Chapter 6

Design of Lean Bernoulli Lines

Motivation: Designers of manufacturing systems select machines for production lines based on their technological characteristics. Material handling devices are selected based on the nature of parts produced, available space, type of production, etc. Their capacity, in their function as in-process buffers, is typically selected as small as possible, i.e., lean. But how lean can lean buffers be? In other words, what is the smallest buffer capacity, which is necessary and sufficient to ensure the desired production rate of a system? This is the question addressed in this chapter.

Overview: Closed formulas for lean buffering in systems with identical Bernoulli machines are derived. These formulas are exact for two- and three-machine lines and approximate for longer lines. For the case of non-identical Bernoulli machines, both closed-form expressions and recursive approaches are developed.

6.1 Parametrization and Problem Formulation

Consider a serial production line with Bernoulli machines defined by conventions (a)-(e) of Subsection 4.2.1. Introduce the following notions:

Line efficiency (E) – production rate of the line, PR, in units of the largest possible production rate of the system.

As it is clear from Chapter 4, the largest production rate is obtained when all buffers are infinite. Denote this production rate as PR_∞. Then, the line efficiency can be expressed as

$$E = \frac{PR}{PR_\infty} = \frac{PR}{\min\{p_1, \ldots, p_M\}}. \tag{6.1}$$

Obviously, $0 < E < 1$.

Lean buffer capacity (LBC) – the sequence

$$N_{1,E}, \dots, N_{M-1,E} \tag{6.2}$$

such that the desired line efficiency E is achieved while $\sum_{i=1}^{M-1} N_{i,E}$ is minimized. In other words, LBC is the buffer capacity, which is necessary and sufficient to obtain the desired PR, quantified by E.

The problem addressed in this chapter is to develop analytical methods for calculating LBC as a function of machine efficiencies p_i, $i = 1, \dots, M$, line efficiency E, and the number of machines in the system M. For the case of identical machines, i.e., $p_i =: p$, $i = 1, \dots, M$, this is carried out in Section 6.2, while the case of non-identical machines, $p_i \neq p_j$, $i, j = 1, \dots, M$, is addressed in Section 6.3.

6.2 Lean Buffering in Bernoulli Lines with Identical Machines

In this section, we assume that all machines have identical efficiency,

$$p_i =: p, \qquad i = 1, \dots, M, \tag{6.3}$$

and, in addition, all buffers are of identical capacity,

$$N_i =: N, \qquad i = 1, \dots, M-1. \tag{6.4}$$

Assumption (6.4) is introduced in order to obtain a compact representation of the results. It should be pointed out that a more efficient buffer capacity allocation in systems satisfying (6.3) is the inverted bowl pattern (see Chapter 5). However, this leads to just a small improvement of the production rate in comparison to the uniform allocation (6.4) (typically, within 1%) and, therefore, is not considered here.

6.2.1 Two-machine lines

In the case of two identical machines, expression (4.19) for the production rate becomes

$$PR = p[1 - Q(p, N)], \tag{6.5}$$

where, as it follows from (4.14),

$$Q(p, N) = \frac{1-p}{N+1-p}. \tag{6.6}$$

Thus, using (6.1), we obtain

$$PR = PR_\infty E = p[1 - Q(p, N_E)], \tag{6.7}$$

where N_E is the buffer capacity necessary and sufficient to ensure E. Taking into account that $PR_\infty = p$, (6.6) and (6.7) result in the following equation:

$$E = 1 - \frac{1-p}{N_E + 1 - p}. \tag{6.8}$$

Solving for N_E and taking into account that N_E is an integer, we obtain

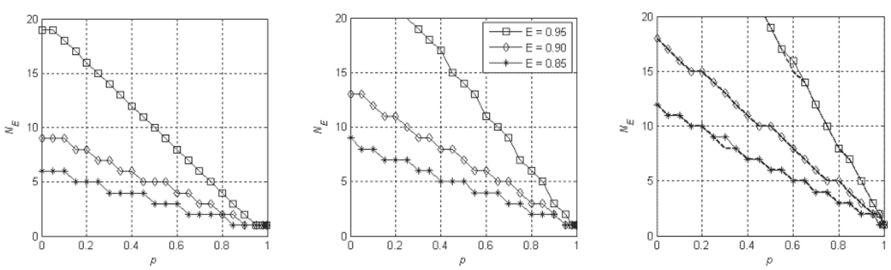
Theorem 6.1 *The lean buffer capacity in Bernoulli lines defined by assumptions (a)-(e) of Subsection 4.2.1 with $M = 2$ and $p_1 = p_2 =: p$ is given by*

$$N_E(M=2) = \left\lceil \frac{E(1-p)}{1-E} \right\rceil, \tag{6.9}$$

where $\lceil x \rceil$ denotes the smallest integer not less than or equal to x.

Note that according to (6.9), N_E cannot be less than 1. Under the blocked before service assumption, buffering $N_E = 1$ implies that the machine itself stores a part being processed and no additional buffering between the machines is required. This can be interpreted as Just-in-Time (JIT) operation.

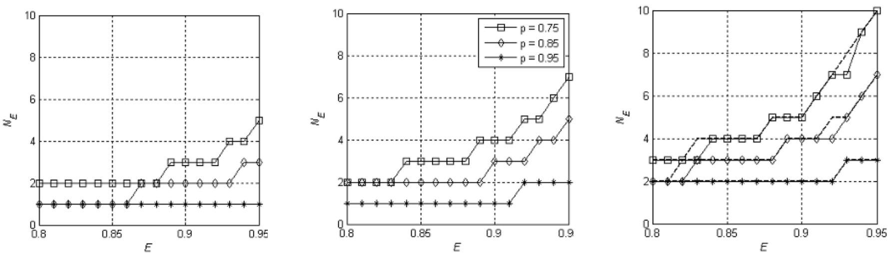
Figures 6.1(a) and 6.2(a) illustrate the behavior of the lean buffer capacity as a function of machine efficiency p and line efficiency E, respectively. From these figures and expression (6.9), we observe:



(a) $M = 2$ (b) $M = 3$ (c) $M = 10$

Figure 6.1: Lean buffering as a function of machine efficiency

- For each E, LBC is a monotonically decreasing function of p, with a practically constant slope.
- For each p, LBC is a monotonically increasing function of E, exhibiting a hyperbolic behavior in $1 - E$.
- JIT operation is acceptable only if p's are sufficiently large. For instance, if the desired line efficiency is 0.85, JIT can be used only if $p > 0.83$, while for $E = 0.95$, p must be larger than 0.95.
- In a practical range of p's, e.g., $0.7 < p < 0.98$, relatively small buffers are required to achieve a large E. For instance, $N_{0.95} = 6$ if $p = 0.7$; if $p = 0.9$, $N_{0.95} = 2$.



(a) $M = 2$ (b) $M = 3$ (c) $M = 10$

Figure 6.2: Lean buffering as a function of line efficiency

6.2.2 Three-machine lines

For a three-machine line, using the aggregation procedure (4.30), the following can be derived:

Theorem 6.2 *The lean buffer capacity in Bernoulli lines defined by conventions (a)-(e) of Subsection 4.2.1 with* $M = 3$ *and* $p_1 = p_2 = p_3 =: p$ *is given by*

$$N_E(M = 3) = \left\lceil \frac{\ln\left\{\frac{1-\sqrt{E}}{1-E}\right\}}{\ln\left\{\frac{(1-p)\sqrt{E}}{1-p\sqrt{E}}\right\}} \right\rceil . \tag{6.10}$$

Proof: See Section 20.1.

The behavior of this N_E is illustrated in Figures 6.1(b) and 6.2(b). Obviously, for most values of p, the lean buffer capacity is increased, as compared with the case of $M = 2$, and the range of p's where JIT is possible is decreased. For instance, if $p = 0.95$, JIT is acceptable for $E < 0.91$, while for $p = 0.85$ it is acceptable for $E < 0.77$.

6.2.3 $M > 3$-machine lines

Theorem 6.3 *The lean buffer capacity in Bernoulli lines defined by conventions (a)-(e) of Subsection 4.2.1 with* $M > 3$ *and* $p_i =: p, i = 1, \dots, M$, *is given by*

$$N_E(M > 3) = \left\lceil \frac{\ln\left\{\frac{1-E-Q(p^f_{M-2}, p^b_{M-1}, N_{M-2})}{(1-E)(1-Q(p^f_{M-2}, p^b_{M-1}, N_{M-2}))}\right\}}{\ln\left\{\frac{(1-p)(1-Q(p^f_{M-2}, p^b_{M-1}, N_{M-2}))}{1-p(1-Q(p^f_{M-2}, p^b_{M-1}, N_{M-2}))}\right\}} \right\rceil , \tag{6.11}$$

where p^f_{M-2} *and* p^b_{M-1} *are the steady states of aggregation procedure (4.30) and function* Q *is defined by (4.14).*

Proof: See Section 20.1.

Although a closed-form expression for Q cannot been derived (since N's are unknown and, therefore, p_i^f and p_i^b cannot be calculated), its estimate can be given as follows:

$$\widehat{Q} = 1 - E^{\frac{1}{2}[1+(\frac{M-3}{M-1})^{M/4}]} + \left(E^{\frac{1}{2}[1+(\frac{M-3}{M-1})^{M/4}]} - E^{(\frac{M-2}{M-1})}\right)\exp\left\{-\frac{E^{\frac{1}{M-1}} - p}{(1-E)^{(1/E)^{2E}}}\right\}. \tag{6.12}$$

Thus, an estimate of LBC for $M > 3$ is defined as

$$\widehat{N}_E(M > 3) = \left\lceil \frac{\ln\left\{\frac{1-E-\widehat{Q}}{(1-E)(1-\widehat{Q})}\right\}}{\ln\left\{\frac{(1-p)(1-\widehat{Q})}{1-p(1-\widehat{Q})}\right\}} \right\rceil, \tag{6.13}$$

where $\widehat{Q}$ is given in (6.12).

The accuracy of estimate (6.12), (6.13) has been evaluated numerically by calculating the exact value of N_E (using aggregation procedure (4.30) to evaluate Q) and comparing it with $\widehat{N}_E$ as follows:

$$\Delta_E = \frac{\widehat{N}_E - N_E}{N_E} \cdot 100\%. \tag{6.14}$$

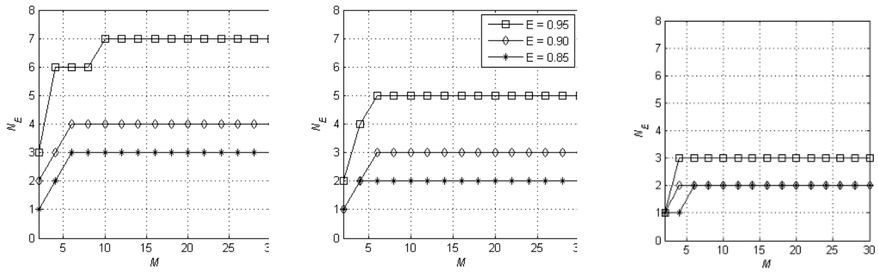
The values of Δ_E have been calculated for $p \in \{0.85, 0.9, 0.95\}$, $M \in \{5, 10, 15, 20, 25, 30\}$, and $E \in \{0.85, 0.9, 0.95\}$. It turned out that $\Delta_E = 0$ for all combinations of these parameters except when $[p = 0.85,\ M = 5,\ E = 0.85]$, where it is 50%. (Such a large error is due to the integer nature of N_E and $\widehat{N}_E$.) Thus, we conclude that in most cases $\widehat{N}_E$ provides a sufficiently accurate estimate of N_E.

The behavior of $\widehat{N}_E$ for $M = 10$ is illustrated in Figures 6.1(c) and 6.2(c). Clearly, the buffer capacity is increased as compared to $M = 3$, and JIT operation becomes unacceptable for all values of p and E considered.

Using (6.12), (6.13), the behavior of lean buffering as a function of M can be investigated. This is illustrated in Figure 6.3. Interestingly, and to a certain degree unexpectedly, $\widehat{N}_E$ is constant for all $M \geq 10$. This implies that the lean buffering, appropriate for lines with 10 machines, is also appropriate for lines with any larger number of machines. Based on this observation, the following can be formulated:

Rule-of-thumb for selecting lean buffering: *In Bernoulli lines defined by conventions (a)-(e) of Subsection 4.2.1 with $M \geq 10$, the capacity of the lean buffering can be selected as shown in Table 6.1.*

Interestingly, the diagonal elements of this "matrix" are all identical, implying that buffer capacity 3 is necessary and sufficient to ensure line efficiency E if $p = E$. The lower triangle of this matrix ($p > E$) has all elements smaller than 3 and the upper ($p < E$) – larger than 3.



(a) $p = 0.85$ (b) $p = 0.90$ (c) $p = 0.95$

Figure 6.3: Lean buffering $\widehat{N}_E$ as a function of the number of machines in the system

Table 6.1: Rule-of-thumb for selecting lean buffer capacity in Bernoulli lines with $M \geq 10$

	$E = 0.85$	$E = 0.90$	$E = 0.95$
$p = 0.85$	3	4	7
$p = 0.90$	2	3	5
$p = 0.95$	2	2	3

6.3 Lean Buffering in Serial Lines with Non-identical Bernoulli Machines

6.3.1 Two-machine lines

In the case of two-machine lines with non-identical machines, as it follows from (6.1) and (4.19), the equation that defines N_E becomes

$$PR = PR_\infty E = p_2[1 - Q(p_1, p_2, N_E)] = p_2\left[1 - \frac{(1-p_1)(1-\alpha)}{1 - \frac{p_1}{p_2}\alpha^{N_E}}\right], \qquad (6.15)$$

where

$$PR_\infty = \min(p_1, p_2) \qquad (6.16)$$

and

$$\alpha = \frac{p_1(1-p_2)}{p_2(1-p_1)}. \qquad (6.17)$$

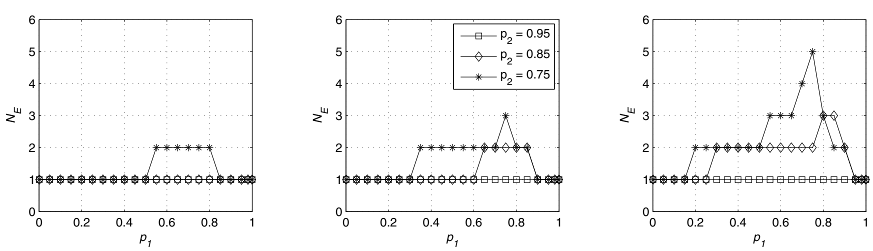
Solving (6.15) for N_E, we obtain:

Theorem 6.4 *The lean buffer capacity in Bernoulli lines defined by conventions (a)-(e) of Subsection 4.2.1 with $M = 2$ is given by*

$$N_E(p_1, p_2) = \left\lceil \frac{\ln\left\{\frac{p_2}{p_1}\left[\frac{p_1 - PR_\infty E}{p_2 - PR_\infty E}\right]\right\}}{\ln \alpha} \right\rceil, \qquad (6.18)$$

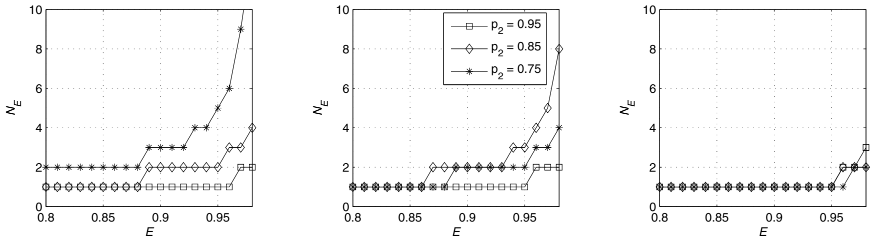
where PR_∞ and α are defined by (6.16) and (6.17), respectively.

Figure 6.4 illustrates the behavior of N_E as a function of p_1 for various values of p_2 and E, while Figure 6.5 shows N_E as a function of E for various p_1 and p_2. From these figures, we conclude:



(a) $E = 0.85$ (b) $E = 0.90$ (c) $E = 0.95$

Figure 6.4: Lean buffering in two-machine lines as a function of the first machine efficiency



(a) $p_1 = 0.75$ (b) $p_1 = 0.85$ (c) $p_1 = 0.95$

Figure 6.5: Lean buffering in two-machine lines as a function of line efficiency

- For p_2 sufficiently large, JIT operation is acceptable for all values of p_1 and E.
- For small p_2, JIT is acceptable only when p_1 is sufficiently large. For instance, if $p_2 = 0.75$, JIT represents LBC only if $p_1 > 0.88$ for $E = 0.9$ and $p_1 > 0.94$ for $E > 0.95$.
- The maximum of N_E tends to take place when $p_1 = p_2$.

Intuitively, it is expected that the lean buffering in a line $\{p_1, p_2\}$ is the same as in the reversed line, i.e., $\{p_2, p_1\}$. It turns out that this is indeed true as stated below.

Theorem 6.5 *Lean buffer capacity has the property of reversibility, i.e.,*

$$N_E(p_1, p_2) = N_E(p_2, p_1). \tag{6.19}$$

Proof: See Section 20.1.

6.3.2 $M > 2$-machine lines

Exact formulas for LBC in the case of $M > 2$ are all but impossible to derive. Therefore, we limit our attention to estimates of $N_{i,E}$. These estimates are obtained based both on closed formulas (6.9)-(6.12), (6.18) and on recursive calculations. Each of these approaches is described below.

Closed formula approaches: The following four methods have been investigated:

I. *Local pair-wise approach.* Consider every pair of consecutive machines, m_i and m_{i+1}, $i = 1, \ldots, M-1$, and select LBC using formula (6.18). This results in the sequence of buffer capacities denoted as

$$N^I_{1,E}, \ldots, N^I_{M-1,E}.$$

II. *Global pair-wise approach.* It is based on applying formula (6.18) to all possible pairs of machines (not necessarily consecutive) and then selecting the capacity of each buffer equal to the largest buffer obtained by this procedure. Clearly, this results in buffers of equal capacity, which is denoted as N_E^{II}.

III. *Local upper bound approach.* Consider all pairs of consecutive machines, m_i and m_{i+1}, $i = 1, \ldots, M-1$, substitute each of them by a two-machine line with identical machines defined by

$$\widehat{p}_i := \min\{p_i, p_{i+1}\}, i = 1, \ldots, M-1,$$

and select LBC using formula (6.9) with $p = \widehat{p}_i$. This results in the sequence of buffer capacities

$$N_{1,E}^{III}, \ldots, N_{M-1,E}^{III}.$$

IV. *Global upper bound approach.* Instead of the original line, consider a line with all identical machines specified by

$$\widehat{p} := \min\{p_1, p_2, \ldots, p_M\}$$

and select the buffer capacity, denoted as N_E^{IV}, using expressions (6.12) and (6.13). Due to the monotonicity of PR with respect to machine efficiency and buffer capacity (see Chapter 4), this approach provides an upper bound of LBC:

$$N_{i,E} \leq N_E^{IV}, i = 1, \ldots, M-1.$$

If the desired line efficiency for two-machine lines, involved in approaches I - III, were selected as E, the resulting efficiency of the M-machine line would be certainly less than E. To avoid this, the efficiency, E', of each of the two-machine lines is calculated as follows: For a given M-machine line, find the buffer capacity using approach IV. Then consider a two-machine line with identical machines, where each machine is defined by $\widehat{p} = \min\{p_1, \ldots, p_M\}$, and the buffer with the capacity as found above. Finally, calculate the production rate and the efficiency of this two-machine line and use it as E' in approaches I - III.

To analyze the performance of approaches I - IV, we consider $100,000$ lines formed by selecting M and p_i randomly and equiprobably from the sets

$$M \in \{4, 5, \ldots, 30\}, \tag{6.20}$$

$$0.70 \leq p \leq 0.97. \tag{6.21}$$

The desired efficiency for each of these lines is also selected randomly and equiprobably from the set

$$0.80 \leq E \leq 0.98. \tag{6.22}$$

For each k-th line, thus formed, we calculate the vector of buffer capacities

$$\mathbf{N}_k^j = \begin{bmatrix} N_{1,k}^j \\ N_{2,k}^j \\ \dots \\ N_{M-1,k}^j \end{bmatrix}, \quad k = 1, \dots, 100,000, \quad j = I, II, III, IV, \tag{6.23}$$

using the four approaches introduced above. The subscripts of $N_{i,k}^j$ represent i-th buffer, $i = 1, \dots, M_k - 1$, of the k-th line, $k = 1, 2, \dots, 100,000$; the superscript $j = I, II, III, IV$ represents the approach used for this calculation. In addition, we calculate the production rate, PR_k^j, and the efficiency, E_k^j, using expressions (4.30)-(4.36) and (6.1), respectively.

The efficacy of approaches I - IV is characterized by the following two metrics:

1. The average buffer capacity per machine among all systems analyzed:

$$N_{ave}^j = \frac{1}{K} \sum_{k=1}^{K} N_k^j \tag{6.24}$$

   where $K = 100,000$ and

$$N_k^j = \frac{1}{M_k - 1} \sum_{i=1}^{M_k - 1} N_{i,k}^j.$$

2. The percent of systems whose E_k^j turns out to be less than the desired efficiency E_k:

$$\Delta^j = \frac{1}{K} \sum_{k=1}^{K} Sg(E_k - E_k^j) \cdot 100\%, \tag{6.25}$$

   where $K = 100,000$ and

$$Sg(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

The results are given in Table 6.2. Clearly, approach I leads to the smallest average buffer capacity but, unfortunately, almost always results in line efficiency less than desired. Thus, a "local" selection of LBC (i.e., based on the two machines surrounding the buffer) is, practically always, unacceptable. Approaches II and III provide line efficiency less than desired in only a small fraction of cases and result in average buffer capacity 2 - 3 times larger than approach I. Approach IV, as expected, always guarantees the desired performance but requires the largest buffering.

To further differentiate between the four approaches, consider their performance as a function of M. To accomplish this, we formed 1000 lines for each $M \in \{4, 6, 8, 10, 15, 20, 25, 30, 80\}$ by selecting p_i's and E's randomly and

Table 6.2: Performance characteristics of approaches I - IV

Approach	I	II	III	IV
N_{ave}^{j}	2.0	6.2	5.3	7.2
Δ^{j}	97.3	0.1	0.1	0.0

equiprobably from sets (6.21) and (6.22), respectively. For each of these lines, we calculated buffer capacities using approaches I - IV and evaluated the performance metrics (6.24) and (6.25). The results are shown in Table 6.3. Examining these data, we conclude the following:

- The local pair-wise approach, in most cases, leads to a lower line efficiency than desired.
- The global pair-wise approach results in good performance from the point of view of both N_{ave} and Δ. For $M \leq 10$, it outperforms approach III from the point of view of N_{ave}. However, it is quite sensitive to M: N_{ave} increases substantially with M.
- The local upper bound approach is less sensitive to M and outperforms approach II for $M > 10$.
- The global upper bound approach substantially overestimates the LBC.

Based on the above, *it is recommended to use the global pair-wise approach in systems with $M \leq 10$ and local upper bound approach in systems with $M > 10$.*

Table 6.3: Effect of M on the performance of approaches I - IV

M	4	6	8	10	15	20	25	30	80
N_{ave}^{I}	1.7	1.9	2.0	2.0	2.0	2.1	2.0	2.1	2.1
N_{ave}^{II}	3.0	4.2	4.8	5.2	5.9	6.4	6.5	7.1	7.6
N_{ave}^{III}	4.2	5.0	5.2	5.2	5.2	5.4	5.3	5.7	5.6
N_{ave}^{IV}	5.0	6.3	6.7	6.8	7.0	7.4	7.3	7.9	7.9
Δ^{I}	88.7	92.8	95.6	97.1	98.4	98.0	99.2	99.3	100.0
Δ^{II}	3.1	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Δ^{III}	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.3
Δ^{IV}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Recursive approaches: The following two recursive methods have been investigated:

V. *Full search approach.* Start from all buffers of capacity 1. Increase the capacity of the first buffer by 1 and, using the aggregation procedure (4.30), calculate the production rate of the system. Return the first buffer capacity to its initial value, increase the second buffer capacity by 1, and calculate the resulting production rate. Repeat the same procedure for all buffers, determine the buffer that leads to the largest production rate, and permanently increase its capacity by 1. Repeat the same procedure until the desired line efficiency is reached. This results in the sequence of buffer capacities

$$N_{1,E}^{V}, \ldots, N_{M-1,E}^{V}.$$

VI. *Bottleneck-based approach.* Consider a production line with buffer capacity calculated according to approach I but rounded down in formula (6.18) rather than up. Although being relatively small, this buffering often leads, as it follows from Tables 6.2 and 6.3, to line efficiency less than desired. Therefore, to improve the line efficiency, increase the buffering according to the following procedure: Using Bottleneck Indicator 5.1, identify the bottleneck machine (or, when applicable, primary bottleneck machine) and increase the capacity of both buffers surrounding this machine by 1. Repeat this procedure until the desired line efficiency is reached. This results in the sequence of buffer capacities denoted as

$$N_{1,E}^{VI}, \ldots, N_{M-1,E}^{VI}.$$

Clearly, approach V gives a smaller buffer capacity than approach VI. However, the latter might require a shorter computation time. Therefore, in order to compare V and VI, the computation time should be taken into account. This additional performance metric is defined as the total computer time necessary to carry out the calculation:

$$t^{j} = t_{end}^{j} - t_{start}^{j}, \tag{6.26}$$

where t_{start}^{j} and t_{end}^{j}, $j \in \{I, \ldots, VI\}$, are the times (in seconds) of the beginning and the end of the computation.

Based on metrics (6.24), (6.25), and (6.26), we compared approaches I-VI using the production systems generated by selecting p_i's randomly and equiprobably from set (6.21). The results are shown in Tables 6.4-6.6. Specifically, Tables 6.4 and 6.5 present the results obtained using 5000 randomly generated 5- and 10-machine lines, respectively, while Table 6.6 is based on the analysis of 2000 randomly generated 15-machine lines. Examining these data, we conclude the following:

- Full search approach, as expected, results in the smallest buffer capacity and the longest calculation time.
- Approaches II - IV, being based on closed-form expressions, are much faster than V (up to two orders of magnitude for long lines) but lead to average buffering 2 - 3 times larger than that of V.
- Approach VI provides a good tradeoff between the calculation time and buffer capacity. In long lines, it is about 20 times faster than V and results

in average buffering almost the same as V (about 10% difference). Also, it is about 6 times slower than I - IV but gives buffering 2 - 3 times smaller than II - IV.

Table 6.4: Performance characteristics of approaches I - VI in five-machine lines

(a) $E = 0.80$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	1.5	2.2	2.4	2.9	1.4	1.5
Δ^j	27.8	0.0	0.0	0.2	0.0	0.0
t^j	8	8	8	8	59	28

(b) $E = 0.85$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	1.7	2.7	3.0	3.6	1.7	1.8
Δ^j	36.0	0.0	0.0	0.2	0.0	0.0
t^j	8	8	8	8	91	38

(c) $E = 0.90$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	2.2	3.6	4.2	5.0	2.0	2.3
Δ^j	33.8	0.0	0.0	0.0	0.0	0.0
t^j	6	6	6	6	91	27

(d) $E = 0.95$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	3.2	6.0	7.9	9.5	2.8	3.3
Δ^j	25.5	0.0	0.0	0.0	0.0	0.0
t^j	6	6	6	6	150	29

Thus, if a recursive approach is to be used, *the bottleneck-based one is recommended*.

Illustrative examples: To illustrate particular cases of lean buffering designed using approaches I - VI, we provide several examples.

Consider the four production lines with machines specified in Table 6.7 along with the desired line efficiency. The estimates of LBC for each of these lines, calculated using approaches I - VI, are shown in Tables 6.8 - 6.11, along with resulting line efficiency. These examples clearly indicate that:

- Approach VI is almost as good as the full search approach V.
- Approach II in most cases outperforms approach III (since $M = 5$); however, it still leads to buffer capacity 2 - 3 times larger than approach V.

Table 6.5: Performance characteristics of approaches I - VI in 10-machine lines

(a) $E \in [0.80, 0.89]$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	1.8	3.5	3.2	4.1	1.8	1.9
Δ^j	66.7	0.0	0.0	0.0	0.0	0.0
t^j	15	15	15	15	757	76

(b) $E \in [0.89, 0.98]$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	3.2	8.0	8.2	10.6	2.7	3.2
Δ^j	48.5	0.0	0.0	0.0	0.0	0.0
t^j	26	26	26	26	2339	124

Table 6.6: Performance characteristics of approaches I - VI in 15-machine lines

(a) $E \in [0.80, 0.89]$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	1.8	3.8	3.3	4.2	1.8	2.0
Δ^j	81.0	0.0	0.0	0.0	0.0	0.0
t^j	18	18	18	18	2452	107

(b) $E \in [0.89, 0.98]$

Approaches	I	II	III	IV	V	VI
N^j_{ave}	3.2	9.0	8.2	10.9	2.6	3.2
Δ^j	57.6	0.0	0.0	0.0	0.0	0.0
t^j	24	24	24	24	5837	138

Table 6.7: Machine parameters and desired line efficiencies

Line	m_1	m_2	m_3	m_4	m_5	E
1	0.78	0.88	0.75	0.91	0.83	0.80
2	0.79	0.84	0.85	0.94	0.76	0.85
3	0.72	0.85	0.74	0.82	0.84	0.90
4	0.77	0.87	0.90	0.90	0.72	0.95

Table 6.8: LBC estimates and resulting efficiency for Line 1 (desired $E = 0.80$)

Buffer	b_1^j	b_2^j	b_3^j	b_4^j	E^j
Desired					0.80
N_i^I	1	1	1	1	0.71
N_i^{II}	2	2	2	2	0.90
N_i^{III}	3	3	3	2	0.96
N_i^{IV}	3	3	3	3	0.96
N_i^V	1	2	2	1	0.83
N_i^{VI}	1	2	2	1	0.83

Table 6.9: LBC estimates and resulting efficiency for Line 2 (desired $E = 0.85$)

Buffer	b_1^j	b_2^j	b_3^j	b_4^j	E^j
Desired					0.85
N_i^I	2	2	1	1	0.84
N_i^{II}	3	3	3	3	0.99
N_i^{III}	3	2	2	3	0.96
N_i^{IV}	3	3	3	3	0.99
N_i^V	1	2	2	1	0.85
N_i^{VI}	1	2	2	1	0.85

Table 6.10: LBC estimates and resulting efficiency for Line 3 (desired $E = 0.90$)

Buffer	b_1^j	b_2^j	b_3^j	b_4^j	E^j
Desired					0.90
N_i^I	2	2	3	3	0.90
N_i^{II}	4	4	4	4	0.97
N_i^{III}	5	5	5	4	0.99
N_i^{IV}	5	5	5	5	0.99
N_i^V	2	3	2	2	0.90
N_i^{VI}	2	3	3	2	0.92

Table 6.11: LBC estimates and resulting efficiency for Line 4 (desired $E = 0.95$)

Buffer	b_1^j	b_2^j	b_3^j	b_4^j	E^j
Desired					0.95
N_i^I	3	3	4	2	0.98
N_i^{II}	5	5	5	5	1.00
N_i^{III}	9	5	4	10	1.00
N_i^{IV}	10	10	10	10	1.00
N_i^V	2	2	2	2	0.96
N_i^{VI}	2	2	3	2	0.97

PSE Toolbox: The six methods of lean buffering calculation for Bernoulli lines are implemented in the **Lean Buffer Design** function of the toolbox. For a description and illustration of these tools, see Subsections 19.6.1 and 19.6.2.

6.4 Case Studies

6.4.1 Automotive ignition coil processing system

The Bernoulli model of this system is given in Figure 3.31 (for Periods 1 and 2). Below, its lean buffering is evaluated using the model for Period 2 (since it corresponds to system conditions during and after this case study).

Rule-of-thumb approach: The rule-of-thumb for selecting lean buffering in Bernoulli lines with ten or more identical machines is given in Table 6.1. Although the system at hand consists of non-identical machines, we use this rule to obtain a quick (but, clearly, not tight) upper bound for the lean buffering.

To accomplish this, "convert" the coil processing system into a virtual production line with identical Bernoulli machines defined by

$$\widehat{p} = \min\{p_1, \ldots, p_{13}\} = 0.8825,$$

where p_i, $i = 1, \ldots, 13$, is the efficiency of the i-th machine in Figure 3.31(b). Since the cycle time in the original line is selected to ensure the nominal throughput of 593.07 parts/hour, the maximum throughput of the virtual line is $593.07 \times 0.8825 = 523.38$ parts/hour.

Assume that the desired line efficiency $E = 0.9$ (i.e., the desired throughput is $523.38 \times 0.9 = 471.05$ parts/hour). Then, taking into account that $\widehat{p}$ is close to 0.9 and using Table 6.1, we infer that the lean buffer capacity in the virtual line is

$$N_i^{Ber} = 3, \quad i = 1, \ldots, 12. \tag{6.27}$$

Due to the monotonicity of PR with respect to machine and buffer parameters (see Chapter 4), we conclude that (6.27) is an upper bound of lean buffering for the original Bernoulli line. Clearly, this buffering is 2 - 3 times larger than that in the model of Figure 3.31(b). (Note that this configuration gives throughput of 521.35 parts/hour.)

Formula-based approach: Since the coil processing line contains more than ten machines, the local upper bound approach (i.e., Approach III of Subsection 6.3.2) is recommended for selecting lean buffering. Using this approach and the machine parameters of Figure 3.31(b), we obtain the following vector of lean buffering estimates for $E = 0.9$, i.e., 471.05 parts/hour:

$$N^{Ber} = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2].$$

This buffering ensures a throughput of 517.29 parts/hour, which is much higher than the desired one, i.e., the buffering is not lean.

Recursive approach: Applying the bottleneck-based approach (i.e., approach VI of Subsection 6.3.2) with $E = 0.9$, we obtain a tighter estimate:

$$N^{Ber} = [1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1].$$

Finally, using the full search approach, we arrive at a still tighter estimate:

$$N^{Ber} = [1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1],$$

and this buffering leads to the line efficiency $E = 0.923$, i.e., the throughput of 483.13 parts/hour.

6.4.2 Automotive paint shop production system

The Bernoulli model of this system is given in Tables 3.11 and 3.12 (for months 1-5). To evaluate lean buffering, we average machine parameters over all five monthly periods and obtain the machine parameter values given in Table 6.12. Assuming that empty carriers are managed appropriately so that m_3 is never starved (see Chapter 5) and neglecting machines 9 - 11 (since they operate perfectly), we use the model of Table 6.12 to characterize lean buffering for the paint shop.

Table 6.12: Five-month averages of machine parameters

p_3	p_4	p_5	p_6	p_7	p_8
0.9705	0.9542	0.9905	0.9987	0.9681	0.9707

Rule-of-thumb approach: Convert the system into a virtual line with identical Bernoulli machines defined by

$$\widehat{p} = \min\{p_3, \dots, p_8\} = 0.954,$$

where p_i, $i = 3, \dots, 8$, is given in Table 6.12. This implies that the maximum throughput is $63 \times 0.954 = 60.116$ jobs/hour. Assuming that the desired efficiency is $E = 0.95$ (which ensures the production of $60.116 \times 0.95 = 57.110$ jobs/hour) we obtain the lean buffering for the virtual system given by

$$N_i = 3, \qquad i = 3, \dots, 7.$$

This results in throughput of 59.76 jobs/hour. Clearly, this buffering is far below that of Table 3.12, implying that the paint shop system is far from being lean.

Recursive approach: Applying the bottleneck-based approach (i.e., approach VI of Subsection 6.3.2) with $E = 0.95$, we obtain a tighter estimate of lean buffering:

$$N = [2, 2, 1, 1, 1],$$

which leads to throughput of 57.56 jobs/hour. This results in over 90% reduction in buffer capacities.

6.5 Summary

- Lean buffer capacity (LBC) is the smallest buffering necessary and sufficient to ensure the desired efficiency of a production line. Thus, LBC is the "just-right" rather than "just-in-time" buffering.
- Lean buffer capacity in serial lines with identical Bernoulli machines can be evaluated using closed-form expressions (6.9)-(6.13) or the rule-of-thumb given in Table 6.1.
- In the case of non-identical machines, a closed formula is available only for two-machine lines (6.18).
- For longer lines, estimates of lean buffer capacity can be obtained using either closed-form expressions or recursive calculations.
- If closed formulas are used, the global pair-wise approach and the local upper bound approach are recommended when $M \leq 10$ and $M > 10$, respectively.
- If recursive calculations are used, the bottleneck-based approach is recommended.

6.6 Problems

Problem 6.1 Consider a Bernoulli serial line with two identical machines. Assume that it operates under the assumption of symmetric blocking described in Problem 4.5.

(a) Derive the formula for lean buffer capacity in such a system.
(b) Plot the lean buffer capacity as a function of machine efficiency.
(c) Plot the lean buffer capacity as a function of line efficiency.
(d) Does the symmetric blocking require larger or smaller lean buffering than that defined by the blocked before service assumption (see Subsection 6.2.1)?

Problem 6.2 Consider the two-machine Bernoulli serial line with defective parts produced as described in Problem 4.4.

(a) Derive the formula for lean buffer capacity in such a system.
(b) Plot the lean buffer capacity as a function of machine efficiency.
(c) Plot the lean buffer capacity as a function of line efficiency.
(d) Compare the resulting graphs with those obtained in Subsection 6.3.1 and comment on the effect of the defectives on the lean buffer capacity.

Problem 6.3 Investigate the precision of the rule-of-thumb given in Table 6.1. Specifically, consider the Bernoulli serial line with 15 identical machines having $p = 0.87$. Assume that the desired line efficiency is 0.93.

(a) Using expressions (6.12) and (6.13), calculate N_E.

(b) Using the rule-of-thumb, determine an estimate of the lean buffer capacity for this system.

(c) Compare the two results and comment on the applicability of the rule-of-thumb for p and E, which are not included in Table 6.1.

Problem 6.4 Consider the ignition device production line of Problem 3.3.

(a) Using any approach of Section 6.3, design the lean buffering for this system with $E = 0.90$.

(b) Using the same approach as in (a), design the lean buffering for this system with $E = 0.97$.

(c) Compare the two results and recommend which line efficiency should be used for practical implementation.

Problem 6.5 Consider the five-machine serial line of Problem 3.4. Assume that the desired line efficiency is 0.97. Using the Bernoulli model of this line:

(a) Determine the estimate of the lean buffer capacity using the global upper bound approach of Subsection 6.3.2.

(b) Determine the estimate of the lean buffer capacity using the bottleneck-based approach.

(c) Compare the two results and state which of these approaches is preferable (from all points of view you find important).

Problem 6.6 Consider a production line with $p = [0.88, 0.95, 0.9, 0.92]$. Select the lean buffering capacity for $E = 0.95$ using the following approaches:

(a) Global pair-wise approach.

(b) Local upper bound approach.

(c) Bottleneck-based recursive approach.

(d) Full search recursive approach.

Compare the results and comment on the advantages and disadvantages of each method.

6.7 Annotated Bibliography

The quantitative notion of lean buffering has been introduced and analyzed in

[**6.1**] E. Enginarlar, J. Li and S.M. Meerkov, "How Lean Can Lean Buffers Be?" *IIE Transactions*, vol. 37, pp. 333-342, 2005.

The lean buffering in Bernoulli lines, has been investigated in

[**6.2**] A.B. Hu and S.M. Meerkov, "Lean Buffering in Serial Production Lines with Bernoulli Machines," *Mathematical Problems in Engineering*, vol. 2006, Article ID 17105, 2006.

It is the basis for the material of this chapter.

Additional results on lean buffering (for the exponential and non-exponential reliability models) can be found in

[**6.3**] E. Enginarlar, J. Li and S.M. Meerkov, "Lean Buffering in Serial Production Lines with Non-exponential Machines," *OR Spectrum*, vol. 27, pp. 195-219, 2005.

[**6.4**] S.-Y. Chiang, A.B. Hu and S.M. Meerkov, "Lean Buffering in Serial Production Lines with Non-identical Exponential Machines," *IEEE Transactions on Automation Science and Engineering*, vol. 5, pp. 298-306, 2008.