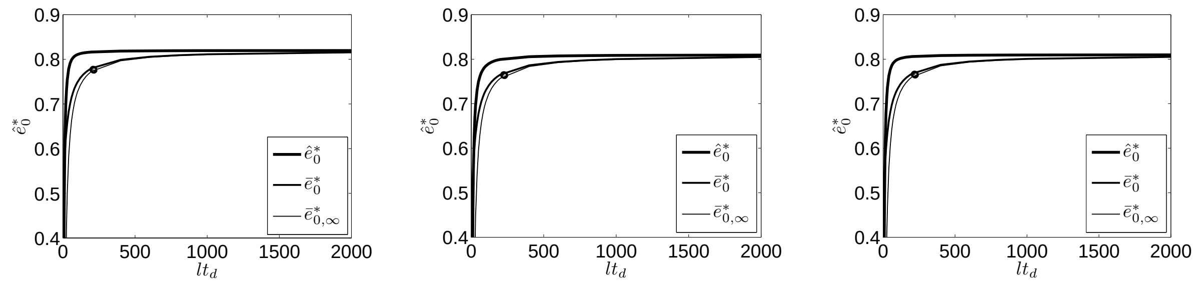
$$p = e_{min}\left[1 - \frac{(2M-1)\frac{e_{max}}{e_{min}}}{M\mu_{min}\tau(lt_d - 1)}(1 - e_{min})\right], \tag{6.17}$$

or, for long lines, with $p = \bar{e}_{0,\infty}^*$ given by (6.15).

The behavior of $\hat{e}_0^*$, $\bar{e}_0^*$, and $\bar{e}_{0,\infty}^*$ is illustrated in Figure 6.2 (where black dots indicate the knee) as a function of lt_d for the three lines given in (6.7). From this figure, we conclude that, similar to the identical machine case, raw material release with rates beyond the knee is not only unnecessary, but detrimental as well.

6.3 Closed-loop control

The control law (5.3) remains applicable to the case of non-identical machines as well, with the only difference that $\hat{e}_0^*$ in (4.14) and (5.2) is calculated according to either (6.12) or (6.14) or (6.15). The closed-loop performance under this law remains roughly the same as for the identical machine case; a detailed quantification of this performance is omitted due to space limitations.



(a) L_1 (b) L_2 (c) L_3

Figure 6.2: Release rates, $\hat{e}_0^*$, $\bar{e}_0^*$, and $\bar{e}_{0,\infty}^*$, as a function of the desired relative lead time, lt_d (for $M = 10$)

7 Extensions

7.1 Asynchronous exponential lines

Although analytical methods for performance evaluation of asynchronous exponential lines are readily available (see, for instance, [43]), and the lead time in these lines can be analyzed in full details, due to space limitations, we discuss here only upper bounds on this performance measure.

Consider an asynchronous exponential line defined by assumptions (i)-(v). Assume that

$$\tau_0 = \min_{1\leqslant i\leqslant M} \tau_i, \quad TP_{isol,0} < \min_{1\leqslant i\leqslant M} TP_{isol,i}, \tag{7.1}$$

and, introduce the relative load factor as follows:

$$\rho_{async} := \frac{TP_{isol,0}}{\min\limits_{1\leqslant i\leqslant M} TP_{isol,i}}. \tag{7.2}$$

Along with this asynchronous line, consider an auxiliary synchronous line with the same release machine and the producing machines defined as follows:

$$\check{\tau}_i = \tau_0, \quad \check{\mu}_i = \mu_i, \quad \check{TP}_{isol,i} = TP_{isol,i}, \ i = 1, 2, \ldots, M. \tag{7.3}$$

From these expressions, we obtain:

$$\check{e}_i = \frac{\tau_0}{\tau_i} e_i, \quad \check{\lambda}_i = \frac{\mu_i}{\check{e}_i}(1 - \check{e}_i), \ i = 0, 1, \ldots, M. \tag{7.4}$$

With these parameters, the relative load factor for the auxiliary line, ρ_{sync}, is the same as for the asynchronous one, i.e., given by (7.2). For the sake of brevity, we omit the subscript of both load factors and denote them as ρ.

Let LT_{async} and LT_{sync} denote the lead times of the original asynchronous line and the auxiliary synchronous one, respectively. We analyze the relationship between LT_{async} and LT_{sync} by simulations. To accomplish this, we form 1000 asynchronous lines with parameters selected randomly and equiprobably from the following sets:

$$\begin{aligned} &M \in [3, 10],\ \tau_i \in [0.8\text{min}, 1.2\text{min}],\ e_i \in [0.7, 0.99],\ i = 1, 2, \ldots, M, \\ &TP_{isol,0} \in [0.7 \min_{1 \leqslant i \leqslant M} TP_{isol,i},\ 0.99 \min_{1 \leqslant i \leqslant M} TP_{isol,i}], \\ &T_{down,i} \in [10\text{min}, 100\text{min}],\ i = 0, 1, \ldots, M. \end{aligned} \tag{7.5}$$

For each of these asynchronous lines, we form an auxiliary synchronous one according to (7.3) and simulate the resulting 1000 pairs of lines using the procedure described in Subsection 3.1. As a result, we obtain:

Observation 7.1 *For all* 1000 *pairs of lines analyzed,* $LT_{async} < LT_{sync}$*. The tightness of this upper bound is* $0.03 \leqslant \Delta_{LT} \leqslant 1.7$*, where* $\Delta_{LT} := \frac{LT_{sync} - LT_{async}}{LT_{async}}$.

7.2 Non-exponential lines

Since analytical methods for performance evaluation of non-exponential lines are not available, the only tool for LT investigation is simulation. Using simulations, this subsection provides upper bounds of LT for both synchronous and asynchronous cases.

7.2.1 Synchronous non-exponential lines

Consider non-exponential lines with ten identical machines having $\tau = 1$min, $T_{dwon} = 5$min, and other parameters selected as combinations from the following sets:

$$\begin{aligned}
&e \in \{0.7, 0.8, 0.9\},\ \rho \in \{0.7, 0.8, 0.9, 0.99\},\\
&\text{Reliability model} \in \{\text{Weibull, gamma, log-normal}\},\\
&CV \in \{0.01, 0.1, 0.25, 0.5, 0.75, 1\},
\end{aligned} \tag{7.6}$$

where CV is the coefficient of variation of the machine reliability model. We confine the analysis to $CV \leqslant 1$ since, as it is shown in an empirical study [45], most of manufacturing equipment in the automotive industry has CV's of up- and downtimes less than 1, and an analytical study [46] proves that if the breakdown and repair rates are increasing in time, the respective CV's are again less than 1.

Simulating the 216 lines, thus formed, we evaluate their lt shown in Table 7.1 by solid curves. As one can see, lt is a monotonically increasing function of CV and, more importantly, it is *practically independent of the machine reliability model*.

Similar results have been obtained for synchronous lines with non-identical non-exponential machines. An example is shown in Figure 7.1 for the three ten-machine lines with e_i and $T_{down,i}$ given in (6.7), $\tau = 1$min, $T_{down,0} = \min_{1\leqslant i\leqslant M} T_{down,i}$, $\rho_{max} = 0.9$, and with the reliability models and CV's given in (7.6).

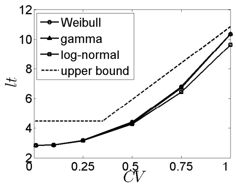
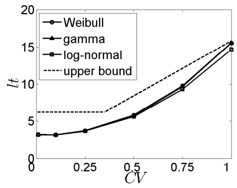
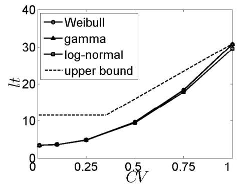
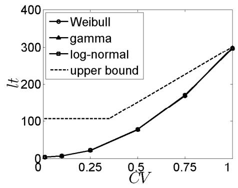
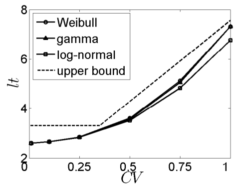
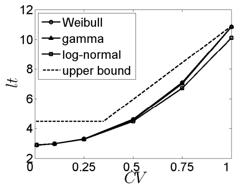
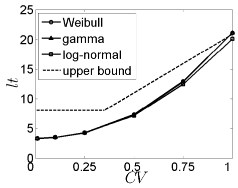
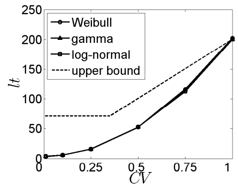
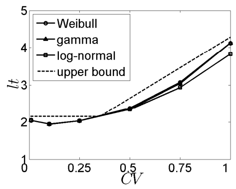
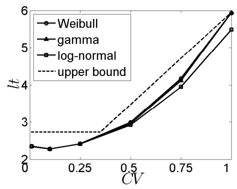
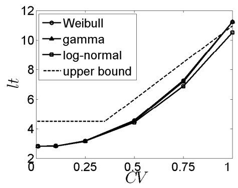
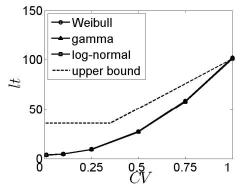
The curves of Table 7.1 and Figure 7.1 offer a possibility to introduce an upper bound on lt for systems under consideration. Indeed, since lt for $CV = 1$ equals that of exponential lines, and the smallest lt in any line is 1, the above mentioned simulation results lead to the following:

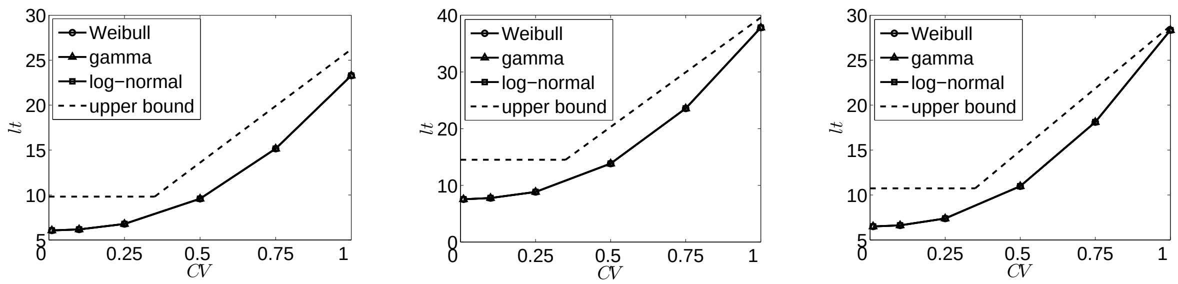
Observation 7.2 *For all synchronous non-exponential lines analyzed in this subsection, lt is upper-bounded by the following empirical formula* (see the broken lines in Table 7.1 and Figure 7.1):

$$lt \leqslant \begin{cases} (\widehat{lt}_{exp} - 1)CV + 1, & for\ 0.35 < CV \leqslant 1, \\ 0.35(\widehat{lt}_{exp} - 1) + 1, & for\ 0 < CV \leqslant 0.35, \end{cases} \tag{7.7}$$

where $\widehat{lt}_{exp}$ is the lt for the corresponding exponential line calculated according to (3.4) for identical machines or (3.3) and (6.1) for non-identical ones.

Table 7.1: Lead time for synchronous non-exponential lines with identical producing machines ($M = 10$)

e	ρ			
	0.7	0.8	0.9	0.99
0.7				
0.8				
0.9				



(a) L_1 (b) L_2 (c) L_3

Figure 7.1: Lead time for synchronous non-exponential lines with non-identical producing machines ($M = 10$)

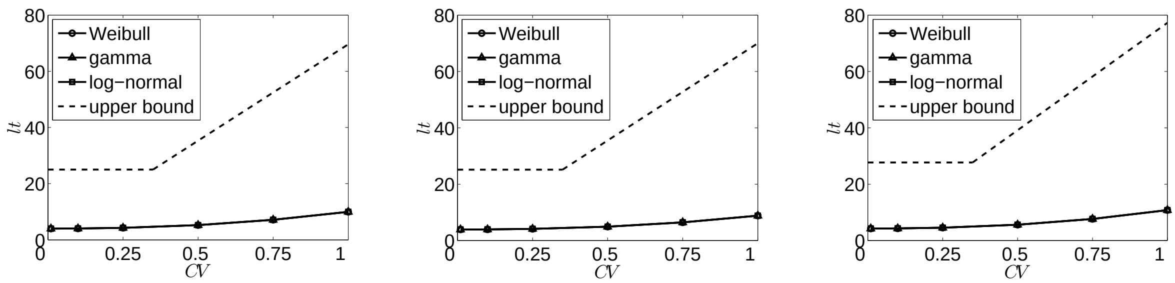
7.2.2 Asynchronous non-exponential lines

As in Subsection 7.1, we analyze asynchronous non-exponential lines by forming an auxiliary synchronous exponential line and showing, by simulations, that lt of the latter can be used as $\widehat{lt}_{exp}$ in the upper bound (7.7) for the former.

To accomplish this, we "de-synchronize" the synchronous lines (6.7) by selecting τ_i randomly and equiprobably from the set [0.8min, 1.2min], and choosing $\tau_0 = \min\limits_{1 \leqslant i \leqslant M} \tau_i$, $T_{down,0} = \min\limits_{1 \leqslant i \leqslant M} T_{down,i}$, and $\rho_{async} = 0.9$. Simulating these lines, we evaluate their lt's shown in Figure 7.2 by solid curves. For each of these lines, we form a synchronous exponential line according to (7.3), evaluate its lt and use it in expression (7.7) as $\widehat{lt}_{exp}$. The resulting upper bound is shown in Figure 7.2 by the broken lines. As one can see, the upper bound still holds, albeit with a lesser tightness. Note, however, that if τ_i's vary around their average value in a ±10% domain (rather than in the ±20% range, as analyzed above), bound (7.7) is more than twice tighter.

While Observation 7.2 refers to synchronous lines with Weibull, gamma, and log-normal reliability models, motivated by the above result, we expect that it has a wider applicability. Specifically, we formulate:

Conjecture 7.1 *Upper bound (7.7) holds for synchronous and asynchronous lines with any machine reliability model having a unique maximum* (the so-called unimodal pdf's) *as long as* $CV \leqslant 1$.



(a) L_1 (b) L_2 (c) L_3

Figure 7.2: Lead time for asynchronous non-exponential lines ($M = 10$)

8 Conclusions and Future Work

This paper provides methods for evaluation and control of production lead time in serial lines. Detailed results are obtained for synchronous exponential lines. Asynchronous exponential and non-exponential lines are investigated in fewer details – only upper bounds and empirical formulas for their LT's are provided. Numerous problems in this area, however, remain open. These include:

- Detailed analysis of LT in asynchronous exponential lines. This can be accomplished using, for instance, the performance analysis technique developed in [43].
- Detailed investigation of LT in synchronous and asynchronous lines with non-exponential machines. This can be accomplished by extensive simulations of systems at hand and investigating the tightness of the upper bounds on their LT's (based on auxiliary exponential lines). This research may lead to validation or repudiation of Conjecture 7.1.
- Detailed analysis of feedback control laws for raw material release in asynchronous exponential and non-exponential lines. The relay-type control law introduced in this paper could be a starting point for this investigation. It seems likely that this control law may be applicable to serial lines with arbitrary unimodal machine reliability model.
- Extension of the results obtained to re-entrant lines, where the lead time is known to be exceptionally long, and the issue of limiting the lead time is of particular importance.
- Extension of the results obtained to assembly system, where the main line is fed by several component lines, and the total lead time must be coordinated and controlled.

Solutions of these problems will result in a relatively complete and practical theory of lead time in manufacturing systems and enable a novel paradigm for production management – maximize the throughput under the constraint that the lead time takes the desired value.

Appendix

The analysis of $\widehat{LT}$ for synchronous exponential lines is based on the recursive aggregation procedure described in [43]. For serial lines with $M+1$ synchronous exponential machines defined by $(\lambda_0, \mu_0), (\lambda_1, \mu_1), \dots, (\lambda_M, \mu_M)$ and M buffers with capacity $N_0, N_1, \dots, N_{M-1}$, the steady state of this procedure, λ_i^f, μ_i^f, $i = 1, 2, \dots, M$, and λ_i^b, μ_i^b, $i = 0, 1, \dots, M-1$, is the unique solution of the following system of transcendental equations:

$$
\begin{aligned}
\mu_i^f &= \mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, N_{i-1}),\ 1 \leqslant i \leqslant M,\\
\lambda_i^f &= \lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, N_{i-1}),\ 1 \leqslant i \leqslant M,\\
\mu_i^b &= \mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, N_i),\ 0 \leqslant i \leqslant M-1,\\
\lambda_i^b &= \lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, N_i),\ 0 \leqslant i \leqslant M-1,
\end{aligned}
\tag{A.1}
$$

with the boundary conditions $\lambda_0^f = \lambda_0, \mu_0^f = \mu_0$ and $\lambda_M^b = \lambda_M, \mu_M^b = \mu_M$ and

$$
Q(x_1, y_1, x_2, y_2, N) = \begin{cases} \frac{(1-e_1)(1-\phi)}{1-\phi\exp(-\beta N)}, & \text{if } \frac{x_1}{y_1} \neq \frac{x_2}{y_2},\\ \frac{x_1(x_1+x_2)(y_1+y_2)}{(x_1+y_1)[(x_1+x_2)(y_1+y_2)+x_2y_1(x_1+x_2+y_1+y_2)N]}, & \text{if } \frac{x_1}{y_1} = \frac{x_2}{y_2}, \end{cases}
\tag{A.2}
$$

where

$$
\begin{aligned}
e_i &= \frac{y_i}{x_i+y_i},\ i = 1, 2,\\
\phi &= \frac{e_1(1-e_2)}{e_2(1-e_1)},\\
\beta &= \frac{(x_1+x_2+y_1+y_2)(x_1y_2 - x_2y_1)}{(x_1+x_2)(y_1+y_2)}.
\end{aligned}
\tag{A.3}
$$

The proofs of Propositions 3.1 and 6.1 are based on the following three lemmas, which extend (A.1)-(A.3) to the case of $N_i = \infty$.

Lemma A.1 *Function $Q(x_1, y_1, x_2, y_2, N)$, defined by (A.2) and (A.3), has the following limit:*

$$
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) = \begin{cases} 0, & \textit{if } \frac{x_1}{y_1} \leqslant \frac{x_2}{y_2},\\ 1 - \frac{e_1}{e_2}, & \textit{if } \frac{x_1}{y_1} > \frac{x_2}{y_2}, \end{cases}
\tag{A.4}
$$

where $e_i = \frac{y_i}{x_i+y_i}$, $i = 1, 2$.

Proof: From (A.2),

- if $\frac{x_1}{y_1} = \frac{x_2}{y_2}$,

$$
\begin{aligned}
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) &= \lim_{N\to\infty} \frac{x_1(x_1+x_2)(y_1+y_2)}{(x_1+y_1)[(x_1+x_2)(y_1+y_2)+x_2y_1(x_1+x_2+y_1+y_2)N]}\\
&= 0;
\end{aligned}
\tag{A.5}
$$

- if $\frac{x_1}{y_1} < \frac{x_2}{y_2}$, then

$$\beta = \frac{(x_1 + x_2 + y_1 + y_2)(x_1 y_2 - x_2 y_1)}{(x_1 + x_2)(y_1 + y_2)} < 0, \tag{A.6}$$

  and, thus,

$$\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) = \lim_{N\to\infty} \frac{(1 - e_1)(1 - \phi)}{1 - \phi \exp(-\beta N)} = 0; \tag{A.7}$$

- if $\frac{x_1}{y_1} > \frac{x_2}{y_2}$, then $\beta > 0$, and, therefore,

$$\begin{aligned}\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) &= \lim_{N\to\infty} \frac{(1 - e_1)(1 - \phi)}{1 - \phi \exp(-\beta N)} \\ &= (1 - e_1)(1 - \phi) \\ &= (1 - e_1)\Big(1 - \frac{e_1(1 - e_2)}{e_2(1 - e_1)}\Big) \\ &= 1 - \frac{e_1}{e_2}.\end{aligned} \tag{A.8}$$

■

Lemma A.2 *Let $e_j = \min\{e_1, e_2, \ldots, e_M\}$, and j is the smallest index at which the minimum is achieved. Let $e_i^f := \frac{\mu_i^f}{\lambda_i^f + \mu_i^f}$ and $e_i^b := \frac{\mu_i^b}{\lambda_i^b + \mu_i^b}$. Then, for $N_i = \infty$, $i = 0, 1, \ldots, M - 1$,*

$$e_i^f = \begin{cases} e_i, & \textit{if } i < j, \\ e_j, & \textit{if } i \geqslant j, \end{cases} \qquad e_i^b = \begin{cases} e_j, & \textit{if } i \leqslant j, \\ e_i, & \textit{if } i > j, \end{cases} \tag{A.9}$$

and the unique solution of (A.1) is

$$\begin{aligned} \lambda_i^f &= (\lambda_i + \mu_i)(1 - e_i^f), \quad \mu_i^f = (\lambda_i + \mu_i)e_i^f, \\ \lambda_i^b &= (\lambda_i + \mu_i)(1 - e_i^b), \quad \mu_i^b = (\lambda_i + \mu_i)e_i^b. \end{aligned} \tag{A.10}$$

Proof: We show that (A.9), (A.10) is the solution of (A.1) and then comment its uniqueness.

Since $e_j = \min_{1\leqslant i\leqslant M} e_i$, i.e., $e_j \leqslant e_i$, $\forall i = 0, 1, \ldots, M$, we have

$$\frac{1 - e_j}{e_j} \geqslant \frac{1 - e_i}{e_i}, \ \forall i = 0, 1, \ldots, M. \tag{A.11}$$

- If $i < j$, then based on (A.4), (A.9), (A.10), and (A.11), we have

$$\begin{aligned} & Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) \\ &= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty] \\ &= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}), (\lambda_{i-1} + \mu_{i-1})e_{i-1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty] \\ &= 0 \end{aligned} \tag{A.12}$$

and

$$\begin{aligned} & Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) \\ &= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty] \\ &= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_j), (\lambda_{i+1} + \mu_{i+1})e_j, (\lambda_i + \mu_i)(1 - e_i), (\lambda_i + \mu_i)e_i, \infty] \\ &= 1 - \frac{e_j}{e_i}. \end{aligned} \tag{A.13}$$

Thus, for the left- and right-hand sides of the first equation of (A.1), we have, respectively,

$$\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_i = \mu_i \tag{A.14}$$

and

$$\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i, \tag{A.15}$$

implying that (A.9) and (A.10) solve the first equation of (A.1) for $i < j$. Similarly, for the left- and right-hand sides of the second equation of (A.1), we have,

$$\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i \tag{A.16}$$

and

$$\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i, \tag{A.17}$$

implying that (A.9) and (A.10) solve the second equation of (A.1) for $i < j$. For the third equation of (A.1), the left- and right-hand sides are respectively

$$\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_j \tag{A.18}$$

and

$$\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i \frac{e_j}{e_i} = (\lambda_i + \mu_i)e_j, \tag{A.19}$$

implying that (A.9) and (A.10) solve the third equation of (A.1) for $i < j$. As for the last equation of (A.1), the left- and right-hand sides are

$$\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_j) \tag{A.20}$$

and

$$\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i + \mu_i\Big(1 - \frac{e_j}{e_i}\Big) = (\lambda_i + \mu_i)(1 - e_j), \tag{A.21}$$

implying that (A.9) and (A.10) solve the last equation of (A.1) for $i < j$.

- If $i = j$, the two Q-functions in (A.1) are respectively

$$\begin{aligned}
&Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) \\
&= Q\big[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty\big] \\
&= Q\big[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}), (\lambda_{i-1} + \mu_{i-1})e_{i-1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty\big] \\
&= 0
\end{aligned} \tag{A.22}$$

and

$$\begin{aligned}
&Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) \\
&= Q\big[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty\big] \\
&= Q\big[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}), (\lambda_{i+1} + \mu_{i+1})e_{i+1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty\big] \\
&= 0.
\end{aligned} \tag{A.23}$$

Thus, the left- and right-hand sides of (A.1) are

$$\begin{aligned}
&\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_j = (\lambda_i + \mu_i)e_i = \mu_i, \\
&\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i, \\
&\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_j) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i, \\
&\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i, \\
&\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_j = (\lambda_i + \mu_i)e_i = \mu_i, \\
&\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i, \\
&\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_j) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i, \\
&\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i,
\end{aligned} \tag{A.24}$$

implying that (A.1) is solved for $i = j$.

- If $i > j$, the two Q-functions in (A.1) are

$$\begin{aligned}
&Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) \\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty] \\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_j), (\lambda_{i-1} + \mu_{i-1})e_j, (\lambda_i + \mu_i)(1 - e_i), (\lambda_i + \mu_i)e_i, \infty] \\
&= 1 - \frac{e_j}{e_i}
\end{aligned} \tag{A.25}$$

and

$$\begin{aligned}
&Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) \\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty] \\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}), (\lambda_{i+1} + \mu_{i+1})e_{i+1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty] \\
&= 0.
\end{aligned} \tag{A.26}$$

Thus, the left- and right-hand sides of (A.1) are

$$\begin{aligned}
&\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_j,\\
&\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i \frac{e_j}{e_i} = (\lambda_i + \mu_i)e_j,\\
&\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_j),\\
&\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i + \mu_i\Big(1 - \frac{e_j}{e_i}\Big) = (\lambda_i + \mu_i)(1 - e_j),\\
&\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_i = \mu_i,\\
&\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i,\\
&\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i,\\
&\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i,
\end{aligned} \tag{A.27}$$

which also implies that (A.1) is solved.

As far as the uniqueness of (A.9) and (A.10) is concerned, it follows directly from Theorem 11.4 of [43]. ■

Lemma A.3 *In synchronous exponential two-machine lines with $e_1 < e_2$,*

$$\lim_{N\to\infty} WIP = \frac{e_1}{\tau}\Big(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\Big)\Big(\frac{1 - e_2}{e_2 - e_1}\Big). \tag{A.28}$$

Proof: From the proof of Theorem 11.3 in [43], we know that

$$\lim_{N\to\infty} WIP = \lim_{N\to\infty} \frac{D_5}{D_2 + D_3} = \lim_{N\to\infty} \frac{\frac{D_2}{-K}}{D_2 + D_3}, \tag{A.29}$$

where $\lim\limits_{N\to\infty} D_2 = -\frac{2 + D_1 + \frac{1}{D_1}}{K}$, $D_1 = \frac{\mu_1+\mu_2}{\lambda_1+\lambda_2}$, $K = \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}$, $D_3 = \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}$. Thus, we have

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{WIP} &= \lim_{N\to\infty} -K(\frac{D_3}{D_2}+1) = K^2\frac{D_3}{2+D_1+\frac{1}{D_1}} - K \\
&= \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)^2(\lambda_2\mu_1-\lambda_1\mu_2)^2}{(\lambda_1+\lambda_2)^2(\mu_1+\mu_2)^2}\frac{\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}}{2+\frac{\mu_1+\mu_2}{\lambda_1+\lambda_2}+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}} \\
&\quad - \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)} \\
&= \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)^2(\lambda_2\mu_1-\lambda_1\mu_2)^2\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)^2} \\
&\quad - \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)} \\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)} \\
&\quad -(\lambda_1+\lambda_2+\mu_1+\mu_2)\Big] \\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1) \\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)^2\Big] \\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1) \\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1) \\
&\quad -\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_1+\mu_2)\Big] \\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[-\lambda_1\mu_2(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1) \\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_1+\mu_2)\Big] \\
&= \frac{\lambda_1\mu_2-\lambda_2\mu_1}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_1+\mu_2)+\frac{\lambda_1\mu_2(\lambda_2+\mu_1)}{\lambda_2\mu_1}+\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big] \\
&= \frac{(\lambda_1+\mu_1)(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_1+\mu_2)+\frac{\lambda_1\mu_2(\lambda_2+\mu_1)}{\lambda_2\mu_1}+\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big] \\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_2+\mu_2)(1-e_2)e_1}\Big[\lambda_2\mu_1(\lambda_1+\mu_2)+\lambda_1\mu_2(\lambda_2+\mu_1) \\
&\quad +\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_1+\lambda_2+\mu_1+\mu_2}\Big] \\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{e_1(1-e_2)}\Big[1+\frac{1}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_2+\mu_2)}\Big(\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_1+\lambda_2+\mu_1+\mu_2}-\lambda_2^2(\mu_1+\mu_2) \\
&\quad -\mu_2^2(\lambda_1+\lambda_2)\Big)\Big] \\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{e_1(1-e_2)}\Big[1-\frac{\lambda_2+\mu_2}{\lambda_1+\lambda_2+\mu_1+\mu_2}\Big] = \frac{(\lambda_1+\mu_1)(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2+\mu_1+\mu_2)e_1(1-e_2)} \\
&= \frac{\mu_1\mu_2(e_2-e_1)}{(\mu_1e_2+\mu_2e_1)e_1(1-e_2)}.
\end{aligned}
\tag{A.30}
$$

Therefore,

$$\lim_{N\to\infty} WIP = \frac{(\mu_1 e_2 + \mu_2 e_1)e_1(1-e_2)}{\mu_1\mu_2(e_2-e_1)} = e_1\Big(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\Big)\Big(\frac{1-e_2}{e_2-e_1}\Big). \tag{A.31}$$

In [43], (A.29) is derived for $\tau = 1$. For general τ, (A.28) follows (see proofs for Chapter 11 in Chapter 20 of [43]). ■

Proof of Proposition 3.1: For the synchronous exponential production line defined by assumptions (i)-(v) with $\lambda_i = \lambda$, $\mu_i = \mu$, $i = 1, 2, \dots, M$ and $e_0 < e$, based on Lemma A.2 we obtain

$$\begin{aligned}
&e_0^f = e_0,\ \mu_0^f = (\lambda_0 + \mu_0)e_0^f = (\lambda_0 + \mu_0)e_0 = \mu_0,\\
&e_i^f = e_0,\ \mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda + \mu)e_0,\ i = 1, 2, \dots, M-1,\\
&e_i^b = e,\ \mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda + \mu)e = \mu,\ i = 1, 2, \dots, M,
\end{aligned} \tag{A.32}$$

which, using Lemma A.3 and [43] (Section 11.1.2), implies that the occupancy of each buffer is

$$\begin{aligned}
\widehat{WIP}_0 &= \frac{e_0^f}{\tau}\Big(\frac{e_0^f}{\mu_0^f} + \frac{e_1^b}{\mu_1^b}\Big)\Big(\frac{1-e_1^b}{e_1^b - e_0^f}\Big)\\
&= \frac{e_0}{\tau}\Big(\frac{e_0}{\mu_0} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-e_0}\Big)
\end{aligned} \tag{A.33}$$

and

$$\begin{aligned}
\widehat{WIP}_i &= \frac{e_i^f}{\tau}\Big(\frac{e_i^f}{\mu_i^f} + \frac{e_{i+1}^b}{\mu_{i+1}^b}\Big)\Big(\frac{1-e_{i+1}^b}{e_{i+1}^b - e_i^f}\Big)\\
&= \frac{e_0}{\tau}\Big(\frac{1}{\lambda+\mu} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-e_0}\Big)\\
&= \frac{2e_0 e}{\mu\tau}\Big(\frac{1-e}{e-e_0}\Big),\ i = 1, 2, \dots, M-1.
\end{aligned} \tag{A.34}$$

Thus, taking into account that

$$\widehat{TP} = \frac{e_0}{\tau}, \tag{A.35}$$

and using Little's law, from (A.33) and (A.34) we obtain the estimate of the waiting time in all buffers as $\frac{\sum_{i=0}^{M-1}\widehat{WIP}_i}{\widehat{TP}} = \Big[\frac{e_0}{\mu_0} + (2M-1)\frac{e}{\mu}\Big]\Big(\frac{1-e}{e-e_0}\Big)$. Since the processing time in all producing machines is $M\tau$, adding the last two quantities results in (3.1). ■

Proof of Proposition 3.2: Let

$$f(\rho) := \alpha\widehat{lt}(\rho) = \alpha\Big(1 + \frac{1}{M\tau}\Big(\frac{\rho}{\mu_0} + \frac{2M-1}{\mu}\Big)\Big(\frac{1-e}{1-\rho}\Big)\Big). \tag{A.36}$$

Then

$$f'(\rho) = \frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)\frac{1-e}{(1-\rho)^2},$$
$$f''(\rho) = \frac{2\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)\frac{1-e}{(1-\rho)^3},$$

and, therefore,

$$\kappa(f(\rho)) = \frac{\left|f''_{\rho\rho}\right|}{(1+f'^2_{\rho})^{\frac{3}{2}}} = \frac{\frac{2\alpha}{M\tau}(\frac{1}{\mu_0} + \frac{2M-1}{\mu})(1-e)}{\Big[(1-\rho)^2 + \frac{\alpha^2}{M^2\tau^2}(\frac{1}{\mu_0} + \frac{2M-1}{\mu})^2\frac{(1-e)^2}{(1-\rho)^2}\Big]^{\frac{3}{2}}}. \tag{A.37}$$

Since $\rho_{knee} = \arg\max_{\rho} \kappa(f(\rho))$, differentiating (A.37) and solving for ρ, we obtain (3.9). Clearly, when M tends to infinity, (3.9) becomes (3.10). ■

Proof of Proposition 4.1: As it follows from (3.4), $\widehat{lt}$ is an increasing function of ρ. Since $0 < \rho < 1$, this implies that

$$\widehat{lt} > 1 + (1-e)\frac{2M-1}{M}\frac{T_{down}}{\tau}. \tag{A.38}$$

For $M \to \infty$ the above inequality becomes

$$\widehat{lt}_{\infty} > 1 + 2(1-e)\frac{T_{down}}{\tau}. \tag{A.39}$$

■

Proof of Proposition 4.2: From (3.4) it follows that

$$\rho^* = 1 - \frac{\mu + (2M-1)\mu_0}{M\mu\mu_0\tau(lt_d - 1) + \mu(1-e)}(1-e), \tag{A.40}$$

which implies that (4.2) holds. As for (4.3), it follows immediately from the proof of Proposition 3.1. ■

Proof of Proposition 6.1: Similar to the proof of Proposition 3.1, with the only difference that, instead of (A.32), we have

$$
\begin{aligned}
&e_i^f = e_0,\ \mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_0,\ i = 0, 1, \ldots, M-1, \\
&e_i^b = e_i,\ \mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_i = \mu_i,\ i = 1, 2, \ldots, M,
\end{aligned}
\tag{A.41}
$$

and, therefore,

$$
\begin{aligned}
\widehat{WIP}_0 &= \frac{e_0^f}{\tau}\Big(\frac{e_0^f}{\mu_0^f} + \frac{e_1^b}{\mu_1^b}\Big)\Big(\frac{1-e_1^b}{e_1^b - e_0^f}\Big) \\
&= \frac{e_0}{\tau}\Big(\frac{e_0}{\mu_0} + \frac{e_1}{\mu_1}\Big)\Big(\frac{1-e_1}{e_1 - e_0}\Big),
\end{aligned}
\tag{A.42}
$$

$$
\begin{aligned}
\widehat{WIP}_i &= \frac{e_i^f}{\tau}\Big(\frac{e_i^f}{\mu_i^f} + \frac{e_{i+1}^b}{\mu_{i+1}^b}\Big)\Big(\frac{1-e_{i+1}^b}{e_{i+1}^b - e_i^f}\Big) \\
&= \frac{e_0}{\tau}\Big(\frac{1}{\lambda_i + \mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1} - e_0}\Big) \\
&= \frac{e_0}{\tau}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1} - e_0}\Big),\ i = 1, 2, \ldots, M-1.
\end{aligned}
\tag{A.43}
$$

■

Proof of Proposition 6.2: From (6.1) and (3.3), we obtain

$$
\widehat{lt} = 1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1} - e_0}\Big). \tag{A.44}
$$

Thus,

$$
\begin{aligned}
\widehat{lt} &\leqslant 1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{min}}{e_{min} - e_0}\Big) \\
&\leqslant 1 + \frac{1}{M\tau}\Big(\frac{e_0}{\mu_0} + (2M-1)\frac{e_{max}}{\mu_{min}}\Big)\Big(\frac{1-e_{min}}{e_{min} - e_0}\Big) \\
&= 1 + \frac{1}{\tau}\Big(\frac{\rho_{max}}{M\mu_0} + \frac{2M-1}{M\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\Big(\frac{1-e_{min}}{1-\rho_{max}}\Big).
\end{aligned}
\tag{A.45}
$$

■

Proof of Proposition 6.3: Similar to the proof of Proposition 3.2, with the only difference that

$$
\begin{aligned}
f(\rho_{max}) :&= \alpha\widehat{lt}(\rho_{max}) = \alpha\Big(1 + \frac{1}{M\tau}\Big(\frac{\rho_{max}}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\Big(\frac{1-e_{min}}{1-\rho_{max}}\Big)\Big),\\
f'(\rho_{max}) &= \frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\frac{1-e_{min}}{(1-\rho_{max})^2},\\
f''(\rho_{max}) &= \frac{2\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\frac{1-e_{min}}{(1-\rho_{max})^3},
\end{aligned}
$$

and, therefore, we obtain (6.8). When M tends to infinity, (6.8) becomes (6.9).

In the following, we focus on proving (6.10), Let

$$
f(\rho_{max}) := \alpha\Big\{1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big)\Big\}. \tag{A.46}
$$

Since $\rho_{knee}(\widehat{lt})$ is the ρ_{max} that maximize the curvature of $f(\rho_{max})$, to prove (6.10), we need to prove that

$$
\kappa(f(\rho_{max})) = \frac{|f''(\rho_{max})|}{(1+f'^2(\rho_{max}))^{\frac{3}{2}}} \tag{A.47}
$$

is an increasing function of $\rho_{max} \in (0, \bar{\rho}_{\infty,knee}]$, where $\bar{\rho}_{\infty,knee}$ is defined in (6.9). In other words, we need to prove

$$
\kappa'(f(\rho_{max})) = \frac{f'''(\rho_{max})(1+f'^2(\rho_{max})) - 3f'(\rho_{max})f''^2(\rho_{max})}{(1+f'^2(\rho_{max}))^{\frac{5}{2}}} \geqslant 0 \tag{A.48}
$$

for all $\rho_{max} \in (0, \bar{\rho}_{\infty,knee}]$, where

$$
\begin{aligned}
f'(\rho_{max}) &= \frac{\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^2} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^2}\Big],\\
f''(\rho_{max}) &= \frac{2\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^3} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^3}\Big],\\
f'''(\rho_{max}) &= \frac{6\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^4} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^4}\Big].
\end{aligned} \tag{A.49}
$$

Let

$$\begin{aligned}
&\gamma_1 = \Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)(1 - e_1),\ \gamma_{i+1} = \Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)(1 - e_{i+1}),\ i = 1, 2, \ldots, M - 1, \\
&\eta_i = \frac{1}{\frac{e_i}{e_{min}} - \rho_{max}},\ i = 1, 2, \ldots, M.
\end{aligned} \tag{A.50}$$

Then, proving (A.48) implies proving

$$\frac{1}{2}\Big(\sum_{i=1}^{M} \gamma_i \eta_i^4\Big)\Big[\Big(\frac{M\tau e_{min}}{\alpha}\Big)^2 + \Big(\sum_{i=1}^{M} \gamma_i \eta_i^2\Big)^2\Big] \geqslant \Big(\sum_{i=1}^{M} \gamma_i \eta_i^2\Big)\Big(\sum_{i=1}^{M} \gamma_i \eta_i^3\Big)^2. \tag{A.51}$$

Based on (6.9) and considering that $\mu_0 \geqslant \mu_{min}$, we obtain

$$\frac{M\tau e_{min}}{\alpha} = \frac{2Me_{max}(1 - e_{min})}{\mu_{min}(1 - \bar{\rho}_{\infty,knee})^2} \geqslant \sum_{i=1}^{M} \frac{\gamma_i}{(1 - \bar{\rho}_{\infty,knee})^2} \geqslant \sum_{i=1}^{M} \gamma_i \eta_i^2,\ \forall \rho_{max} \in (0, \bar{\rho}_{\infty,knee}]. \tag{A.52}$$

Thus, (A.51) becomes

$$\Big(\sum_{i=1}^{M} \gamma_i \eta_i^4\Big)\Big(\sum_{i=1}^{M} \gamma_i \eta_i^2\Big) \geqslant \Big(\sum_{i=1}^{M} \gamma_i \eta_i^3\Big)^2, \tag{A.53}$$

i.e., Cauchy-Schwarz inequality, which completes the proof. ∎

Proof of Proposition 6.4: Rewriting (6.1) as

$$\begin{aligned}
\widehat{LT} - M\tau &= \frac{e_0}{\mu_0}\Big(\frac{1 - e_1}{e_1 - e_0}\Big) + \frac{e_1}{\mu_1}\Big(\frac{1 - e_1}{e_1 - e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1 - e_{i+1}}{e_{i+1} - e_0}\Big) \\
&= \frac{1 - e_1}{\mu_0}\Big(\frac{e_1}{e_1 - e_0} - 1\Big) + \frac{e_1}{\mu_1}\Big(\frac{1 - e_1}{e_1 - e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1 - e_{i+1}}{e_{i+1} - e_0}\Big),
\end{aligned} \tag{A.54}$$

and taking into account that $0 < e_0 < \min_{1 \leqslant i \leqslant M} e_i$, we observe that the right-hand side of (A.54) is a monotonically increasing function of e_0. Thus,

$$\widehat{LT} - M\tau > \sum_{i=1}^{M} \frac{1 - e_i}{\mu_i} + \sum_{i=1}^{M-1} \frac{e_i(1 - e_{i+1})}{\mu_i e_{i+1}}, \tag{A.55}$$

i.e., the first inequality in (6.11) holds. The second and the third inequalities are derived similarly using (6.4) and (6.5), respectively. ∎

Proof of Proposition 6.5: Under the assumptions of Proposition 6.1, for any desired lead time LT_d satisfying (6.11), the release rate $\hat{e}_0^*$ that ensures this lead time is a real root less than $\min_{1\leqslant i\leqslant M} e_i$ of the equation

$$LT_d = M\tau + \sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big). \tag{A.56}$$

Rewrite the above equation as follows:

$$\begin{aligned} LT_d - M\tau = {} & \frac{1-e_1}{\mu_0}\Big(\frac{e_1}{e_1-e_0} - 1\Big) \\ & + \frac{e_1}{\mu_1}\Big(\frac{1-e_1}{e_1-e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big). \end{aligned} \tag{A.57}$$

Since the right-hand side of (A.57) is a monotonically increasing function of e_0 when $0 < e_0 < \min_{1\leqslant i\leqslant M} e_i$, equation (A.57) (or (A.56)) has a unique real solution less than $\min_{1\leqslant i\leqslant M} e_i$ ensuring the desired lead time LT_d. Multiplying (A.56) by $\prod_{j=0}^{M-1}(e_{j+1} - e_0)$ and re-arranging the terms, we obtain (6.12). In other words, for any desired lead time LT_d satisfying (6.11), the release rate $\hat{e}_0^*$ that ensures this lead time is the unique real root less than $\min_{1\leqslant i\leqslant M} e_i$ of the M-th order polynomial equation (6.12).

The statements on $\widehat{TP}^*$ and $\widehat{WIP}^*$ follow from the proof of Proposition 6.1. ■

Proof of Proposition 6.6: Solving the equation in (6.4) with $lt = lt_d$, we obtain

$$\bar{\rho}^*_{max} = 1 - \frac{\mu_{min} + (2M-1)\mu_0\frac{e_{max}}{e_{min}}}{M\mu_{min}\mu_0\tau(lt_d - 1) + \mu_{min}(1-e_{min})}(1-e_{min}). \tag{A.58}$$

This, taking into account (6.3), results in (6.14). Expression (6.15) is obtained similarly using (6.5).

Clearly, if $\mu_0 \geqslant \mu_{min}$, then (6.6) holds, which implies that (6.16) holds as well. ■

References

[1] J. H. Blackstone Jr., D. T. Phillips, and G. L. Hogg, "A state-of-the-art survey of dispatching rules for manufacturing job shop operations," *International Journal of Production Research*,

vol. 20, no. 1, pp. 27–45, 1982.

[2] R. T. Barrett and S. N. Kadipasaoglu, "Dispatching rules for a dynamic flow shop," *Production and Inventory Management Journal*, vol. 31, no. 1, pp. 54–58, 1990.

[3] A. M. Waikar, B. R. Sarker, and A. M. Lal, "A comparative study of some priority dispatching rules under different shop loads," *Production Planning & Control*, vol. 6, no. 4, pp. 301–310, 1995.

[4] Y.-F. Hung and I.-R. Chen, "A simulation study of dispatch rules for reducing flow times in semiconductor wafer fabrication," *Production Planning & Control*, vol. 9, no. 7, pp. 714–722, 1998.

[5] C. Rajendran and O. Holthaus, "A comparative study of dispatching rules in dynamic flowshops and jobshops," *European Journal of Operational Research*, vol. 116, no. 1, pp. 156–170, 1999.

[6] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 4th ed. Springer, 2012.

[7] A. Hübl, H. Jodlbauer, and K. Altendorfer, "Influence of dispatching rules on average production lead time for multi-stage production systems," *International Journal of Production Economics*, vol. 144, no. 2, pp. 479–484, 2013.

[8] Y. Sugimori, K. Kusunoki, F. Cho, and S. Uchikawa, "Toyota production system and kanban system materialization of just-in-time and respect-for-human system," *International Journal of Production Research*, vol. 15, no. 6, pp. 553–564, 1977.

[9] G. R. Bitran and L. Chang, "A mathematical programming approach to a deterministic kanban system," *Management Science*, vol. 33, no. 4, pp. 427–441, 1987.

[10] L. J. Krajewski, B. E. King, L. P. Ritzman, and D. S. Wong, "Kanban, MRP, and shaping the manufacturing environment," *Management Science*, vol. 33, no. 1, pp. 39–57, 1987.

[11] D. Mitra and I. Mitrani, "Analysis of a kanban discipline for cell coordination in production lines. I," *Management Science*, vol. 36, no. 12, pp. 1548–1566, 1990.

[12] B. J. Berkley, "A review of the kanban production control research literature," *Production and Operations Management*, vol. 1, no. 4, pp. 393–411, 1992.

[13] S. R. Tayur, "Structural properties and a heuristic for kanban-controlled serial lines," *Management Science*, vol. 39, no. 11, pp. 1347–1368, 1993.

[14] J. A. Muckstadt and R. T. Sridhar, "A comparison of alternative kanban control mechanisms. I: Background and structural results," *IIE Transactions*, vol. 27, no. 2, pp. 140–150, 1995.

[15] M. Di Mascolo, Y. Frein, and Y. Dallery, "An analytical method for performance evaluation of kanban controlled production systems," *Operations Research*, vol. 44, no. 1, pp. 50–64, 1996.

[16] M. S. Akturk and F. Erhun, "An overview of design and operational issues of kanban systems," *International Journal of Production Research*, vol. 37, no. 17, pp. 3859–3881, 1999.

[17] M. Lage Jr. and M. G. Filho, "Variations of the kanban system: Literature review and classification," *International Journal of Production Economics*, vol. 125, no. 1, pp. 13–21, 2010.

[18] E. Topan and Z. M. Avsar, "An approximation for kanban controlled assembly systems," *Annals of Operations Research*, vol. 182, no. 1, pp. 133–162, 2011.

[19] J. Matzka, M. Di Mascolo, and K. Furmans, "Buffer sizing of a heijunka kanban system," *Journal of Intelligent Manufacturing*, vol. 23, no. 1, pp. 49–60, 2012.

[20] R. Sato and Y. Khojasteh-Ghamari, "An integrated framework for card-based production control systems," *Journal of Intelligent Manufacturing*, vol. 23, no. 3, pp. 717–731, 2012.

[21] M. L. Spearman, D. L. Woodruff, and W. J. Hopp, "CONWIP: A pull alternative to kanban," *International Journal of Production Research*, vol. 28, no. 5, pp. 879–894, 1990.

[22] I. Duenyas and W. J. Hopp, "Estimating the throughput of an exponential CONWIP assembly system," *Queueing Systems*, vol. 14, no. 1, pp. 135–157, 1993.

[23] W. J. Hopp and M. L. Roof, "Setting WIP levels with statistical throughput control (STC) in CONWIP production lines," *International Journal of Production Research*, vol. 36, no. 4, pp. 867–882, 1998.

[24] A. M. Bonvik, Y. Dallery, and S. B. Gershwin, "Approximate analysis of production systems operated by a CONWIP/finite buffer hybrid control policy," *International Journal of Production Research*, vol. 38, no. 13, pp. 2845–2869, 2000.

[25] W. G. Gilland, "A simulation study comparing performance of CONWIP and bottleneck-based release rules," *Production Planning & Control*, vol. 13, no. 2, pp. 211–219, 2002.

[26] J. M. Framinan, P. L. Gonzalez, and R. Ruiz-Usano, "The CONWIP production control system: Review and research issues," *Production Planning & Control*, vol. 14, no. 3, pp. 255–265, 2003.

[27] J. Geraghty and C. Heavey, "A comparison of hybrid push/pull and CONWIP/pull production inventory control policies," *International Journal of Production Economics*, vol. 91, no. 1, pp. 75–90, 2004.

[28] S. Gstettner and H. Kuhn, "Analysis of production control systems kanban and CONWIP," *International Journal of Production Research*, vol. 34, no. 11, pp. 3253–3273, 1996.

[29] K. K. Yang, "Managing a flow line with single-kanban, dual-kanban or CONWIP," *Production and Operations Management*, vol. 9, no. 4, pp. 349–366, 2000.

[30] H. Jodlbauer and A. Huber, "Service-level performance of MRP, kanban, CONWIP and DBR due to parameter stability and environmental robustness," *International Journal of Production Research*, vol. 46, no. 8, pp. 2179–2195, 2008.

[31] Y. Khojasteh-Ghamari, "A performance comparison between kanban and CONWIP controlled assembly systems," *Journal of Intelligent Manufacturing*, vol. 20, no. 6, pp. 751–760, 2009.

[32] ——, "Developing a framework for performance analysis of a production process controlled by kanban and CONWIP," *Journal of Intelligent Manufacturing*, vol. 23, no. 1, pp. 61–71, 2012.

[33] L. Zhang, "Kanban-controlled exponential produciton lines: Analysis and design," *Journal of Manufacturing Technology Management*, vol. 24, no. 3, pp. 358–383, 2013.

[34] A. Azaron, H. Katagiri, K. Kato, and M. Sakawa, "Modelling complex assemblies as a queueing network for lead time control," *European Journal of Operational Research*, vol. 174, no. 1, pp. 150–168, 2006.

[35] A. Azaron and F. Kianfar, "A multi-objective lead time control problem in multistage assembly systems using an interactive method," *Applied Mathematics and Computation*, vol. 176, no. 2, pp. 609–620, 2006.

[36] C. Perkgoz, A. Azaron, H. Katagiri, K. Kato, and M. Sakawa, "A multi-objective lead time control problem in multi-stage assembly systems using genetic algorithms," *European Journal of Operational Research*, vol. 180, no. 1, pp. 292–308, 2007.

[37] S. Yaghoubi, S. Noori, and F. Pourdadashi-Komachali, "Lead time control in multi-server multi-stage assembly system," *The International Journal of Advanced Manufacturing Technology*, vol. 61, no. 1-4, pp. 351–368, 2012.

[38] S. Yaghoubi, S. Noori, and A. Azaron, "Lead time control in multi-class multi-stage assembly systems with finite capacity," *Computers & Industrial Engineering*, vol. 66, no. 4, pp. 808–817, 2013.

[39] C.-J. Kuo, C.-F. Chien, and J.-D. Chen, "Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 103–111, 2011.

[40] G. Chen, L. Zhang, J. Arinez, and G. Xiao, "Analysis of production completion time in Bernoulli serial lines with finite buffers," in *IEEE International Conference on Automation Science and Engineering*, 2013, pp. 770–775.

[41] S. Biller, S. M. Meerkov, and C.-B. Yan, "Raw material release rates to ensure desired production lead time in Bernoulli serial lines," *International Journal of Production Research*, vol. 51, no. 14, pp. 4349–4364, 2013.

[42] S. M. Meerkov and C.-B. Yan, "Cellular production lines with asymptotically reliable Bernoulli machines: Lead time analysis and control," 2014, accepted for publication in *IEEE Transactions on Automation Science and Engineering*.

[43] J. Li and S. M. Meerkov, *Production Systems Engineering*. New York: Springer, 2009.

[44] R. Courant and F. John, *Introduction to Calculus and Analysis*. Springer, 1999, vol. 1.

[45] R. R. Inman, "Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems," *Production and Operations Management*, vol. 8, no. 4, pp. 409–432, 1999.

[46] J. Li and S. M. Meerkov, "On the coefficients of variation of uptime and downtime in manufacturing equipment," *Mathematical Problems in Engineering*, vol. 2005, no. 1, pp. 1–6, 2005.