

A Method for “Sweet Point” Operation of Re-entrant Lines

Semyon M. Meerkov*, Chao-Bo Yan*,**

* *Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA
(e-mail: smm@umich.edu, chaobo.yan@gmail.com)*

** *State Key Laboratory for Manufacturing Systems Engineering,
Xi'an Jiaotong University, Xi'an, Shaanxi 710054, China*

Abstract: Production systems with infinite buffers typically have the lead time (LT) vs. throughput (TP) characteristic curve (CC) with a knee-type shape. Operating the system below the knee is not efficient, since TP can be increased without an appreciable increase in LT . Operating the system above the knee is also counterproductive, since LT becomes extremely large without a significant increase of TP . Thus, the desirable operating point is at the knee, which is why it is often referred to in practice as the “sweet point”. In this paper, using an empirical/analytical approach, CCs of re-entrant lines described by the bottleneck workcenter-based model with exponential machines are analyzed, the positions of the sweet point are quantified, and an open- and closed-loop job release strategies that ensure the operation at the knee are developed. The development is carried out in terms of the First Buffer First Served and Last Buffer First Served dispatch policies, although CCs for other policies can be investigated in a similar manner.

Keywords: Re-entrant lines, Characteristic curves, Lead time, Throughput, Exponential machines, Open- and closed-loop job release, Sweet point operation

1. INTRODUCTION

1.1 Motivation and goal

Production systems with infinite buffers typically have the lead time (LT) vs. throughput (TP) characteristic curve (CC) with a knee-type shape as illustrated in Fig. 1. Here LT is the average time a job (e.g., lot) spends in the system, being processed or waiting for processing (LT is sometimes referred to as the production cycle-time or flow-time), and TP is the average number of jobs produced by the system per unit of time. In the steady state of system operation, $TP = RR$, where RR is the Release Rate, i.e., the number of jobs released into the system per unit of time. So, the CC of Fig. 1 can be understood as LT vs. RR or, in different notations, as $LT = F(RR)$, where $F(\cdot)$ is a function defined by the production system, i.e., its structure (e.g., serial, cellular, or re-entrant), the number of machines in the system, their reliability models (e.g., Bernoulli, exponential, Weibull, etc.), and, in the case of re-entrant lines, dispatch policies.

The shape of $F(RR)$ implies that operating the system below the knee (indicated by the black dot in Fig. 1) is not efficient since RR (and, thus, TP) can be increased without an appreciable increase of LT . Operating the system above the knee is also counterproductive, since LT (and, thus, work-in-process, WIP) becomes extremely large without a significant increase of TP . Thus, the desirable operating point is at the knee, which is why it is often referred to in practice as the “sweet point”. In this paper, we use the terms “sweet point” and “knee” interchangeably.

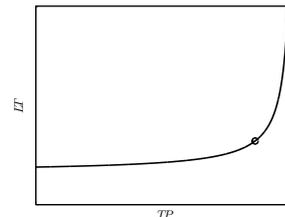


Fig. 1. Characteristic curve of a production system with infinite buffers

Operating a production system at or close to the knee requires the knowledge of function $F(RR)$. For *serial* production lines, this function has been investigated analytically in Biller et al. (2013) and Meerkov and Yan (2014b) for machines having the Bernoulli and exponential reliability models, respectively. *Cellular* lines with Bernoulli machines have been analyzed in Meerkov and Yan (2014a). The goal of the current paper is to investigate this function for *re-entrant* lines with exponential machines and, on this basis, develop open- and closed-loop job release policies, which ensure the “sweet point” operation. Note that re-entrant lines in semiconductor industry are particularly known for excessively long LT ; hence, providing a method for its evaluation and control is an important industrial problem. Clearly, it is of theoretical interest as well.

1.2 System considered and problems addressed

The development reported in this paper is carried out using the bottleneck workcenter model of a re-entrant line

(BNWC model) introduced in Rose (1998) and further explored in Yan et al. (2012). This model (see Fig. 2) considers in details BNWC (consisting of M machines and N buffers), while all other workcenters are modeled as fixed time delays, T_0, T_1, \dots, T_N . In addition, the BNWC model is defined by dispatch policies employed. The current paper considers two of such policies: First Buffer First Served (FBFS) and Last Buffer First Served (LBFS), although other dispatch policies can be explored in a similar manner.

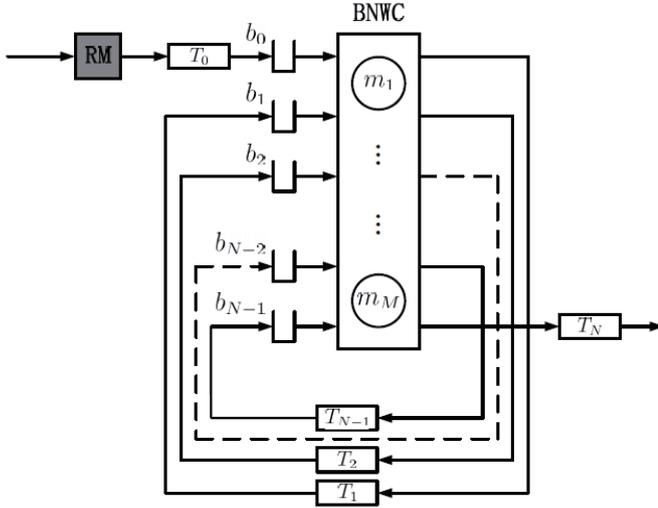


Fig. 2. BNWC-based re-entrant line with release machine and delays

To model the job release, we associate with this system a job release mechanism (RM, shown in Fig. 2 in gray). Various modes of RM operation can be considered. Initially, we assume that RM is another machine (referred to as the *release machine*), which releases a job during a machine cycle-time with a certain probability. Then we generalize the results to deterministic, once-per-hour or once-per-shift, release policies.

Given this system, the problems addressed in this paper are to investigate its *CC*, i.e., *LT* vs. *RR*; quantify the position of the knee; determine the release rate to enable the operation at the knee (this is interpreted as open-loop control of *LT*); and develop a closed-loop job release policy, which ensures the desired *LT*, while maximizing *TP*, even if the parameters of the producing machines are not known precisely.

1.3 Related literature

The literature on performance analysis and design of re-entrant lines contains hundreds of publications, many of which are summarized in the recent monograph by Mönch et al. (2012). Below, we offer a few remarks on this literature, mostly in order to place the current paper in the perspective of the existing results.

The current literature on analytical investigation of re-entrant lines can be classified into three groups. The first one is devoted to stability analysis under various dispatch policies, see, for instance, Kumar (1993); Yan et al. (2012). The main results here are that for all practical dispatch

policies re-entrant lines are stable, while some contrived ones may lead to instability, even if the release rate is below the system capacity. In these publications, stability was understood as buffers occupancy being bounded. In Yan et al. (2012), the notion of Lyapunov stability was applied to the BNWC-based model of re-entrant lines, and it has been shown that some of the practical dispatch policies are, in fact, unstable in the sense of Lyapunov, implying that under strong perturbations (e.g., long downtime of the BNWC), the system may leave the equilibrium point and lock into a periodic or even chaotic regime. In particular, it was shown that from this point of view, FBFS is preferable to LBFS.

The second group addresses the issue of efficacy of various dispatch policies from the point of view of system throughput (see, for instance, Pfund et al. (2006); Chen et al. (2010); Sarin et al. (2011)). Here, numerous dispatch rules have been analyzed, and it has been shown that those favoring almost completed lots or lots with the smallest slack time lead to a better *WIP* performance (e.g., LBFS is preferable to FBFS).

Finally, the last group addresses the issue of job release to ensure a sufficiently small *WIP*, while maximizing *TP* (Fowler et al. (2002); Qi et al. (2009)). The main approaches here are kanban and CONWIP. It is shown, both theoretically and in applications, that each of these approaches may lead to a substantial improvement of the *TP* vs. *WIP* behavior.

In spite of these considerable achievements, the current literature does not offer a method for operating a re-entrant line at the sweet point of the *LT* vs. *RR* characteristic. The current paper is intended to contribute to this end.

1.4 Abbreviations and notations

Abbreviations: BNWC - bottleneck workcenter, C (subscript) - release per cycle, *CC* - characteristic curve, FBFS - First Buffer First Served, LBFS - Last Buffer First Served, r (subscript) - re-entrant line, RM - release mechanism.

Notations: a and b - constants, e - machine efficiency, E - deterministic release, *LT* - lead time, M - number of machines in the BNWC, N - number of re-entrant paths, *RI* - release interval, *RR* - release rate, τ - machine cycle-time, T_{down} - machine downtime, T_i - time delays, T_{up} - machine uptime, *TP* - throughput, *WIP* - work-in-process

1.5 Paper outline

The remainder of this paper is structured as follows: Section 2 defines formally the BNWC-based model of re-entrant lines. Section 3 quantifies the characteristic curves of re-entrant lines described by this model, investigates the position of their sweet points, and provides a comparative investigation of *CC*s under FBFS and LBFS dispatch policies. Sections 4 and 5 are devoted to open- and closed-loop control of job release to ensure the operation at the sweet point or, for that matter, at any desired point of *CC*. Conclusions and topics for future work are provided in Section 6. The proofs are included in the Appendix.

2. BNWC MODEL

Consider the re-entrant line of Fig. 2. Let LT_{total} be the total average time that a job spends in the system with delays T_0, T_1, \dots, T_N , and LT the average time a job spends in the system without the delays. Then, in the steady state of operation,

$$LT_{total} = LT + \sum_{i=0}^N T_i. \quad (1)$$

Below and in the subsequent section, we consider the system without the delays, evaluate LT , and specify LT_{total} using (1). We define this system (see Fig. 3) by the following assumptions:

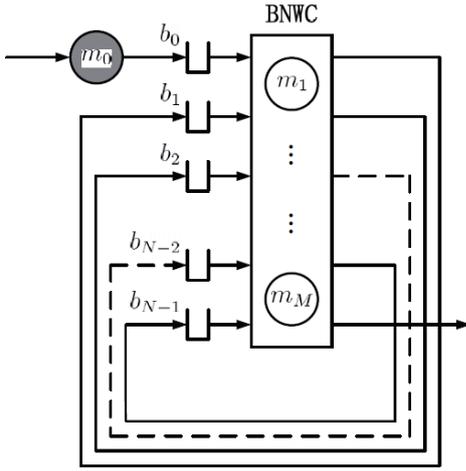


Fig. 3. BNWC-based re-entrant line with release machine

- (i) The system consists of the BNWC with M producing machines, m_1, m_2, \dots, m_M , a release machine, m_0 , and N buffers, b_0, b_1, \dots, b_{N-1} , storing jobs at various stages of their processing.
- (ii) All producing machines in the BNWC are identical. The release machine, m_0 , has the same cycle time, τ (in min), as the producing machines. (This assumption is introduced to simplify the presentation.)
- (iii) Each machine obeys the exponential reliability model defined by breakdown and repair rates λ and μ for producing machines and λ_0 and μ_0 for the release machine, both in 1/min. While λ and μ are fixed, λ_0 and μ_0 are design parameters to be selected at will. (This assumption is also introduced for simplification.)
- (iv) Each buffer is of infinite capacity.
- (v) The flow model is assumed, i.e., infinitesimal quantity of jobs, produced during an infinitesimal time interval, are transferred to and from the buffers. BNWC is starved if the total occupancy of all N buffers is less than M ; m_0 is never starved. Machine failures are time-dependent, i.e., a machine can be down even if it is starved. (These are standard assumptions in Production Systems Engineering, Li and Meerkov (2009).)

This model (i)-(v) is used throughout this paper.

3. QUANTIFICATION OF CHARACTERISTIC CURVE AND SWEET POINTS

3.1 Approach

The approach to analyses of CCs and sweet points of model (i)-(v) is based on the results of Meerkov and Yan (2014b), where function $F(RR)$ and its knee are analyzed for serial lines in terms of two dimensionless quantities. The first one, referred to as the *relative lead time*, is given by

$$lt = \frac{LT}{M\tau}, \quad (2)$$

where M is the number of machines in the serial line and τ is the machine cycle-time (assuming that all machines have the same τ ; reference Meerkov and Yan (2014b) provides also a generalization for non-identical τ 's). Obviously, lt quantifies LT in units of its smallest possible value, i.e., $M\tau$. For instance, $lt = 5$ implies that LT is 5 times longer than the total processing time.

The second quantity, referred to as the *relative workload*, is defined as

$$\rho = \frac{e_0}{e}, \quad (3)$$

where e is the producing machine efficiency, i.e.,

$$e = \frac{T_{up}}{T_{up} + T_{down}} \quad (4)$$

($T_{up} = \frac{1}{\lambda}$ and $T_{down} = \frac{1}{\mu}$ are the average up- and downtime of the producing machines), and e_0 is the release machine efficiency, which can be interpreted as the probability to release a job during the machine cycle-time, τ .

In terms of these normalizations and under the assumption

$$e_0 < e \quad (5)$$

(to ensure the existence of a steady state), the following expression, which defines the function $F(RR)$, has been derived in Meerkov and Yan (2014b):

$$\hat{lt} = \frac{a}{1 - \rho} + b, \quad (6)$$

where \hat{lt} is a sufficiently accurate estimate of lt (see Meerkov and Yan (2014b) for details), and a and b are constants given by

$$a = \frac{1 - e}{M\tau} [T_{down,0} + (2M - 1)T_{down}], \quad (7)$$

$$b = 1 - \frac{1 - e}{M\tau} T_{down,0}$$

($T_{down,0} = \frac{1}{\mu_0}$ is the average downtime of the release machine).

The characteristic curve $\hat{lt}(\rho)$, specified by (6), (7), is shown in Fig. 4 for the serial line with the parameters indicated in the figure caption. Using the definition of the knee as a point of CC with the largest curvature, it is shown in Meerkov and Yan (2014b) that the knee of $\hat{lt}(\rho)$ satisfies the relationship:

$$\alpha \frac{d(\hat{lt})}{d\rho} \Big|_{\rho=\rho_{knee}} = 1 \quad (8)$$

or, taking into account (6),

$$\rho_{knee} = 1 - \sqrt{\alpha a}. \quad (9)$$

In (8) and (9), α is the scaling ratio defined by the operating regime of the system at hand. Specifically, assume $\rho \in [\rho_{min}, 1)$ and $\hat{lt} \in (0, \hat{lt}_{max}]$, where ρ_{min} is the smallest relative load factor of interest and \hat{lt}_{max} is the largest acceptable relative lead time. Then

$$\alpha = \frac{1 - \rho_{min}}{\hat{lt}_{max}}. \quad (10)$$

In Fig. 4, $\alpha = 1/4000$ and the knee, defined by (9), is shown by the black dot (with the coordinates indicated). Thus, the position of the knee depends on the regime of system operation (i.e., α) and the parameter a of CC, while it is independent of b .

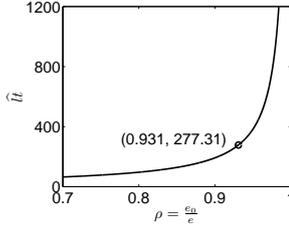


Fig. 4. Characteristic curve and sweet point of a serial line with $M = 10$, $\tau = 1\text{min}$, $e = 0.9$, $T_{down,0} = 10\text{min}$, $T_{down} = 100\text{min}$

Unfortunately, expression (6) for re-entrant lines defined by assumptions (i)-(v) proved to be all but impossible to derive analytically. Therefore, we resort to an empirical/analytical approach, whereby the constants a and b are determined empirically and the validity of (6) for re-entrant lines is justified by simulations, while the position of the knee as well as the open- and closed-loop release control policies are quantified analytically.

More specifically, consider the system of Fig. 3 and assume that $M < N$, i.e., the number of machines in the bottleneck workcenter is less than the number of processing stages. Then, the approach of this paper consists of the following:

- (a) Introduce the relative lead time and the relative workload for re-entrant lines:

$$\begin{aligned} lt_r &= \frac{LT}{N\tau}, \\ \rho_r &= \frac{N}{M} \frac{e_0}{e}. \end{aligned} \quad (11)$$

To ensure stability, assume $\rho_r < 1$, i.e.,

$$e_0 < \frac{Me}{N}. \quad (12)$$

- (b) Identify empirically the two constants involved in expression (6). This is accomplished by evaluating LT by either simulations or operating the system under two workload factors, say, ρ_{r1} and ρ_{r2} . Using (6), this leads to two equations with two unknowns. Solving these equations, we obtain:

$$\begin{aligned} a_r &= \frac{lt_r(\rho_{r1}) - lt_r(\rho_{r2})}{\rho_{r1} - \rho_{r2}} (1 - \rho_{r1})(1 - \rho_{r2}), \\ b_r &= \frac{lt_r(\rho_{r2})(1 - \rho_{r2}) - lt_r(\rho_{r1})(1 - \rho_{r1})}{\rho_{r1} - \rho_{r2}}, \end{aligned} \quad (13)$$

where $lt_r(\rho_{ri})$, is the empirically determined relative lead time under the relative load factor, ρ_{ri} , $i = 1, 2$.

- (c) Next, we justify that expression (6) with constants (13) holds for any $\rho_r \in [\rho_{r,min}, 1)$. This is accomplished by simulating 100 systems with randomly selected parameters (see Subsection 3.2 for details). Thus, the characteristic curve for re-entrant lines is established empirically as

$$\hat{lt}_r(\rho_r) = \frac{a_r}{1 - \rho_r} + b_r. \quad (14)$$

- (d) The analytical part of the approach is based on (14). Specifically, we quantify the position of the knee by (9) and develop open- and closed-loop job release policies that ensure the operation at the sweet point or at any other desired point of the characteristic curve.

3.2 Accuracy

In this subsection, we investigate the accuracy of $\hat{lt}_r(\rho_r)$ specified by (14). This is carried out in two steps. First, we construct 100 re-entrant lines obeying the BNWC-based model, identify empirically their constants a_r and b_r , and, thus, specify their characteristic curves. Second, we evaluate the accuracy of the resulting expressions by comparing them with the true characteristic curves, $lt_r(\rho_r)$, identified by simulations.

Specifically, the 100 re-entrant lines satisfying assumptions (i)-(v) are constructed by selecting the parameters randomly and equiprobably from the following sets:

$$\begin{aligned} M &\in [2, 5], \quad N \in [M + 1, 10], \quad e \in [0.7, 0.99], \\ T_{down,0} &= T_{down} \in [1\text{min}, 10\text{min}], \end{aligned} \quad (15)$$

and, without loss of generality, use $\tau = 1\text{min}$. For each of these lines, ρ_{r1} is selected randomly and equiprobably from the interval $\rho_{r1} \in [0.8, 0.97]$ and ρ_{r2} is assumed to be $0.9\rho_{r1}$. To evaluate $lt_r(\rho_{ri})$, $i = 1, 2$, we employ the following simulation procedure (also used in all subsequent sections): In addition to a warm-up period of 2,000,000min, the simulation runs for 22,000,000min, and 20 repetitions of this procedure are carried out. This procedure results in a 95% confidence interval of $\pm 2.5\%$ of lt_r . Using $lt_r(\rho_{r1})$ and $lt_r(\rho_{r2})$, thus evaluated, the constants a_r and b_r are calculated according to (13). Thus, the characteristic curve $\hat{lt}_r(\rho_r)$ is specified analytically by (14) for all $\rho_{r1} \in (0, 1)$.

To investigate the accuracy of $\hat{lt}_r(\rho_r)$, we simulate the 100 systems discussed above for

$$\rho_r \in P = \{0.8, 0.85, 0.9, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97\}, \quad (16)$$

evaluate $lt_r(\rho_r)$, and quantify the relationship between $lt_r(\rho_r)$ and $\hat{lt}_r(\rho_r)$ by

$$\epsilon = \max_{\rho_r \in P} \frac{|\hat{lt}_r(\rho_r) - lt_r(\rho_r)|}{lt_r(\rho_r)} \times 100\%. \quad (17)$$

These analyses are carried out under FBFS and under LBFS dispatch policies. As a results, we obtain:

- Under FBFS dispatch, the smallest and the largest ϵ 's are 1.45% and 33.10%, respectively; the average error (over the 100 systems analyzed) is 12.19%.
- Under LBFS dispatch, the smallest and the largest ϵ 's are 0.05% and 1.93%, respectively; the average error is 0.60%.

Typical relationships between the curves $\widehat{lt}_r(\rho_r)$ and $lt_r(\rho_r)$ are illustrated in Figs. 5 and 6 for FBFS and LBFS, respectively, along with the position of the sweet point (calculated using (9) with $\alpha = 1/750$). Obviously, the accuracy of the approximation (14) for LBFS is very high, while for FBFS the errors could be relatively significant. However, recognizing that machine parameters on the factory floor are rarely known with accuracy better than $\pm 5\%$, we conclude that the estimate (14) is precise enough for the lead time analysis and control.

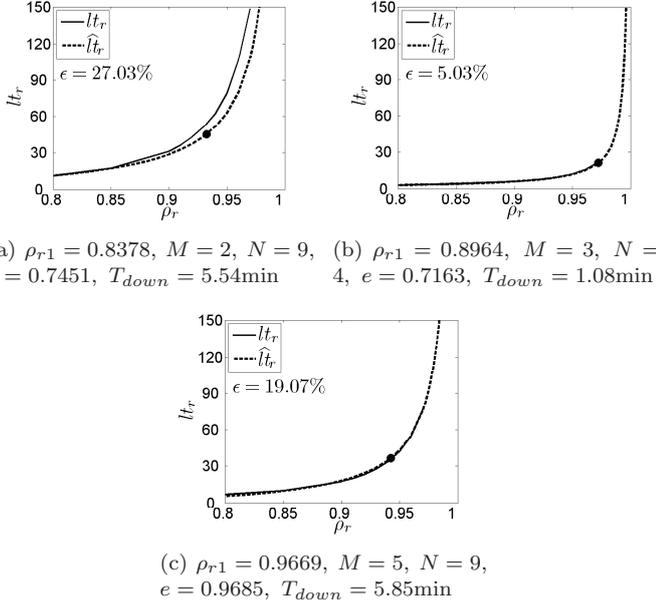


Fig. 5. Typical relationship between $\widehat{lt}_r(\rho_r)$ and $lt_r(\rho_r)$ under FBFS dispatch

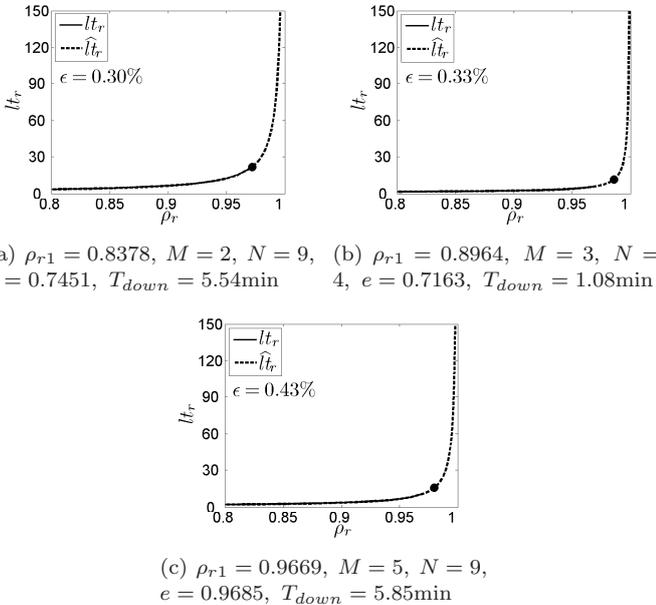


Fig. 6. Typical relationship between $\widehat{lt}_r(\rho_r)$ and $lt_r(\rho_r)$ under LBFS dispatch

3.3 Comparison of lead time under FBFS and LBFS dispatch

To illustrate relative properties of lt under FBFS and LBFS, consider the re-entrant line defined by assumptions (i)-(v) with the following parameters:

$$\begin{aligned} M = 3, N = 10, e = 0.9, \\ T_{down,0} = T_{down} = 5\text{min}, \tau = 1\text{min}. \end{aligned} \quad (18)$$

Using the procedure of Subsection 3.2 with $\rho_{r1} = 0.9$, evaluate the constants involved in (14). The results are:

- under FBFS, $a_r = 2.7526$, $b_r = -6.0343$;
- under LBFS, $a_r = 0.3807$, $b_r = 0.6366$.

The corresponding $\widehat{lt}_r(\rho_r)$'s are shown in Fig. 7, along with the position of the knees (with $\alpha = 1/400$). Clearly, LBFS outperforms FBFS. As one can see, operating both systems at the sweet points leads to \widehat{TP}^{LBFS} , which is 5.7% larger than \widehat{TP}^{FBFS} , and lt_r 52% smaller than \widehat{lt}_r^{FBFS} .

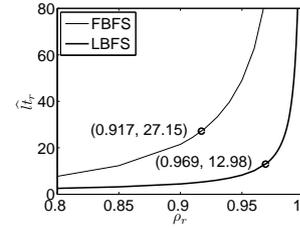


Fig. 7. Characteristic curves of system (18) under FBFS and LBFS

In addition, based on the 100 systems analyzed, we arrive at the following:

Observation 3.1.

$$a_r^{LBFS} < a_r^{FBFS}. \quad (19)$$

Thus, according to (9), $\rho_{r,knee}^{FBFS} < \rho_{r,knee}^{LBFS}$, i.e., $\widehat{TP}_{knee}^{LBFS} > \widehat{TP}_{knee}^{FBFS}$.

Observation 3.2.

$$\frac{\widehat{lt}_r^{FBFS}(\rho_r)}{\widehat{lt}_r^{LBFS}(\rho_r)} \xrightarrow{\rho_r \rightarrow 1^-} N. \quad (20)$$

This implies, for instance, that the lead time of a system with 9 re-entrant paths under FBFS is up to 10 times larger than that under LBFS, if the relative workload is sufficiently high.

Observation 3.3.

$$\frac{\widehat{lt}_r^{FBFS}(\rho_r)}{\widehat{lt}_r^{LBFS}(\rho_r)} \xrightarrow{\rho_r \rightarrow 0^+} 1. \quad (21)$$

Thus, the lead times of the system under FBFS and LBFS are practically the same, if the workload is sufficiently low.

While this subsection shows that LBFS outperforms FBFS, reference Yan et al. (2012) arrived at the opposite conclusion. However, in Yan et al. (2012), the perturbation was a catastrophic breakdown of all machines in the BNWC. Under this perturbation, Yan et al. (2012) showed that FBFS leads to a faster recovery than LBFS. Thus, the relative advantages/disadvantages of each dispatch policy

depend on the model of perturbations considered. Assuming that a statistical (e.g., exponential) reliability model is more prevalent in practice than catastrophic breakdowns, LBFS could be viewed as superior to FBFS.

4. OPEN-LOOP CONTROL OF LT

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

4.1 Random job release

Calculating $\rho_{r,knee}$ using (9) allows us to evaluate job release rate $\hat{e}_{0,knee}$, which ensures the operation at the sweet point. Indeed, as it follows from (11),

$$\hat{e}_{0,knee} = \frac{Me}{N} \rho_{r,knee}. \quad (22)$$

This can be implemented as releasing a job once per machine cycle-time with probability $\hat{e}_{0,knee}$. The release rate to ensure operation at any other point of CC is specified by:

Proposition 4.1. Consider the re-entrant lines defined by assumptions (i)-(v) with $M < N$ and either FBFS or LBFS dispatch. Then, the set of feasible lead times, $\mathcal{F}_{\hat{lt}}$, is given by

$$\hat{lt}_r > a_r + b_r. \quad (23)$$

For any feasible desired lead time, $lt_d \in \mathcal{F}_{\hat{lt}}$, the release rate is given by

$$\hat{e}_0^* = \frac{Me}{N} \left(1 - \frac{a_r}{lt_d - b_r} \right). \quad (24)$$

For this release rate,

$$\widehat{TP}^* = \frac{\hat{e}_0^*}{\tau}, \quad \widehat{WIP}^* = N\hat{e}_0^* \left(\frac{Ma_r e}{Me - N\hat{e}_0^*} + b_r - 1 \right). \quad (25)$$

Note that if $lt_d = \hat{lt}_{r,knee}$, (24) reduces to (22). The behavior of \hat{e}_0^* as a function of lt_d under both FBFS and LBFS is illustrated in Fig. 8 for the re-entrant line (18), with black dots indicating $(\hat{lt}_{r,knee}, \hat{e}_0^*(\hat{lt}_{r,knee}))$. From this figure follows:

Observation 4.1. Under both FBFS and LBFS, for $lt_d < \hat{lt}_{r,knee}$, the optimal release rate \hat{e}_0^* (and, therefore, \widehat{TP}^*) is a rapidly increasing function of lt_d . For $lt_d > \hat{lt}_{r,knee}$, \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 $\widehat{WIP}^* = \widehat{TP}^* (LT_d - N\tau)$).

4.2 Deterministic job release

The random, once-per-cycle, job 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.

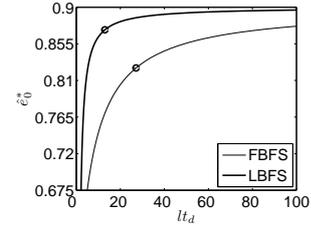


Fig. 8. Release rate, \hat{e}_0^* , as a function of the desired lead time, lt_d

Let $\hat{e}_0^*(lt_d)$ be the once-per-cycle release rate calculated using (24). Then, within a release interval, RI (in hour), the deterministic release, \hat{E}_{RI}^* (jobs/release interval), is defined as:

$$\hat{E}_{RI}^* = \lfloor H\hat{e}_0^*(lt_d) \rfloor, \quad (26)$$

where $\lfloor x \rfloor$ denotes the largest integer not greater than x , and H is the number of cycles in a release interval, i.e., $H = \frac{60RI}{\tau}$.

While release according to (26) results in the obvious inequality

$$\widehat{LT}(\hat{E}_{RI}^*) < \widehat{LT}(\hat{e}_0^*) + 60RI, \quad (27)$$

where $\widehat{LT}(\hat{E}_{RI}^*)$ and $\widehat{LT}(\hat{e}_0^*)$ are the lead times under (26) and (24), respectively, the losses of the throughput under deterministic release (26) are not obvious and must be evaluated. We carry out this evaluation by simulating re-entrant line (18) under both FBFS and LBFS. Based on lt_d selected, \hat{e}_0^* and \hat{E}_{RI}^* are evaluated using (24) and (26), respectively. For the systems considered, we ran the simulations with once-per-cycle and once-per-release-interval release and evaluated the resulting throughputs, TP_C and TP_{RI} (both in jobs/min), where the subscripts “C” and “RI” denote cycle and release interval, respectively. Based on these measurements, we quantified losses in TP by

$$TP_{loss} = \frac{TP_C - TP_{RI}}{TP_C} \times 100\%. \quad (28)$$

The results are shown in Table 1 for $RI = 1$ hour and for $RI = 8$ hour shift. As one can see, under both FBFS and LBFS, throughput losses for once-per-hour release are significant (due to relatively small \hat{E}_{RI}^*), while for once-per-shift release, the losses are negligible.

5. CLOSED-LOOP CONTROL OF LT

5.1 Scenario

The previous section provides methods for estimating job 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 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

$$\hat{e}_0^*(lt_d) > \frac{M}{N} \min_{1 \leq i \leq M} e_{real,i}, \quad (29)$$

Table 1. Throughput losses due to once-per- RI release

(a) FBFS, once-per-hour release					
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	TP_C	TP_{RI}	TP_{loss} (%)
6	0.2082	12	0.2083	0.2000	3.97
8	0.2170	13	0.2171	0.2167	0.18
20	0.2415	14	0.2414	0.2333	3.36
50	0.2567	15	0.2567	0.2500	2.61
100	0.2630	15	0.2630	0.2500	4.95

(b) LBFS, once-per-hour release					
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	TP_C	TP_{RI}	TP_{loss} (%)
2.4	0.2117	12	0.2117	0.2000	5.54
3.2	0.2299	13	0.2299	0.2167	5.75
8	0.2560	15	0.2560	0.2500	2.35
20	0.2647	15	0.2647	0.2500	5.55
40	0.2674	16	0.2674	0.2667	0.27

(c) FBFS, once-per-shift release					
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	TP_C	TP_{RI}	TP_{loss} (%)
6	0.2082	99	0.2083	0.2063	0.97
8	0.2170	104	0.2171	0.2167	0.18
20	0.2415	115	0.2414	0.2396	0.77
50	0.2567	123	0.2567	0.2563	0.18
100	0.2630	126	0.2630	0.2625	0.20

(d) LBFS, once-per-shift release					
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	TP_C	TP_{RI}	TP_{loss} (%)
2.4	0.2117	101	0.2117	0.2104	0.63
3.2	0.2299	110	0.2299	0.2292	0.31
8	0.2560	122	0.2560	0.2542	0.72
20	0.2647	127	0.2647	0.2646	0.04
40	0.2674	128	0.2674	0.2667	0.27

resulting in an arbitrarily large lead time.

To prevent this situation, feedback control may be used to throttle the job release if the work-in-process in the systems exceeds a certain limit. A number of such control strategies can be proposed. Here, we investigate the one proposed for serial lines in Meerkov and Yan (2014b), 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 release takes place; otherwise it does not. In Subsection 5.2 below we formally introduce this control law and in Subsection 5.3 investigate its performance using simulations.

5.2 Control law

Consider a re-entrant line defined by assumptions (i)-(v) with the nominal breakdown and repair rates λ and μ . Let LT_d be the desired lead time (in min). Based on this information, calculate \hat{e}_0^* and \hat{E}_{RI}^* using (24) and (26), respectively. Also, calculate the nominal total work-in-process using Little's law: since \widehat{TP}^* is given by the first formula in (25), and the total waiting time in all buffers is $LT_d - N\tau$, we obtain:

$$\widehat{WIP}_{nominal} = \frac{\hat{e}_0^*}{\tau} (LT_d - N\tau). \quad (30)$$

Using these data, introduce the following control law:

$$E(s+1) = \begin{cases} \hat{E}_{RI}^*, & \text{if } WIP_{total}(s) \leq \widehat{WIP}_{nominal}, \\ 0, & \text{otherwise,} \end{cases} \quad (31)$$

where $s = 0, 1, \dots$, is the index of the release interval; $E(s+1)$ is the number of job 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 .

5.3 Performance evaluation

To evaluate the performance of feedback law (31), we use the re-entrant line (18) as the nominal one and form a real one for it. The real line is formed by increasing or decreasing machine up- and downtimes randomly and equiprobably within $\pm 50\%$ of their nominal values. Producing machine parameters of the resulting line are as follows:

$$e = [0.9154, 0.7678, 0.8708], \quad \mathbf{T}_{down} = [3.02, 7.15, 4.09] \text{min.}$$

We simulate this system with and without feedback control (31) under both FBFS and LBFS. The desired lead time, lt_d , for FBFS is selected 10 times larger than that for LBFS (see Observation 3.2). Based on these simulations, we evaluate the lead times in open- and closed-loop cases (denoted as lt_{OL} and lt_{CL} , respectively). The results, shown in Table 2, lead to the following:

Observation 5.1. Under both FBFS and LBFS, closed-loop job release according to (31) maintains the lead time close to the desired, while the open-loop release results, in some cases, in an unbounded lt .

Table 2. Lead time under control law (31)

(a) FBFS, once-per-hour release				
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	lt_{OL}	lt_{CL}
100	0.2630	15	15.52	15.45
200	0.2664	15	15.53	15.53
300	0.2676	16	∞	288.77
400	0.2682	16	∞	394.55

(b) LBFS, once-per-hour release				
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	lt_{OL}	lt_{CL}
10	0.2590	15	4.87	4.86
20	0.2647	15	4.87	4.87
30	0.2665	15	4.87	4.87
40	0.2674	16	∞	41.7

(c) FBFS, once-per-shift release				
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	lt_{OL}	lt_{CL}
30	0.2494	119	47.94	47.16
50	0.2567	123	∞	61.26
80	0.2614	125	∞	80.53
100	0.2630	126	∞	91.82

(d) LBFS, once-per-shift release				
lt_d	\hat{e}_0^*	\hat{E}_{RI}^*	lt_{OL}	lt_{CL}
3	0.2265	108	22.17	22.17
5	0.2464	118	24.27	24.26
8	0.2560	122	29.25	26.51
10	0.2590	124	∞	28.90

This paper provides a method for operating re-entrant lines (described by BNWC-based model with identical exponential machines) at the sweet point of the $\widehat{lt}_r(\rho_r)$ -curve, or, for that matter, at any other desired point of this characteristic curve. The method consists of determining empirically two constants and using them in the formula for the characteristic curve, $\widehat{lt}_r(\rho_r)$. Based on this formula, one can calculate the job release rate, which ensures the desired mode of operation. This can be implemented either in open-loop format (if the machine parameters are known precisely) or in closed-loop (if they are not).

Numerous problems, however, remain opened. They can be classified into two groups: extensions of the results obtained and novel problems related to this subject matter.

Extension problems:

- Investigation of the accuracy of (14) under dispatch policies other than FBFS and LBFS.
- Analysis of $\widehat{lt}_r(\rho_r)$ -curve for non-identical exponential machines.
- Analysis of $\widehat{lt}_r(\rho_r)$ -curve for non-exponential machines.
- Investigation of the accuracy of the characteristic curve (14) as a function of ρ_{r1} and ρ_{r2} .

New problems:

- Analytical derivation of a formula for $\widehat{lt}_r(\rho_r)$ for BNWC-based model. As mentioned in Subsection 3.1, we have worked on this problem, but we were not able to solve it. Nevertheless, we believe that this problem is solvable using the “right” simplification technique.
- Analysis of $\widehat{lt}_r(\rho_r)$ -curve for re-entrant lines modeled as multiple workcenters with a single or multiple bottlenecks.

Solutions of these problems will lead to a relatively complete theory for sweet point operation of re-entrant lines.

APPENDIX

Proof of Proposition 4.1: As it follows from (14), $\widehat{lt}_r(\rho_r)$ is an increasing function of ρ_r . Since $0 < \rho_r < 1$, this implies (23).

As for the optimal release rate, from (14) it follows that

$$\rho_r^* = 1 - \frac{a_r}{lt_d - b_r}, \quad (32)$$

which, using (11), leads to (24). As far as (25) is concerned, clearly, $\widehat{TP}^* = \frac{\widehat{e}_0^*}{\tau}$, and, based on Little’s law and (11), we obtain:

$$\begin{aligned} \widehat{WIP}^* &= \widehat{TP}^* (LT_d - N\tau) = \frac{\widehat{e}_0^*}{\tau} (lt_d - 1)N\tau \\ &= \frac{\widehat{e}_0^*}{\tau} \left(\frac{a_r}{1 - \rho_r^*} + b_r - 1 \right) N\tau \\ &= N\widehat{e}_0^* \left(\frac{Ma_r e}{Me - N\widehat{e}_0^*} + b_r - 1 \right). \end{aligned} \quad (33)$$

■

- Billar, S., Meerkov, S.M., and Yan, C.B. (2013). Raw material release rates to ensure desired production lead time in Bernoulli serial lines. *International Journal of Production Research*, 51(14), 4349–4364.
- Chen, J.Y., Pfund, M.E., Fowler, J.W., Montgomery, D.C., and Callarman, T.E. (2010). Robust scaling parameters for composite dispatching rules. *IIE Transactions*, 42(11), 842–853.
- Fowler, J.W., Hogg, G.L., and Mason, S.J. (2002). Workload control in the semiconductor industry. *Production Planning & Control*, 13(7), 568–578.
- Kumar, P.R. (1993). Re-entrant lines. *Queueing Systems*, 13(1-3), 87–110.
- Li, J. and Meerkov, S.M. (2009). *Production Systems Engineering*. New York: Springer.
- Meerkov, S.M. and Yan, C.B. (2014a). Cellular production lines with asymptotically reliable Bernoulli machines: Lead time analysis and control. Accepted for publication in *IEEE Transactions on Automation Science and Engineering*.
- Meerkov, S.M. and Yan, C.B. (2014b). Production lead time in serial lines: Evaluation, analysis, and control. Accepted for publication in *IEEE Transactions on Automation Science and Engineering* (posted at <http://www.eecs.umich.edu/smm/publications>).
- Mönch, L., Fowler, J.W., and Mason, S.J. (2012). *Production planning and control for semiconductor wafer fabrication facilities: modeling, analysis, and systems*, volume 52. Springer.
- Pfund, M.E., Mason, S.J., and Fowler, J.W. (2006). Semiconductor manufacturing scheduling and dispatching. In *Handbook of Production Scheduling*, 213–241. Springer.
- Qi, C., Sivakumar, A.I., and Gershwin, S.B. (2009). An efficient new job release control methodology. *International Journal of Production Research*, 47(3), 703–731.
- Rose, O. (1998). WIP evolution of a semiconductor factory after a bottleneck workcenter breakdown. In *1998 Winter Simulation Conference*, 997–1003.
- Rose, O. (1999). Estimation of the cycle time distribution of a wafer fab by a simple simulation model. In *The International Conference on Semiconductor Manufacturing Operational Modeling and Simulation*, 133–138.
- Rose, O. (2000). Why do simple wafer fab models fail in certain scenarios. In *2000 Winter Simulation Conference*, 1481–1490.
- Rose, O. (2007). Improving the accuracy of simple simulation models for complex production systems. In *2007 INFORMS Simulation Society Research Workshop*, 5–8.
- Sarin, S.C., Varadarajan, A., and Wang, L. (2011). A survey of dispatching rules for operational control in wafer fabrication. *Production Planning and Control*, 22(1), 4–24.
- Yan, C.B., Hassoun, M., and Meerkov, S.M. (2012). Equilibria, stability, and transients in re-entrant lines under FBFS and LBFS dispatch and constant release. *IEEE Transactions on Semiconductor Manufacturing*, 25(2), 211–229.