SYSTEM-THEORETIC PROPERTIES OF PRODUCTION LINES

by

Liang Zhang

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2009

Doctoral Committee:
Professor Semyon. M. Meerkov, Chair
Professor Jessy W. Grizzle
Professor Shixin J. Hu
Professor Pierre T. Kabamba
一万年太久，只争朝夕！
To my family.
ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Professor Semyon M. Meerkov for his guidance, insights, encouragement, and “push” throughout my entire study. It has been a great pleasure and honor to work with him during my years at the University of Michigan. I also would like to express my appreciation to the other members in my committee, Professors Jessy Grizzle, Jack Hu, and Pierre Kabamba.

I am particularly grateful to Professor Jingshan Li from the University of Kentucky. He is the one who introduced this great research group to me and offered me great help during this research.

I would like to thank Drs. S. Biller, S. P. Marin and J. Arinez from General Motors R&D Center and the management and production personnel from GM Hamtramck Assembly Plant and Orion Assembly Plant for their help with the case studies. Also, I would like to thank Professor N. Shimkin from Technion and my colleagues at the University of Michigan for many valuable discussions. These include Dr. S. Ching, Mr. A. B. Hu, Mr. J. Z. Huang, Mr. B. Rumao, Mr. H. Shih and Mr. F. Xu.

Constant love and care from my family have been a continuous source of my strength to finish this work. I would like to thank my wife Chenfei Ma, my father Baolin Zhang, and my mother Xiufen Cui. They have been supporting and believing in me even during the most difficult times. This volume is devoted to you. Without your love, this work would not be possible.

Finally, I would like to acknowledge the financial aid from National Science Foundation, Department of EECS, Department of ME, and General Motors Corporation.
# TABLE OF CONTENTS

DEDICATION ................................................................. ii

ACKNOWLEDGEMENTS .................................................. iii

LIST OF FIGURES .......................................................... viii

LIST OF TABLES ............................................................ xi

LIST OF APPENDICES ...................................................... xiii

CHAPTER

I. INTRODUCTION .............................................................. 1

1.1 Motivation .............................................................. 1

1.2 System-Theoretic Properties Addressed ......................... 2

1.2.1 Reversibility ......................................................... 2

1.2.2 Monotonicity ......................................................... 3

1.2.3 System-wide effects of up- and downtime .................... 3

1.2.4 Improvability ......................................................... 4

1.2.5 Bottlenecks .......................................................... 4

1.2.6 Transients ......................................................... 5

1.3 Approach of This Research .......................................... 5

1.4 Related Literature .................................................... 6

1.4.1 Serial lines ......................................................... 6

1.4.2 Closed lines ......................................................... 7

1.4.3 Product quality ................................................... 8

1.4.4 Summary of the related literature and topics of this research ......................................................... 10

1.5 Outline .............................................................. 10

II. SYSTEM MODELS AND PROBLEMS .................................. 12

2.1 Introduction .......................................................... 12

2.2 Terminology .......................................................... 12

2.3 Machine Reliability Models ......................................... 13

2.3.1 Slotted time case .................................................. 14

2.3.2 Continuous time case ............................................. 15
2.4 Systems Considered ........................................ 17
  2.4.1 Slotted time case .................................... 17
  2.4.2 Continuous time case ................................ 20
2.5 Performance Measures .................................... 21
2.6 Problems Addressed ...................................... 22

III. SYSTEM-THEORETIC PROPERTIES OF SERIAL LINES 24
  3.1 Review of Existing Results on System-Theoretic Properties . . 24
    3.1.1 Serial Bernoulli lines ............................ 24
    3.1.2 Serial exponential lines .......................... 30
  3.2 Novel Facts on System-Theoretic Properties of Serial Lines . . 31
    3.2.1 Reversibility .................................... 31
    3.2.2 Monotonicity .................................... 32
    3.2.3 Effects of up- and downtime ....................... 33
    3.2.4 Improvability ................................... 35
    3.2.5 Bottlenecks ..................................... 37
  3.3 Summary ................................................ 38

IV. SYSTEM-THEORETIC PROPERTIES OF CLOSED LINES 39
  4.1 Introduction .......................................... 39
  4.2 Closed Bernoulli Lines ................................ 40
    4.2.1 Reversibility .................................... 40
    4.2.2 Monotonicity .................................... 41
    4.2.3 Unimpediment .................................... 42
    4.2.4 Improvability ................................... 44
    4.2.5 Bottlenecks ..................................... 51
  4.3 Closed Continuous Lines ................................ 54
    4.3.1 Reversibility .................................... 54
    4.3.2 Monotonicity .................................... 55
    4.3.3 Unimpediment .................................... 55
    4.3.4 Improvability ................................... 56
  4.4 Case Study: Automotive Paint Shop Production System .......... 57
    4.4.1 System description and layout .................... 57
    4.4.2 Structural modeling ................................ 59
    4.4.3 Modeling and identification of the machines and buffers 60
    4.4.4 Overall system model .............................. 61
    4.4.5 S-improvability .................................. 62
  4.5 Summary ................................................ 62

V. SYSTEM-THEORETIC PROPERTIES OF PRODUCTION LINES WITH QUALITY ISSUES 64
  5.1 Introduction .......................................... 64
  5.2 Bernoulli Lines with Non-perfect Quality Machines ............. 65
    5.2.1 Model ............................................. 65
    5.2.2 Reversibility .................................... 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.3</td>
<td>Monotonicity</td>
<td>67</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Bottlenecks</td>
<td>67</td>
</tr>
<tr>
<td>5.2.5</td>
<td>Inspection machine positioning</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Bernoulli Lines with Quality-Quantity Coupling Machines</td>
<td>79</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Model</td>
<td>79</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Monotonicity</td>
<td>79</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Bottlenecks</td>
<td>81</td>
</tr>
<tr>
<td>5.4</td>
<td>Bernoulli Lines with Rework</td>
<td>87</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Model and approach</td>
<td>87</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Reversibility</td>
<td>89</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Monotonicity</td>
<td>91</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Bottlenecks</td>
<td>91</td>
</tr>
<tr>
<td>5.5</td>
<td>Case Study: Automotive Paint Shop Production System</td>
<td>101</td>
</tr>
<tr>
<td>5.5.1</td>
<td>System model and problem formulation</td>
<td>101</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Improvement project: Stage 1</td>
<td>110</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Improvement project: Stage 2</td>
<td>113</td>
</tr>
<tr>
<td>5.6</td>
<td>Summary</td>
<td>116</td>
</tr>
<tr>
<td>VI</td>
<td>TRANSIENT PROPERTIES OF SERIAL LINES</td>
<td>118</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>118</td>
</tr>
<tr>
<td>6.2</td>
<td>Bernoulli Lines with Identical Machines</td>
<td>119</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Problem formulation</td>
<td>119</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Analysis of the second largest eigenvalue</td>
<td>124</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Analysis of the pre-exponential factors</td>
<td>130</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Analysis of the settling time</td>
<td>132</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Analysis of the production losses</td>
<td>140</td>
</tr>
<tr>
<td>6.3</td>
<td>Bernoulli Lines with Non-identical Machines</td>
<td>146</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Problem formulation</td>
<td>146</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Transients in the constrained case</td>
<td>147</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Transients in the unconstrained case</td>
<td>153</td>
</tr>
<tr>
<td>6.4</td>
<td>Geometric Lines with Identical Machines</td>
<td>157</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Problem formulation</td>
<td>157</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Transients of individual machines</td>
<td>158</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Transients of two-machine lines with (N = 1)</td>
<td>161</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Transients of two-machine lines with (N \geq 2)</td>
<td>166</td>
</tr>
<tr>
<td>6.5</td>
<td>Application: Float-based Production Systems</td>
<td>169</td>
</tr>
<tr>
<td>6.5.1</td>
<td>System description</td>
<td>169</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Terminal performance measures and problem addressed</td>
<td>171</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Design of lean float</td>
<td>173</td>
</tr>
<tr>
<td>6.6</td>
<td>Summary</td>
<td>180</td>
</tr>
<tr>
<td>VII</td>
<td>CONCLUSIONS AND FUTURE WORK</td>
<td>182</td>
</tr>
<tr>
<td>7.1</td>
<td>Conclusions</td>
<td>182</td>
</tr>
<tr>
<td>7.2</td>
<td>Future work</td>
<td>184</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

1.1 Examples of production systems ........................................... 2  
2.1 Geometric reliability model ............................................... 14  
2.2 Exponential reliability model ............................................. 15  
2.3 Closed production line ..................................................... 18  
2.4 Serial production line with perfect quality, non-perfect quality, QQC, and inspection machines ............................................. 19  
2.5 Serial production line with rework ...................................... 20  
3.1 $M$-machine Bernoulli line and its reverse ............................ 24  
3.2 BN identification in $M$-machine lines .................................. 29  
4.1 Production rate of closed lines as a function of the number of carriers ......................................................... 40  
4.2 Example of $S$-Continuous Improvement Procedure 4.1 .............. 48  
4.3 Example of $N_0$-Continuous Improvement Procedure 4.1 .......... 49  
4.4 Towards bottleneck identification in closed serial lines ............... 51  
4.5 BN identification in a two-machine closed line ......................... 52  
4.6 BN identification in a five-machine closed line ......................... 53  
4.7 Layout of the automotive paint shop system ............................ 58  
4.8 Structural model of the automotive paint shop system ................. 60  
4.9 Closed line model of the automotive paint shop system ............... 61  
5.1 Bernoulli serial line with quantity inspection machines ............... 65  
5.2 Example 5.1 .................................................................. 69  
5.3 Example 5.2 .................................................................. 72  
5.4 Segmentations ................................................................ 73  
5.5 Example 5.3 .................................................................. 74  
5.6 Example 5.4 .................................................................. 75  
5.7 Example 5.5 .................................................................. 77  
5.8 $\hat{PR}$ as a function of $p_{insp}$ and $I_{insp}$ for Example 5.5 ........ 77  
5.9 Example 5.6 .................................................................. 78  
5.10 $\hat{PR}$ as a function of $p_{insp}$ and $I_{insp}$ for Example 5.6 ......... 78  
5.11 Serial line with perfect, non-perfect and QQC machines ............... 79  
5.12 Example 5.7 .................................................................. 80  
5.13 $PR$ as a function of $p_1$ and $p_2$ in Example 5.7 ................. 80  
5.14 Example 5.8 .................................................................. 81
6.19 SLE and PEFs as functions of machine efficiencies for the unconstrained case with $N = 10$ ........................................ 154
6.20 SLE and PEFs as functions of machine efficiencies for the unconstrained case with $N = 2$ ........................................ 155
6.21 Two-machine Bernoulli line and its reverse .......................... 155
6.22 Illustration of reversibility ............................................. 157
6.23 Two-machine geometric line ........................................... 157
6.24 Behavior of $|\lambda_1|$ as a function of $R$ ................................ 159
6.25 Partitioning of the simplex $0 < P < R < 1$ according to SLE .... 164
6.26 Transients of $PR$ for $e = 0.9$ ........................................ 165
6.27 Transients of $PR$ with increased uptime or decreased downtime for $e = 0.7$, $R = 0.1$ and $e' = 0.9$ ................................. 166
6.28 Illustration of the accuracy of expression (6.88) for $e = 0.9$ and $R = 0.1168$ ......................................................... 169
6.29 Average accuracy of expression (6.88) ................................ 170
6.30 Float-based production system with two departments ............... 173
6.31 Transient performance measures .................................... 174
6.32 Terminal performance measures ..................................... 174
6.33 Comparison of Approaches I, II and III .............................. 180

B.1 Accuracy of Numerical Fact 4.3 ...................................... 229
B.2 Accuracy of Numerical Fact 4.4 ...................................... 231
B.3 Accuracy of Bottleneck Indicator 4.1 ................................. 233
B.4 Dependence of practically unimpeding $(N_0, S)$ on machine parameters .......................... 235
B.5 Closed and equivalent open two-machine lines ...................... 236

C.1 QQC functions for the justification of Numerical Fact 5.4 ........ 251
## LIST OF TABLES

4.1 Example of $S$-Continuous Improvement Procedure 4.1 (starting from $S = 2$) ................................................................. 48
4.2 Example of $S$-Continuous Improvement Procedure 4.1 (starting from $S = 21$) ............................................................... 49
4.3 Example of $N_0$-Continuous Improvement Procedure 4.1, $PR_o = 0.6665$ ................................................................. 49
4.4 Comparison of the $S$-Continuous Improvement Procedure 4.1 and equation (4.9) ......................................................... 50
4.5 Capacity of the machines (jobs/hour) ......................................................... 59
4.6 System performance (jobs/hour) ......................................................... 59
4.7 Average production losses (jobs/hour) ......................................................... 59
4.8 Machine parameters ($p_i$) ................................................................. 60
4.9 Buffer capacity ................................................................. 61
4.10 Closed line model validation data: $TP$ ......................................................... 61
4.11 Open line performance and ($N_0, S$)-impediment ......................................................... 62
4.12 $S$-Continuous Improvement Procedure 4.1 for Month 1 ......................................................... 62
5.1 $\hat{PR}$ and $\hat{SR}$ of Lines 1-3 in Example 5.2 ......................................................... 71
5.2 Operation downtime $T_{\text{down},i}$ ................................................................. 105
5.3 Operation efficiency $e_i$ ................................................................. 105
5.4 Operation capacity $c_i$ ................................................................. 106
5.5 Operation productivity parameters $p_i$ of Weeks 1-4 ......................................................... 107
5.6 Buffer parameters ................................................................. 107
5.7 Model validation for Weeks 1-4 ................................................................. 109
5.8 Operation productivity parameters $p_i$ of Weeks 5-8 ......................................................... 114
5.9 Model validation for Weeks 5-8 ................................................................. 114
5.10 TP-BN continuous improvement procedure ................................................................. 116
6.1 Accuracy of approximation $\hat{PR}(n)$ ................................................................. 134
6.2 Accuracy of approximation $\hat{WTP}(n)$ ................................................................. 135
6.3 Accuracy of estimate $\hat{t}_{sPR}$ ................................................................. 140
6.4 Accuracy of estimate $\hat{t}_{sWIP}$ ................................................................. 141
6.5 Percent of loss $\Lambda_T^0$ for ten-machine lines ................................................................. 143
6.6 Float capacity ................................................................. 177
6.7 Production deficit ................................................................. 178
6.8 Service level ................................................................. 178
LIST OF APPENDICES

A. Proofs and Justifications for Chapter III .................................................. 186
B. Proofs and Justifications for Chapter IV .................................................. 220
C. Proofs and Justifications for Chapter V .................................................. 241
D. Proofs and Justifications for Chapter VI .................................................. 259
CHAPTER I

INTRODUCTION

1.1 Motivation

Production systems are sets of machines and material handing devices (buffers) arranged so as to produce desired products. The earliest publications on quantitative analysis of production systems appeared in [1]-[3]. During the past 50 years, production systems have been studied extensively and numerous results have been reported (see monographs [4]-[8] and reviews [9], [10]). Among these studies, performance analysis and optimization of production systems are mostly investigated. In contrast, system-theoretic properties of production systems have rarely been discussed in the literature. These properties, however, are of importance because they reveal the fundamental principles that characterize the behavior of such systems. This dissertation is intended to provide a contribution in this direction.

In this work, we consider production systems with unreliable machines and finite buffers. Examples of such systems are shown in Figure 1.1, where the circles represent machines and rectangles are buffers.

Due to random machine breakdowns and finiteness of buffers, such systems can be viewed as a class of stochastic nonlinear dynamical systems. In Control Theory, controllability, observability, and stability are often used to characterize a dynamical system’s property. For production systems with Markovian dynamics, the stability is ensured by the properties of the eigenvalues of the state transition matrices. As for
the controllability and observability, since active feedback control is practically never used to manage production systems, they are not of immediate importance at this time. Thus, new system-theoretic properties should be defined to uncover the fundamental principles that govern production systems behavior. To accomplish this, six system-theoretic properties of production lines are addressed in this dissertation and informally described next. Later in this document (Chapters III-VI), these properties are defined and analyzed rigorously.

1.2 System-Theoretic Properties Addressed

1.2.1 Reversibility

The first system-theoretic property addressed in this work is reversibility. Specifically, a production line is called reversible if the throughput of the system (expected number of parts produced per unit of time) remains the same when the parts flow is reversed [11]. The reversibility property has many practical implications and can facilitate continuous improvement and design of production systems. Analyzing the reversibility of a production line can give answers to questions such as:

- Should larger buffers and more efficient machines be placed at its beginning, at
the end, or in the middle of a line so that the system throughput is maximized?

- In a serial line, if the machines are identical and only one buffer is available, where should it be placed within the line so that the throughput is maximized?

- If all machines and buffers in a serial line are identical and one machine can be improved, which one should it be, so that the throughput is maximized?

1.2.2 Monotonicity

Monotonicity property of a production line refers to the monotonic or non-monotonic behavior of its performance measures as functions of machine, buffer, and other system parameters. Although these functions can rarely be expressed in closed form, understanding their monotonicity property is necessary, because it characterizes the direction of change of parameters to improve the system’s performance. Indeed, if the system performance is non-monotonic with respect to a certain parameter, one should be very careful in continuous improvement and design regarding this particular argument, since erroneously selected value may result in performance degradation.

1.2.3 System-wide effects of up- and downtime

Machine reliability is usually characterized by up- and downtime distributions. Let $T_{up}$ and $T_{down}$ denote the average up- and downtime. Then, the machine efficiency, $e$, i.e., the expected value of the number of parts produced during a cycle time (which is the time necessary to process a part), can be calculated as follows:

$$e = \frac{T_{up}}{T_{up} + T_{down}} = \frac{1}{1 + \frac{T_{down}}{T_{up}}}.$$  \hfill (1.1)

Clearly, for a single machine its throughput in isolation remains the same for either shorter or longer up- and downtimes as long as $T_{up}/T_{down}$ is fixed. In addition, it is obvious that machine efficiency can by improved by either increasing $T_{up}$ or decreasing $T_{down}$ and increasing $T_{up}$ or decreasing $T_{down}$ by the same factor $(1 + \alpha)$, $\alpha > 0$, has
the same effect since

\[ e' = \frac{1}{1 + \frac{T_{\text{down}}}{(1+\alpha)T_{\text{up}}}}. \] (1.2)

Therefore, for an isolated machine, up- and downtime have the same effect on its performance.

For a production system with machines and buffers interacting with each other, the effects of up- and downtime on the system’s performance as a whole may be different. Indeed, to characterize the effects of up- and downtime, the following two questions are to be answered:

1. To maximize the throughput, should machines with shorter up- and downtime be used or machines with longer up- and downtimes?

2. To maximize the throughput, is it more beneficial to increase machine uptime or decrease machine downtime?

### 1.2.4 Improvability

A production system is called improvable if the limited resources available can be re-allocated so that a performance measure is improved. Improvability is related to optimality. Indeed, an unimprovable system is optimal. However, the term improvable is used to indicate that the objective is not necessarily to render the system optimal but rather to determine whether it can or cannot be improved. In addition, given the lack of precise information on the factory floor, optimality may not be practically achievable whereas improvability may still be characterized by simple indicators which are robust with respect to imprecise information.

### 1.2.5 Bottlenecks

Bottleneck (BN) is an important system-theoretic property of production systems. Indeed, bottleneck identification and elimination are considered as one of the most
effective way in continuous improvement of production systems. In this research, we define the bottlenecks as *the machine and the buffer, which impede the system performance in the strongest manner*. Under this definition, bottleneck identification in serial open lines, closed lines, and production lines with quality issues are studied.

### 1.2.6 Transients

Dynamic systems are characterized by both steady state and *transient* behavior. In Control Theory, the transient response is often characterized by damping, rise time, overshoot, settling time, etc. For production systems, in contrast to the numerous results on steady state performance, transient properties remain practically unexplored. Transient properties of production systems, however, are of practical importance. Indeed, if the steady state is reached after a relatively long period of time, the system may suffer substantial production losses. In addition, for systems operating mostly in the transient period, the steady state analysis techniques developed may become inapplicable. Thus, investigation of the transient properties of production systems is necessary and is carried out in this work.

### 1.3 Approach of This Research

To analyze the system-theoretic properties of production lines mentioned above, we use the following approaches:

- For systems that are characterized by Markov processes, we derive the equations, which describe the flow of parts through the system. Then, system-theoretic properties are analyzed using these equations.

- For non-Markovian systems, hypotheses are formulated guided by the results obtained for the Markovian ones. Then, numerical investigations are carried out to justify these hypotheses.
Finally, based on the results of the Markovian and non-Markovian systems, conjectures are made for general cases.

1.4 Related Literature

1.4.1 Serial lines

In comparison with numerous publications on performance analysis of production lines, rigorous investigations of system-theoretic properties exist only in a small body of literature. It has been proved in [11]-[14] that serial lines possess the reversibility property. Monotonicity of serial lines was investigated [14]-[16]. It has been proved that the throughput of a serial line is monotonically increasing in machine efficiency and buffer capacity. The effects of up- and downtime on system performance are discussed in [4].

The concept of *improvability* of production systems was first introduced in [14]. Detailed study of improvability under different resource constraints has been carried out in [17]. Specifically, it studied the improvability properties with respect to workforce and buffer capacity and developed indicators for continuous improvement procedures.

Bottlenecks identification and elimination have been a central topic in control and improvement of production systems and several notions of bottleneck have been proposed in the literature, for instance, [18], [19]. Rigorous study of bottleneck identification in production lines was initiated in [20], which developed an effective arrow-based method to identify the bottleneck in Bernoulli serial lines using the probabilities of machine blockages and starvations. This method is then extended to serial lines with exponential machines in [21]-[23].

As it was mentioned in Subsection 1.2.6, steady state analysis has been the focus of most studies on manufacturing systems. However, properties of transients in production systems have not been investigated systematically. The importance of
transient properties of production systems was first enunciated in [24], which pointed out several problems in manufacturing that require transient analysis. Inspired by [24], [25] analyzed the transient behavior of a flexible machine center that processes several classes of jobs with significant setup times and priority scheduling. In addition, there is a vast body of literature concerned with fluid models of stochastic systems (see, for example, [26]-[30]). While [26]-[28] address communication networks, [29] and [30] use production systems terminology and consider the transient behavior of a buffer with random input and output flows. The latest of these publications, [30], develops an algorithm for a numerical solution of a partial differential equation, which describes the evolution of the probability density function for a buffer with Markov modulated input and output flows. Despite the results obtained, no specific properties of transients in throughput and work-in-process have been analyzed.

1.4.2 Closed lines

Studies on closed production lines have been carried out in the literature for almost five decades (see a review article [31]) and a number of important results have been obtained. Specifically, methods for evaluating closed line throughput have been developed in [32]-[34]. Reversibility property of closed line was established in [35] and various system-theoretic properties of closed line throughput have been investigated in [36]-[40]. It was discovered that for closed queueing systems with infinite buffers the system throughput is monotonically increasing as a function of total number of jobs in the system [36]-[38], while for closed systems with finite buffers, the system throughput is non-monotonic and concave with respect to total job population [39], [40]. The equivalence of closed and appropriately defined open two-machine lines has been established in [41] and [42]. Important results concerning the selection of the number of carriers that maximizes the throughput have been reported in [43]. Performance evaluation and control of multi-loop production systems are discussed
Finally, references [16] and [45] provide results on buffer capacity allocation in closed lines.

### 1.4.3 Product quality

Product quality is a critical problem in manufacturing systems. In production lines with quality issues, inspection devices are often included to identify and remove defective parts. In some cases, a defective product is sent to repair and then returns to an appropriate operation for subsequent re-processing. For such variations of production lines, the knowledge of the conventional serial lines become inapplicable.

The existing publications on quantitative analysis of product quality inspection in manufacturing systems can be classified into two groups. The first one is concerned with using optimization techniques for selecting and positioning inspection devices so that cost-related performance measures are optimized. For these purposes, linear, nonlinear, dynamic, and stochastic programming techniques have been used along with Pontryagin’s maximum principle, genetic algorithms and simulated annealing. One of the first publications on this topic is [46]. Examples of recent articles include [47]-[52]. Two comprehensive reviews of this literature can be found in [53] and [54]. In addition, the relationship of quality with time, cost, and flexibility in manufacturing systems is discussed in [55] and economic optimization of off-line product inspection is investigated in [56], [57], and [58].

The second group is concerned with interaction between quality and quantity issues in production systems with unreliable machines and finite buffers. The main problem here is analysis of the production rate for systems, which include non-perfect machines and inspections devices. Perhaps, the first publication on quality/quantity interaction in serial production lines is [59], where a simple system consisting of two asymptotically reliable machines and a quality control device has been investigated. A Markov chain method for performance analysis has been used, and the issue of in-
spection device placement has been analyzed. This direction of research continued in [60] and [61]. Specifically, based on Taylor series expansions, [60] extended the performance evaluation results to serial lines with finite number of asymptotically reliable machines obeying the geometric reliability model. Paper [61] considered long serial lines with Bernoulli machines and inspection devices and developed an aggregation procedure for their performance evaluation. In addition, production lines with rework have been studied in a few recent publications [62]-[67]. Mostly performance analysis problems have been addressed. Specifically, analytical techniques for evaluating their production rates and probabilities of machine blockages and starvations have been derived under the assumption that the machines obey the exponential reliability model.

Another direction of research on the interplay between the quality and the structure of a production system was initiated in [68], where it was argued that the structure of a production system affects the quality of parts produced. In addition, [68] included extensive discussions of various quality/quantity issues and cited a comprehensive list of publication on this subject. This paper created a substantial following, and in the last few years a number of related articles have been published. Namely, [69] used the Markov chain approach to developed a method for performance evaluation of production systems operating under the Statistical Process Control; [70] proposed a new model for machines with both quality and operational failures and developed a Markov process-based method for performance analysis of production systems consisting of such machines; [71] explored quality/quantity issues in Flexible Manufacturing Systems; serial lines with quality-quantity coupling (QQC) machines have been addressed in [72], where a model of QQC has been proposed and the trade-off between operating speed and throughput in several basic production systems has been analyzed; finally, a taxonomy of quality/quantity issues is described in [73].
1.4.4 Summary of the related literature and topics of this research

As pointed out above, reversibility, monotonicity, and bottleneck properties of serial lines have been discussed in the literature for a number of formulations. Improvability properties were discussed only in the case of Bernoulli serial lines. In contrast, transient properties of serial lines, as well as effects of up- and downtime and improvability of serial lines with continuous time model of machine reliability still remain practically unexplored. Similar to serial lines, reversibility and monotonicity in closed lines have been analyzed in several publications. The improvability property and bottlenecks of closed lines, however, have not been investigated. For production lines with quality issues, system-theoretic properties have not yet been discussed in the literature.

Thus, to contribute to the understanding of system-theoretic properties of production lines, this work addresses the following topics:

- Reversibility, monotonicity, effects of up- and downtime, and improvability in serial lines with continuous model of machine reliability;
- Reversibility, monotonicity, improvability, and bottlenecks properties of closed lines;
- Reversibility, monotonicity, and bottleneck properties of production lines with quality issues;
- Transient properties of serial lines.

1.5 Outline

The outline of this dissertation is as follows: Chapter II introduces the system models considered and formulates the problems. System-theoretic properties of serial open lines and closed lines are investigated in Chapters III and IV, respectively. Production
lines with quality issues are discussed in Chapter V. Transient properties of serial lines are studied in Chapter VI. Finally, the conclusions and topics for future research are provided in Chapter VII. All proofs and numerical justifications are given in the appendices.
CHAPTER II

SYSTEM MODELS AND PROBLEMS

2.1 Introduction

As mentioned in Chapter I, the purpose of this research is to study system-theoretic properties of production lines with unreliable machine and finite buffers. Since there are various notions and conventions on production system used in the literature (see review paper [9]), to avoid confusion and to formalize the presentation, this chapter is devoted to define a set of standard vocabulary used throughout this work.

2.2 Terminology

**Cycle time** $\tau$: the time necessary to process a part by a machine. The cycle time may be constant, variable, or random. In large volume production systems, $\tau$, is practically always constant or close to being constant. This is the case in most production systems of the automotive, electronics, appliance, and other industries. Variable or random cycle time takes place in job-shop environments where each part may have different processing specifications. In this work, we consider only machines with a constant cycle time; similar developments, however, can be carried out for the case of random (e.g., exponentially distributed) processing time.

**Machine capacity** $c$: the number of parts produced by a machine per unit of time when the machine is up. Clearly, in the case of constant $\tau$, 
\[ c = \frac{1}{\tau}. \]

Machines in a production system may have identical or different cycle times. In the case of identical cycle time, the time axis may be considered as slotted or unslotted.

**Slotted time:** the time axis is slotted with the slot duration equal to the cycle time. In this case, all transitions – changes of machines’ status (up or down) and changes of buffers’ occupancy – are considered as taking place only at the beginning or the end of the time slots.

**Unslotted time or continuous time:** changes of machines’ status (up or down) and changes of buffers’ occupancy may occur at any time moment. If the cycle times of all machines are identical, such a system is referred to as synchronous. If the cycle times are not identical, the system is called asynchronous. Production systems with machines having different cycle times are typically considered as operating in unslotted time.

**Machine reliability model:** the probability mass functions (pmf’s) or the probability density functions (pdf’s) of the up- and downtime of the machine in the slotted or unslotted time, respectively. In addition, the expected value and coefficient of variation of up- and downtime are denoted as \( T_{\text{up}}, T_{\text{down}}, CV_{\text{up}} \) and \( CV_{\text{down}} \), respectively. In the next section, the machine reliability models used in this research are described.

### 2.3 Machine Reliability Models

In this work, the following machine reliability models are used:
2.3.1 Slotted time case

**Bernoulli reliability model (B):** Assume the status of a machine can be modeled using a Bernoulli random variable

\[ P[\{\text{machine is up during a time slot}\}] = p, \]
\[ P[\{\text{machine is down during a time slot}\}] = 1 - p. \]

Then, we refer to this machine as a *Bernoulli machine*, i.e., obeying the *Bernoulli reliability model*. Clearly, the status of a Bernoulli machine during a time slot is independent of all other time slots. In addition, parameter \( p \) is the efficiency of a Bernoulli machine.

This reliability model is simple but practical. Indeed, it is applicable to operations where the unscheduled downtime is, on the average, comparable to the machine cycle time. This often happens in automotive painting and assembly operations, where the downtime is primarily due to quality problems rather than machine breakdowns.

**Geometric reliability model (geo):** Assume that the status of a machine during each time slot depends on its status in the previous time slot as shown in the transition diagram of Figure 2.1.

![Figure 2.1: Geometric reliability model](image)

It can be shown that the up- and downtime of this machine, denoted as \( t_{up} \) and \( t_{down} \) are characterized by the following distributions:

\[ P[t_{up} = t] = P(1 - P)^{t-1}, \quad t = 1, 2, ... \]
Clearly, \( t_{up} \) and \( t_{down} \) are geometric random variables and we refer to such a machine as a geometric machine, i.e., obeying the geometric reliability model. In addition, it is easy to show that for a geometric machine

\[
P[t_{down} = t] = R(1-R)^{t-1}, \quad t = 1, 2, \ldots
\]

2.3.2 Continuous time case

Exponential reliability model (\( \text{exp} \)): Consider a machine in Figure 2.2, which is a continuous time analogue of the geometric machine. Namely, if it is up (respectively, down) at time \( t \), it goes down (respectively, up) during an infinitesimal time \( \delta t \) with probability \( \lambda \delta t \) (respectively, \( \mu \delta t \)). The parameters \( \lambda \) and \( \mu \) are called the breakdown and repair rates, respectively.

It can be shown that the pdf’s of the up- and downtime of this machine, denoted as \( t_{up} \) and \( t_{down} \), are as follows:

\[
f_{t_{up}}(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \\
f_{t_{down}}(t) = \mu e^{-\mu t}, \quad t \geq 0.
\]

Clearly, \( t_{up} \) and \( t_{down} \) are exponential random variables and we refer to such a machine as an exponential machine, i.e., obeying the exponential reliability model. In addition,
it is easy to show that for an exponential machine

\[ T_{up} = \frac{1}{\lambda}, \quad T_{down} = \frac{1}{\mu}, \]
\[ CV_{up} = 1, \quad CV_{down} = 1, \]
\[ e = \frac{\mu}{\lambda + \mu}. \]

**Weibull reliability model (W):** Weibull distribution is widely used in Reliability Theory. For a machine obeying Weibull reliability model, its up- and downtime pdf’s are given by

\[ f_{t_{up}}(t) = \lambda^{\Lambda}e^{-(\lambda t)\Lambda t^{\Lambda-1}}, \quad t \geq 0, \]
\[ f_{t_{down}}(t) = \mu^{M}e^{-(\mu t)M t^{M-1}}, \quad t \geq 0. \]

where \( \Lambda \) and \( M \) are positive numbers. It can be calculated that for a Weibull machine

\[ T_{up} = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\Lambda}\right), \quad T_{down} = \frac{1}{\mu} \Gamma\left(1 + \frac{1}{M}\right), \]
\[ CV_{up} = \frac{\sqrt{\Gamma(1 + \frac{2}{\Lambda}) - \Gamma^2(1 + \frac{1}{\Lambda})}}{\sqrt{\Gamma(1 + \frac{1}{\Lambda})}}, \quad CV_{down} = \frac{\sqrt{\Gamma(1 + \frac{2}{M}) - \Gamma^2(1 + \frac{1}{M})}}{\sqrt{\Gamma(1 + \frac{1}{M})}}. \]

**Gamma reliability model (ga):** For a machine obeying the gamma reliability model, its up- and downtime pdf’s are given by gamma distribution, i.e.,

\[ f_{t_{up}}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{\Lambda-1}}{\Gamma(\Lambda)}, \quad t \geq 0, \]
\[ f_{t_{down}}(t) = \mu e^{-\mu t} \frac{(\mu t)^{M-1}}{\Gamma(M)}, \quad t \geq 0, \]

where

\[ \Gamma(x) = \int_0^\infty s^{x-1}e^{-s}ds, \]
and $\Lambda$, $M$ are positive numbers. In addition, it can be calculated that for a gamma machine

$$T_{up} = \frac{\Lambda}{\lambda}, \quad T_{down} = \frac{M}{\mu},$$

$$CV_{up} = \frac{1}{\sqrt{\Lambda}}, \quad CV_{down} = \frac{1}{\sqrt{M}}.$$  

**Log-normal reliability model (LN):** For a machine obeying the log-normal reliability model, its up- and downtime pdf’s are given by

$$f_{t_{up}}(t) = \frac{1}{\sqrt{2\pi\Lambda t}} e^{-\frac{(\ln t - \lambda)^2}{2\Lambda^2}}, \quad t \geq 0,$$

$$f_{t_{down}}(t) = \frac{1}{\sqrt{2\pi M t}} e^{-\frac{(\ln t - \mu)^2}{2M^2}}, \quad t \geq 0,$$

where $\Lambda$ and $M$ are positive numbers. In addition, it can be calculated that for a log-normal machine

$$T_{up} = e^{\lambda + \frac{\Lambda^2}{\tau}}, \quad T_{down} = e^{\mu + \frac{M^2}{\tau}},$$

$$CV_{up} = \sqrt{e^{\Lambda^2} - 1}, \quad CV_{down} = \sqrt{e^{M^2} - 1}.$$  

2.4 Systems Considered

2.4.1 Slotted time case

**Serial lines:** In this dissertation, serial production lines, as shown in Figure 1.1(a), are defined by the following assumptions:

(a) The production line consists of $M$ machines, $m_1, \ldots, m_M$, and $M - 1$ buffers, $b_1, \ldots, b_{M-1}$.

(b) All machines have identical cycle time $\tau$. The time axis is slotted with the slot duration $\tau$. The status of the machines is determined at the beginning of each time slot according to their reliability models.
(c) Each buffer $b_i$, $i = 1, \ldots, M - 1$, is characterized by its capacity, $N_i$, where $1 \leq N_i < \infty$. The state of the buffer (i.e., the number of parts in it) is determined at the end of each time slot.

(d) Machine $m_i$, $i = 2, \ldots, M$, is starved during a time slot if it is up and buffer $b_{i-1}$ is empty at the beginning of the time slot. Machine $m_1$ is never starved.

(e) Machine $m_i$, $i = 1, \ldots, M - 1$, is blocked during a time slot if it is up, buffer $b_i$ has $N_i$ parts at the beginning of the time slot and machine $m_{i+1}$ fails to take a part during this time slot. Machine $m_M$ is never blocked.

**Closed lines:** Closed lines, as shown in Figure 2.3, are widely used in large volume manufacturing environment where parts are transported within the system on carriers. In addition to the above assumptions (a)-(e) for serial lines, closed lines operate under the following extra assumptions:

![Figure 2.3: Closed production line](image)

(f) The parts are transported within the system on carriers. The total number of carriers is $S$ and the capacity of empty carrier buffer $b_0$ is $1 \leq N_0 < \infty$.

(g) The parts are placed on carriers at the input of machine $m_1$. It is assumed that the parts are always available so that $m_1$ is not starved for parts but can be starved for carriers (when $b_0$ is empty).

(h) The parts are removed from carriers at the output of machine $m_M$. It is assumed that $m_M$ is not blocked by a subsequent operation but can be blocked by carriers
(when \( b_0 \) is full and \( m_1 \) is either down or blocked).

**Production lines with quality issues:** To investigate fundamental properties of production lines with quality issues, we study Bernoulli serial lines with machines of the following types:

- *perfect quality machines*, i.e., producing parts with no defects;

- *non-perfect quality machines* obeying the *Bernoulli quality model*, i.e., producing a good (non-defective) part with probability \( g \) and a defective with probability \( 1 - g \);

- *non-perfect quality machines with quality-quantity coupling* (QQC), i.e., having their quality parameters \( g \) as a monotonically decreasing function of their efficiency, \( p \);

- *inspection machines*, i.e., inspecting parts quality, perhaps along with carrying out some technological operations.

Such a production line is shown in Figure 2.4, where the perfect, non-perfect, and QQC machines are represented by white, shaded, and double-shaded circles, respectively, while the inspection machines are black ones; the arrows under the inspection machines indicate scrapping defective parts.

![Figure 2.4: Serial production line with perfect quality, non-perfect quality, QQC, and inspection machines](image)

In some production systems, it is more economical to have defective parts repaired and reworked, rather than scrapped. In these cases, the system has a structure illustrated in Figure 2.5. Here, machines \( m_1, \ldots, m_M \) constitute the main line,
$m_{r1}, \ldots, m_{rM_r}$ represent the repair line, and $m_k$ and $m_j$ are the split and merge machines, respectively; clearly, the split machine carries out the inspection operation. The probability that a good part emerges at the output of $m_k$, denoted as $q$, is referred to as the quality buy rate.

In Chapter V, we provide detailed descriptions of these systems and analyze their system-theoretic properties.

### 2.4.2 Continuous time case

**Serial lines:** In continuous time case, serial production lines operate according to the following assumptions:

(a) The system consists of $M$ machines $m_i$, $i = 1, \ldots, M$, and $M - 1$ buffers, $b_i$, $i = 1, \ldots, M - 1$.

(b) Each machine $m_i$, $i = 1, \ldots, M$, has two states: up and down. When up, the machine is capable of producing with rate $c_i$ (parts/unit of time); when down, no production takes place.

(c) The up- and downtime of each machine are continuous random variables, $t_{\text{up},i}$ and $t_{\text{down},i}$, $i = 1, \ldots, M$, and are determined by its reliability model. It is assumed that these random variables are mutually independent.

(d) Each in-process buffer $b_i$, $i = 1, \ldots, M - 1$, is characterized by its capacity, $0 < N_i < \infty$. 

![Figure 2.5: Serial production line with rework](image)
(e) Machine $m_i, i = 2, \ldots, M$, is starved at time $t$ if it is up at time $t$ and buffer $b_{i-1}$ is empty at time $t$.

(f) Machine $m_i, i = 1, \ldots, M - 1$, is blocked at time $t$ if it is up at time $t$, buffer $b_i$ is full at time $t$ and machine $m_{i+1}$ fails to take any work from this buffer at time $t$.

Closed lines: In continuous time case, closed production lines operate according to assumptions (a)-(f) of Subsection 2.4.2 and the following additional assumptions:

(g) The parts are transported within the system on carriers. The total number of carriers is $S$ and the capacity of empty carrier buffer $b_0$ is $0 < N_0 < \infty$.

(h) The parts are placed on carriers at the input of machine $m_1$. It is assumed that the parts are always available so that $m_1$ is not starved for parts but can be starved for carriers (when $b_0$ is empty).

(i) The parts are removed from carriers at the output of machine $m_M$. It is assumed that $m_M$ is not blocked by a subsequent operation but can be blocked by carriers (when $b_0$ is full and $m_1$ is either down or blocked).

2.5 Performance Measures

In this work, the following performance measures are considered:

Production rate ($PR$): average number of parts produced by the last machine of a production system per cycle time in the steady state of system operation.

This metric is appropriate for production systems with all machines having identical cycle times. In the asynchronous case, this metric is referred to as
Throughput \((TP)\): average number of parts produced by the last machine of a productions system per unit of time in the steady state of system operation. Clearly, \(TP\) can be used in the synchronous case as well; in this case,

\[
TP = c \cdot PR.
\]

Work-in-process of buffer \(i\) \((WIP_i)\): average number of parts contained in the buffer \(i\) of a production system in the steady state of its operation.

Blockage of machine \(i\) \((BL_i)\): steady state probability that machine \(i\) is up, buffer \(i\) is full, and machine \(i+1\) does not take a part from the buffer.

Starvation of machine \(i\) \((ST_i)\): steady state probability that machine \(i\) is up and buffer \(i-1\) is empty.

For production lines with quality issues, \(PR\) and \(TP\) refer to the production rate and throughput of good (i.e., non-defective) parts, respectively. In this case, two additional performance measures are considered:

Scrap rate \((SR_i)\): average number of defective parts scrapped by inspection machine \(m_i\), per cycle time;

Consumption rate \((CR)\): average number of raw parts consumed by the first machine per cycle time.

2.6 Problems Addressed

In the framework of the above models, this dissertation considers the following problems:
**Reversibility problem:** Given a production line, determine if the throughput remains the same when the parts flow is reversed and investigate the relationships of the performance measures between the original and reverse lines.

**Monotonicity problem:** Consider the throughput of a production line as a function of machine, buffer, and other system parameters, determine the monotonic property of this function with respect to each parameter.

**Effects of up- and downtime problem:** 1) Fixing the efficiency of a machine determine if machines with shorter up- and downtime or longer ones are desirable. 2) Determine if increasing the uptime of a machine in a production line can result in better performance than decreasing the downtime of this machine by the same factor.

**Improvability problem:** Consider a production line constrained by certain resource, determine criteria for re-allocating the resource so that the throughput is maximized.

**Bottleneck problem:** Given a production line, determine the BN of the system, i.e., the machine and/or buffer that affect the throughput in the strongest manner.

**Transient problem:** Given a production line, investigate its transient behavior, i.e., how fast it reaches its steady state.
CHAPTER III

SYSTEM-THEORETIC PROPERTIES OF SERIAL LINES

3.1 Review of Existing Results on System-Theoretic Properties

3.1.1 Serial Bernoulli lines

Numerous results on system-theoretic properties have been obtained for serial lines defined by assumptions (a)-(e) with Bernoulli machines. In this subsection, we review the reversibility, monotonicity, and improvability properties investigated in [14] and bottleneck properties studied in [20].

Reversibility: Consider a Bernoulli line $L$, defined by assumptions (a)-(e) of Subsection 2.4.1, and its reverse $L_r$ (see Figure 3.1).

![Diagram of M-machine Bernoulli line and its reverse](image.png)

Figure 3.1: $M$-machine Bernoulli line and its reverse
Theorem 3.1 [14] The performance measures of a serial Bernoulli line, \( L \), and its reverse, \( L_r \), are related as follows:

\[
PR^L = PR^{L_r},
\]
\[
BL_i^L = ST_{(M-i+1)r}^{L_r}, \quad i = 1, \ldots, M - 1.
\]

Monotonicity:

Theorem 3.2 [14] The production rate of a serial Bernoulli line, \( PR \), is

- strictly monotonically increasing in \( N_i \), \( i = 1, \ldots, M - 1 \);
- strictly monotonically increasing in \( p_i \), \( i = 1, \ldots, M \).

Thus, the production rate of a serial Bernoulli line can be improved by increasing machine efficiency and buffer capacity.

Improvability: Consider a serial Bernoulli line defined by assumptions (a)-(e) of Subsection 2.4.1. Assume that \( N_i \)’s and \( p_i \)’s are constrained as follows:

\[
\sum_{i=1}^{M-1} N_i = N^*, \tag{3.1}
\]
\[
\prod_{i=1}^{M} p_i = p^*, \tag{3.2}
\]

where \( N^* \) and \( p^* \) are positive numbers with \( p^* \) satisfying \( p^* < 1 \). Constraint (3.1) implies that the total buffer capacity cannot exceed \( N^* \). Constraint (3.2) can be interpreted as a bound on the machine efficiency or workforce. Indeed, in many systems, assignment of the workforce (both machine operators and skilled trades for repair and maintenance) defines the machine efficiency and, thus, \( p_i \)’s. Therefore, we
refer to (3.1) and (3.2) as the buffer capacity (BC) and workforce (WF) constraints, respectively.

Let $PR = PR(p_1, \ldots, p_M, N_1, \ldots, N_{M-1})$ denote the production rate of the system at hand.

**Definition 3.1** A serial production line with Bernoulli machines is:

- **improvable with respect to BC** if there exists a sequence $N'_1, \ldots, N'_{M-1}$ such that

$$\sum_{i=1}^{M-1} N'_i = N^*\quad (3.3)$$

and

$$PR(p_1, \ldots, p_M, N'_1, \ldots, N'_{M-1}) > PR(p_1, \ldots, p_M, N_1, \ldots, N_{M-1});\quad (3.4)$$

otherwise, it is unimprovable with respect to BC;

- **improvable with respect to WF** if there exists a sequence $p'_1, \ldots, p'_M$ such that

$$\prod_{i=1}^{M} p'_i = p^*\quad (3.5)$$

and

$$PR(p'_1, \ldots, p'_M, N_1, \ldots, N_{M-1}) > PR(p_1, \ldots, p_M, N_1, \ldots, N_{M-1});\quad (3.6)$$

otherwise, it is unimprovable with respect to WF;

- **improvable with respect to BC and WF simultaneously** if there exist sequences $N'_1, \ldots, N'_{M-1}$ and $p'_1, \ldots, p'_M$ such that

$$\sum_{i=1}^{M-1} N'_i = N^*,\quad \prod_{i=1}^{M} p'_i = p^*\quad (3.7)$$
and

\[
PR(p_1', \ldots, p_M', N_1', \ldots, N_{M-1}') > PR(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}); \quad (3.8)
\]

otherwise, it is unimprovable with respect to BC and WF simultaneously.

Criteria for improvability in terms of each of these definitions have been derived in [14] and are summarized below.

**WF-Improvability Indicator 3.1:** A Bernoulli line is practically unimprovable with respect to WF if each buffer is, on the average, close to being half full.

**WF and BC-Improvability Indicator 3.1:** A Bernoulli line is unimprovable with respect to WF and BC simultaneously if all internal machines are starved and blocked with equal frequency, the first machine is blocked as frequently as the last machine is starved, all buffers are of equal capacity and have equal steady state occupancy, close to being half full.

**BC-Improvability Indicator 3.1:** A Bernoulli line is practically unimprovable with respect to BC if the average occupancy of each buffer is as close to the average availability of its downstream buffer as possible.

**Bottlenecks:** The bottlenecks of a serial Bernoulli line are defined as follows:

**Definition 3.2** Machine \( m_i, i \in \{1, \ldots, M\}, \) is the bottleneck machine (BN-m) of a Bernoulli line if

\[
\frac{\partial PR}{\partial p_i} > \frac{\partial PR}{\partial p_j}, \quad \forall j \neq i. \quad (3.9)
\]

**Definition 3.3** Buffer \( b_i, i \in \{1, \ldots, M - 1\}, \) is the bottleneck buffer (BN-b) of
Thus, Definitions 3.2 and 3.3 imply that $m_i$ is the BN-m if its infinitesimal improvement leads to the largest increase of the production rate, as compared with a similar improvement of any other machine in the system and BN-b is the buffer, which leads to the largest increase of the $PR$ if its capacity is increased by 1, as compared with increasing any other buffer in the system. Moreover, a machine with the smallest $p_i$ is not necessarily the BN-m and a buffer with the smallest capacity is not necessarily the BN-b.

While Definitions 3.2 and 3.3 provide formal characterizations of the BN-m and BN-b, they cannot be applied directly because the partial derivatives involved could be neither measured on the factory floor nor calculated analytically with any acceptable accuracy. To make these definitions practical, [20] reformulated them in terms of quantities, which are either available through measurements on the factory floor or through analytical calculations or both:

**Theorem 3.3** [20] For two-machine Bernoulli lines, the inequality

$$\frac{\partial PR}{\partial p_1} > \frac{\partial PR}{\partial p_2} \quad \text{(respectively, } \frac{\partial PR}{\partial p_1} < \frac{\partial PR}{\partial p_2} \text{)}$$

(3.11)

takes place if and only if

$$BL_1 < ST_2 \quad \text{(respectively, } BL_1 > ST_2).$$

This result relates the “non-measurable” and “non-calculable” partial derivatives of $PR$ with the “measurable” and “calculable” probabilities of blockages and starva-
tions. In addition, it states that the BN-m can be identified without even knowing parameters of the machines and buffer, but just by measuring $ST_2$ and $BL_1$.

Inspired by this theorem, an arrow-based method has been developed to identify the BN in longer lines: arrange the probabilities of starvations ($ST_i$) and blockages ($BL_i$) under each machine as shown in Figure 3.2 and place arrows directed from one machine to another according to the following rule

$$ST_i \quad 0 \quad x \quad x \quad x$$

$$BL_i \quad x \quad x \quad x \quad 0$$

Figure 3.2: BN identification in $M$-machine lines

**Arrow Assignment Rule 3.1:** If $BL_i > ST_{i+1}$, assign the arrow pointing form $m_i$ to $m_{i+1}$. If $BL_i < ST_{i+1}$, assign the arrow pointing from $m_{i+1}$ to $m_i$.

**Bottleneck Indicator 3.1:** In a Bernoulli line with $M > 2$ machines,

- if there is a single machine with no emanating arrows, it is the BN-m;
- if there are multiple machines with no emanating arrows, the one with the largest severity is the Primary BN-m (PBN-m), where the severity of each (local) BN-m is defined by

$$S_i = |ST_{i+1} - BL_i| + |ST_i - BL_{i-1}|, \quad i = 2, \ldots, M - 1,$$

$$S_1 = |ST_2 - BL_1|, \quad S_M = |ST_M - BL_{M-1}|; \quad (3.12)$$

- the BN-b is the buffer immediately upstream of the BN-m (or PBN-m) if it is more often starved than blocked, or immediately downstream of the BN-m (or PBN-m) if it is more often blocked than starved.
3.1.2 Serial exponential lines

Bottleneck identification in serial lines defined by assumptions (a)-(f) of Subsection 2.4.2 with exponential machines has been studied in [21] and [23]. Here, we review the results on the most practical and important one – c-BN:

**Definition 3.4** Machine $m_i$, $i \in \{1, \ldots, M\}$, is the c-bottleneck (c-BN) of a serial line defined by assumptions (a)-(f) of Subsection 2.4.2 if

$$\frac{\partial TP}{\partial c_i} > \frac{\partial TP}{\partial c_j}, \quad \forall j \neq i. \quad (3.13)$$

Thus, $m_i$ is the c-BN if increasing its capacity leads to the largest increase of $TP$, compared to increasing the capacity of any other machine in the system.

Bottleneck buffers in serial lines with machines having continuous models of reliability are defined similarly to those in the Bernoulli case:

**Definition 3.5** Buffer $b_i$, $i \in \{1, \ldots, M - 1\}$, is the bottleneck buffer (BN-b) of a serial line if

$$TP(N_1, \ldots, N_i + 1, \ldots, N_{M-1}) > TP(N_1, \ldots, N_j + 1, \ldots, N_{M-1}), \quad \forall j \neq i. \quad (3.14)$$

It can be shown that machines with the smallest $T_{up}$ or largest $T_{down}$ are not necessarily the bottlenecks (in the sense of Definition 3.4). Also, machines with the smallest efficiency,

$$e_i = \frac{T_{up,i}}{T_{up,i} + T_{down,i}},$$

or the smallest capacity, $c_i$, or the smallest throughput in isolation, $TP_i = c_ie_i$, are not necessarily the bottlenecks. In addition, the buffer with the smallest capacity is not necessarily BN-b.

To identify the c-BN in serial continuous lines, Bottleneck Indicator 3.1 was ex-
tended to exponential lines in [23]. Specifically, assign arrows pointing from one machine to another using Arrow Assignment Rule 3.1. Then,

**c-BN Indicator 3.1:** In serial lines with exponential machines,

- if there is a single machine with no emanating arrows, it is the c-BN of the system;
- if there are multiple machines with no emanating arrows, the primary c-BN (i.e., $P_c$-BN) is the one with the largest severity;
- the BN-$b$ is one of the buffers surrounding $P_c$-BN; specifically, if $P_c$-BN is machine $m_i$, then BN-$b$ is $b_i$ if $BL_i > ST_i$ or $b_{i-1}$ if $BL_i < ST_i$.

Next, we present new results on system-theoretic properties of serial production lines.

### 3.2 Novel Facts on System-Theoretic Properties of Serial Lines

#### 3.2.1 Reversibility

It has been shown in Subsection 3.1.1 that Bernoulli lines possess the property of reversibility. For serial lines with exponential machines, the following can be proved:

**Theorem 3.4** The performance measures of a serial exponential line, $L$, and its reverse, $L_r$, are related as follows:

\[
TP^L = TP^{L_r},
\]

\[
WIP^L_i = N_i - WIP^{L_r}_{(M-i)r}, \quad i = 1, \ldots, M - 1,
\]

\[
BL^L_i = ST^{L_r}_{(M-i+1)r}, \quad i = 1, \ldots, M - 1.
\]

**Proof:** See Appendix A.
Using simulation, we show that Theorem 3.4 holds for serial lines with non-exponential machines as well:

**Numerical Fact 3.1** The performance measures of a serial line, $L$, and its reverse, $L_r$, with machines having their reliability models in the set $\{\text{exp, W, ga, LN}\}$, are related as follows:

\[
TP_L = TP_{L_r}, \quad (3.15)
\]

\[
WIP_{L_r}^{i_r} = N_{M-i} - WIP_{M-i}^L, \quad i = 1, \ldots, M-1, \quad (3.16)
\]

\[
BL_i^L = ST_{(M-i+1)r}, \quad i = 1, \ldots, M-1. \quad (3.17)
\]

**Justification:** See Appendix A.

Therefore, serial lines with continuous time models of machine reliability are also reversible. To extend this property to general lines, we formulate:

**Conjecture 3.1** Numerical Fact 3.1 holds for serial lines with any unimodal distribution of up- and downtime.

### 3.2.2 Monotonicity

**Numerical Fact 3.2** For serial lines defined by assumptions (a)-(f) of Subsection 2.4.2 with machines having their reliability models in the set $\{\text{exp, W, ga, LN}\}$, $TP$ is

- strictly monotonically increasing in $T_{up,i}$, $c_i$, $i = 1, \ldots, M$, and $N_i$, $i = 1, \ldots, M-1$;

- strictly monotonically decreasing in $T_{down,i}$, $CV_{up,i}$ and $CV_{down,i}$, $i = 1, \ldots, M$. 

32
Justification: See Appendix A.

Thus, the throughput of a serial line defined by assumptions (a)-(f) of Subsection 2.4.2 can be improved by increasing machine capacity, uptime, and buffer capacity. In addition, decreasing machine downtime or decreasing the coefficients of variation of machine up- and downtime also lead to improved throughput of the system.

For the general case, we formulate the following conjecture:

Conjecture 3.2 Numerical Fact 3.2 holds for serial lines with any unimodal distribution of up- and downtime.

3.2.3 Effects of up- and downtime

Consider two synchronous exponential two-machine lines, $l_1$ and $l_2$, with

$$e_{1}^{l_1} = e_{2}^{l_2}, \quad e_{1}^{l_2} = e_{2}^{l_2}, \quad N_{l_1}^{i} = N_{l_2}^{i},$$

(3.18)

i.e., with machines of identical efficiency and identical buffer capacity. Assume also that

$$T_{down,i}^{l_1} < T_{down,i}^{l_2}, \quad i = 1, 2,$$

(3.19)

i.e., the machines in $l_1$ have shorter up- and downtime than the machines in $l_2$.

Theorem 3.5 In synchronous exponential two-machine lines, under assumptions (3.18), (3.19),

$$PR_{l_1}^{i} > PR_{l_2}^{i}.$$

Proof: See Appendix A.

Thus, shorter up- and downtime lead to a higher throughput than longer ones, even if the machines’ efficiency remains the same. This phenomenon takes place
because finite buffers protect against shorter downtime better than against longer ones. Mathematically, this phenomenon is due to the fact that the probabilities of buffer being empty and full are larger for machines with longer up- and downtime.

Another characterization of the effects of up- and downtime can be given as follows:

Clearly, $PR$ can be improved by either increasing the uptime of a machine or decreasing its downtime. Is it more beneficial to increase the uptime, say by a factor $1 + \alpha$, $\alpha > 0$, or decrease its downtime by the same factor? When the production line consists of a single machine (in fact, with an arbitrary continuous time reliability model), the answer is obvious: the efficiency of the machine with either increased uptime or decreased downtime remains the same since in either case

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

Thus, for an isolated machine, increasing uptime or decreasing downtime by the same factor has the same effect. The situation is different in the case of more than one-machine systems. Specifically,

**Theorem 3.6** In synchronous exponential two-machine lines, $PR$ has a larger increase when the downtime of a machine is decreased by a factor $(1 + \alpha)$, $\alpha > 0$, than when the uptime is increased by the same factor.

**Proof:** See Appendix A.

Unfortunately, no analytical proof is available at present for Theorems 3.5 and 3.6 in $M > 2$-machine lines or in asynchronous lines. However, the following statements can be formulated based on simulations:

**Numerical Fact 3.3** For serial lines defined by assumptions (a)-(f) of Subsection 2.4.2 with machines having their reliability models in the set \{exp, W, ga, LN\},

(1) If the machines’ efficiency remains the same, then shorter up- and downtime lead to a higher throughput than longer ones.
The throughput has a larger increase when the downtime of a machine is decreased by a factor \((1 + \alpha)\), \(\alpha > 0\), than when the uptime is increased by the same factor.

**Justification:** See Appendix A.

For lines with general model of machine reliability, we formulate:

**Conjecture 3.3** Numerical Fact 3.3 holds for serial lines with any unimodal distribution of up- and downtime.

### 3.2.4 Improvability

For serial lines with continuous models of machine reliability, in addition to BC constraint (3.1), *cycle time constraint* is considered, which is defined as

\[
\sum_{i=1}^{M} \tau_i = \tau^*,
\]

where \(\tau^*\) is a given constant. This implies that the total processing time of a part by all machines in the system is \(\tau^*\). Improvability with respect to the cycle time would mean, then, that some of the machines could be sped-up and others slowed-down so that the performance of the system as a whole is improved, while the total processing time remains constant. In practice, such a re-allocation is often possible since the speed of most machining and assembly operations can be adjusted relatively easily, typically within \(\pm 10\%\) of their nominal values. In the case of synchronous lines, this would mean that they may be “de-synchronized”, if this leads to improved performance.

Based on the above, this subsection considers improvability with respect to buffer capacity (BC) and cycle time (CT) defined as follows:
Definition 3.6 A serial production line with a continuous time model of machine reliability is:

- improvable with respect to BC if there exists a sequence $N'_1, \ldots, N'_{M-1}$ such that

$$
\sum_{i=1}^{M-1} N'_i = N^*
$$

and

$$
TP(N'_1, \ldots, N'_{M-1}) > TP(N_1, \ldots, N_{M-1}); \quad (3.21)
$$

- improvable with respect to CT if there exists a sequence $\tau'_1, \ldots, \tau'_M$ such that

$$
\sum_{i=1}^{M} \tau'_i = \tau^*
$$

and

$$
TP(\tau'_1, \ldots, \tau'_M) > TP(\tau_1, \ldots, \tau_M). \quad (3.22)
$$

(For the sake of brevity, expressions (3.21) and (3.22) include only the variables that are being adjusted for improvability.)

Analytical proofs of the criteria for improvability of serial lines with continuous time models of machine reliability are all but impossible to derive (due to the complexity of the expressions for their $TP$). Therefore, motivated by similar results for Bernoulli lines described above, we introduce heuristic formulations of these criteria and then justify them by calculations and simulations.

Improvability with respect to CT:

CT-Improvability Indicator 3.1: A serial production line with machines having their reliability models in the set \{exp, W, ga, LN\} is unimprovable with respect to CT if each buffer is, on average, close to being half full.
Justification: See Appendix A.

As far as the general machine reliability model is concerned, the following conjecture is formulated:

**Conjecture 3.4** CT-Improvability Indicator 3.1 can be used for serial production lines with any unimodal distribution of up- and downtime.

**Improvability with respect to BC:**

**BC-Improvability Indicator 3.2:** A serial production line with machines having their reliability models in the set \{exp, W, ga, LN\} is unimprovable with respect to BC if the buffer capacity allocation is such that

\[
\max_i |WIP_{i-1} - (N_i - WIP_i)|
\]

is minimized over all sequences \(N_1, \ldots, N_{M-1}\) such that \(\sum_{i=1}^{M} N_i = N^*\).

Justification: See Appendix A.

To extend the above conclusion to the general machine reliability model, we formulate the following conjecture:

**Conjecture 3.5** BC-Improvability Indicator 3.2 can be used for serial production lines with any unimodal distribution of up- and downtime.

### 3.2.5 Bottlenecks

The \(c\)-BN indicator developed in [23] has been applied to serial lines with non-exponential machines. The results of this simulation study show that

**Numerical Fact 3.4** \(c\)-BN Indicator 3.1 can be used to identify \(c\)-BNs in serial lines with machines having their reliability models in the set \{exp, W, ga, LN\}.
Justification: See Appendix A.

Based on the above, we formulate:

**Conjecture 3.6** *c-BN Indicator 3.1 can be used for serial lines with machines having arbitrary distributions of up- and downtime.*

### 3.3 Summary

- Serial lines possess the reversibility property.

- The throughput of a serial line is a monotonic function of machine and buffer parameters.

- Shorter up- and downtime lead to a higher production rate (or throughput) than longer ones, even if machine efficiency remains constant.

- A decrease in downtime leads to higher throughput than a similar increase in uptime.

- Serial lines are unimprovable with respect to WF or CT re-allocation if each buffer is, on the average, close to being half full.

- Serial lines are unimprovable with respect to BC re-allocation if the average occupancy of each buffer is on the average close to the average availability of its immediate downstream buffer.

- The bottlenecks in a serial line can be identified by an arrow-based method using machine blockages and starvations.
CHAPTER IV

SYSTEM-THEORETIC PROPERTIES OF CLOSED LINES

4.1 Introduction

Production lines in large volume manufacturing environments often have parts transported from one operation to another on carriers (sometimes referred to as pallets, skids, etc.). Since in these situations the number of parts in the system is bounded by the number of available carriers, these lines are called closed with respect to carriers (or just closed).

Since in a closed line the first machine can be starved for carriers and the last blocked by $b_0$, the production rate of the closed line, $PR_{cl}$, is, at best, equal to that of the corresponding open line, $PR_o$. If, however, either $N_0$ or $S$ or both are chosen inappropriately, the closed nature of the line impedes system performance and, as a result, $PR_{cl}$ can be substantially lower than $PR_o$. An illustration is given in Figure 4.1, where $PR_{cl}$ is shown as a function of $S$ for various values of $N_0$; $PR_o$ is also indicated by the broken lines. An interpretation of these graphs is as follows: For the system of Figure 4.1(a), the empty carrier buffer capacity $N_0 = 2$ is too small, since $PR_{cl} < PR_o$ for any $S$. With $N_0 = 4$, there is a single value of $S$ (specifically, $S = 4$) that guarantees $PR_{cl} = PR_o$. When $N_0 = 6$, the equality $PR_{cl} = PR_o$ holds on the set $S = \{4, 5, 6\}$. Finally, when $N_0 = 10$, the set of “non-impeding” $S$’s becomes even larger ($S = \{4, 5, 6, 7, 8, 9, 10\}$). Clearly, the drop of $PR_{cl}$ for small and for large values of $S$ is due to starvation of $m_1$ for carriers and blockage of $m_2$ by $b_0$,
respectively, and, as Figure 4.1(a) shows, $PR_{cl}$ is practically (however, not exactly) symmetric in $S$. A similar interpretation can be given for Figure 4.1(b) as well.

![Graphs of production rate for two-machine and five-machine lines](image)

*Figure 4.1: Production rate of closed lines as a function of the number of carriers*

Given the above, a question arises: How should $N_0$ and $S$ be selected so that, on one hand, the closed nature of the line does not impede the open line performance and, on the other hand, $N_0$ and $S$ are sufficiently small so that the closed line is “lean?” In this chapter, we also investigate a unique property, outside the six basic ones addressed in Chapter I, of closed lines: unimpediment of $(N_0, S)$.

### 4.2 Closed Bernoulli Lines

In this section, we study closed lines defined by assumptions (a)-(h) of Subsection 2.4.1 with all machines obeying the Bernoulli reliability model.

#### 4.2.1 Reversibility

**Theorem 4.1** Closed Bernoulli two-machine lines defined by assumptions (a)-(h) of Subsection 2.4.1 are reversible in the sense that

$$PR_{cl}^L = PR_{cl}^{L_r},$$

where $L$ and $L_r$ are a closed line and its reverse, respectively.
Proof: See Appendix B.

For $M > 2$-machine lines, analytical investigation is all but impossible due to the complexity of the Markov chain involved. Therefore, simulations are used to show the following result:

**Numerical Fact 4.1** Closed Bernoulli lines defined by assumptions (a)-(h) of Subsection 2.4.1 with $M > 2$ machines are reversible in the sense that

$$PR_{cl}^L = PR_{cl}^{L_r},$$  \hspace{1cm} (4.1)

where $L$ and $L_r$ are a closed line and its reverse, respectively.

Justification: See Appendix B.

Thus, closed Bernoulli lines are indeed reversible.

### 4.2.2 Monotonicity

**Theorem 4.2** For a two-machine closed Bernoulli line defined by assumptions (a)-(h) of Subsection 2.4.1, function $PR_{cl}(p_1, p_2, N_1, N_0, S)$ is

- strictly increasing in $p_1$ and $p_2$;
- non-strictly increasing in $N_1$ and $N_0$;
- nonmonotonic concave in $S$.

Proof: See Appendix B.

**Numerical Fact 4.2** For a closed Bernoulli line defined by assumptions (a)-(h) of Subsection 2.4.1 with $M > 2$ machines, function $PR_{cl}(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}, N_0, S)$ is
• strictly increasing in $p_i$, $i = 1, \ldots, M$;

• non-strictly increasing in $N_i$, $i = 0, \ldots, M - 1$;

• nonmonotonic concave in $S$.

**Justification:** See Appendix B.

These properties of closed Bernoulli lines are of practical importance. First, it states that, similar to open lines, increasing $p_i$’s always leads to an increased production rate in closed lines as well. Second, it states that, unlike open lines, increasing buffer capacity does not always lead to improved performance. Finally, it reaffirms the evidence of Figure 4.1 that $PR_{cl}$ is nonmonotonic concave in $S$.

### 4.2.3 Unimpediment

**Definition 4.1** A pair $(N_0, S)$ is unimpeding if

$$PR_{cl}(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}, N_0, S) = PR_o(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}); \quad (4.2)$$

otherwise, $(N_0, S)$ is called impeding.

**Theorem 4.3** For a closed Bernoulli two-machine line, the pair $(N_0, S)$ is unimpeding if and only if

$$N_1 < S \leq N_0. \quad (4.3)$$

**Proof:** See Appendix B.

Along with characterizing the unimpeding values of $N_0$ and $S$, this theorem has another important implication. It states that, in fact, unimpeding $N_0$ and $S$ are
independent of the machine efficiencies $p_1$ and $p_2$: as long as (4.3) is observed, the closed nature of the line does not impede the open line behavior, i.e.,

$$PR_{cl}(p_1, p_2, N_1, N_0, S) = PR_o(p_1, p_2, N_1),$$

no matter what $p_1$ and $p_2$ are. Thus, changing $p_i$’s cannot change an unimpeding pair $(N_0, S)$ into an impeding one, and vise versa.

Similar to the two-machine case, the behavior of closed Bernoulli lines with $M > 2$ machines can be described by ergodic Markov chains. However, unlike the two-machine case, closed-form expressions for the performance measures are all but impossible to derive. Therefore, we resort to a more restrictive statement:

**Theorem 4.4** Assume that a closed Bernoulli line defined by assumptions (a)-(h) of Subsection 2.4.1 satisfies the condition:

$$\sum_{i=1}^{M-1} N_i < S \leq N_0.$$  \hspace{1cm} (4.4)

Then the pair $(N_0, S)$ is unimpeding and, therefore, all performance characteristics of this line coincide with those of the corresponding open line, i.e.,

$$PR_{cl}(p_1, \ldots, p_M, N_1, \ldots, N_0, S) = PR_o(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}),$$

$$WIP_{cl}^i(p_1, \ldots, p_M, N_1, \ldots, N_0, S) = WIP_o^i(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}),$$

$$i = 1, \ldots, M - 1,$$

$$BL_{cl}^i(p_1, \ldots, p_M, N_1, \ldots, N_0, S) = BL_o^i(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}),$$

$$ST_{cl}^i(p_1, \ldots, p_M, N_1, \ldots, N_0, S) = ST_o^i(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}),$$

$$i = 1, \ldots, M.$$  \hspace{1cm} (4.5)

**Proof**: See Appendix B.

Theorem 4.4 implies, in particular, that under condition (4.4), the unimpeding
pair \((N_0, S)\) is independent of \(p_i\)'s. We show below that a similar property holds even when (4.4) does not take place. Since we show this by simulations, the notion of unimpediment is modified to account for the fact that \(PR_{cl}\) and \(PR_o\) are determined with finite accuracy and, therefore, equality (4.2) may hold only approximately.

**Definition 4.2** A pair \((N_0, S)\) is practically unimpeding if

\[
\frac{|PR_{cl}(N_0, S) - PR_o|}{PR_o} \leq \delta \ll 1.
\]

**Numerical Fact 4.3** In closed Bernoulli lines defined by assumptions (a)-(h) of Subsection 2.4.1, a practically unimpeding pair \((N_0, S)\) remains practically unimpeding for all values of \(p_i, i = 1, \ldots, M\), as long as \(N_0, N_1, \ldots, N_{M-1}\) remain the same.

**Justification:** See Appendix B.

### 4.2.4 Improvability

Assuming that the machine and in-process buffer parameters are fixed, i.e., \(p_i, i = 1, \ldots, M\), and \(N_i, i = 1, \ldots, M - 1\), are given, consider the function \(PR_{cl}(N_0, S)\).

**Definition 4.3** A closed line is:

- **\(S^+\)-improvable if**
  \[PR_{cl}(N_0, S + 1) > PR_{cl}(N_0, S);\]

- **\(S^-\)-improvable if**
  \[PR_{cl}(N_0, S - 1) > PR_{cl}(N_0, S);\]

- **\(S\)-unimprovable if \(PR_{cl}(N_0, S^*)\) cannot be increased for any other \(S\), i.e.,**
  \[PR_{cl}(N_0, S^*) \geq PR_{cl}(N_0, S), \quad \forall S \neq S^*.\]
Note that $S^*$ may be non-unique, as illustrated in Figure 4.1. Clearly, the unimprovable $S^*$ is a function of $N_0$, and is denoted throughout this chapter as $S^* = S^*(N_0)$.

**Definition 4.4** A closed line is $N_0$-improvable if

$$PR_{cl}(N_0 + 1, S^*(N_0 + 1)) > PR_{cl}(N_0, S^*(N_0));$$

otherwise, it is unimprovable and the pair $(N^*_0, S^*(N^*_0))$ is called $(N_0, S)$-unimprovable.

Below we provide methods for identifying whether a line is improvable in an appropriate sense or not. We denote $S^*_{\text{min}}$ and $S^*_{\text{max}}$ as the smallest and largest unimprovable $S$. Clearly, in some systems $S^*_{\text{min}} = S^*_{\text{max}}$ (see Figure 4.1).

**Two-machine case:**

**Theorem 4.5** For $S \notin [S^*_{\text{min}}, S^*_{\text{max}}]$, a closed Bernoulli two-machine line defined by assumptions (a)-(h) of Subsection 2.4.1 is

- $S^+$-improvable if
  $$\sum_{i=1}^{2} ST_i > \sum_{i=1}^{2} BL_i;$$

- $S^-$-improvable if
  $$\sum_{i=1}^{2} ST_i < \sum_{i=1}^{2} BL_i.$$

**Proof:** See Appendix B.

For $S \in [S^*_{\text{min}}, S^*_{\text{max}}]$, increasing or decreasing $S$ leads to a limit cycle, i.e., “oscillations” between $S$ and $S - 1$ or $S + 1$. In this case, the best $S$ (i.e., the one resulting in the largest $PR_{cl}$) must be selected from the limit cycle.
It is convenient to introduce the notation:

\[ I = \sum_{i=1}^{M} ST_i - \sum_{i=1}^{M} BL_i; \quad (4.6) \]

and refer to \( I \) as the \( S \)-Improvability Indicator. Thus, positive (respectively, negative) \( I \)'s imply \( S^+ \)- (respectively, \( S^- \)) improvability.

In addition to its direct value as a tool for \( S \)-improvability identification, the utility of this theorem (and the subsequent similar statement for \( M > 2 \)) is in the fact that \( S \)-improvability can be identified without knowing the machine and buffer parameters but just by measuring the frequency of blockages and starvations of the machines during normal system operation.

Finally, we formulate:

**Theorem 4.6** If a system is \( S \)-unimprovable and \( ST_1 \) or \( BL_2 \) is non-zero, then the system is \( N_0 \)-improvable.

**Proof**: See Appendix B.

\( M > 2 \)-machine case: An extension of Theorem 4.5 for the case of \( M > 2 \) is all but impossible to derive analytically. However, based on simulations, we conclude that it takes place for any \( M \). Specifically,

**Numerical Fact 4.4** For \( S \notin [S_{\text{min}}^*, S_{\text{max}}^*] \), a closed Bernoulli line defined by assumptions (a)-(h) of Subsection 2.4.1 with \( M > 2 \) is

- \( S^+ \)-improvable if
  \[ \sum_{i=1}^{M} ST_i > \sum_{i=1}^{M} BL_i; \]
- \( S^- \)-improvable if
  \[ \sum_{i=1}^{M} ST_i < \sum_{i=1}^{M} BL_i. \]
Justification: See Appendix B.

Thus, $I$ of (4.6) is still the indicator of improvability: $S^+$ if $I$ is positive and $S^-$ if $I$ is negative.

Based on this, we formulate:

**S-Continuous Improvement Procedure 4.1:**

1. Evaluate $ST_i$ and $BL_i$ for all machines in the system.
2. Calculate the $S$-improvability indicator $I = \sum_{i=1}^{M} ST_i - \sum_{i=1}^{M} BL_i$.
3. If $I > 0$, increase $S$ by one; if $I < 0$, decrease $S$ by one.
4. Return to (1) and continue until a limit cycle is reached.
5. Select the $S$ from the limit cycle, which gives the largest $PR_{cl}$; this $S$ is unimprovable and is denoted as $S^*(N_0)$.

Clearly, if for the above $S^*(N_0)$,

$$PR_{cl}(N_0, S^*(N_0)) < PR_{cl}(N_0 + 1, S^*(N_0 + 1)),$$

the pair $(N_0, S^*(N_0))$ is impeding and, therefore, is improvable with respect to $N_0$. This improvement can be carried out using

**$N_0$-Continuous Improvement Procedure 4.1:**

1. For a given $N_0$, carry out S-Continuous Improvement Procedure 4.1 and determine $S^*(N_0)$ and $S^*(N_0 + 1)$.
2. If $|PR_{cl}(N_0, S^*(N_0)) - PR_{cl}(N_0 + 1, S^*(N_0 + 1))| > \delta PR_{cl}(N_0, S^*(N_0))$, $\delta \ll 1$, increase $N_0$ by 1, return to (1).
(3) If $|PR_{cl}(N_0, S^*(N_0)) - PR_{cl}(N_0 + 1, S^*(N_0 + 1))| \leq \delta PR_{cl}(N_0, S^*(N_0)), \delta \ll 1$, the system is unimprovable with respect to $N_0$; this $N_0$ and the resulting $S$ is an unimpeding pair and is denoted as $(N_0^*, S^*(N_0^*))$.

Below, two examples illustrating $S$- and $N_0$-improvement procedures are given. In the first example, the system of Figure 4.2 is considered and $S$-Continuous Improvement Procedure 4.1 is carried out starting from $S = 2$ and $S = 21$. The results are given in Tables 4.1 and 4.2, respectively. In both cases, the unimprovable number of carriers is 10.

In the second example, $N_0$-Continuous Improvement Procedure 4.1 (with $\delta = 0.01$) is applied to the system of Figure 4.3. As a result, an unimprovable pair $(N_0^*, S^*(N_0^*))$ is obtained with $N_0^* = 5$ and $S^*(N_0^*) = 11$ (see Table 4.3).

![Figure 4.2: Example of $S$-Continuous Improvement Procedure 4.1](image-url)

![Figure 4.3: Example of $N_0$-Continuous Improvement Procedure 4.1](image-url)

<table>
<thead>
<tr>
<th>Step #</th>
<th>$S$</th>
<th>$I$</th>
<th>$PR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2.3126</td>
<td>0.4189</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.4071</td>
<td>0.5877</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.7802</td>
<td>0.7006</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.4940</td>
<td>0.7447</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.3302</td>
<td>0.7623</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0.2572</td>
<td>0.7664</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0.1922</td>
<td>0.7673</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>0.0841</td>
<td>0.7673</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>$-0.0577$</td>
<td>$0.7674$</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0.0841</td>
<td>0.7673</td>
</tr>
</tbody>
</table>

Table 4.1: Example of $S$-Continuous Improvement Procedure 4.1 (starting from $S = 2$)
Table 4.2: Example of $S$-Continuous Improvement Procedure 4.1 (starting from $S = 21$)

<table>
<thead>
<tr>
<th>Step #</th>
<th>$S$</th>
<th>$I$</th>
<th>$PR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
<td>$-0.4005$</td>
<td>0.7596</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>$-0.1886$</td>
<td>0.7665</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>$-0.1111$</td>
<td>0.7671</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>$-0.0936$</td>
<td>0.7672</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>$-0.0927$</td>
<td>0.7673</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>$-0.0940$</td>
<td>0.7673</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>$-0.0916$</td>
<td>0.7672</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>$-0.0934$</td>
<td>0.7672</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>$-0.0917$</td>
<td>0.7666</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>$-0.0921$</td>
<td>0.7672</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>$-0.0888$</td>
<td>0.7672</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>$-0.0577$</td>
<td>0.7674</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
<td>0.0841</td>
<td>0.7673</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>$-0.0577$</td>
<td>0.7674</td>
</tr>
</tbody>
</table>

Figure 4.3: Example of $N_0$-Continuous Improvement Procedure 4.1

Table 4.3: Example of $N_0$-Continuous Improvement Procedure 4.1, $PR_0 = 0.6665$

<table>
<thead>
<tr>
<th>Step #</th>
<th>$N_0$</th>
<th>$S$</th>
<th>$I$</th>
<th>$PR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0.7551</td>
<td>0.5823</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0.5114</td>
<td>0.6116</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>7</td>
<td>0.3380</td>
<td>0.6287</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>8</td>
<td>0.2045</td>
<td>0.6386</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>9</td>
<td>0.0693</td>
<td>0.6425</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>10</td>
<td>$-0.1017$</td>
<td>0.6439</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>10</td>
<td>0.0367</td>
<td>0.6531</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>11</td>
<td>$-0.1200$</td>
<td>0.6530</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>10</td>
<td>0.1345</td>
<td>0.6580</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>11</td>
<td>0.0094</td>
<td>0.6597</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>12</td>
<td>$-0.1337$</td>
<td>0.6575</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>11</td>
<td>0.1055</td>
<td>0.6627</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>12</td>
<td>$-0.0081$</td>
<td>0.6618</td>
</tr>
</tbody>
</table>
Comparisons: The existing literature offers an interesting formula for selecting $S$ in closed lines with machines having random processing time and with blocked after service (BAS) convention (which implies that even if the downstream buffer is full, a machine can process a part). This formula is [40]:

$$\hat{S} = M + \left\lceil \frac{\sum_{i=0}^{M-1} N_{i}^{BAS}}{2} \right\rceil,$$  \hspace{1cm} (4.7)

where $N_{i}^{BAS}, i = 0, 1, \ldots, M - 1$, is the $i$-th buffer capacity under the BAS convention and $[x]$ denotes the smallest integer greater than or equal to $x$. The blocked before service (BBS) convention, used in this chapter, implies that the machine itself is a unit of buffer capacity; therefore,

$$N_{i}^{BAS} = N_{i} - 1,$$  \hspace{1cm} (4.8)

where $N_{i}$ is the $i$-th buffer capacity under the BBS convention. Thus, formula (4.7) for systems under the BBS convention becomes

$$\hat{S} = M + \left\lceil \frac{\sum_{i=0}^{M-1} (N_{i} - 1)}{2} \right\rceil = \left\lceil \frac{M + \sum_{i=0}^{M-1} N_{i}}{2} \right\rceil.$$  \hspace{1cm} (4.9)

To investigate the relationship between $S^*$ obtained by $S$-Continuous Improvement Procedure 4.1 and the $\hat{S}$ provided in expression (4.9), we use the examples of the previous subsection. The results are as follows: In the first example, $S^* = 10$, while $\hat{S} = 15$. This results in $PR(S^*) = 0.6439$ and $PR(\hat{S}) = 0.7674$. In the second example, the results are summarized in Table 4.4. As one can see, both approaches lead to similar results with $S^*$ being somewhat smaller than $\hat{S}$.

Table 4.4: Comparison of the $S$-Continuous Improvement Procedure 4.1 and equation (4.9)

<table>
<thead>
<tr>
<th>$N_0$</th>
<th>$S_{opt}$</th>
<th>$PR(S_{opt})$</th>
<th>$S^*$</th>
<th>$PR(S^*)$</th>
<th>$\hat{S}$</th>
<th>$PR(\hat{S})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>0.6439</td>
<td>10</td>
<td>0.6439</td>
<td>11</td>
<td>0.6438</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0.6531</td>
<td>10</td>
<td>0.6531</td>
<td>12</td>
<td>0.6526</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>0.6600</td>
<td>11</td>
<td>0.6597</td>
<td>12</td>
<td>0.6600</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>0.6630</td>
<td>11</td>
<td>0.6627</td>
<td>13</td>
<td>0.6629</td>
</tr>
</tbody>
</table>
4.2.5 Bottlenecks

The bottleneck (BN) machine of a closed line is defined in the same manner as that of an open line (see Definition 3.2), i.e., \( m_i, i \in \{1, \ldots, M\} \), is the BN if

\[
\frac{\partial P_{R_{cl}}}{\partial p_i} > \frac{\partial P_{R_{cl}}}{\partial p_j}, \quad \forall j \neq i.
\] (4.10)

Unfortunately, Bottleneck Indicator 3.1 is not directly applicable to closed lines. Indeed, while in open lines there always exists at least one machine with no emanating arrows (because \( ST_1 = BL_M = 0 \)), in closed lines there may be none. This is illustrated in Figure 4.4, where, in addition to the usual arrows, the arrow between \( m_1 \) and \( m_M \) is assigned according to the same rule: if \( ST_1 > BL_M \), the arrow is directed to the left; if \( ST_1 < BL_M \), it is directed to the right. To resolve this problem, the usual arrow-based method is modified as described next.

![Figure 4.4: Towards bottleneck identification in closed serial lines](image)

**Two-machine lines:**

**Theorem 4.7** In closed Bernoulli two-machine lines defined by assumptions (a)-(h) of Subsection 2.4.1, machine \( m_1 \) (respectively, machine \( m_2 \)) is the BN if and only if

\[
ST_1 + BL_1 < ST_2 + BL_2, \quad (\text{respectively, } ST_1 + BL_1 > ST_2 + BL_2).
\] (4.11)
Proof: See Appendix B.

This theorem can be interpreted as follows: Refer to $ST_1 + BL_1$ as the virtual blockage of $m_1$ and to $ST_2 + BL_2$ as the virtual starvation of $m_2$, i.e.,

$$BL_{1,v} = ST_1 + BL_1,$$
$$ST_{2,v} = ST_2 + BL_2,$$

and assume that the virtual starvation of $m_1$ and virtual blockage of $m_2$ are 0, i.e.,

$$ST_{1,v} = BL_{2,v} = 0.$$  \hspace{1cm} (4.12)

Assign arrows between $m_1$ and $m_2$ according to the same rule as in the case of open lines but using virtual blockages and starvations of $m_1$ and $m_2$. Then the machine with no emanating arrows is the BN of the closed line. This is illustrated in Figure 4.5. Thus, using virtual, rather than real, blockages and starvations allows us to extend the open line BN identification technique for two-machine lines to closed ones. As it is shown below, this can be done for $M > 2$ as well.

![Figure 4.5: BN identification in a two-machine closed line](image-url)
Consider an $M > 2$-machine closed Bernoulli line and assume that $ST_i$ and $BL_i$, $i = 1, \ldots, M$, are identified during normal system operation. Similar to the case $M = 2$, introduce the virtual blockages and starvations of the machines as follows:

$$BL_{1,v} := ST_1 + BL_1,$$

$$ST_{1,v} := 0,$$

$$BL_{i,v} := BL_i, \quad i = 2, \ldots, M - 1,$$

$$ST_{i,v} := ST_i, \quad i = 2, \ldots, M - 1,$$

$$BL_{M,v} := 0,$$

$$ST_{M,v} := ST_M + BL_M.$$

Using $ST_{i,v}$ and $BL_{i,v}$, assign arrows between $m_i$ and $m_{i+1}$ according to Arrow Assignment Rule 3.1, i.e., an arrow is directed from $m_i$ to $m_{i+1}$ if $BL_{i,v} > ST_{i+1,v}$ and from $m_{i+1}$ to $m_i$ if $BL_{i,v} < ST_{i+1,v}$. Since $ST_{1,v} = BL_{M,v} = 0$, there is at least one machine with no emanating arrow (see Figure 4.6).

![Figure 4.6: BN identification in a five-machine closed line](image-url)
**Bottleneck Indicator 4.1:** In closed Bernoulli lines defined by assumptions (a)-(h) of Subsection 2.4.1,

- if there is a single machine with no emanating arrows, it is the BN-m;

- if there are multiple machines with no emanating arrows, the one with the largest severity is the Primary BN (PBN), where the severity of each (local) BN is defined by

\[
S_{i,v} = \max(|ST_{i+1,v} - BL_{i,v}|, |ST_i - BL_{i-1,v}|), \quad i = 2, \ldots, M - 1,
\]

\[
S_{1,v} = \max(|ST_{2,v} - BL_{1,v}|, |ST_1 - BL_{M,v}|),
\]

\[
S_{M,v} = \max(|ST_{M,v} - BL_{M-1,v}|, |ST_1 - BL_{M,v}|).
\]

**Justification:** See Appendix B.

Thus, the bottleneck of a closed Bernoulli line can be identified using the arrow-based method based on virtual, rather than real, blockages and starvations.

### 4.3 Closed Continuous Lines

#### 4.3.1 Reversibility

**Numerical Fact 4.5** For closed lines defined by assumptions (a)-(i) of Subsection 2.4.2 with machines having their reliability models in the set \{exp, W, ga, LN\}, the throughputs of \(L\), and its reverse, \(L_r\), are related as follows:

\[
TP^L = TP^{L_r}.
\]

**Justification:** See Appendix B.

Based on the above, we formulate that:
Conjecture 4.1 Numerical Fact 4.5 holds for closed lines with any unimodal distribution of up- and downtime.

4.3.2 Monotonicity

Numerical Fact 4.6 For closed lines defined by assumptions (a)-(i) of Subsection 2.4.2 with machines having their reliability models in the set \{\text{exp, W, ga, LN}\}, function \(TP_{cl}(c_i, T_{up,i}, T_{down,i}, CV_{up,i}, CV_{down,i}, N_i, S)\) is

- strictly increasing in \(T_{up,i}\) and \(c_i\), \(i = 1, \ldots, M\);
- strictly decreasing in \(T_{down,i}\), \(CV_{up,i}\) and \(CV_{down,i}\), \(i = 1, \ldots, M\);
- non-strictly increasing in \(N_i\), \(i = 0, \ldots, M - 1\);
- non-monotonic concave in \(S\).

Justification: See Appendix B.

Clearly, throughputs of closed lines and open lines share the same monotonicity property with respect to machine parameters, while strict monotonicity with respect to buffer capacity in open lines becomes non-strict in closed lines due to the \((N_0, S)\) pair. For closed lines with general machine reliability, we formulate that:

Conjecture 4.2 Numerical Fact 4.6 holds for closed lines with any unimodal distribution of up- and downtime.

4.3.3 Unimpediment

Similar to the Bernoulli case, it can be proved that:
Theorem 4.8 The pair \((N_0, S)\) is unimpeding if

\[
\sum_{i=1}^{M-1} N_i \leq S \leq N_0.
\] (4.14)

Proof: See Appendix B.

Numerical Fact 4.7 For closed lines defined by assumptions (a)-(i) of Subsection 2.4.2 with machines having their reliability models in the set \{exp, W, ga, LN\}, a practically unimpeding pair \((N_0, S)\) remains practically unimpeding as long as \(N_1, \ldots, N_{M-1}\) remain the same.

Justification: See Appendix B.

For closed lines with general machine reliability, we formulate that:

Conjecture 4.3 Numerical Fact 4.7 holds for closed lines with any unimodal distribution of up- and downtime.

4.3.4 Improvability

Theorem 4.9 The S-Improvability Indicator holds for closed lines defined by assumptions (a)-(i) of Subsection 2.4.2 with \(M = 2\).

Proof: See Appendix B.

Numerical Fact 4.8 The S-Improvability Indicator holds for closed lines defined by assumptions (a)-(h) of Subsection 2.4.2 with \(M > 2\) machines having their reliability models in the set \{exp, W, ga, LN\}.
Justification: See Appendix B.

Based on the above, we formulate that:

Conjecture 4.4 The $S$-Improvability Indicator holds for closed lines with any unimodal distribution of up- and downtime.

4.4 Case Study: Automotive Paint Shop Production System

4.4.1 System description and layout

In this section, we apply the results obtained for closed lines to a paint shop production system at an automotive assembly plant. The layout of the system is shown in Figure 4.7. It consists of 11 operations in which the car bodies (referred to as jobs) are cleaned (chemically or physically), sealed (against water leaks), painted, and, finally finessed. Operations 5, 6, and 8 consist of two parallel lines (due to capacity considerations). Operation 10 consists of five parallel painting booths (for both capacity reasons and to ensure color variety).

The jobs within the system are transported by conveyors on two types of carriers. The transfer from one carrier to another occurs after Op. 3. Thus, carriers of type 1 are used in Ops. 1-3 and type 2 in Ops. 4-11.

The technological operations are performed while the jobs are moving on their carriers. That is why the parts of the conveyor where the work is being carried out are called operational conveyors, while other parts are referred to as accumulators.

The conveyor as a whole has a modular structure in the sense that each operational conveyor can be stopped without stopping other operational conveyors. Workers typically use push-buttons and stop their operational conveyors in order to complete
respective operations with the desired quality. Thus, the downtime is typically due to quality issues rather than machine breakdowns.

The numbers in the circles of Figure 4.7 indicate the number of jobs within operational conveyors necessary to ensure continuous production. The numbers in rectangles show the minimal occupancy of accumulators to ensure continuous production and the maximal number of jobs that could be contained within an accumulator. Thus, the difference between these two numbers is the buffering capacity of the accumulator.

This system was designed to produce 63 jobs/hour (see the capacity, $c_i$, of each operation given in Table 4.5). In reality, however, the throughput was much lower, averaging 52.1 jobs/hour (see Table 4.6 where the measured average throughput for five consecutive months is shown). The goal of this case study was to determine reasons for the production losses and to provide recommendations for their elimination.
Table 4.5: Capacity of the machines (jobs/hour)

<table>
<thead>
<tr>
<th>Ops.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_i$</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>72</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4.6: System performance (jobs/hour)

<table>
<thead>
<tr>
<th>Period</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>53.5</td>
<td>43.81</td>
<td>51.27</td>
<td>54.28</td>
<td>55.89</td>
</tr>
</tbody>
</table>

4.4.2 Structural modeling

The average production losses in Ops. 1-11 due to internal reasons (i.e., excluding blockages and starvations) are shown in Table 4.7. Clearly, Ops. 1 and 2 have very low or no losses and, therefore, can be excluded. To accomplish this, we conceptually transfer the common point of the two loops of Figure 4.7 from the output of Op. 3 to its input. This transformation does not lead to reduced accuracy since Op. 3 operates in so-called no-gap mode (i.e., no empty space between consecutive jobs on the operational conveyor is allowed). Therefore, after aggregating the parallel machines of Ops. 5, 6, 8 and 10, we represent the system as shown in Figure 4.8.

Table 4.7: Average production losses (jobs/hour)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 1</td>
<td>0</td>
<td>0.05</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Op. 2</td>
<td>0.05</td>
<td>0.45</td>
<td>0.07</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Op. 3</td>
<td>2.88</td>
<td>2.15</td>
<td>0.64</td>
<td>2.26</td>
<td>1.35</td>
</tr>
<tr>
<td>Op. 4</td>
<td>2.77</td>
<td>2.60</td>
<td>4.45</td>
<td>2.00</td>
<td>2.60</td>
</tr>
<tr>
<td>Op. 5</td>
<td>0.23</td>
<td>0.01</td>
<td>0.04</td>
<td>1.07</td>
<td>1.64</td>
</tr>
<tr>
<td>Op. 6</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0.02</td>
<td>0.39</td>
</tr>
<tr>
<td>Op. 7</td>
<td>1.09</td>
<td>3.13</td>
<td>1.09</td>
<td>2.68</td>
<td>2.05</td>
</tr>
<tr>
<td>Op. 8</td>
<td>1.39</td>
<td>3.42</td>
<td>1.28</td>
<td>2.73</td>
<td>0.41</td>
</tr>
<tr>
<td>Op. 9</td>
<td>6.18</td>
<td>7.38</td>
<td>7.01</td>
<td>6.59</td>
<td>6.14</td>
</tr>
<tr>
<td>Op. 10</td>
<td>0.35</td>
<td>1.40</td>
<td>1.66</td>
<td>3.09</td>
<td>3.63</td>
</tr>
<tr>
<td>Op. 11</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>
4.4.3 Modeling and identification of the machines and buffers

**Machines:** Since, according to the factory floor measurement data, the downtime of each operation is mostly of the same order of magnitude as the cycle time, we adopt the Bernoulli model of machine reliability. The parameters \( p_i \) are calculated as follows:

\[
p_i = \min \left\{ 1, \frac{c_i - L_i}{63} \right\}, \quad i = 3, \ldots, 11,
\]

where \( c_i \) and \( L_i \) are the capacity and the losses of the \( i \)-th operation given in Tables 4.5 and 4.7, respectively. These parameters are summarized in Table 4.8.

<table>
<thead>
<tr>
<th>Operations</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 1</td>
<td>0.9543</td>
<td>0.9560</td>
<td>0.9963</td>
<td>1</td>
<td>0.9827</td>
<td>0.9779</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Month 2</td>
<td>0.9659</td>
<td>0.9587</td>
<td>0.9998</td>
<td>1</td>
<td>0.9503</td>
<td>0.9457</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Month 3</td>
<td>0.9898</td>
<td>0.9294</td>
<td>0.9994</td>
<td>1</td>
<td>0.9827</td>
<td>0.9797</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Month 4</td>
<td>0.9641</td>
<td>0.9683</td>
<td>0.9830</td>
<td>0.9997</td>
<td>0.9575</td>
<td>0.9567</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Month 5</td>
<td>0.9786</td>
<td>0.9587</td>
<td>0.9740</td>
<td>0.9938</td>
<td>0.9675</td>
<td>0.9935</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Buffers:** As it is indicated above, the buffering capacity of each accumulator is the difference between its maximal and minimal occupancy. The capacity of the buffers after Ops. 6 and 10 is assumed to be the sum of the capacities of the parallel buffers. The buffers within Op. 5 are omitted. The resulting data on buffer capacity are summarized in Table 4.9.
Table 4.9: Buffer capacity

<table>
<thead>
<tr>
<th>Operations</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_i$</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>60</td>
<td>5</td>
<td>26</td>
<td>25</td>
<td>88</td>
</tr>
</tbody>
</table>

### 4.4.4 Overall system model

As it follows from above, the paint shop system can be modeled as a closed Bernoulli line (see Figure 4.9). The machine and buffer parameters are given in Tables 4.8 and 4.9, respectively. The minimum number of carriers to ensure continuous operation is 401, while the total number of carries available in the system is 409. Therefore, the effective number of carriers, $S$, is 8.

![Figure 4.9: Closed line model of the automotive paint shop system](image)

To validate the closed line model, we evaluate the throughput of the system in Figure 4.9 using simulation and compare it to that measured on the factory floor. The results are summarized in Table 4.10. Clearly, the estimated and measured quantities are close to each other, except for Month 2. This discrepancy is attributed to the fact that during this period a new car model was introduced and, perhaps, some transient phenomena played a substantial role. Thus, omitting Month 2, we assume that the model is validated.

Table 4.10: Closed line model validation data: $TP$

<table>
<thead>
<tr>
<th></th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est. $TP$</td>
<td>52.59</td>
<td>51.54</td>
<td>52.77</td>
<td>51.92</td>
<td>52.64</td>
</tr>
<tr>
<td>Actual $TP$</td>
<td>53.50</td>
<td>43.81</td>
<td>51.27</td>
<td>54.28</td>
<td>55.89</td>
</tr>
<tr>
<td>Error</td>
<td>$-1.70%$</td>
<td>$17.64%$</td>
<td>$2.93%$</td>
<td>$-4.35%$</td>
<td>$-5.81%$</td>
</tr>
</tbody>
</table>
4.4.5 $S$-improvability

The open line throughput and $S$-Improvability Indicator under the current $S$ of the five months are shown in Table 4.11. Obviously, the system is $S^+$-improvable. Therefore, we carry out $S$-Continuous Improvement Procedure 4.1 in order to determine the smallest number of carriers, which ensures the desired performance. The results, based on Month 1 data, are shown in Table 4.12. Clearly, with $S = 16$ the system attains the open line throughput and, thus, is unimpeded by its closed nature. The same property holds for all other months, since, as it is stated in Numerical Fact 4.2, the unimpeding pair $(N_0, S)$ is independent of machine parameters. Thus, the performance of the paint shop with $S = 16$ is equal to its open line performance, which is about a 10% improvement over the performance ensured by $S = 8$.

Table 4.11: Open line performance and $(N_0, S)$-impediment

<table>
<thead>
<tr>
<th></th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TP_o$</td>
<td>59.29</td>
<td>58.55</td>
<td>59.79</td>
<td>59.77</td>
<td>60.10</td>
</tr>
<tr>
<td>$TP_{cl}$</td>
<td>52.59</td>
<td>51.54</td>
<td>52.77</td>
<td>51.92</td>
<td>52.64</td>
</tr>
<tr>
<td>$I$</td>
<td>1.3535</td>
<td>1.4575</td>
<td>1.3412</td>
<td>1.4259</td>
<td>1.3463</td>
</tr>
</tbody>
</table>

Table 4.12: $S$-Continuous Improvement Procedure 4.1 for Month 1

<table>
<thead>
<tr>
<th>Step #</th>
<th>$S$</th>
<th>$I$</th>
<th>$TP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>1.3534</td>
<td>52.59</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>0.7217</td>
<td>57.00</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0.5216</td>
<td>58.36</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>0.4297</td>
<td>58.93</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>0.3926</td>
<td>59.16</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>0.3780</td>
<td>59.24</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>0.3735</td>
<td>59.27</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0.3711</td>
<td>59.28</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>0.3707</td>
<td>59.29</td>
</tr>
</tbody>
</table>

4.5 Summary

- Closed production lines are reversible.
• Throughput of a closed line is monotonic with respect to machine and buffer parameters and non-monotonic concave with respect to the number of carriers in the system.

• A criterion of improvability with respect to the number of carriers is established. Specifically, if \( \sum_{i=1}^{M} ST_i > \sum_{i=1}^{M} BL_i \), (respectively, \( \sum_{i=1}^{M} ST_i < \sum_{i=1}^{M} BL_i \)), the system is \( S^+ \)- (respectively, \( S^- \))-improvable, i.e., \( PR \) can be increased by adding (respectively, removing) a carrier.

• A criterion of improvability with respect to the empty carrier buffer capacity is derived, and a corresponding continuous improvement procedure is proposed.

• A method for identifying bottleneck machines in closed Bernoulli lines is suggested. Specifically, it is shown that bottlenecks in closed lines can be identified based on the same procedure as that for open lines but using the so-called virtual, rather than real, probabilities of blockages and starvations.

• The methods developed are applied to an automotive paint shop production system. \( S^- \)-improvability analysis shows that the current system is \( S^+ \)-improvable. Numerical results indicate that about 10% improvement in throughput can be achieved by increasing the number of carriers in the system.
CHAPTER V

SYSTEM-THEORETIC PROPERTIES OF PRODUCTION LINES WITH QUALITY ISSUES

5.1 Introduction

Along with productivity, quality is of central importance in modern manufacturing. The issue of product quality has three aspects: design for quality, build for quality, and inspect for quality. Although the first two, if carried out successfully, may eliminate the necessity of the third one, quality inspection is still a part of most manufacturing operations. Recently, studies on production lines with quality issues have received enormous attention. However, system-theoretic properties of production lines with quality issues have not been studied. In this chapter, we investigate fundamental system-theoretic properties of such systems. Specifically, we study Bernoulli serial lines with machines of the following types:

- perfect quality machines;
- non-perfect quality machines with the Bernoulli quality model;
- non-perfect quality machines with quality-quantity coupling (QQC);
- inspection machines.

The outline of this chapter is as follows: Section 5.2 is devoted to Bernoulli lines with non-perfect quality machines but without QQC machines. Section 5.3 considers
systems with QQC. Section 5.4 addresses Bernoulli lines with rework. Finally, a case study is described in Section 5.5.

5.2 Bernoulli Lines with Non-perfect Quality Machines

5.2.1 Model

Consider a production line shown in Figure 5.1. Suppose it operates according to assumptions (a)-(e) of Subsection 2.4.1 along with:

(f) The machines are of three types: perfect quality machines \( m_i, i \in I_p \); non-perfect quality machines \( m_i, i \in I_{np} \); and inspection machines \( m_i, i \in I_{insp} \). Here \( I_p \), \( I_{np} \) and \( I_{insp} \) represent a partitioning of \( \{1, \ldots, M\} \) into the sets of indices, which indicate the positions of perfect, non-perfect and inspection machines, respectively. Both perfect and non-perfect quality machines are referred to as producing machines.

(g) The non-perfect machines obey the Bernoulli quality model, i.e., each part produced by a non-perfect quality machine \( m_i, i \in I_{np} \), is good with probability \( g_i \) and defective with probability \( 1 - g_i \). Parameter \( g_i \) is referred to as the quality of \( m_i, i \in I_{np} \). For convenience, the expression \( g_i, i \in \{1, \ldots, M\} \), is used to indicate the quality parameters of all machines, where \( g_i = 1 \) for \( i \notin I_{np} \).

(h) For each inspection machine \( m_i, i \in I_{insp} \), “producing a part” implies that the part quality is identified. The defects are assumed to be identified perfectly, i.e., no defects are missed and no perfect parts are identified as defectives. Defective
parts are discarded (i.e, scrapped), while non-defective parts are placed in the 
downstream buffer or, in the case of $m_M$, shipped to the customer.

Note that both the Bernoulli reliability and Bernoulli quality models are applicable
to automotive assembly and painting operations where the downtime of the machines
is comparable to their cycle time and the defects are due to independent random
events (e.g., dust, scratches, etc.).

This model can represent production lines without inspection machines as well.
In this case, the last machine, $m_M$, can be viewed, conceptually, as performing, along
with its technological operation, a quality inspection and separation of good and
defective parts produced.

As mentioned in Section 2.5, in the framework of the above model, two perfor-
mance measures are considered in addition to $PR, WIP_i, ST_i$ and $BL_i$: scrap rate
($SR_i$) and consumption rate ($CR$). An aggregation-based recursive procedure has
been developed to approximate these performance measures with high accuracy (see
Appendix C). We denote these estimates as $\hat{PR}, \hat{CR}, \hat{SR}_i, \hat{WIP}_i, \hat{ST}_i$ and $\hat{BL}_i$, and
derive system-theoretic properties of such systems based on this technique.

5.2.2 Reversibility

Consider a production line, $L$, with perfect and non-perfect quality machines but
without inspection machines (i.e., with the last machine separating good and defective
parts) and its reverse, $L_r$.

**Theorem 5.1** The production rates of lines $L$ and $L_r$ are related as follows:

$$\hat{PR}_L = \hat{PR}_r.$$  \hspace{1cm} (5.1)

**Proof:** See Appendix C.
5.2.3 Monotonicity

It has been shown in Subsection 3.1.2 that $PR$ in serial Bernoulli lines is monotonically increasing in $p_i$ and $N_i$. A question arises: What are the monotonicity properties of $PR$, as well as those of $SR$ and $CR$, in production lines defined by assumptions (a)-(h)? The answer is as follows.

**Theorem 5.2** In Bernoulli lines defined by assumptions (a)-(h) of Subsection 5.2.1,

- $\hat{PR}$, $\hat{SR}$ and $\hat{CR}$ are monotonically increasing in $p_i$, $i = 1, \ldots, M$, and $N_i$, $i = 1, \ldots, M - 1$.
- $\hat{PR}$ is monotonically increasing in $g_i$, $i \in I_{np}$.
- $\hat{SR}$ is monotonically decreasing in $g_i$, $i \in I_{np}$.
- $\hat{CR}$ is monotonically decreasing in $g_i$, $i \in I_{np}$, unless the only inspection machine is the last one, in which case, $\hat{CR}$ is independent of $g_i$.

**Proof:** See Appendix C.

5.2.4 Bottlenecks

**Definitions:** Production lines with non-perfect quality machines have bottlenecks of two types: those with largest effect on $PR$ through their efficiency, $p_i$, $i \in \{1, \ldots, M\}$, and those with largest effect on $PR$ through their quality, $g_i$, $i \in I_{np}$. Accordingly, we introduce:

**Definition 5.1** Machine $m_i$, $i \in \{1, \ldots, M\}$, is the production rate bottleneck ($PR$-BN) of a line defined by assumptions (a)-(h) of Subsection 5.2.1 if

$$\frac{\partial PR}{\partial p_i} > \frac{\partial PR}{\partial p_j}, \quad \forall j \neq i. \quad (5.2)$$
Definition 5.2 Machine $m_i, \ i \in \{1, \ldots, M\}$, is the quality bottleneck (Q-BN) of a line defined by assumptions (a)-(h) of Subsection 5.2.1 if

$$\frac{\partial PR}{\partial g_i} > \frac{\partial PR}{\partial g_j}, \ \forall \ j \neq i.$$  \ \ (5.3)

Methods of PR- and Q-BNs identification are developed next.

**Bottlenecks in production lines with a single inspection machine $m_M$: Production rate bottlenecks:** Consider two production lines – L1 defined by assumptions (a)-(h) of Subsection 5.2.1 and L2 defined by assumptions (a)-(e) of Subsection 2.4.1 and having all machines of the same efficiency and all buffers of the same capacity as the corresponding machines and buffers in L1. Let $\hat{PR}(1)$ and $\hat{PR}(2)$ be the production rates of L1 and L2 obtained using calculation with $q_M = 1$ in the case of L2 (see Appendix C for details of the performance evaluation method).

**Theorem 5.3** Under the above assumptions,

$$\frac{\partial \hat{PR}(2)}{\partial p_i} > \frac{\partial \hat{PR}(2)}{\partial p_j}, \ \forall \ j \neq i \iff \frac{\partial \hat{PR}(1)}{\partial p_i} > \frac{\partial \hat{PR}(1)}{\partial p_j}, \ \forall \ j \neq i,$$ \ \ (5.4)

i.e., $m_i$ is the PR-BN of L1 if and only if it is the bottleneck of L2.

**Proof:** See Appendix C.

Based on Theorem 5.3, PR-BN of a serial line with one inspection machine $m_M$ can be identified using Bottleneck Indicator 3.1. Note that since the PR-BN is independent of $g_i$’s, it follows from Theorem 5.3 that the PR- and Q-BNs are decoupled. Thus, the PR-BN remains the same no matter how $g_i$’s are changed.

**Example 5.1:** Consider the production line shown in Figure 5.2. The numbers above the machines and buffers represent their parameters; for each non-perfect machine, the first number is its efficiency and the second its quality. The first three rows
of numbers under the machines represent their starvations, blockages and severity obtained by calculation (see Appendix C), respectively. The last three rows represent the same performance measures evaluated using simulation. Based on these data and Bottleneck Indicator 3.1, we conclude that the primary bottleneck in this system is \( m_4 \). Note that it is not the worst machine, as far as \( p_i \)'s are concerned, and it remains PR-PBN no matter how \( g_i \)'s, \( i \in I_{np} \), are changed.

Figure 5.2: Example 5.1

Quality bottlenecks: The following theorem identifies Q-BNs.

**Theorem 5.4** In Bernoulli lines defined by assumptions (a)-(h) with the only inspection machine \( m_M \), the inequality

\[
\frac{\partial \hat{PR}}{\partial g_i} > \frac{\partial \hat{PR}}{\partial g_j}, \quad \forall j \neq i, \tag{5.5}
\]

takes place if and only if

\[
g_i = \min_{l \in I_{np}} g_l, \tag{5.6}
\]

i.e., the Q-BN is the machine with the worst quality; if \( \min g_l \) is achieved at more than one machine, each is the Q-BN.

**Proof:** See Appendix C.
Thus, in the production line of Figure 5.2, \( m_9 \) is the Q-BN and remains the Q-BN, no matter how \( p_9 \) or any other \( p_i \)'s are changed.

**Bottlenecks in production lines with multiple inspection machines or a single inspection machine other than \( m_M \):** Unfortunately, none of the above properties of PR- and Q-BNs hold for lines with multiple inspection machines or with one inspection machine, which is not the last one. Therefore, bottlenecks are identified directly based on Definitions 5.1 and 5.2, using finite differences instead of derivatives in (5.2) and (5.3), i.e., \( m_i \) is the PR-BN if

\[
\frac{\Delta PR^i(p_1, \ldots, p_M)}{\Delta p_i} > \frac{\Delta PR^j(p_1, \ldots, p_M)}{\Delta p_j}, \quad \forall j \neq i,
\]

and \( m_i \) is the Q-BN if

\[
\frac{\Delta PR^i(g_1, \ldots, g_M)}{\Delta g_i} > \frac{\Delta PR^j(g_1, \ldots, g_M)}{\Delta g_j}, \quad \forall j \neq i,
\]

where

\[
\Delta PR^i(p_1, \ldots, p_M) = PR(p_1, \ldots, p_i + \Delta p_i, \ldots, p_M) - PR(p_1, \ldots, p_i, \ldots, p_M),
\]

\[
\Delta PR^i(g_1, \ldots, g_M) = PR(g_1, \ldots, g_i + \Delta g_i, \ldots, g_M) - PR(g_1, \ldots, g_i, \ldots, g_M),
\]

and

\[
\Delta p_i \ll 1, \quad \Delta g_i \ll 1.
\]

Since neither \( \Delta PR^i(p_1, \ldots, p_M) \) nor \( \Delta PR^i(g_1, \ldots, g_M) \) can be evaluated on the factory floor during normal system operation, we replace them by

\[
\Delta \overline{PR}^i(p_1, \ldots, p_M) = \overline{PR}(p_1, \ldots, p_i + \Delta p_i, \ldots, p_M) - \overline{PR}(p_1, \ldots, p_i, \ldots, p_M),
\]

\[
\Delta \overline{PR}^i(g_1, \ldots, g_M) = \overline{PR}(g_1, \ldots, g_i + \Delta g_i, \ldots, g_M) - \overline{PR}(g_1, \ldots, g_i, \ldots, g_M),
\]

where the right hand sides of (5.12) and (5.13) are obtained by calculations
Appendix C). A question arises whether (5.7) and (5.8) imply and are implied by

\[
\frac{\Delta \hat{PR}_i(p_1, \ldots, p_M)}{\Delta p_i} > \frac{\Delta \hat{PR}_j(p_1, \ldots, p_M)}{\Delta p_j}, \quad \forall j \neq i, \quad (5.14)
\]

\[
\frac{\Delta \hat{PR}_i(g_1, \ldots, g_M)}{\Delta g_i} > \frac{\Delta \hat{PR}_j(g_1, \ldots, g_M)}{\Delta g_j}, \quad \forall j \neq i. \quad (5.15)
\]

This question arises because, although $PR$ and $\hat{PR}$ are close to each other, their partial derivatives, in general, may be quite different. The answer is given below.

**Numerical Fact 5.1** For $\Delta p_i = 0.01$ and $\Delta g_i = 0.01$, (5.14) and (5.15) practically always imply and are implied by (5.7) and (5.8).

**Justification:** See Appendix C.

**Example 5.2:** Consider Line 1 shown in Figure 5.3(a). Based on (5.14) and (5.15), we determine that $m_2$ is both PR- and Q-BN. Thus, the most effective way of productivity improvement is by increasing either $p_2$ or $g_2$. Increasing $p_2$ from 0.8 to 0.9, shifts PR-BN to $m_3$ and Q-BN to $m_4$ (see Figure 5.3(b)); note that the Q-BN is not the worst quality machine in the system. If, instead of increasing $p_2$, the value of $g_2$ is increased from 0.82 to 0.9, the bottlenecks, shown in Figure 5.3(c), are $m_2$ (PR-BN) and $m_4$ (Q-BN); note that although the line is perfectly symmetric and $m_2$ and $m_4$ are identical, their effect on the production rate is different.

The performance characteristics of Lines 1-3 are summarized in Table 5.1. It quantifies the advantages of increasing $g_i$’s, rather than increasing $p_i$’s, for system improvement from the point of view of both the production and scrap rates.

<table>
<thead>
<tr>
<th>Line</th>
<th>$\hat{PR}$</th>
<th>$\hat{SR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 1</td>
<td>0.5494</td>
<td>0.1951</td>
</tr>
<tr>
<td>Line 2</td>
<td>0.5886</td>
<td>0.2090</td>
</tr>
<tr>
<td>Line 3</td>
<td>0.5967</td>
<td>0.1400</td>
</tr>
</tbody>
</table>
\[ \frac{\Delta \bar{F}_R}{\Delta \eta_i} = 0.164 \quad \frac{\Delta \bar{F}_R}{\Delta \sigma_i} = -0.614 \]

(a) Line 1

\[ \frac{\Delta \bar{F}_R}{\Delta \eta_i} = 0.167 \quad \frac{\Delta \bar{F}_R}{\Delta \sigma_i} = -0.613 \]

(b) Line 2

\[ \frac{\Delta \bar{F}_R}{\Delta \eta_i} = 0.153 \quad \frac{\Delta \bar{F}_R}{\Delta \sigma_i} = -0.570 \]

(c) Line 3

Figure 5.3: Example 5.2
**Local bottlenecks:** As described above, identification of bottlenecks in the general case requires evaluation of all partial derivatives involved in Definitions 5.1 and 5.2. Below, we introduce the notions of local bottlenecks and show that they can be identified using the same techniques as those in the case of a single inspection machine \( m_M \). In addition, we show that one of the local bottlenecks is, practically always, the global one.

Consider the production line defined by assumptions (a)-(h) of Subsection 5.2.1 and its segmentations shown in Figures 5.4. Note that the segments in Figure 5.4(a) are not overlapping and consist of producing machines located between each pair of consecutive inspection machines (or the producing machines before the first inspection machine); they are referred to as *Q-segments*. In contrast, the segments of Figure 5.4(b) are overlapping and consist of a pair of consecutive inspection machines and all the producing machines between them (or the producing machines before the first and after the last inspection machine); they are referred to as *PR-segments*.

**Definition 5.3** Machine \( m^l_i \) is the local PR-BN \((LPR-BN)\) of segment \( l \) if

\[
\frac{\partial PR}{\partial p^l_i} > \frac{\partial PR}{\partial p^l_j}, \quad \forall j \neq i, i, j \in l\text{-th PR-segment.} \tag{5.16}
\]
**Definition 5.4** Machine $m_i$ is the local Q-BN (LQ-BN) of segment $l$ if

$$\frac{\partial PR}{\partial g_i} > \frac{\partial PR}{\partial g_j}, \quad \forall j \neq i, \ i, j \in l-th \ Q-segment. \tag{5.17}$$

**Theorem 5.5** The LQ-BN of each Q-segment is the machine with the smallest $g_i$. One of the LQ-BNs is the Q-BN of the line.

**Proof**: See Appendix C.

**Example 5.3**: Consider the production line shown in Figure 5.5. According to Theorem 5.5, its LQ-BNs are $m_2$ and $m_7$. The sensitivities of $\hat{PR}$ and $PR$ to $g_2$ and $g_7$ are also shown in Figure 5.5 (for $\Delta g_i = 0.01$). Thus, $m_7$ is the Q-BN. Note that it is not the worst quality machine in the system. Note also that based on LQ-BNs, two, rather than six, sensitivities must be evaluated to identify Q-BN.

**Numerical Fact 5.2** The LPR-BNs can be can be identified, practically always, using Bottleneck Indicator 3.1 applied to each PR-segment, under the assumption that its first machine is not starved and the last is not blocked. One of the LPR-BNs is, practically always, the PR-BN of the line.
**Justification:** See Appendix C.

**Example 5.4:** Consider the production line shown in Figure 5.6 along with its PR-segments 1-3. The LPR-BNs, identified using Numerical Fact 5.2 by both simulations and calculations, are $m_1$, $m_6$ and $m_8$. The sensitivities of $\hat{PR}$ and $PR$ to $p_i$’s are also shown in Figure 5.6(a) (for $\Delta p_i = 0.01$). Thus, $m_1$ is the PR-BN, and it can be identified, using local bottlenecks, based on three, rather than ten, sensitivities.

![Diagram of production line and PR-segments](image)

**5.2.5 Inspection machine positioning**

In a production line with quality issues, the throughput is affected by the position of the inspection devices. A question arises: Given the number of inspection machines and the parameters $p_i$, $g_i$ and $N_i$, of all machines and buffers, where should the inspection machines be placed so that:
• no defective products are shipped to the customer;

• $PR$ is maximized?

Clearly, if the last producing machine is of non-perfect quality and only one inspection machine is available, the solution of this problem is trivial: the inspection machine should be placed at the end of the line. However, when the last machine produces no defectives, or two or more inspection machines are available, the design problem is non-trivial. A solution of this problem is given below.

**Motivating examples**  As it is illustrated below, the solution of the problem of positioning inspection devices is quite sensitive to the parameters of the machines involved.

**Example 5.5:** Consider the production line with no inspection machines shown in Figure 5.7(a). Its bottleneck (BN), identified using Bottleneck Indicator 3.1, is $m_3$. Consider also the inspection module, shown in Figure 5.7(b), with the inspection machine of efficiency $p_{\text{insp}}$ and buffer of capacity 3. The problem is to position the inspection module so that $\hat{PR}$ is maximized. To accomplish this, we carry out the full search, i.e., we place the inspection module in all possible positions, $I_{\text{insp}} = \{i\}$, $i = 2, 3, 4, 5$, and, using recursive procedure (C.15), calculate $\hat{PR}$ as a function of $p_{\text{insp}}$ for $g_1 = 0.98, 0.90$ and $0.80$. The results are shown in Figure 5.8. From these results, the following conclusions can be made:

• For $g_1 = 0.98$: If $p_{\text{insp}}$ is low, the optimal position of $m_{\text{insp}}$ is at the end of the line, i.e., $m_5$; if $p_{\text{insp}}$ is mid-range, it is immediately after the non-perfect machine, i.e., $m_2$; if $p_{\text{insp}}$ is high, the optimal position of $m_{\text{insp}}$ is immediately before the BN, i.e., $m_3$.

• For $g_1 = 0.90$: The situation is quite similar but with different ranges of $p_{\text{insp}}$ leading to various positions of the inspection machine.
• For $g_1 = 0.80$: For all values of $p_{\text{insp}}$, the optimal position of the inspection machine is immediately after the non-perfect machine, i.e., $m_2$.

![Figure 5.7: Example 5.5](image1)

Example 5.6: Consider the production line of Figure 5.9 with $m_2$ as its BN. Placing the inspection module in all possible positions (i.e., $I_{\text{insp}} = \{i\}$, $i = 4, 5, 6, 7$), we obtain the results shown in Figure 5.10. Although they are quite similar to those of Figure 5.8, there is an important difference: for large $p_{\text{insp}}$, the optimal position of the inspection module is immediately after the non-perfect quality machine, no matter what the value of $g_3$ is.

Based on the above observations and taking into account that in most cases inspection machines are of high efficiency, we limit below our attention to the case when $p_{\text{insp}}$ is sufficiently close to 1 and formulate an empirical rule for a single inspection machine positioning.
Empirical design rule: Let $m_{last}$ denote the last non-perfect quality machine in the line defined by assumptions (a)-(h) of Subsection 5.2.1 with $|I_{insp}| = 0$, i.e., the line with perfect and non-perfect quality machines but without an inspection system. Refer to such a line as the original line. The position of a single inspection machine, which results in the maximum $\hat{PR}$, is referred to as optimal. Let BNs be the bottlenecks of the original line identified using Bottleneck Indicator 3.1.

Numerical Fact 5.3 There exists a $p^*$ such that for all $p_{insp} \geq p^*$, the optimal position of the inspection module is, practically always, either immediately after $m_{last}$ or either before or after a BN located downstream of $m_{last}$.

Justification: See Appendix C.

Numerical Fact 5.3 provides the following empirical rule: If there are no BNs of the original line downstream of $m_{last}$, the optimal position of an efficient inspection machine is immediately after $m_{last}$. If there are BNs downstream of $m_{last}$, there
are several candidates that must be explored before the optimal position of \( m_{\text{insp}} \) is determined: immediately after \( m_{\text{last}} \) and immediately after and before each BN of the original line located downstream of \( m_{\text{last}} \).

This fact allows for the optimal placement of a single, sufficiently efficient inspection machine without evaluating the efficacy of all possible placements. No results of this nature are available for placing multiple inspection machines. The recommendation that can be given at this time is to carry out the full search of all possible placements, evaluating the efficacy of each one.

5.3 Bernoulli Lines with Quality-Quantity Coupling Machines

5.3.1 Model

**Model:** We consider here the same model as in the previous section, i.e., defined by conventions (a)-(h) of Subsection 5.2.1, along with an additional assumption:

(i) The non-perfect quality machines are of two types: with and without quality-quantity coupling (QQC). In the former case, \( g_i = g_i(p_i), i \in I_{\text{QQC}} \subseteq I_{\text{np}}, \) where \( g_i(p_i) \) is a differentiable monotonically decreasing function. Here \( I_{\text{QQC}} \) is a set of indices representing the positions of the QQC machines.

Also, to simplify the presentation, we assume that there is no inspection machines, and the last producing machine, \( m_M \), separates good and defective parts. Such a system is illustrated in Figure 5.11.

![Figure 5.11: Serial line with perfect, non-perfect and QQC machines](image)

5.3.2 Monotonicity

**Two-machine case:**
Theorem 5.6 Consider a Bernoulli line defined by assumptions (a)-(i) of Subsection 5.3.1 with \( M = 2 \) and \( N_1 \geq 2 \). Then, there exist \( p_i^* \) and \( p_i^{**} \), \( 0 < p_i^* \leq p_i^{**} < 1 \), \( i \in I_{QC} \), such that \( PR \), as a function of \( p_i \), is monotonically increasing for \( p_i \in (0, p_i^*) \) and monotonically decreasing for \( p_i \in (p_i^{**}, 1) \).

Proof: See Appendix C.

Example 5.7: Consider the production line shown in Figure 5.12. Assume

\[
g_1(p_1) = 1 - 0.4p_1. \tag{5.18}
\]

The production rate of this system is illustrated in Figure 5.13. Clearly for every \( p_2 \), \( PR \) is a concave function of \( p_1 \), i.e., \( p_1^* = p_1^{**} \).

\[
| p_1, g_1(p_1) | \quad 2 \quad | p_2, 0.95 |
\]

Figure 5.12: Example 5.7

\[
| \begin{array}{c}
| 0.35 | \\
| 0.4 \quad 0.45 \quad 0.5 \quad 0.55 |
\end{array} | \quad | p_1 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \quad 1 |
\]

Figure 5.13: \( PR \) as a function of \( p_1 \) and \( p_2 \) in Example 5.7

\( M > 2 \)-machine case: Due to the complexity of the production rate evaluation, an analytical extension of Theorem 5.6 for \( M > 2 \) is all but impossible to derive. However, based on numerical calculations, the following can be established.
Numerical Fact 5.4 For a Bernoulli line defined by assumptions (a)-(i) of Subsection 5.3.1 with $M > 2$ and $N_i \geq 2$, $i = 1, \ldots, M - 1$, there exist $0 < p_i^* \leq p_i^{**} < 1$, $i \in I_{QQC}$, such that $\hat{PR}$, as a function of $p_i$, is monotonically increasing for $p_i \in (0, p^*)$ and monotonically decreasing for $p_i \in (p_i^{**}, 1)$.

Justification: See Appendix C.

Example 5.8: Consider the production line shown in Figure 5.14. Assume that

$$g_3(p_3) = \frac{1}{0.6p_3 + 1} \quad (5.19)$$

The production rate of the system as a function of $p_3$ and $p_4$ is illustrated in Figure 5.15. Clearly, $\hat{PR}$ exhibits a similar non-monotonic behavior with respect to $p_i$, $i \in I_{QQC}$, as in two-machine lines.

5.3.3 Bottlenecks

Definition: Introduce the following
Definition 5.5 In a Bernoulli line defined by assumptions (a)-(i) of Subsection 5.3.1, machine $m_i$ is:

- $p^+\text{-BN}$ if
  \[
  \frac{\partial PR}{\partial p_i} > \frac{\partial PR}{\partial p_j}, \quad \forall j \neq i, \tag{5.20}
  \]

and

\[
\frac{\partial PR}{\partial p_i} > 0;
\]

- $p^-\text{-BN}$ if (5.20) holds and

\[
\frac{\partial PR}{\partial p_i} < 0.
\]

Clearly, if the BN is a $p^+$-BN, the largest improvement of $PR$ is obtained by increasing the efficiency of the BN machine; if the BN is a $p^-$-BN, decreasing its efficiency leads to the largest increase of $PR$. The properties of $p^+$-BN and $p^-$-BN in serial lines with QQC machines are investigated below.

Two-machine case: Given the machine and buffer parameters, the $p^+$- and $p^-$-BNs of two-machine lines defined by assumptions (a)-(i) of Subsection 5.3.1 can be identified directly using expressions for \(\frac{\partial PR}{\partial p_1}\) and \(\frac{\partial PR}{\partial p_2}\) derived in the proof of Theorem 5.6 (see Appendix C). Specifically, these expressions are

\[
\frac{\partial PR}{\partial p_1} = p_2 \left( \frac{dg_1(p_1)}{dp_1} (1 - Q(p_1, p_2, N_1)) - \frac{\partial Q(p_1, p_2, N_1)}{\partial p_1} g_1(p_1) \right) g_2(p_2), \tag{5.21}
\]

\[
\frac{\partial PR}{\partial p_2} = p_1 \left( \frac{dg_2(p_2)}{dp_2} (1 - Q(p_2, p_1, N_1)) - \frac{\partial Q(p_2, p_1, N_1)}{\partial p_2} g_2(p_2) \right) g_1(p_1), \tag{5.22}
\]
where
\[
\frac{\partial Q(x, y, N)}{\partial x} = \begin{cases} 
\frac{y(1-x)(\alpha^N(x,y)-1)+(y-x)N\alpha^N(x,y)}{(y-x\alpha^N(x,y))^\alpha(1-x)} & \text{if } x \neq y, \\
-\frac{N(N+1)}{2p(N+1-p)^2} & \text{if } x = y = p.
\end{cases}
\] (5.23)

Based on these expressions, the properties of the bottlenecks are illustrated below.

**Example 5.9:** Consider the production line shown in Figure 5.16 with
\[
g_i(p_i) = \frac{1}{0.6p_i^3 + 1}, \quad p_i \in (0, 1), \quad i = 1, 2.
\]

Its bottlenecks, identified using (5.21) and (5.22), are shown in Figure 5.17(a). Note that if both machines were of perfect quality, \( m_1 \) would be \( p^+\)-BN in the lower triangle of Figure 5.17(a) and \( m_2 \) in the upper one. Thus, having QQC machines preserves the same bottlenecks for \( p_i \)'s sufficiently small and reverses them for larger \( p_i \)'s.

\[
\begin{array}{cc}
p_1, g_1(p_1) & 2 \quad p_2, g_2(p_2) \\
m_1 & b_1 & m_2
\end{array}
\]

**Figure 5.16: Example 5.9**

A similar situation takes place when the quality functions, \( g_i(p_i) \), are not identical. This is illustrated in Figure 5.17(b) for
\[
g_1(p_1) = \frac{1}{0.6p_1^3 + 1}, \quad p_1 \in (0, 1),
\]
\[ g_2(p_2) = \frac{1}{0.4p_2^2 + 1}, \quad p_2 \in (0, 1). \]

Finally, the bottlenecks in systems with only one QQC machine are illustrated in Figure 5.17(c) for
\[ g_1(p_1) = \frac{1}{0.6p_1^2 + 1}, \quad p_1 \in (0, 1), \]
\[ g_2(p_2) = 0.9. \]

Thus, when both machines are QQC, the bottleneck can be eliminated by increasing its efficiency if \( p_i \)'s are sufficiently small and decreasing its efficiency if \( p_i \)'s are large. When only one machine is QQC, if the non-QQC machine is the bottleneck, it is always \( p^+\)-BN, while if the QQC machine is the bottleneck, it is \( p^+\)-BN for small \( p_i \)'s and \( p^-\)-BN for large ones.

**M > 2-machine case:** Unfortunately, the partial derivatives involved in Definition 5.5 cannot be calculated using closed-form expressions for \( M > 2 \). Therefore, the BNs are identified using finite difference estimates instead of the partial derivatives in (5.20), i.e., \( m_i \) is the \( p^+\)-BN (respectively, \( p^-\)-BN) if
\[
\left| \frac{\Delta PR^i(p_1, \ldots, p_M)}{\Delta p_i} \right| > \left| \frac{\Delta PR^j(p_1, \ldots, p_M)}{\Delta p_j} \right|, \quad \forall j \neq i, \quad (5.24)
\]
\[
\frac{\Delta PR^i}{\Delta p_i} > 0, \quad \text{(respectively, } \frac{\Delta PR^i}{\Delta p_i} < 0 \text{)}, \quad (5.25)
\]
where \( \Delta PR^i(p_1, \ldots, p_M) \) is given in (5.9) and \( \Delta p_i \ll 1 \).

Since \( \Delta PR^i(p_1, \ldots, p_M) \) cannot be evaluated on the factory floor during normal system operation, we replace it by \( \hat{\Delta PR}^i(p_1, \ldots, p_M) \) defined in (5.12). Thus, inequalities (5.24) and (5.25) become
\[
\left| \frac{\hat{\Delta PR}^i(p_1, \ldots, p_M)}{\Delta p_i} \right| > \left| \frac{\hat{\Delta PR}^j(p_1, \ldots, p_M)}{\Delta p_j} \right|, \quad \forall j \neq i, \quad (5.26)
\]
\[
\frac{\hat{\Delta PR}^i}{\Delta p_i} > 0, \quad \text{(respectively, } \frac{\hat{\Delta PR}^i}{\Delta p_i} < 0 \text{)}. \quad (5.27)
\]
For production lines defined by assumptions (a)-(h) of Subsection 5.2.1, i.e., lines with no QQC machines, the equivalence of (5.24) and (5.26) has been justified in the previous section. A similar result is obtained for systems considered here.

**Numerical Fact 5.5** For $\Delta p_i = 0.01$, (5.24) and (5.25) practically always imply and are implied by (5.26) and (5.27), respectively.

**Justification**: See Appendix C.

**Example 5.10**: Consider the production lines shown in Figures 5.18(a) with

\[
\begin{align*}
g_3(p_3) &= \frac{1}{0.34p_3^3 + 1}, \\
g_4(p_4) &= \frac{1}{0.43p_4^3 + 1}.
\end{align*}
\]

Using simulations and calculations to identify its BN (with $\Delta p_i = 0.01$), we determine that $m_4$ is the $p^-$-BN. Decreasing its efficiency by 0.05, we observe that $m_2$ becomes the $p^+$-BN (Figure 5.18(b)). With its efficiency increased by 0.05, the bottleneck shifts to $m_3$, which is the $p^+$-BN (Figure 5.18(c)). Increasing now its efficiency by 0.1, we determine that it remains the bottleneck, however, now it is $p^-$-BN (Figure 5.18(d)). Decreasing its efficiency by 0.05, results in $m_1$ as the $p^+$-BN (Figure 5.18(e)).

This example illustrates that in serial lines with QQC machines the process of bottleneck identification and elimination is more involved than that for other serial lines.
Figure 5.18: Bottlenecks in Example 5.10
5.4 Bernoulli Lines with Rework

5.4.1 Model and approach

Model: We consider the same model as in the previous section, i.e., defined by conventions (a)-(i) of Subsection 5.3.1, along with the following additional assumptions:

(j) The merge machine $m_j$ is starved if both $b_{j-1}$ and $b_{r_{Mr}}$ are empty. The split machine $m_k$ is blocked by the main line if $b_k$ is full and $m_{k+1}$ does not take a part during this time slot; it is blocked by the rework loop if $b_{r_0}$ is full and $m_{r_1}$ does not take a part during this time slot.

(k) Parts from the repair line have a higher priority than those from the main line. In other words, $m_j$ does not take a part from $b_{j-1}$ unless $b_{r_{Mr}}$ is empty.

To simplify the presentation, we assume that $m_k$ is the only inspection machine in the system (i.e., $I_{insp} = \{k\}$) and no QQC machines are present (i.e., $|I_{QQC}| = 0$).

Clearly, in the line defined by assumptions (a)-(k), the quality buy rate $q$, i.e., the probability that a good part emerges at the output of $m_k$, depends on the quality parameters of the machines $g_i$, $i \in I_{np}$. Indeed, introduce

$$g_I = \prod_{i=1}^{j-1} g_i, \quad g_{II} = \prod_{i=j}^{k} g_i, \quad g_{III} = \prod_{i=k+1}^{M} g_i = 1, \quad g_{IV} = \prod_{i=r_1}^{r_{Mr}} g_i.$$

Figure 5.19: Serial production line with rework
Then, \( q \) can be expressed as

\[
q = g_{II} [g_I + (1 - q)g_{IV}],
\]

(5.30)
i.e.,

\[
q = \frac{g_{II} g_{IV}}{1 + (g_{IV} - g_I) g_{II}}.
\]

(5.31)

In the framework of model (a)-(k), since \( m_j \) has two upstream buffers, \( b_{j-1} \) and \( b_{r_M_r} \), and \( m_k \) has two downstream buffers, \( b_{k+1} \) and \( b_{r_0} \), their starvations and blockages are denoted as follows:

\[
ST_{j_1}^{lwr} = P[m_j \text{ is starved by } b_{j-1}],
\]

\[
ST_{j_2}^{lwr} = P[m_j \text{ is starved by } b_{r_M_r}],
\]

\[
BL_{k_1}^{lwr} = P[m_k \text{ is blocked by } b_k],
\]

\[
BL_{k_2}^{lwr} = P[m_k \text{ is blocked by } b_{r_0}] .
\]

**Approach:** Due to the complexity of Markov chains involved in their description, a direct analysis of Bernoulli lines with rework does not seem feasible. Therefore, we use a simplification technique based on overlapping decomposition, which represents the line with rework as four *overlapping* serial lines and four *virtual* serial lines shown in Figure 5.20(a) and (b), respectively. The overlapping lines include the *overlapping* machines \( m_j \) and \( m_k \), each belonging to three lines. The virtual lines do not contain overlapping machines; instead, they include six virtual machines, \( m_j^1, m_j^2, m_j^4 \) and \( m_k^2, m_k^3, m_k^4 \), the efficiencies of which are selected so as to represent the effect of the rest of the system on a particular virtual line. Calculating appropriately the efficiencies, \( p_j^1, p_j^2, p_j^4 \) and \( p_k^2, p_k^3, p_k^4 \), of the virtual machines, allows us to analyze the performance
of the original line with rework (see Appendix C).

\[ p_1, g_1, N_1 \rightarrow \ldots \rightarrow p_j, g_j, N_j \rightarrow \ldots \rightarrow p_k, g_k, N_k \rightarrow \ldots \rightarrow p_{M-1}, g_{M-1}, N_{M-1} \rightarrow p_M, N_M \]

(a) Overlapping serial lines

\[ \hat{PR}_{lwr}, \text{ see Appendix C for details of the performance evaluation method} \]

\[ PR_{lwr} \]

(b) Virtual serial lines

5.4.2 Reversibility

As it is shown in Subsection 5.2.2, serial production lines with no rework observe the property of reversibility: the production rates of a serial line and its reverse are the same. However, as we show below, reversibility does not hold for lines with rework.

Indeed, consider the serial line and its reverse shown in Figure 5.21(a) and (b), respectively. The production rates of the original and reverse lines are evaluated using both calculation (denoted as \( \hat{PR}_{lwr} \), see Appendix C for details of the performance evaluation method) and simulation (denoted as \( PR_{lwr} \)). It is obvious that the production rate of the reverse line is not the same as that of the original one both in
calculation and in simulation results (see the data of Figure 5.21). Thus, reversibility is violated. The lack of reversibility constitutes a fundamental difference between the usual (i.e., serial open) Bernoulli lines and those with rework.

\[ m_1 b_1 m_2 b_2 m_3 b_3 m_4 b_4 m_5 b_5 m_6 b_6 m_7 b_7 m_8 b_8 m_9 b_9 b_{10} \]

\[ 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.7 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad q = 0.80 \]

(a) Production System 1: \( \hat{PR}_{lwr} = 0.4779 \), \( PR_{lwr} = 0.4657 \)

\[ m_1 b_1 m_2 b_2 m_3 b_3 m_4 b_4 m_5 b_5 m_6 b_6 b_{10} \]

\[ 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.7 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad 0.9 \quad 3 \quad q = 0.80 \]

(b) Reverse of Production System 1: \( \hat{PR}_{lwr} = 0.5330 \), \( PR_{lwr} = 0.4903 \)

Figure 5.21: Production system with rework and its reverse

In addition, comparing the data of Figure 5.21, we observe that placing more efficient machines toward the end of the line yields a higher production rate than placing them upstream. This is also qualitatively different from serial lines with no rework where the position of a machine does not indicate its importance for the performance of the system.

The reasons for the loss of reversibility in Bernoulli lines with rework can be explained by the asymmetric routing of jobs at the merge and the split machines. Indeed, at the input of the merge machine, the repaired jobs have higher priority than the new ones. Also, after the split machine, the routing of the jobs is based on the quality buy rate. Finally, the merge machine is starved when both \( b_{j-1} \) and \( b_{rM} \) are empty, while, due to the blocked before service convention, the split machine does not produce a part when it is blocked either by \( b_k \) or by \( b_{r0} \). As a result, the flow of jobs in Bernoulli lines with rework is no longer reversible, and the downstream machines have a larger effect on the production rate than do the upstream ones.
Note that the above reasons do not hold if the split and the merge machines are $m_M$ and $m_1$, respectively. Thus, lines with rework, where the repaired parts are processed by all machines in the system, are reversible.

### 5.4.3 Monotonicity

**Numerical Fact 5.6** The production rate of a Bernoulli line with rework is monotonically increasing in machine and buffer parameters.

**Justification:** See Appendix C.

Thus, increasing machine efficiency and quality, as well as enlarging buffer capacity, always lead to improved throughput.

### 5.4.4 Bottlenecks

The bottleneck machine can be defined as in Definition 3.2: $m_i$ is the bottleneck (BN) if

$$\frac{\partial PR_{	ext{ler}}}{\partial p_i} > \frac{\partial PR_{	ext{ler}}}{\partial p_l}, \quad \forall l \neq i.$$  \hspace{1em} (5.32)

However, due to the split and merge operations, Bottleneck Indicator 3.1 is not directly applicable. An extension of this indicator to lines with rework is described next.

**Approach:** Using overlapping decomposition, a Bernoulli line with rework is reduced to four usual serial lines (see Figure 5.20). Their BNs can be identified using $BL_i^l$ and $ST_i^l$, $l = 1, 2, 3, 4$. This, however, would identify not the BN of the line with rework but the BNs of the serial lines, where the rest of the system is represented by virtual machines $m_1^1, m_2^2, m_3^3, m_4^4$ and $m_1^2, m_2^3, m_3^4$. In other words, each of these BNs would be a machine with the strongest effect on the overall production rate from the point of view of an individual virtual line. Although these bottlenecks may be
of interest in some applications, our goal is to identify the true BN of the line with rework in the sense of (5.32).

Therefore, rather than using the virtual lines, at the first stage of the identification procedure, we determine BNs of the four overlapping lines of Figure 5.20(a); we refer to them as local bottlenecks (LBNs). Then, at the second stage, we identify the overall bottleneck of the line with rework, which is referred to as the global bottleneck (GBN).

We describe below how LBNs can be identified (using either calculated or measured data) and show that one of them is, practically always, the GBN.

**Local bottlenecks identification:** Consider a line with rework and represent its overlapping lines as shown in Figure 5.22. Clearly, each overlapping machine enters three of them. Assume that these lines are in isolation, i.e., the first machines are not starved and the last ones are not blocked. The probabilities of blockages and starvations of all other machines can be either calculated or measured during normal system operation.

Specifically, when the parameters of the machines and buffers are known, the estimates of machine starvation and blockages, $\tilde{ST}^{lwr}_i$ and $\tilde{BL}^{lwr}_i$, can be evaluated analytically (see Appendix C). Based on these data, the BN of each overlapping line can be identified using Bottleneck Indicator 3.1.

When the parameters of the machines and buffers or the quality buy rate are not available, $ST^{lwr}_i$ and $BL^{lwr}_i$ can be evaluated during normal system operation, keeping in mind that $m_j$ and $m_k$ can be starved and blocked, respectively, by two buffers. Based on these data, again the four LBNs can be identified using Bottleneck Indicator 3.1. Clearly, the same can be carried out when a simulation model is available.

Thus, based on the data of machine blockages and starvations, one can identify four local bottlenecks of a Bernoulli line with rework.
Global bottleneck identification: The GBN identification is based on the following

**Numerical Fact 5.7** In Bernoulli lines with rework defined by assumptions (a)-(k) of Subsection 5.4.1, the GBN is, practically always, one of the four LBNs.

**Justification:** See Appendix C.

The application of this numerical fact is clear: To identify the GBN, one must test the effect of each LBN on the production rate of the system; the LBN with the largest effect is the GBN. While this process is somewhat involved, it can be facilitated by the following:

**Numerical Fact 5.8** For Bernoulli lines with rework defined by assumptions (a)-(k) of Subsection 5.4.1,

(α) if an overlapping machine is the LBN in three overlapping lines, it is practically always the GBN;
(β) if an overlapping machine is the LBN in only one overlapping line, it is practically never the GBN.

**Justification:** See Appendix C.

**Numerical Fact 5.9** In a Bernoulli line with rework defined by assumptions (a)-(k) of Subsection 5.4.1 and with quality buy rate $q^*$,

(α) if GBN is a non-overlapping machine of line 1, then this machine is practically always the GBN for all $q > q^*$;

(β) if GBN is a non-overlapping machine of line 3, then this machine is practically always the GBN for all $q > q^*$;

(γ) if GBN is a non-overlapping machine of line 4, then this machine is practically always the GBN for all $q < q^*$;

(δ) if GBN is a non-overlapping machine of line 2 or line 4, then the GBN is practically always in either line 2 or line 4 for all $q < q^*$.

**Justification:** See Appendix C.

**Example 5.11:** Consider the line with rework shown in Figure 5.23 along with corresponding overlapping lines 1-4. Their local bottlenecks, identified both by calculation and measurement-based approaches (using simulations), are also indicated. Since $m_4$ is the LBN in only one of the overlapping lines, according to Numerical Fact 5.8(β) it is not the GBN. Thus, the candidates are $m_7$ and $m_{r1}$. Increasing their efficiencies by 0.01, we determine that $m_{r1}$ is the GBN. According to Numerical Fact 5.9(γ), this machine remains the GBN for all quality buy rate less than 0.8.

Note that in this case (as well as in all other cases considered below), the two-stage procedure identifies correctly the GBN (verified by increasing each $p_i$, $i \in I_m$, by 0.01 and evaluating the resulting effect on the PR).
Figure 5.23: GBN identification of Example 5.11 for $q = 0.8$
Assume now that the quality buy rate in this system increases to 0.85. Then, as it follows from Figure 5.24 and Numerical Fact 5.8(β), the GBN is either $m_7$ or $m_6$. Increasing their efficiencies by 0.01 leads to the conclusion that $m_6$ is the GBN. Note

that it is not the worst machine of the main line.

Further, Figures 5.25 and 5.26 show that the GBNs are $m_8$ and $m_1$, if the quality buy rates are 0.9 and 0.95, respectively.

Figure 5.24: GBN identification of Example 5.11 for $q = 0.85$
Figure 5.25: GBN identification of Example 5.11 for $q = 0.9$
Figure 5.26: GBN identification of Example 5.11 for $q = 0.95$. 

(a) LBN identification of overlapping line 1

(b) LBN identification of overlapping line 2

(c) LBN identification of overlapping line 3

(d) LBN identification of overlapping line 4
This example indicates that *bottlenecks may be shifting not only because of changes in the machine and buffer parameters but also due to changes in the quality buy rates.* Thus, BNs must be tracked when the product quality is fluctuating.

**Example 5.12:** Consider the system shown in Figure 5.27. Assume first that the quality buy rate is 1, i.e., the rework loop is not activated. Then, since the machine efficiencies are allocated symmetrically and the buffers are identical, machines \( m_2 \) and \( m_9 \) are the bottlenecks, and they have an identical effect on the system’s production rate.

On the other hand, when \( q < 1 \), for instance, \( q = 0.95 \), the data of Figure 5.27 show that the GBN is unique: machine \( m_9 \). This difference between lines with and without the rework loop is due to the lack of reversibility in the former and the fact that downstream machines have a larger effect on the production rate than those upstream.
Figure 5.27: GBN identification of Example 5.12

(a) LBN identification of overlapping line 1

(b) LBN identification of overlapping line 2

(c) LBN identification of overlapping line 3

(d) LBN identification of overlapping line 4

\[
\Delta P_{R_{\text{bar}}}/\Delta p = 0.18, \quad \Delta P_{R_{\text{bar}}}/\Delta p = 0.12, \quad \Delta P_{R_{\text{bar}}}/\Delta p = 0.23
\]

\[
\Delta P_{R_{\text{bar}}}/\Delta p = 0.28, \quad \Delta P_{R_{\text{bar}}}/\Delta p = 0.13, \quad \Delta P_{R_{\text{bar}}}/\Delta p = 0.31
\]
5.5 Case Study: Automotive Paint Shop Production System

5.5.1 System model and problem formulation

System layout: In this case study, we apply the method developed above to a paint shop at a different assembly plant than that considered in Section 4.4. The layout of this paint shop is illustrated in Figure 5.28. The system consists of 20 operations, where the jobs are cleaned, sealed, sanded, painted, and finessed. The quality of jobs is inspected at Ops. 6, 15, and 17. The defective jobs are repaired at Ops. 18, 19, and 20. Due to capacity reasons, Ops. 12-14 consist of two parallel lines, referred to as east and west. The west line processes only the first time jobs (i.e., the jobs that did not undergo repairs), while the east line processes both first time and repaired jobs.

![Figure 5.28: Layout of the paint shop system](image)

All jobs are transported by conveyors on three types of carriers: the body shop carriers, paint shop carriers, and general assembly carriers. The transfer from the body shop carriers to the paint shop carriers occurs in front of Op. 5, while the transfer from the paint shop carriers to the general assembly carriers takes place after Op. 17. Thus, the paint shop is a closed serial line with parallel operations and rework.
Structural modeling: The system under consideration operates under a control policy, which ensures that Op. 5 and Op. 17 are never starved for or blocked by paint shop carriers, respectively. Therefore, the closed loop with respect to paint shop carriers does not impede the system performance and, thus, can be ignored. In addition, since, according to the factory floor measurements, the probability of Op. 4 blockage by body shop carriers and the probability of Op. 17 starvation by general assembly carriers are less than 0.001, the effects of two other loops can also be ignored. Thus, we model the paint shop as an open, rather than closed, system.

Further, we simplify the model by combining the rework loops and aggregating parallel operations. Specifically, based on the factory floor measurements, the number of defective jobs identified at Ops. 6 and 17 is substantially less than that at Op. 15. Therefore, the repair loops of Ops. 19 and 20 are omitted, and all defectives are assumed to be identified at Op. 15. Thus, after aggregating parallel machines of Ops. 12-14, the structural model of the system takes the form shown in Figure 5.29, which is an open serial line with rework.

![Figure 5.29: Structural model of the paint shop system](image)

The quality issues in this system are modeled as follows: According to the factory floor data, defective jobs are produced mainly at Ops. 10 (manual wet sand), 12 (automated paint booth), and 18 (repair). Therefore, these operations are modeled as non-perfect quality machines indicated in Figure 5.29 by shaded circles. In addition, the manual wet sanding of dust and scratches, carried out at Op. 10, results in low quality when the speed of the operational conveyor is increased and high quality when the speed is decreased. Therefore, we model Op. 10 as a QQC machine, although its QQC characteristics are not available from the factory floor measurements. (Note that
these characteristics are identified in Subsection 5.5.3 based on the improved system operation, see Figure 5.32.) To reflect the QQC property, Op. 10 is represented in Figure 5.29 by additional shading (i.e., the circle is double shaded).

The structural model of Figure 5.29 is the basis for the continuous improvement project described in this application.

**Parameters identification:** Below, the process of identification of productivity and quality parameters for all operations is described.

*Productivity parameters:* To determine the productivity parameters of each operation, three characteristics have been measured on the factory floor during one month of system operation (referred to as Weeks 1-4):

- The operation *efficiency* \(e_i\), i.e., the fraction of time Op. \(i\) is up during the total time of system operation, \(i = 1, \ldots, 18\).

- The *cycle time* \(\tau_i\), i.e., the time (in hours) necessary to process a job by Op. \(i\); the inverse of \(\tau_i\) is referred to as the *capacity* \(c_i\) (jobs/hour) of Op. \(i\), \(i = 1, \ldots, 18\).

- The *downtime* \(T_{\text{down},i}\), i.e., the average number of cycle times Op. \(i\) is down when it fails, \(i = 1, \ldots, 18\).

The operation downtimes evaluated over Weeks 1-4 are summarized in Table 5.2, while the operation efficiencies and capacities for each week are given in Tables 5.3 and 5.4. As one can see, the average downtime of most operations is of the order of magnitude of their cycle times. This is because the downtime in this paint shop system is not due to mechanical breakdowns, but rather due to quality reasons, whereby an operator stops a part of the operational conveyor in order to finish the job with the highest possible quality. Therefore, the Bernoulli reliability model, according to which each operation produces a part during a cycle time with probability \(p_i\) and fails to
do so with probability $1 - p_i$, is applicable and used in this case study. Since each operation has a different cycle time, to obtain the Bernoulli model of the system as a whole, we calculate $p_i$’s by normalizing the capacities by the highest capacity among all operations, i.e., by $c_{14}$. Specifically:

- For Ops. 1-11, 15-18, the Bernoulli parameters are calculated as

$$p_i = \frac{c_i}{c_{14}} e_i, \quad i = 1, \ldots, 11, 15, \ldots, 18,$$

where

$$c_{14} = c_{14,e} + c_{14,w}.$$  \hspace{1cm} (5.34)

- For Ops. 12-14, the parallel operations are aggregated and the Bernoulli parameters for the resulting operations are calculated using

$$p_i = \frac{c_{i,e} e_{i,e} + c_{i,w} e_{i,w}}{c_{14}}, \quad i = 12, \ldots, 14.$$  \hspace{1cm} (5.35)

- For Op. 17, since it can be blocked by General Assembly, the Bernoulli parameter is calculated as

$$p_{17} = \frac{c_{17}}{c_{14}} e_{17}(1 - P_{BL}),$$

where $P_{BL}$ is the probability that Op. 17 is blocked by General Assembly. During Weeks 1-4, this probability was evaluated as

- Week 1: $P_{BL} = 0.0253$,
- Week 2: $P_{BL} = 0.0143$,
- Week 3: $P_{BL} = 0.0288$,
- Week 4: $P_{BL} = 0.0208$.

The $p_i$’s thus obtained and the buffer capacities, $N_i$’s, identified on the factory floor, are summarized in Tables 5.5 and 5.6, respectively. The data of Tables 5.5 and
<table>
<thead>
<tr>
<th>Operation</th>
<th>$T_{down,i}$</th>
<th>Operation</th>
<th>$T_{down,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 1</td>
<td>7.26</td>
<td>Op. 12e</td>
<td>0.51</td>
</tr>
<tr>
<td>Op. 2</td>
<td>5.32</td>
<td>Op. 12w</td>
<td>0.75</td>
</tr>
<tr>
<td>Op. 3</td>
<td>6.52</td>
<td>Op. 13e</td>
<td>0.45</td>
</tr>
<tr>
<td>Op. 4</td>
<td>1.02</td>
<td>Op. 13w</td>
<td>0.43</td>
</tr>
<tr>
<td>Op. 5</td>
<td>3.18</td>
<td>Op. 14e</td>
<td>1.17</td>
</tr>
<tr>
<td>Op. 6</td>
<td>0.36</td>
<td>Op. 14w</td>
<td>0.88</td>
</tr>
<tr>
<td>Op. 7</td>
<td>1.73</td>
<td>Op. 15</td>
<td>7.48</td>
</tr>
<tr>
<td>Op. 8</td>
<td>0.88</td>
<td>Op. 16</td>
<td>0.80</td>
</tr>
<tr>
<td>Op. 9</td>
<td>0.79</td>
<td>Op. 17</td>
<td>1.13</td>
</tr>
<tr>
<td>Op. 10</td>
<td>0.58</td>
<td>Op. 18</td>
<td>0.62</td>
</tr>
<tr>
<td>Op. 11</td>
<td>1.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th>$e_i$</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 1</td>
<td>0.9997</td>
<td>0.9999</td>
<td>1.0000</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td>Op. 2</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9995</td>
<td>0.9744</td>
<td></td>
</tr>
<tr>
<td>Op. 3</td>
<td>0.9958</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9777</td>
<td></td>
</tr>
<tr>
<td>Op. 4</td>
<td>0.9352</td>
<td>0.9594</td>
<td>0.9264</td>
<td>0.9285</td>
<td></td>
</tr>
<tr>
<td>Op. 5</td>
<td>0.9945</td>
<td>0.9916</td>
<td>0.9935</td>
<td>0.9837</td>
<td></td>
</tr>
<tr>
<td>Op. 6</td>
<td>0.9999</td>
<td>0.9995</td>
<td>0.9998</td>
<td>0.9993</td>
<td></td>
</tr>
<tr>
<td>Op. 7</td>
<td>0.9565</td>
<td>0.9609</td>
<td>0.9333</td>
<td>0.9401</td>
<td></td>
</tr>
<tr>
<td>Op. 8</td>
<td>0.9688</td>
<td>0.9617</td>
<td>0.9757</td>
<td>0.9829</td>
<td></td>
</tr>
<tr>
<td>Op. 9</td>
<td>0.9974</td>
<td>0.9835</td>
<td>0.9854</td>
<td>0.9815</td>
<td></td>
</tr>
<tr>
<td>Op. 10</td>
<td>0.8846</td>
<td>0.9112</td>
<td>0.8951</td>
<td>0.8894</td>
<td></td>
</tr>
<tr>
<td>Op. 11</td>
<td>0.9896</td>
<td>0.9686</td>
<td>0.9963</td>
<td>0.9981</td>
<td></td>
</tr>
<tr>
<td>Op. 12e</td>
<td>0.9834</td>
<td>0.9836</td>
<td>0.9756</td>
<td>0.9754</td>
<td></td>
</tr>
<tr>
<td>Op. 12w</td>
<td>0.9271</td>
<td>0.9727</td>
<td>0.9739</td>
<td>0.9682</td>
<td></td>
</tr>
<tr>
<td>Op. 13e</td>
<td>0.9971</td>
<td>0.9982</td>
<td>0.9998</td>
<td>0.9981</td>
<td></td>
</tr>
<tr>
<td>Op. 13w</td>
<td>0.9957</td>
<td>0.9980</td>
<td>0.9979</td>
<td>0.9973</td>
<td></td>
</tr>
<tr>
<td>Op. 14e</td>
<td>0.8923</td>
<td>0.8690</td>
<td>0.8722</td>
<td>0.8681</td>
<td></td>
</tr>
<tr>
<td>Op. 14w</td>
<td>0.9092</td>
<td>0.8947</td>
<td>0.8620</td>
<td>0.9116</td>
<td></td>
</tr>
<tr>
<td>Op. 15</td>
<td>0.9961</td>
<td>0.9948</td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Op. 16</td>
<td>0.9087</td>
<td>0.9059</td>
<td>0.8850</td>
<td>0.8780</td>
<td></td>
</tr>
<tr>
<td>Op. 17</td>
<td>0.9909</td>
<td>0.9877</td>
<td>0.9903</td>
<td>0.9930</td>
<td></td>
</tr>
<tr>
<td>Op. 18</td>
<td>0.7068</td>
<td>0.8039</td>
<td>0.7692</td>
<td>0.8069</td>
<td></td>
</tr>
<tr>
<td>Operation</td>
<td>Week 1</td>
<td>Week 2</td>
<td>Week 3</td>
<td>Week 4</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Op. 1</td>
<td>75.22</td>
<td>75.25</td>
<td>75.27</td>
<td>75.31</td>
<td></td>
</tr>
<tr>
<td>Op. 2</td>
<td>77.42</td>
<td>76.76</td>
<td>76.47</td>
<td>76.53</td>
<td></td>
</tr>
<tr>
<td>Op. 3</td>
<td>74.23</td>
<td>75.03</td>
<td>74.69</td>
<td>74.97</td>
<td></td>
</tr>
<tr>
<td>Op. 4</td>
<td>73.11</td>
<td>73.08</td>
<td>73.13</td>
<td>73.10</td>
<td></td>
</tr>
<tr>
<td>Op. 5</td>
<td>84.07</td>
<td>82.85</td>
<td>83.75</td>
<td>84.17</td>
<td></td>
</tr>
<tr>
<td>Op. 6</td>
<td>75.60</td>
<td>75.31</td>
<td>76.35</td>
<td>76.96</td>
<td></td>
</tr>
<tr>
<td>Op. 7</td>
<td>74.97</td>
<td>75.31</td>
<td>75.04</td>
<td>76.11</td>
<td></td>
</tr>
<tr>
<td>Op. 8</td>
<td>74.17</td>
<td>74.35</td>
<td>73.88</td>
<td>74.29</td>
<td></td>
</tr>
<tr>
<td>Op. 9</td>
<td>73.92</td>
<td>74.23</td>
<td>73.77</td>
<td>74.19</td>
<td></td>
</tr>
<tr>
<td>Op. 10</td>
<td>74.20</td>
<td>74.17</td>
<td>74.15</td>
<td>74.17</td>
<td></td>
</tr>
<tr>
<td>Op. 11</td>
<td>46.82</td>
<td>46.62</td>
<td>46.72</td>
<td>46.81</td>
<td></td>
</tr>
<tr>
<td>Op. 12e</td>
<td>45.06</td>
<td>45.07</td>
<td>45.06</td>
<td>45.06</td>
<td></td>
</tr>
<tr>
<td>Op. 12w</td>
<td>45.07</td>
<td>45.07</td>
<td>45.01</td>
<td>45.02</td>
<td></td>
</tr>
<tr>
<td>Op. 13e</td>
<td>45.10</td>
<td>45.08</td>
<td>45.11</td>
<td>45.08</td>
<td></td>
</tr>
<tr>
<td>Op. 13w</td>
<td>45.03</td>
<td>45.06</td>
<td>45.06</td>
<td>45.06</td>
<td></td>
</tr>
<tr>
<td>Op. 14e</td>
<td>47.53</td>
<td>47.59</td>
<td>47.56</td>
<td>47.54</td>
<td></td>
</tr>
<tr>
<td>Op. 14w</td>
<td>47.49</td>
<td>46.97</td>
<td>47.49</td>
<td>47.44</td>
<td></td>
</tr>
<tr>
<td>Op. 15</td>
<td>89.00</td>
<td>87.83</td>
<td>87.59</td>
<td>88.45</td>
<td></td>
</tr>
<tr>
<td>Op. 16</td>
<td>75.22</td>
<td>75.28</td>
<td>75.39</td>
<td>75.03</td>
<td></td>
</tr>
<tr>
<td>Op. 17</td>
<td>68.19</td>
<td>68.42</td>
<td>68.72</td>
<td>68.75</td>
<td></td>
</tr>
<tr>
<td>Op. 18</td>
<td>33.83</td>
<td>33.75</td>
<td>33.83</td>
<td>33.87</td>
<td></td>
</tr>
</tbody>
</table>
5.6 constitute the Bernoulli productivity model of the paint shop system.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 1</td>
<td>0.7926</td>
<td>0.7929</td>
<td>0.7932</td>
<td>0.7936</td>
</tr>
<tr>
<td>Op. 2</td>
<td>0.8158</td>
<td>0.8088</td>
<td>0.8054</td>
<td>0.7858</td>
</tr>
<tr>
<td>Op. 3</td>
<td>0.7789</td>
<td>0.7906</td>
<td>0.7870</td>
<td>0.7723</td>
</tr>
<tr>
<td>Op. 4</td>
<td>0.7205</td>
<td>0.7388</td>
<td>0.7139</td>
<td>0.7152</td>
</tr>
<tr>
<td>Op. 5</td>
<td>0.8811</td>
<td>0.8657</td>
<td>0.8769</td>
<td>0.8725</td>
</tr>
<tr>
<td>Op. 6</td>
<td>0.7965</td>
<td>0.7932</td>
<td>0.8044</td>
<td>0.8104</td>
</tr>
<tr>
<td>Op. 7</td>
<td>0.7556</td>
<td>0.7626</td>
<td>0.7379</td>
<td>0.7539</td>
</tr>
<tr>
<td>Op. 8</td>
<td>0.7571</td>
<td>0.7535</td>
<td>0.7597</td>
<td>0.7694</td>
</tr>
<tr>
<td>Op. 9</td>
<td>0.7769</td>
<td>0.7693</td>
<td>0.7660</td>
<td>0.7673</td>
</tr>
<tr>
<td>Op. 10</td>
<td>0.6916</td>
<td>0.7121</td>
<td>0.6994</td>
<td>0.6951</td>
</tr>
<tr>
<td>Op. 11</td>
<td>0.9816</td>
<td>0.9671</td>
<td>0.9829</td>
<td>0.9855</td>
</tr>
<tr>
<td>Op. 12</td>
<td>0.9072</td>
<td>0.9290</td>
<td>0.9251</td>
<td>0.9224</td>
</tr>
<tr>
<td>Op. 13</td>
<td>0.9463</td>
<td>0.9480</td>
<td>0.9491</td>
<td>0.9476</td>
</tr>
<tr>
<td>Op. 14</td>
<td>0.9019</td>
<td>0.8787</td>
<td>0.8685</td>
<td>0.8906</td>
</tr>
<tr>
<td>Op. 15</td>
<td>0.9342</td>
<td>0.9207</td>
<td>0.9230</td>
<td>0.9321</td>
</tr>
<tr>
<td>Op. 16</td>
<td>0.7202</td>
<td>0.7186</td>
<td>0.7031</td>
<td>0.6941</td>
</tr>
<tr>
<td>Op. 17</td>
<td>0.6941</td>
<td>0.7019</td>
<td>0.6964</td>
<td>0.7045</td>
</tr>
<tr>
<td>Op. 18</td>
<td>0.2424</td>
<td>0.2859</td>
<td>0.2743</td>
<td>0.2880</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th>$N_i$</th>
<th>Operation</th>
<th>$N_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 2</td>
<td>9</td>
<td>Op. 11</td>
<td>8</td>
</tr>
<tr>
<td>Op. 3</td>
<td>27</td>
<td>Op. 12</td>
<td>68</td>
</tr>
<tr>
<td>Op. 4</td>
<td>2</td>
<td>Op. 13</td>
<td>85</td>
</tr>
<tr>
<td>Op. 6</td>
<td>3</td>
<td>Op. 15</td>
<td>3</td>
</tr>
<tr>
<td>Op. 7</td>
<td>4</td>
<td>Op. 16</td>
<td>4</td>
</tr>
<tr>
<td>Op. 8</td>
<td>11</td>
<td>Op. 17</td>
<td>–</td>
</tr>
<tr>
<td>Op. 9</td>
<td>114</td>
<td>Op. 18</td>
<td>5, 12</td>
</tr>
</tbody>
</table>

**Quality parameters:** To construct the quality model, we note that in automotive paint shops, most uncorrelated defects are due to dust on the factory floor and scratches of car bodies, while correlated defects are relatively rare. Therefore, the defects can be viewed as the results of independent random events, and the Bernoulli
quality model is applicable, i.e., a job produced by a non-perfect quality machine is
good with probability \( g_i \), and defective with probability \( 1 - g_i \), \( i = 10, 12, 18 \). Parameter \( g_i \) is referred to as the quality of Op. \( i \). Unfortunately, these parameters are not available from the factory floor measurements. What is available is the quality buy rate, \( q \), i.e., the frequency with which a job is identified as non-defective at the output of Op. 15. In addition, since the quality of each job is determined at Op. 15 and confirmed at Op. 17 (see Figure 5.28), good jobs are very unlikely to be identified as defectives. Therefore, we assume that the defects are identified perfectly, i.e., no defects are missed and no good jobs are identified as defectives. Then, it follows from the total probability formula that

\[
q = q \cdot g_{10}g_{12} + (1 - q) \cdot g_{12}g_{18}.
\] (5.37)

Therefore, the quality buy rate of the system can be expressed as

\[
q = \frac{g_{12}g_{18}}{1 - g_{10}g_{12} + g_{12}g_{18}}.
\] (5.38)

Based on factory floor measurements, the quality buy rate for Weeks 1-4 has been identified as

- Week 1: \( q = 0.7102 \),
- Week 2: \( q = 0.7132 \),
- Week 3: \( q = 0.7111 \),
- Week 4: \( q = 0.7165 \).

Clearly, this does not allow us to identify the quality parameters of Ops. 10, 12 and 18 or the QQC function of Op. 10. However, as it turns out, this information is sufficient for the model validation and the subsequent development of an improvement project.

**Model validation** Using the performance evaluation technique described in Subsection 5.4.1, we estimate of the production rate, \( \overline{PR} \) of the system considered. Then,
using $\widehat{PR}$, the estimate, $\widehat{TP}$, of the throughput is calculated as follows:

$$\widehat{TP} = c_{14}\widehat{PR}. \quad (5.39)$$

Based on this approach and the data of Tables 5.5 and 5.6, we calculate $\widehat{TP}$ for Weeks 1-4 and compare it with $TP$ measured on the factory floor. To evaluate the accuracy of the model, introduce

$$\epsilon_{TP} = \frac{|TP - \widehat{TP}|}{TP} \cdot 100\%. \quad (5.40)$$

The results are summarized in Table 5.7. Clearly, the fidelity of the model for Weeks 1-4 is sufficiently high. In general, more measurement data would be desirable to validate the model. However, since no such measurements are available, we accept this model and use it for subsequent developments. (Note that in Subsection 5.5.2, when one more month data become available, we validate the model again and find it to be sufficiently precise.)

**Goal of the improvement project:** As it follows from Tables 5.3 and 5.4, the throughput in isolation (i.e., when neither blockages nor starvations take place) of each operation in the direct path of the system of Figure 5.29 is quite high. In fact,
the worst from this point of view is Op. 17, the isolation throughput of which is

Week 1: \( c_{17}e_{17} = 67.58 \) jobs/hour,

Week 2: \( c_{17}e_{17} = 67.58 \) jobs/hour,

Week 3: \( c_{17}e_{17} = 68.05 \) jobs/hour,

Week 4: \( c_{17}e_{17} = 68.27 \) jobs/hour.

Since, the overall system throughput is 54 jobs/hour, the losses are more than 13 jobs/hour, i.e., over 20%.

These losses are due to two reasons. The first one is obvious: the quality buy rate is low \( (q \approx 0.7) \), which implies that about 30% of jobs are processed at least twice. Less obvious are losses due to the blockages and starvations of the operations (by full and empty buffers, respectively). Since the latter cannot be investigated analytically without knowing the quality parameters of Ops. 10, 12 and 18, the goal of the project is formulated in two stages:

**Stage 1:** Suggest and implement an improvement project, which would increase the quality buy rate and, thus, the \( TP \).

**Stage 2:** Based on the results of Stage 1, suggest further improvements that would reduce operation blockages and starvations and, thus, contribute to \( TP \) increase.

The results obtained for each of these stages are reported Subsections 5.5.2 and 5.5.3.

**5.5.2 Improvement project: Stage 1**

**Potential improvement by increasing quality buy rate:** To analyze the potential improvement by increasing the quality buy rate, we calculate \( \hat{TP} \) of the system, based on the data of Weeks 1-4 shown in Table 5.5, for various \( q \)'s. The results are given in Figure 5.30. From this figure, we conclude that increasing \( q \) from 0.7 to
0.9 leads to a \( \hat{TP} \) improvement of about 10 jobs/hour. Below, we address system improvement first through \( q \) and then through blockages and starvations.

![Figure 5.30: Throughput as a function of quality buy rate](image)

**Quality bottleneck identification:** The Q-BN is determined by evaluating \( \partial TP / \partial g_i \), \( i = 10, 12, 18 \):

\[
\begin{align*}
    \frac{\partial TP}{\partial g_{10}} &= \frac{\partial TP}{\partial q} \cdot \frac{\partial q}{\partial g_{10}} = \frac{\partial TP}{\partial q} \cdot g_{12} g_{18} (1 - g_{10} g_{12} + g_{12} g_{18})^2, \\
    \frac{\partial TP}{\partial g_{12}} &= \frac{\partial TP}{\partial q} \cdot \frac{\partial q}{\partial g_{12}} = \frac{\partial TP}{\partial q} \cdot g_{18} (1 - g_{10} g_{12} + g_{12} g_{18})^2, \\
    \frac{\partial TP}{\partial g_{18}} &= \frac{\partial TP}{\partial q} \cdot \frac{\partial q}{\partial g_{18}} = \frac{\partial TP}{\partial q} \cdot g_{12} (1 - g_{10} g_{12}) (1 - g_{10} g_{12} + g_{12} g_{18})^2.
\end{align*}
\]

(5.41)

Since \( 0 < g_i < 1 \), \( i = 10, 12, 18 \), from (5.41), we obtain

\[
    \frac{\partial TP}{\partial g_{10}} < \frac{\partial TP}{\partial g_{12}}.
\]

Moreover, since \( q > 0.5 \), it follows immediately from (5.38) that

\[
    g_{12} g_{18} > 1 - g_{10} g_{12}
\]

and, therefore,

\[
    \frac{\partial TP}{\partial g_{18}} < \frac{\partial TP}{\partial g_{10}} < \frac{\partial TP}{\partial g_{12}}.
\]

(5.42)
Thus, Op. 12 is the Q-BN of the system.

**Suggested improvement:** Although, as it is shown above, the Q-BN is Op. 12 (automated paint booths), its quality improvement is much more difficult and costly than improving the next most constraining operation, i.e., Op. 10 (manual wet sand). Therefore, the quality buy rate improvement efforts of this project have been centered on Op. 10.

Since Op. 10 is viewed as QQC, the approach to its quality improvement is to slow it down, so that the operators could complete each job with a higher quality. Since this would lead to a reduced capacity, it is suggested to add a parallel station to Op. 10 (see Figure 4), so that the total capacity would not be compromised. The new stations, Op. 10w and Op. 10e, are expected to be staffed by the same personnel as the original Op. 10 but distributed among the two stations.

![Figure 5.31: Suggested modification of Op. 10](image)

Since the QQC function of Op. 10 is unknown, specific values of the capacities of Ops. 10w and 10e could not be calculated analytically. However, based on experience and intuition, the management decided to reduce the capacity of Op. 10 from 74 jobs/hour to 45 jobs/hour in both Ops. 10w and 10e. Thus, the combined capacity of Ops. 10w and 10e is increased from 74 to 90 jobs/hour. (Note that if the quality buy rate were to remain the same, this capacity modification would result in just 0.01 jobs/hour improvement of the throughput.)

The modified system has been put in place, and the results of the next month operation are described below (Weeks 5-8).
Implementation results: The quality buy rate and the throughput during Weeks 5-8 of the modified system operation were identified through factory floor measurements as

Week 5: \( q = 0.7933, \quad TP = 58.21 \text{ jobs/hour}, \)
Week 6: \( q = 0.8013, \quad TP = 59.88 \text{ jobs/hour}, \)
Week 7: \( q = 0.7947, \quad TP = 58.12 \text{ jobs/hour}, \)
Week 8: \( q = 0.7987, \quad TP = 59.11 \text{ jobs/hour}. \)

Thus, both quality buy rate and throughput are increased about 10%, as compared to the data of Weeks 1-4 (see Subsection 5.5.1).

To construct and validate the mathematical model of the modified system, the performance characteristics, described in Subsection 5.5.1, have been measured during Weeks 5-8. Based on these measurements, the Bernoulli model has been constructed using expressions (5.33)-(5.36) along with

\[
p_{10} = \frac{c_{10,e}e_{10,e} + c_{10,w}e_{10,w}}{c_{14}}. \tag{5.43}
\]

The results are summarized in Table 5.8. Using these data, we evaluate the accuracy of the model for the modified system (see Table 5.9). Although the estimate \( \hat{TP} \) has the error about 5% in comparison with the \( TP \) measured on the factory floor, we still view the model acceptable, since the measurement data used for its development may also have errors of the same order of magnitude. Below, we use this model to suggest further system improvements and evaluate their efficacy.

5.5.3 Improvement project: Stage 2

QQC function identification: Let \( p'_{10} \) denote the Bernoulli parameter of one station of Op. 10 during Weeks 1-8. Since in Weeks 1-4 Op. 10 consisted of one
Table 5.8: Operation productivity parameters $p_i$ of Weeks 5-8

<table>
<thead>
<tr>
<th>Operation</th>
<th>Week 5</th>
<th>Week 6</th>
<th>Week 7</th>
<th>Week 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Op. 1</td>
<td>0.7832</td>
<td>0.7893</td>
<td>0.7898</td>
<td>0.7739</td>
</tr>
<tr>
<td>Op. 2</td>
<td>0.8117</td>
<td>0.8080</td>
<td>0.7931</td>
<td>0.7698</td>
</tr>
<tr>
<td>Op. 3</td>
<td>0.7801</td>
<td>0.7816</td>
<td>0.7870</td>
<td>0.7726</td>
</tr>
<tr>
<td>Op. 4</td>
<td>0.7059</td>
<td>0.7236</td>
<td>0.7464</td>
<td>0.7231</td>
</tr>
<tr>
<td>Op. 5</td>
<td>0.8735</td>
<td>0.8637</td>
<td>0.8780</td>
<td>0.8725</td>
</tr>
<tr>
<td>Op. 6</td>
<td>0.7612</td>
<td>0.7541</td>
<td>0.7802</td>
<td>0.7843</td>
</tr>
<tr>
<td>Op. 7</td>
<td>0.7588</td>
<td>0.7522</td>
<td>0.7570</td>
<td>0.7512</td>
</tr>
<tr>
<td>Op. 8</td>
<td>0.7675</td>
<td>0.7603</td>
<td>0.7578</td>
<td>0.7605</td>
</tr>
<tr>
<td>Op. 9</td>
<td>0.7787</td>
<td>0.7819</td>
<td>0.7744</td>
<td>0.7801</td>
</tr>
<tr>
<td>Op. 10</td>
<td>0.8203</td>
<td>0.8420</td>
<td>0.7988</td>
<td>0.8145</td>
</tr>
<tr>
<td>Op. 11</td>
<td>0.9827</td>
<td>0.9738</td>
<td>0.9736</td>
<td>0.9850</td>
</tr>
<tr>
<td>Op. 12</td>
<td>0.9173</td>
<td>0.9124</td>
<td>0.9321</td>
<td>0.9094</td>
</tr>
<tr>
<td>Op. 13</td>
<td>0.9479</td>
<td>0.9467</td>
<td>0.9490</td>
<td>0.9456</td>
</tr>
<tr>
<td>Op. 14</td>
<td>0.9495</td>
<td>0.9435</td>
<td>0.9708</td>
<td>0.9269</td>
</tr>
<tr>
<td>Op. 15</td>
<td>0.9378</td>
<td>0.9235</td>
<td>0.9213</td>
<td>0.9321</td>
</tr>
<tr>
<td>Op. 16</td>
<td>0.9804</td>
<td>0.9861</td>
<td>0.9887</td>
<td>0.9884</td>
</tr>
<tr>
<td>Op. 17</td>
<td>0.9462</td>
<td>0.9555</td>
<td>0.9481</td>
<td>0.9393</td>
</tr>
<tr>
<td>Op. 18</td>
<td>0.6839</td>
<td>0.7162</td>
<td>0.7149</td>
<td>0.7220</td>
</tr>
</tbody>
</table>

Table 5.9: Model validation for Weeks 5-8

<table>
<thead>
<tr>
<th></th>
<th>$TP$ (jobs/hour)</th>
<th>$\hat{TP}$ (jobs/hour)</th>
<th>$\epsilon_{TP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 5</td>
<td>58.21</td>
<td>61.32</td>
<td>5.35%</td>
</tr>
<tr>
<td>Week 6</td>
<td>59.88</td>
<td>62.51</td>
<td>4.39%</td>
</tr>
<tr>
<td>Week 7</td>
<td>58.12</td>
<td>61.50</td>
<td>5.82%</td>
</tr>
<tr>
<td>Week 8</td>
<td>59.11</td>
<td>61.59</td>
<td>4.20%</td>
</tr>
</tbody>
</table>

114
station and in Weeks 5-8 of two stations,

\[
p'_{10} = \begin{cases} 
  p_{10}, & \text{Weeks 1-4}, \\
  \frac{p_{10}}{2}, & \text{Weeks 5-8}.
\end{cases}
\]  

(5.44)

Given that the only change between Weeks 1-4 and Weeks 5-8 is the modification of Op. 10, we approximate \( q \) as a linear function of \( p'_{10} \) using least squares method. The results are shown in Figure 5.32 and we obtain the QQC relationship as follows:

\[
q(p'_{10}) = 0.9152 - 0.2891 p'_{10}.
\]  

(5.45)

Using this relationship, we identify below the TP-BNs of the modified system and predict potential improvements through their elimination.

**Throughput bottleneck identification and potential improvements:** Using the Week 8 data and inequality (5.20), we identify TP-BNs by evaluating \( \frac{\Delta TP}{\Delta p_i} \) for all operations with \( \Delta p_i = 0.01 \). All TP-BNs identified are \( p^+\)-BN since, as it turns out, the QQC operation, Op. 10, is never the TP-BN.

Specifically, the TP-BN of the modified paint shop (i.e., the paint shop with two stations in Op. 10) is Op. 17. Increasing its Bernoulli parameter by 10%, we calculate
that the resulting $\hat{T}P$ is 63.73 jobs/hour, and the new TP-BN is Op. 4. Next, we increase the Op. 4 Bernoulli parameter, identify the new TP-BN and so on, until $\hat{T}P$ is sufficiently close to its maximum value. These steps are summarized in Table 4 along with the predicted $TP$. Thus, if carried out, Stage 2 would result in 9.5% additional throughput improvement. This implies that Stages 1 and 2 combined would bring about the throughput improvement of 14 jobs/hour as compared with original system performance.

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Resulting $\hat{T}P$</th>
<th>New TP-BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>−</td>
<td>61.59</td>
<td>Op. 17</td>
</tr>
<tr>
<td>Increase $p_{17}$ by 10%</td>
<td>63.73</td>
<td>Op. 4</td>
</tr>
<tr>
<td>Increase $p_{4}$ by 10%</td>
<td>63.86</td>
<td>Op. 15</td>
</tr>
<tr>
<td>Increase $p_{15}$ by 5%</td>
<td>65.20</td>
<td>Op. 7</td>
</tr>
<tr>
<td>Increase $p_{7}$ by 10%</td>
<td>65.39</td>
<td>Op. 16</td>
</tr>
<tr>
<td>Increase $p_{16}$ by 10%</td>
<td>67.44</td>
<td>Op. 6</td>
</tr>
</tbody>
</table>

5.6 Summary

- Reversibility and monotonicity properties of Bernoulli lines with non-perfect quality machines are similar to those with only perfect quality machines.

- Production lines with non-perfect quality and inspection machines have two types of bottlenecks: production rate bottlenecks (PR-BN) and quality bottlenecks (Q-BN).

- In a special case where the only inspection machine is the last one, PR-BN can be identified using Bottleneck Indicator 3.1 and Q-BN is the machine with the worst quality. Also, PR- and Q-BN are decoupled in the sense that elimination of one does not affect the position of the other.

- In general, PR-BN cannot be identified using Bottleneck Indicator 3.1, Q-BN
is not necessarily the worst quality machine of the system, and PR- and Q-BNs are coupled, i.e., elimination of one affects the position of the other.

- Although in the general case, neither PR- nor Q-BNs can be identified using Bottleneck Indicator 3.1, the notion of local bottlenecks alleviates some of the difficulties in their identification.

- The optimal position of a single sufficiently efficient inspection machine is either immediately after the last non-perfect quality machine or around the PR-BN located downstream of this machine.

- The production rate of good parts in Bernoulli lines with quality-quantity coupling (QQC) machines is a non-monotonic (concave) function of machine efficiency.

- Therefore, a BN can be eliminated either by increasing machine efficiency (for $p^+$-BN) or decreasing machine efficiency (for $p^-$-BN).

- Serial lines with rework do not possess the property of reversibility.

- Monotonicity properties of lines with rework are similar to those without rework.

- In lines with rework, more efficient machines must be placed towards the end of the line to ensure the largest throughput.

- Local bottlenecks of lines with rework can be identified using Bottleneck Indicator 3.1.

- One of the local bottlenecks in lines with rework is, practically always, the global bottleneck.

- The methods developed are applied to an automotive paint shop system and significant productivity and quality improvements are obtained.
CHAPTER VI

TRANSIENT PROPERTIES OF SERIAL LINES

6.1 Introduction

All previous chapters address the steady state properties of production lines. The current one is intended to investigate their transient behavior, i.e., to study how fast they reach their steady states.

The transient properties of production systems are of practical importance. Indeed, if the steady state is reached after a relatively long period of time, the system may suffer substantial production losses. For instance, if the cycle time of a production system is 1 minute and the plant shift is 500 minutes, the system may lose more than 10% of its production due to transients, if at the beginning of the shift all buffers were empty.

Transient properties are also important in systems that operate with the so-called “floats”. Typically, systems with floats are used when it is desired that the production rate of the line be equal or, at least, close to that of the best machine. This is accomplished by extending the operating time of all other machines in order to build-up floats, which would prevent starvations of the best machine during the overall system operation. Thus, in this case, the system operates in a transient regime, having buffers initially full. In this situation, the main technical problem is to select the minimum float that leads to the maximum throughput.

Thus, analysis of transients in production lines is of importance from the practical
point of view. Clearly, it is important from the theoretic perspective as well. In this chapter, we investigate the transient properties of serial lines.

6.2 Bernoulli Lines with Identical Machines

6.2.1 Problem formulation

Consider a serial production line defined by assumptions (a)-(e) of Subsection 2.4.1. Assume that it consists of two identical machines, i.e., \( M = 2 \) and \( p_1 = p_2 =: p \). Then the transition probabilities can be arranged in a transition matrix as follows:

\[
A = \begin{bmatrix}
1 - p & p(1 - p) & 0 & \cdots & 0 & 0 \\
p & p^2 + (1 - p)^2 & p(1 - p) & 0 & \cdots & 0 \\
0 & p(1 - p) & p^2 + (1 - p)^2 & p(1 - p) & \ddots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots \\
\vdots & \vdots & \ddots & p(1 - p) & p^2 + (1 - p)^2 & p(1 - p) \\
0 & 0 & \cdots & 0 & p(1 - p) & p^2 + 1 - p
\end{bmatrix}
\]

(6.1)

Let \( x_i(n) \), \( i = 0, 1, 2, \ldots, N \), \( n = 0, 1, \ldots \), be the probability that there are \( i \) parts in the buffer at the end of time slot \( n \). Then the evolution of vector \( x(n) = [x_0(n), \ldots, x_N(n)]^T \) can be described by the following constrained linear equation:

\[
x(n + 1) = Ax(n), \quad \sum_{i=0}^{N} x_i(n) = 1.
\]

(6.2)

Given this description, \( PR \) and \( WIP \) can be viewed as the outputs \( y(n) \) of the state-space model (6.2), defined as follows:

\[
PR(n) = p \sum_{i=1}^{N} x_i(n), \\
WIP(n) = \sum_{i=1}^{N} ix_i(n).
\]

(6.3) (6.4)
Together, (6.2), (6.3), and (6.4) represent a state space description:

\[ x(n + 1) = Ax(n), \quad \|x(n)\|_1 = 1, \]
\[ y(n) = Cx(n), \]

where \(A\) is given in (6.1),

\[ C = \begin{bmatrix} 0 & p & p & \cdots & p \\ 0 & 1 & 2 & \cdots & N \end{bmatrix}, \]

and

\[ \|x(n)\|_1 = \sum_{i=0}^{N} x_i(n). \]

This description, along with its generalization for \(M > 2\), is the basis for transient analysis of Bernoulli lines. Next, we formulate four problems related to transient analysis.

**Second largest eigenvalue problem:** Since \(A\), defined by (6.1), is a transition matrix of an ergodic Markov chain, it has a unique largest eigenvalue \(\lambda_0 = 1\). Assume that the second largest eigenvalue (SLE), \(\lambda_1\), is real and simple while all other eigenvalues are distinct. Numerical calculations suggest that this assumption holds for matrix (6.1) regardless of the values of \(p\) and \(N\). Arrange all eigenvalues of \(A\) as follows:

\[ 1 = \lambda_0 > \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_N|. \]

Due to (6.8), there exists a nonsingular matrix \(Q \in \mathbb{R}^{(N+1)\times(N+1)}\) such that the substitution

\[ \tilde{x}(n) = Qx(n), \]

transforms (6.5)-(6.7) into

\[ \tilde{x}(n + 1) = \tilde{A}\tilde{x}(n), \]
\[ y(n) = \tilde{C}\tilde{x}(n), \tag{6.10} \]

where

\[ \tilde{A} = QAQ^{-1} = \text{diag}[1 \; \lambda_1 \cdots \lambda_N], \tag{6.11} \]
\[ \tilde{C} = CQ^{-1} = \begin{bmatrix} \tilde{C}_{10} & \tilde{C}_{11} & \cdots & \tilde{C}_{1N} \\ \tilde{C}_{20} & \tilde{C}_{21} & \cdots & \tilde{C}_{2N} \end{bmatrix}. \tag{6.12} \]

For any admissible initial condition \( x(0) \), i.e., an initial condition such that

\[ \sum_{i=0}^{N} x_i(0) = 1, \]

the solution of (6.10) is

\[ \tilde{x}(n) = \text{diag}[1 \; \lambda_1^n \cdots \lambda_N^n]\tilde{x}(0), \tag{6.13} \]

where

\[ \tilde{x}(0) = Qx(0). \tag{6.14} \]

This implies that states \( \tilde{x}(n) \) and, due to (6.9), \( x(n) \), converge to their steady states as exponential functions of time with bases \( \lambda_i \). Hence, the duration of transients is defined by the largest of \( |\lambda_i|, i \neq 0 \), i.e., by the SLE.

In the framework of a Bernoulli line with two identical machines, the SLE is a function of the machine efficiency \( p \) and the buffer capacity \( N \). For \( M > 2 \), it depends also on \( M \), i.e., \( \lambda_1 = \lambda_1(p, N, M) \). The first problem addressed in this section is to analyze the behavior of \( \lambda_1 \) as a function of all three arguments. The results of this analysis should reveal which values of system parameters lead to inherently long and which to inherently short transients of buffer occupancy. This is carried out in Subsection 6.2.2.
Pre-exponential factor problem: As it follows from (6.10)-(6.14), the dynamics of the output \( y(n) \) are described by the equations:

\[
PR(n) = \tilde{C}_{10} + \tilde{C}_{11}\bar{x}_1(0)\lambda_1^n + \cdots + \tilde{C}_{1N}\bar{x}_N(0)\lambda_N^n, \tag{6.15}
\]

\[
WIP(n) = \tilde{C}_{20} + \tilde{C}_{21}\bar{x}_1(0)\lambda_1^n + \cdots + \tilde{C}_{2N}\bar{x}_N(0)\lambda_N^n, \tag{6.16}
\]

where it is taken into account that \( \bar{x}_0(0) = 1 \) (since the first row of \( Q \) is the left eigenvector of \( A \) given by \([1, 1, \ldots, 1]\)).

Clearly,

\[
\lim_{n \to \infty} PR(n) = \tilde{C}_{10} = PR_{ss}, \tag{6.17}
\]

\[
\lim_{n \to \infty} WIP(n) = \tilde{C}_{20} = WIP_{ss}, \tag{6.18}
\]

where \( PR_{ss} \) and \( WIP_{ss} \) are the steady state values of the production rate and work-in-process, respectively, which can be calculated analytically according to [17]. Thus,

\[
PR(n) = PR_{ss}\left[1 + \frac{\tilde{C}_{11}}{\tilde{C}_{10}}\bar{x}_1(0)\lambda_1^n + \cdots + \frac{\tilde{C}_{1N}}{\tilde{C}_{10}}\bar{x}_N(0)\lambda_N^n\right], \tag{6.19}
\]

\[
WIP(n) = WIP_{ss}\left[1 + \frac{\tilde{C}_{21}}{\tilde{C}_{20}}\bar{x}_1(0)\lambda_1^n + \cdots + \frac{\tilde{C}_{2N}}{\tilde{C}_{20}}\bar{x}_N(0)\lambda_N^n\right]. \tag{6.20}
\]

These equations indicate that the transients of \( PR \) and \( WIP \) are characterized not only by the eigenvalues of \( A \) but also by the pre-exponential factors \( \tilde{C}_{ij} \), \( i = 1, 2, \ldots, N \). Since the initial conditions \( \bar{x}_j(0) \) enter (6.19) and (6.20) in a similar manner and since \( \lambda_1 \) is the SLE, the most important pre-exponential factors are \( \frac{\tilde{C}_{11}}{\tilde{C}_{10}} \) and \( \frac{\tilde{C}_{21}}{\tilde{C}_{20}} \). Denote them as

\[
\Psi_{11} = \left| \frac{\tilde{C}_{11}}{\tilde{C}_{10}} \right|, \quad \Psi_{21} = \left| \frac{\tilde{C}_{21}}{\tilde{C}_{20}} \right|. \tag{6.21}
\]

Coefficients \( \Psi_{11} \) and \( \Psi_{21} \) describe to what extent the SLE affects the transients of the outputs. In addition, the relationship between these coefficients shows which of the outputs, \( PR \) or \( WIP \), has faster transients: if \( \Psi_{11} < \Psi_{21} \), \( PR \) converges to its steady
state value faster than \textit{WIP}; if the inequality is reversed, the opposite takes place.

The second problem considered in this section is to \textit{analyze the behavior of the pre-exponential factors} \( \Psi_{11} = \Psi_{11}(p, N, M) \) and \( \Psi_{21} = \Psi_{21}(p, N, M) \). This is carried out in Subsection 6.2.3.

**Settling time problem:** In Control Theory, the transients of feedback systems are often characterized by the settling time, \( t_s \), which is the time necessary for the output to reach and remain within \( \pm 5\% \) of its steady state value, provided that the input is a unit step function and the initial conditions are zero. A similar notion can be used to characterize the transients of production lines as well. Indeed, the zero initial condition could be interpreted as having all buffers initially empty, i.e.,

\[
x^i(0) = [1, 0, \ldots, 0]^T, \quad i = 1, \ldots, M - 1,
\]

where \( x^i(0) \) is the vector of probabilities of initial occupancy of buffer \( i \). The step input, in the framework of production systems, is incorporated in the fact that matrix \( A \) has an eigenvalue equal to 1 and, therefore, the outputs converge to non-zero steady state values, \( PR_{ss} \) and \( WIP_{ss} \). Thus, the production system defined by assumptions (a)-(e) of Subsection 2.4.1 can be characterized by two settling times — with respect to \( PR \) and \( WIP \) — denoted as \( t_{sPR} \) and \( t_{sWIP} \), respectively.

The third problem considered in this section is to \textit{analyze settling times} \( t_{sPR} \) and \( t_{sWIP} \) as \textit{functions of} \( p, N, \) and \( M \). This is carried out in Subsection 6.2.4.

**Production losses problem:** The three problems formulated above are mathematical in nature. The problem of production losses, although based on the results of the previous three, is clearly practical. It addresses the question of how much production is lost due to transients if the initial buffer occupancy is zero, and what is the smallest initial buffer occupancy necessary to guarantee no production losses due to transients.
To formulate this question precisely, introduce the notion of production losses during period $T$:

\[ L_T(x^1(0), \ldots, x^{M-1}(0)) = \sum_{n=0}^{T} [PR_{ss} - PR(n; x^1(0), \ldots, x^{M-1}(0))] \]  

(6.23)

where $x^i(0)$ is the vector of probability of initial occupancy of buffer $i$, i.e., $x^i(0) = [x^i_0(0), \ldots, x^i_{N_i}(0)]^T$, $i = 1, 2, \ldots, M - 1$. Of particular interest are the initial conditions corresponding to all buffers being empty and all buffers having nonzero identical occupancy $h(0)$. The corresponding production losses are denoted as $L^0_T(p, N, M)$ and $L^{h(0)}_T(p, N, M)$, respectively.

A more telling metric of production losses is the percent of loss defined by

\[ \Lambda^0_T(p, N, M) = \frac{L^0_T(p, N, M)}{T \cdot PR_{ss}} \cdot 100\%, \quad \Lambda^{h(0)}_T(p, N, M) = \frac{L^{h(0)}_T(p, N, M)}{T \cdot PR_{ss}} \cdot 100\%. \] 

(6.24)

The fourth problem considered in this section is to analyze the properties of $\Lambda^0_T(p, N, M)$ and to determine the smallest initial buffer occupancy $h^*(0)$ such that $\Lambda^{h^*(0)}_T(p, N, M) = 0$. This problem is discussed in Subsection 6.2.5.

### 6.2.2 Analysis of the second largest eigenvalue

**Two-machine lines:** For $N = 1$ and 2, the characteristic polynomials of matrix $A$ defined by (6.1) (with $(\lambda - 1)$ being factored out) are:

- $N = 1$:  
  \[ \text{det}(\lambda I - A) = (\lambda - 1) \left( \lambda + 2p - p^2 - 1 \right), \]

- $N = 2$:  
  \[ \text{det}(\lambda I - A) = (\lambda - 1) \left[ \lambda^2 + (-2 + 4p - 3p^2) \lambda + (1 - p)^4 \right], \]

leading to the following expressions for the SLE as functions of $p$:

\[ \lambda_1(p, N = 1, M = 2) = (1 - p)^2, \]  

(6.25)

\[ \lambda_1(p, N = 2, M = 2) = \frac{2 + 3p^2 - 4p + p\sqrt{4(1 - p)^2 + p^2}}{2}. \]  

(6.26)
Characteristic polynomials for $N \geq 3$ can also be calculated. For instance,

\[
N = 3 : \quad \det(\lambda I - A) = (\lambda - 1)[\lambda^3 + (-3 - 5p^2 + 6p)\lambda^2 + \\
(-12p + 20p^2 + 3 + 6p^4 - 16p^3)\lambda - (1 - p)^6],
\]

\[
N = 4 : \quad \det(\lambda I - A) = (\lambda - 1)[\lambda^4 + (-4 - 7p^2 + 8p)\lambda^3 + \\
(-24p + 6 + 42p^2 - 36p^3 + 15p^4)\lambda^2 + \\
(24p + 40p^5 - 80p^4 - 4 - 63p^2 + 92p^3 - 10p^6)\lambda \\
+ (1 - p)^8].
\]

However, since their roots, as well as those for $N \geq 5$, cannot be found analytically, we calculate the SLE numerically for various values of $p \in (0, 1)$. The results, along with (6.25) and (6.26), are shown in Figure 6.1.

![Figure 6.1: Second largest eigenvalue as a function of $p$ in two-machine lines](image)

As it follows from this figure, the behavior of the SLE as a function of $p$ is qualitatively different for $N = 1$ and $N \geq 2$: monotonically decreasing and non-monotonic convex, respectively. In other words, increasing machine efficiency speeds up the transients if $N = 1$ and may slow them down if $N \geq 2$. The explanation of this phenomenon is as follows: Due to the blocked before service assumption, $N = 1$ implies that each machine serves as a buffer capable of storing one part and no additional
buffering is present. In other words, \( N = 1 \) represents just-in-time (JIT) operation. If a buffer is present (i.e., \( N \geq 2 \)), the states evolve slowly when the machines operate almost “synchronously”, i.e., are up or down almost simultaneously. Roughly, this synchronism can be characterized by the probability that both machines are up or down simultaneously. Since this probability is \( p^2 + (1 - p)^2 \), the states move slowly when \( p \) is close to either 0 or 1.

As one can see from Figure 6.1, for each \( N \geq 2 \), there exists a unique \( p^* \), which to the smallest SLE, i.e.,

\[
p^*(N) = \arg \min_p \lambda_1(p, N, M = 2).
\]

The behavior of \( p^* \) as a function of \( N \) is shown in Figure 6.2. Interestingly, for all \( N \geq 2 \), \( p^* \) belongs to a relatively narrow interval \((0.5, 0.6)\). This can be explained by the fact that the probability of one machine being up and the other being down is \( 2p(1 - p) \), which reaches its maximum at \( p = 0.5 \).

To illustrate the duration of the transients for \( p = p^* \) and for \( p \)'s close to 1, Figure 6.3(a) shows \( x_5(n) \) for a system with \( N = 5 \) and the initial condition \( x(0) = [1, 0, 0, 0, 0, 0]^T \). Clearly, large \( p \)'s lead to transients that are an order of magnitude slower than those defined by \( p^* \).
Figure 6.3: Transients of buffer occupancy for $M = 2$

Figure 6.4: Second largest eigenvalue as a function of buffer capacity
The graphs of Figure 6.1 indicate that the SLE is a monotonically increasing function of $N$. More explicitly, the behavior of SLE as a function of $N$ is illustrated in Figure 6.4. It shows, in particular, that for $N = 10$, the SLE is close to 1 for any $p$. Thus, large buffers lead to very long transients. This is illustrated in Figure 6.3(b), which shows $x_3(n)$ for three systems with $N = 3, 5$ and $10$ and with $p = 0.9$ and initially empty buffers. Clearly, $N = 10$ results in transients that are an order of magnitude longer than those for $N = 3$.

The above analysis leads to:

**Numerical Fact 6.1** *In Bernoulli lines with identical machines defined by assumptions (a)-(e) of Subsection 2.4.1 with $M = 2$, function $\lambda_1(p,N)$, $N \geq 2$, is monotonically increasing in $N$ and non-monotonic convex in $p$.*

**Justification:** See Appendix D.

**Three-machine lines:** For $M = 3$, the transition matrix of a system defined by assumptions (a)-(e) with identical machines can be explicitly written for $N = 1$ and $N = 2$:

$$A(N = 1) = \begin{bmatrix}
1 - p & p(1 - p) & 0 & 0 \\
0 & (1 - p)^2 & p(1 - p) & p^2(1 - p) \\
p & p^2 & 1 - p & p(1 - p) \\
0 & p(1 - p) & p^2 & (1 - p) + p^3
\end{bmatrix}, \quad (6.27)$$

$$A(N = 2) = \begin{bmatrix}
1 - p & (1 - p)p & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & (1 - p)^2 & (1 - p)p & (1 - p)p & 0 & 0 & 0 & 0 \\
0 & 0 & (1 - p)^2 & (1 - p)p & 0 & 0 & 0 & 0 \\
p & p^2 & (1 - p)^2 & (1 - p)p & 0 & 0 & 0 & 0 \\
0 & (1 - p)p & p^2 & (1 - p)^3 + p^3 & p^3 + (1 - p)^2 & 0 & 0 & 0 \\
0 & 0 & (1 - p)p & (1 - p)^2 + p^3 & (1 - p)p & 0 & 0 & 0 \\
0 & 0 & 0 & (1 - p)p & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}, \quad (6.28)$$
It can be verified analytically that \( \lambda = 1 - p \) is an eigenvalue of (6.27) and numerically that it is the SLE. Thus,

$$\lambda_1(p, N = 1, M = 3) = 1 - p.$$  \hfill (6.29)

The SLE of (6.28) can be evaluated numerically. These SLEs are illustrated in Figure 6.5 (along with their counterparts for \( M = 2 \)). From this figure, we conclude that the SLEs for \( M = 3 \) and \( M = 2 \) are qualitatively similar as functions of \( p \). In addition, \( \lambda_1(p, N, M = 3) > \lambda_1(p, N, M = 2), \) \( N \in \{1, 2\} \), implying that transients for \( M = 3 \) are longer than those for \( M = 2 \).

![Figure 6.5: Second largest eigenvalue as a function of machine efficiency in two- and three-machine lines](image)

Since no results on the behavior of SLE for \( N \geq 3 \) could be derived at this time, based on the above observations, we formulate that:

**Conjecture 6.1** In Bernoulli lines with identical machines defined by assumptions (a)-(e) of Subsection 2.4.1, function \( \lambda_1(p, N, M) \) is monotonically increasing in \( N \) and \( M \) and non-monotonic convex in \( p \).
6.2.3 Analysis of the pre-exponential factors

The effect of the pre-exponential factors on the transients of outputs, $PR$ and $WIP$, is shown in expressions (6.19) and (6.20). In this subsection, we analyze their behavior as functions of system parameters.

**Theorem 6.1** In Bernoulli lines with identical machines defined by assumptions (a)-(e) of Subsection 2.4.1,

$$\Psi_{11}(p, N = 1, M = 2) = \Psi_{21}(p, N = 1, M = 2) = \sqrt{1 + (1 - p)^2}. \quad (6.30)$$

**Proof:** See Appendix D.

![Figure 6.6: Behavior of pre-exponential factors $\Psi_{11}$ and $\Psi_{21}$ as a function of machine efficiency in two-machine lines with $N = 1$](image)

Thus, if $N = 1$, the SLE affects both outputs, $PR$ and $WIP$, identically. In addition, since $\Psi_{11}(p, N = 1, M = 2)$, as a function of $p$, behaves as shown in Figure 6.6, JIT leads not only to a monotonically decreasing SLE but also to a monotonically decreasing pre-exponential factor, implying that the transients of $PR$ and $WIP$ are somewhat faster in a system with $p$ close to 1.

For $N \geq 2$, the situation is qualitatively different.
**Numerical Fact 6.2** In Bernoulli lines with identical machines defined by assumptions (a)-(e) of Subsection 2.4.1 with \( N \geq 2 \),

\[
\Psi_{11}(p, N, M = 2) < \Psi_{21}(p, N, M = 2)
\]

and, in addition,

\[
\Psi_{11}(p, N, M = 2) \to 0 \quad \text{as} \quad p \to 1.
\]

**Justification:** See Appendix D.

The behavior of \( \Psi_{11} \) and \( \Psi_{21} \) for \( N = 2, 3, \) and 5 is shown in Figure 6.7. From this figure and (6.15), (6.16), it can be concluded that:

- Since both pre-exponential factors are monotonically decreasing in \( p \), the effect of the SLE on the transients of \( PR \) and \( WIP \) is decreasing when \( p \) is increasing.
- Since \( \Psi_{11}(p, N, M = 2) < \Psi_{21}(p, N, M = 2) \), the SLE affects \( WIP \) more than \( PR \), i.e., transients of \( PR \) are faster than those of \( WIP \).
- Since for most \( p \)'s, \( \Psi_{11}(p, N, M = 2) < 1 \), the transients of \( PR \) are faster than those of the states (i.e., buffer occupancy).
- Finally, since \( \Psi_{11}(p, N, M = 2) \to 0 \) as \( p \to 1 \), the effect of SLE on the duration of transients of \( PR \) becomes negligible when \( p \) is close to 1.

Unfortunately, no general statement concerning the pre-exponential factors for \( M \geq 3 \) could be derived. However, based on the above observations and numerical evidence, we formulate that:

**Conjecture 6.2** Numerical Fact 6.2 holds for \( M \geq 3 \) as well.
Figure 6.7: Behavior of pre-exponential factors Ψ_{11} and Ψ_{21} as a function of machine efficiency in two-machine lines with $N \geq 2$

6.2.4 Analysis of the settling time

To analyze the settling times, $t_{sPR}$ and $t_{sWIP}$, one has to know the behavior of $PR$ and $WIP$ as a function of $n$. Therefore, in this subsection, we first analyze the trajectories of $PR(n)$ and $WIP(n)$ and then utilize them to evaluate the settling time.

Behavior of $PR(n)$ and $WIP(n)$: Two-machine lines: The trajectories of $PR(n)$ and $WIP(n)$ can be analyzed numerically by solving (6.5)-(6.7). In addition, analytical approximations, based on the SLE, can be constructed. This is carried out below.

Approximation of $PR(n)$: Consider a Bernoulli line with two identical machines defined by assumptions (a)-(e) of Subsection 2.4.1. Assume that at the initial time
the buffer is empty. Then,

\[ PR(0) = P[\text{a part is produced during time slot 1}] = 0, \quad (6.33) \]

\[ PR(1) = P[\text{a part is produced during time slot 2}] = P[\text{machine 1 is up during time slot 1}] \cdot P[\text{machine 2 is up during time slot 2}] = p^2. \quad (6.34) \]

For time slots \( n \geq 2 \), introduce the following approximation:

\[ \hat{PR}(n) = PR_{ss} \left(1 - \beta \lambda_1^{n-1}\right), \quad (6.35) \]

where \( \lambda_1 \) is the SLE and \( PR_{ss} \) is the steady state production rate, i.e.,

\[ PR_{ss} = \frac{Np}{N + 1 - p}. \quad (6.36) \]

Selecting \( \beta \) so that \( PR(1) = \hat{PR}(1) \), we obtain

\[ \beta = 1 - \frac{p(N + 1 - p)}{N}, \quad 0 < p < 1. \quad (6.37) \]

Note that \( \beta \) tends to 0 as \( p \) tends to 1, which makes (6.35)-(6.37) in agreement with conclusion (\( \delta \)) of Subsection 6.2.3: when \( p \) is close to 1 the effect of \( \lambda_1 \) on \( PR \) diminishes to 0.

The accuracy of approximation (6.35)-(6.37) is illustrated in Figure 6.8 by comparing it with \( PR(n) \), obtained by solving (6.5)-(6.7) numerically. This accuracy is quantified in Table 6.1 by

\[ \Delta_{PR} = \max_{n=1,2,...} \left| \frac{PR(n) - \hat{PR}(n)}{PR(n)} \right| \cdot 100\%. \quad (6.38) \]

Expressions (6.35)-(6.37) are utilized later to evaluate \( ts_{PR} \).

Approximation of \( WIP(n) \): Consider a Bernoulli line defined by assumptions (a)-(e)
Figure 6.8: Comparison of $PR(n)$ and $\hat{PR}(n)$ for $M = 2$

Table 6.1: Accuracy of approximation $\hat{PR}(n)$

<table>
<thead>
<tr>
<th></th>
<th>$p = 0.6$</th>
<th>$p = 0.7$</th>
<th>$p = 0.8$</th>
<th>$p = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 2$</td>
<td>1.72</td>
<td>0.65</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>$N = 3$</td>
<td>6.45</td>
<td>3.98</td>
<td>2.10</td>
<td>0.82</td>
</tr>
<tr>
<td>$N = 4$</td>
<td>10.40</td>
<td>6.91</td>
<td>3.99</td>
<td>1.72</td>
</tr>
<tr>
<td>$N = 5$</td>
<td>13.64</td>
<td>9.27</td>
<td>5.55</td>
<td>2.49</td>
</tr>
</tbody>
</table>
of Subsection 2.4.1 with two identical machines and the buffer initially empty. Then,

\[ WIP(0) = 0, \quad (6.39) \]
\[ WIP(1) = P[\text{a part is present in the buffer at the end of time slot 1}] = p. \quad (6.40) \]

For time slots \( n \geq 2 \), the following approximation is introduced:

\[ \hat{WIP}(n) = WIP_{ss} \left( 1 - \gamma \lambda_1^{n-1} \right), \quad (6.41) \]

where, as before, \( \lambda_1 \) is the SLE, \( WIP_{ss} \) is the steady state work-in-process, i.e.,

\[ WIP_{ss} = \frac{N(N+1)}{2(N+1-p)}, \quad (6.42) \]

and \( \gamma \) is selected so that \( WIP(1) = \hat{WIP}(1) \), i.e.,

\[ \gamma = 1 - \frac{2p(N+1-p)}{N(N+1)}. \quad (6.43) \]

Note that \( \gamma \) tends to 0 as \( p \) tends to 1 only for \( N = 1 \), which makes (6.43) in agreement with Numerical Fact 6.2 of Subsection 6.2.3.

The accuracy of approximation (6.41)-(6.43) is illustrated in Figure 6.9, and quantified in Table 6.2 by

\[ \Delta_{WIP} = \max_{n=1,2,...} \frac{|WIP(n) - \hat{WIP}(n)|}{WIP(n)} \cdot 100\%. \quad (6.44) \]

<table>
<thead>
<tr>
<th>( N )</th>
<th>( p = 0.6 )</th>
<th>( p = 0.7 )</th>
<th>( p = 0.8 )</th>
<th>( p = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.71</td>
<td>0.34</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>2.02</td>
<td>1.28</td>
<td>0.72</td>
<td>0.29</td>
</tr>
<tr>
<td>4</td>
<td>5.18</td>
<td>4.18</td>
<td>3.18</td>
<td>2.39</td>
</tr>
<tr>
<td>5</td>
<td>8.97</td>
<td>7.53</td>
<td>6.28</td>
<td>5.17</td>
</tr>
<tr>
<td>$N = 3$</td>
<td>$N = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Graph for $p = 0.7$" /></td>
<td><img src="image2" alt="Graph for $p = 0.7$" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image3" alt="Graph for $p = 0.8$" /></td>
<td><img src="image4" alt="Graph for $p = 0.8$" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image5" alt="Graph for $p = 0.9$" /></td>
<td><img src="image6" alt="Graph for $p = 0.9$" /></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.9: Comparison of $WIP(n)$ and $\hat{WIP}(n)$ for $M = 2$
Expressions (6.39)-(6.43) are used later for $t_{sWIP}$ evaluation.

Comparison of $PR(n)$ and $WIP(n)$: To compare the transients of $PR$ and $WIP$, Figure 6.10 shows the graphs of $\hat{PR}(n)/PR_{ss}$ and $\hat{WIP}(n)/WIP_{ss}$ for various $p$'s and $N$'s. This figure also supports the conclusions of Subsection 6.2.3. Specifically, for $N = 1$, the transients of $PR$ and $WIP$ are identical. As $N$ becomes larger, the difference becomes more pronounced. For large $N$, the transients of $PR$ are orders of magnitude faster than those of $WIP$.

$M > 2$-machine lines: For $M > 2$, the behavior of $PR(n)$ and $WIP(n) = \sum_{i=1}^{M-1} WIP_i(n)$, where $WIP_i(n)$ is the occupancy of buffer $i$, is analyzed using numerical simulations of systems defined by assumptions (a)-(e) of Subsection 2.4.1. A C++ code is constructed to perform the simulations. For each line considered, 5000 runs are carried out. Within each iteration, every buffer is initialized to be empty, and the status of the machines, up or down, is selected with probability $p$ and $1 - p$, respectively. Then the average is calculated over 5000 runs for the output at each time slot, resulting in a 90% confidence intervals of less than 0.01 for $PR(n)$ and 0.05 for $WIP(n)$.

The results of these simulations are shown in Figure 6.11. This figure also supports the conclusions of Subsections 6.2.2 and 6.2.3: the transients of $PR$ and $WIP$ become slower as $N$ and $M$ increase.

Figure 6.11 exhibits one more interesting phenomenon of transients in serial production lines: the transportation delay. Unlike the dynamic delay, the transportation delay is not related to eigenvalues but to the time necessary for the input to reach the output. When $M = 2$ and the buffer is initially empty, the transportation delay for $PR$ and $WIP$ is 1 time slot. This is reflected in (6.33) and (6.39). When $M > 2$, the transportation delay for $WIP(n) = \sum_{i=1}^{M-1} WIP_i(n)$ remains 1, while for $PR(n)$ it becomes $M - 1$. Thus, for large $M$ the transportation delay for $PR$ may be signif-
<table>
<thead>
<tr>
<th>$N$</th>
<th>$p = 0.7$</th>
<th>$p = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>10</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
</tbody>
</table>

Figure 6.10: Comparison of transients of $\hat{PR}(n)/PR_{ss}$ and $\hat{WIP}(n)/WIP_{ss}$ for $M = 2$
<table>
<thead>
<tr>
<th>$N = 1$</th>
<th>$M = 3$</th>
<th>$M = 5$</th>
<th>$M = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
<td></td>
</tr>
<tr>
<td>$N = 3$</td>
<td>$M = 3$</td>
<td>$M = 5$</td>
<td>$M = 10$</td>
</tr>
<tr>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
<td></td>
</tr>
<tr>
<td>$N = 5$</td>
<td>$M = 3$</td>
<td>$M = 5$</td>
<td>$M = 10$</td>
</tr>
<tr>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.11: Transients of $PR(n)/PR_{ss}$ and $WIP(n)/WIP_{ss}$ in lines with $M > 2$ for $p = 0.9$

icant. This is shown in Figure 6.11, where the transients of $PR$ and $WIP$ for $N = 1$ are no longer identical. However, as $N$ increases, the effect of the dynamics becomes dominant, and the transportation delay does not alter the nature of the response in any significant manner (see the case of $M = 10$, $N = 5$ in Figure 6.11).

Analysis of $t_{sPR}$ and $t_{sWIP}$: For $M = 2$, approximations (6.35)-(6.37) and (6.41)-(6.43) lead to:

**Theorem 6.2** In Bernoulli lines defined by assumptions (a)-(c) of Subsection 2.4.1 with $M = 2$, the estimates of $t_{sPR}$ and $t_{sWIP}$ can be given as follows

\[
\hat{t}_{sPR} = 1 - \frac{\ln(20\beta)}{\ln \lambda_1}, \quad (6.45)
\]

\[
\hat{t}_{sWIP} = 1 - \frac{\ln(20\gamma)}{\ln \lambda_1}, \quad (6.46)
\]
where $\beta$ and $\gamma$ are defined in (6.37) and (6.43), respectively. In addition,

$$\hat{t}_{sPR} = \hat{t}_{sWIP}, \quad \text{for } N = 1,$$
(6.47)

and

$$\hat{t}_{sPR} < \hat{t}_{sWIP}, \quad \text{for } N \geq 2.$$
(6.48)

**Proof:** See Appendix D.

The behavior of estimates (6.45) and (6.46), along with the exact values, $t_{sPR}$ and $t_{sWIP}$, is illustrated in Figure 6.12. Clearly, $t_{sPR}$ is monotonically decreasing as a function of $p$ while $t_{sWIP}$ is convex, which is consistent with the behavior of the SLE and PEF analyzed in Sections 6.2.2 and 6.2.3. The accuracy in terms of

$$\Delta_{t_{sPR}} = \frac{t_{sPR} - \hat{t}_{sPR}}{t_{sPR}} \cdot 100\%,$$
(6.49)

$$\Delta_{t_{sWIP}} = \frac{t_{sWIP} - \hat{t}_{sWIP}}{t_{sWIP}} \cdot 100\%,$$
(6.50)

is quantified in Tables 6.3 and 6.4. Obviously, $\hat{t}_{sWIP}$ is more accurate than $\hat{t}_{sPR}$. Also, both estimates become more accurate for larger $p$ and less accurate for larger $N$.

<table>
<thead>
<tr>
<th></th>
<th>$p = 0.6$</th>
<th>$p = 0.7$</th>
<th>$p = 0.8$</th>
<th>$p = 0.9$</th>
<th>$p = 0.95$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$N = 3$</td>
<td>-25.00</td>
<td>-28.57</td>
<td>-14.28</td>
<td>-25.00</td>
<td>0</td>
</tr>
<tr>
<td>$N = 5$</td>
<td>-66.66</td>
<td>-84.61</td>
<td>-100.00</td>
<td>-166.66</td>
<td>0</td>
</tr>
</tbody>
</table>

### 6.2.5 Analysis of the production losses

**Approach:** In this subsection, we analyze production losses during a shift of duration $T$ cycles. It turns out that these losses are relatively insensitive to $T$ as long as
Figure 6.12: Behavior of settling time of PR and WIP as a function of $p$ and $N$ for $M = 2$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$t_{sPR}$ and $t_{sPR}$</th>
<th>$t_{sWIP}$ and $t_{sWIP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Graph for N=1" /></td>
<td><img src="image" alt="Graph for N=1" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Graph for N=3" /></td>
<td><img src="image" alt="Graph for N=3" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image" alt="Graph for N=5" /></td>
<td><img src="image" alt="Graph for N=5" /></td>
</tr>
</tbody>
</table>

Table 6.4: Accuracy of estimate $t_{sWIP}$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$p = 0.6$</th>
<th>$p = 0.7$</th>
<th>$p = 0.8$</th>
<th>$p = 0.9$</th>
<th>$p = 0.95$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-7.6923</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-2.9412</td>
<td>-2.7027</td>
<td>-2.1739</td>
<td>-2.5641</td>
<td>-2.0833</td>
</tr>
</tbody>
</table>

141
$T \gg 1$. Therefore, we assume that $T = 500$ minutes, which is typical for automotive assembly plants where the cycle time is around 1 minute and the shift is 8 hours.

The production losses have been defined in Subsection 6.2.1 as

$$L_T(x^1(0), \ldots, x^{M-1}(0); p, N, M)$$

$$= \sum_{n=0}^{T} [PR_{ss} - PR(n; x^1(0), \ldots, x^{M-1}(0); p, N, M)], \quad (6.51)$$

or, as the percent of loss,

$$\Lambda_T(x^1(0), \ldots, x^{M-1}(0); p, N, M) = \frac{L_T}{T \cdot PR_{ss}} \cdot 100\%.$$  

(6.52)

We evaluate $L_T$ and $\Lambda_T$ as follows: For $M = 2$ and $M = 3$ (with $N = 1$ and 2), $PR(n)$ is calculated numerically by solving (6.5)-(6.7), respectively. For all other $M$ and $N$, $PR(n)$ is evaluated by simulations using the C++ program and the procedures described in Subsection 6.2.4. To evaluate $PR_{ss}$, another 20 runs of simulations are performed with the first 10,000 time slots as the warm-up time and the following 100,000 time slots used for the evaluation, which results in a 95% confidence interval of less than 0.003.

Using this approach, results on $\Lambda_T^0(p, N, M)$ and $\Lambda_T^{h(0)}(p, N, M)$ are presented below.

**Percent of loss when the buffers are empty at the beginning of the shift:**

Figure 6.13 shows the behavior of $\Lambda_T^0$ as a function of $p$, $N$, and $M$. In addition, the losses are quantified in Table 6.5 for $M = 10$. From these data, we conclude:

- $\Lambda_T^0(p, N, M)$ is a monotonically decreasing function of $p$ with an almost constant rate (except for $p$’s close to 1).

- $\Lambda_T^0(p, N, M)$ is a monotonically increasing function of $N$, again with an almost constant rate.
For large \( M \) and \( N \) and small \( p \), \( \Lambda^0_T(p, N, M) \) can be quite large. For instance, if \( M \) and \( N \) are 10 and \( p = 0.6 \), \( \Lambda^0_T(p, N, M) \) is almost 12%.

Table 6.5: Percent of loss \( \Lambda^0_T \) for ten-machine lines

<table>
<thead>
<tr>
<th>( N )</th>
<th>( p = 0.6 )</th>
<th>( p = 0.7 )</th>
<th>( p = 0.8 )</th>
<th>( p = 0.9 )</th>
<th>( p = 0.95 )</th>
<th>( p = 0.99 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.88</td>
<td>2.49</td>
<td>2.18</td>
<td>1.92</td>
<td>1.79</td>
<td>1.71</td>
</tr>
<tr>
<td>2</td>
<td>3.95</td>
<td>3.39</td>
<td>2.85</td>
<td>2.52</td>
<td>2.38</td>
<td>2.25</td>
</tr>
<tr>
<td>3</td>
<td>5.07</td>
<td>4.24</td>
<td>3.70</td>
<td>3.23</td>
<td>3.06</td>
<td>2.76</td>
</tr>
<tr>
<td>4</td>
<td>6.14</td>
<td>5.15</td>
<td>4.45</td>
<td>3.95</td>
<td>3.69</td>
<td>3.01</td>
</tr>
<tr>
<td>5</td>
<td>7.14</td>
<td>6.08</td>
<td>5.31</td>
<td>4.62</td>
<td>4.22</td>
<td>3.17</td>
</tr>
<tr>
<td>6</td>
<td>8.28</td>
<td>7.02</td>
<td>6.10</td>
<td>5.21</td>
<td>4.63</td>
<td>3.25</td>
</tr>
<tr>
<td>7</td>
<td>9.25</td>
<td>7.90</td>
<td>6.78</td>
<td>5.70</td>
<td>4.88</td>
<td>3.30</td>
</tr>
<tr>
<td>8</td>
<td>10.18</td>
<td>8.69</td>
<td>7.39</td>
<td>6.10</td>
<td>5.11</td>
<td>3.31</td>
</tr>
<tr>
<td>9</td>
<td>10.99</td>
<td>9.35</td>
<td>7.89</td>
<td>6.40</td>
<td>5.26</td>
<td>3.32</td>
</tr>
<tr>
<td>10</td>
<td>11.72</td>
<td>9.84</td>
<td>8.33</td>
<td>6.67</td>
<td>5.41</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Percent of loss when the buffers are not empty at the beginning of the shift: Figure 6.14 shows the behavior of \( \Lambda^h_T(0) \) as a function of \( h(0) \), \( p \), \( N \), and \( M \). From this figure we conclude:

- \( \Lambda^h_T(0)(p, N, M) \) is a monotonically decreasing function of \( h(0) \) with a decreasing rate.
- Roughly, \( \Lambda^h_T(0)(p, N, M) = 0 \) if \( h(0) = \lceil N/2 \rceil \), where \( \lceil x \rceil \) is the smallest integer no less than \( x \). Thus,

**Numerical Fact 6.3** Half full buffers provide the smallest initial buffer occupancy, which leads to practically zero losses due to transients.

**Justification:** See Appendix D.
<table>
<thead>
<tr>
<th>$M$ Value</th>
<th>$\Lambda_T^0$ vs $p$</th>
<th>$\Lambda_T^0$ vs $N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M = 2$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$M = 3$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$M = 5$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>$M = 10$</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
</tbody>
</table>

Figure 6.13: Percent of production loss, $\Lambda_T^0$, as functions of $p$, $N$, and $M$
$N = 2$

$N = 5$

$N = 10$

$M = 2$

$M = 3$

$M = 5$

$M = 10$

Figure 6.14: Percent of production loss, $\Lambda_T^{h(0)}$, as functions of $h(0)$, $p$, $N$, and $M$
6.3 Bernoulli Lines with Non-identical Machines

6.3.1 Problem formulation

In the previous section, Bernoulli lines with identical machines have been analyzed. The study of systems with non-identical machines is also important, for both industrial and theoretical reasons. Indeed, many systems of the factory floor consist of non-identical machines; this makes such a study relevant from the industrial point of view. Theoretically, such a study is relevant because the phenomena observed in the identical machine case are based on statistical fluctuations only, whereas in the non-identical case a “drift”, produced by different machine efficiencies, is also of significance. Thus, due to their practical and theoretical relevance, this section is devoted to the study of transients in Bernoulli lines with two non-identical machines.

The state transition matrix of this system is the following \((N + 1) \times (N + 1)\)-dimensional matrix:

\[
A = \begin{bmatrix}
1 - p_1 & p_2(1 - p_1) & 0 & \cdots & 0 \\
p_1 & 1 - p_1 - p_2 + 2p_1p_2 & p_2(1 - p_1) & \cdots & 0 \\
0 & p_1(1 - p_2) & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & 1 - p_1 - p_2 + 2p_1p_2 & p_2(1 - p_1) \\
0 & 0 & \cdots & p_1(1 - p_2) & p_1p_2 + 1 - p_2
\end{bmatrix}.
\] (6.53)

Similar to the identical machine case, such a system can be described by (6.5) but with \(A\) given above and \(C\) defined by

\[
C = \begin{bmatrix}
0 & p_2 & p_2 & \cdots & p_2 \\
0 & 1 & 2 & \cdots & N
\end{bmatrix}.
\] (6.54)

The evolution of the states and outputs can be evaluated by (6.13), (6.19) and (6.20). To quantify the duration of the transients of buffer occupany, \(PR\) and \(WIP\), the properties of SLE and PEFs have been analyzed in the previous section for lines with
identical machines. In the case of non-identical machines, this section considers the following problems:

**Constrained case:** Consider a Bernoulli line defined by assumptions (a)-(e) of Subsection 2.4.1 with a constraint

\[ p_1 p_2 = p^*. \]  
(6.55)

This constraint is introduced to create a one-parametric family of systems for the investigation of the effect of different machine efficiencies on the transients. Under (6.55), we analyze the properties of SLE and PEFs as functions of machine and buffer parameters. This is carried out in Subsection 6.3.2.

**Unconstrained case:** Consider a Bernoulli line defined by assumptions (a)-(e) of Subsection 2.4.1 with \( p_1 \) and \( p_2 \) being arbitrary numbers from \((0, 1)\). In this situation, we have a two-parametric family of systems, suitable for reversibility analysis of transients in serial lines. Given the above, we investigate the SLE and PEFs as functions of machine and buffer parameters and, more importantly, reversibility properties of \( PR \) and \( WIP \) transients. This is carried out in Subsection 6.3.3.

### 6.3.2 Transients in the constrained case

**Transients of states:** For \( N = 1 \) and \( N = 2 \), SLE can be calculated explicitly:

\[
\begin{align*}
\lambda_1(N = 1) &= (1 - p_1)(1 - p_2), \\
\lambda_1(N = 2) &= (1 - p_1)(1 - p_2) + p_1 p_2 + \sqrt{p_1 p_2 [4(1 - p_1)(1 - p_2) + p_1 p_2]} / 2.
\end{align*}
\]  
(6.56, 6.57)

Clearly, \( \lambda_1 \) reaches its maximum at \( p_1 = p_2 = \sqrt{p^*} \). It can be shown that the same holds for \( N > 2 \). Specifically,

**Numerical Fact 6.4** For serial lines defined by assumptions (a)-(e) of Subsection 2.4.1 and (6.55), \( \lambda_1 \) reaches its maximum at \( p_1 = p_2 = \sqrt{p^*} \).
Justification: See Appendix D.

This implies that the state transients are the slowest when the machines are identical. When the machines are not identical, the drift, generated by $|p_1 - p_2|$, leads to faster transients.

The behavior of $\lambda_1$ as a function of $p^*$, $p_2$ and $N$ is illustrated in the first column of Figure 6.15. As one can see, SLE is an increasing functions of $p^*$ and $N$, which implies that the state transients slow down when $p^*$ and $N$ increase. This is in agreement with the results obtained in Section for identical machines, according to which longer buffers and high machine efficiency lead to slower transients.

To illustrate the above conclusions in the time domain, consider three production lines defined by assumption (a)-(e) of Subsection 2.4.1 with $N = 3$ and $p^* = 0.49$:

$$L_1 : p_1 = 0.7, p_2 = 0.7,$$
$$L_2 : p_1 = 0.6125, p_2 = 0.8,$$
$$L_3 : p_1 = 0.5444, p_2 = 0.9.$$  \hspace{1cm} (6.58)

To quantify the transients of the states, we calculate the distance between $x(n)$ and its steady state, $x(\infty)$, as follows:

$$||x(n) - x(\infty)|| = \left( \sum_{i=0}^{N} [x_i(n) - x_i(\infty)]^2 \right)^{1/2},$$  \hspace{1cm} (6.59)

where $x_i(n), i = 0, \ldots, N, n = 1, 2, \ldots$, is evaluated by (6.5) and $x_i(\infty), i = 0, \ldots, N$, defined by (see [17]):

$$x_0(\infty) = \frac{(1 - p_1)(1 - \alpha(p_1, p_2))}{1 - \frac{p_1}{p_2} \alpha^N(p_1, p_2)},$$
$$x_i(\infty) = \frac{\alpha^i(p_1, p_1)}{1 - p_2} x_0(\infty), \quad i = 1, \ldots, N,$$
$$\alpha(p_1, p_2) = \frac{p_1(1 - p_2)}{p_2(1 - p_1)}.$$
Figure 6.15: SLE and PEFs as functions of machine and buffer parameters for the constrained case
Assume that the buffer is initially empty. The evolution of $||x(n) - x(\infty)||$ for $L_1$, $L_2$ and $L_3$ under zero initial conditions is illustrated in Figure 6.16. Clearly, $L_1$ has the slowest transients, while $L_3$ has the fastest, due the largest difference between $p_1$ and $p_2$.

\[ L_1 \quad L_2 \quad L_3 \]

Figure 6.16: Transients of $x(n)$ for lines $L_1$, $L_2$ and $L_3$

**Transients of outputs:** For $N = 1$, if the buffer is initially empty, the transients of the outputs can be expressed as follows:

\[
PR(n) = PR_{ss} (1 - \lambda_1^n), \tag{6.60}
\]
\[
WIP(n) = WIP_{ss} (1 - \lambda_1^n), \tag{6.61}
\]

where, according to [17],

\[
PR_{ss} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2}, \tag{6.62}
\]
\[
WIP_{ss} = \frac{p_1}{p_1 + p_2 - p_1 p_2}. \tag{6.63}
\]

Therefore, pre-exponential factors $\Psi_{11} \tilde{x}_1(0) = \Psi_{21} \tilde{x}_1(0) \equiv 1$ for all $p_1, p_2 \in (0, 1)$, i.e., both outputs have transients with identical duration. This phenomenon has also been observed in the identical machine case.

For $N \geq 2$, the following result is obtained:
Numerical Fact 6.5 For serial lines defined by assumptions (a)-(e) of Subsection 2.4.1 and (6.55) with $N \geq 2$,

- $\Psi_{11}\tilde{x}_1(0)$ reaches its maximum at $p_1 = p_2 = \sqrt{p^*}$;
- $\Psi_{21}\tilde{x}_1(0)$ is a monotonically decreasing (respectively, increasing) function of $p_2$ (respectively, $p_1$);
- $\Psi_{11}\tilde{x}_1(0) \to 0$ as $p_i, i = 1, 2$, tends to 1;
- $\Psi_{21}\tilde{x}_1(0) \to 0$ as $p_2$ tends to 1;
- for all $p_1$ and $p_2$, $\Psi_{11}\tilde{x}_1(0) < \Psi_{21}\tilde{x}_1(0)$.

Justification: See Appendix D.

The behavior of $\Psi_{11}\tilde{x}_1(0)$ and $\Psi_{21}\tilde{x}_1(0)$ is illustrated in the second and third columns of Figure 6.15, respectively; in addition, Figure 6.17 is used to illustrate the last statement of Numerical Fact 6.5.

Numerical Fact 6.5 along with Figures 6.15 and 6.17 lead to the following conclusions:

- The transients of $PR$ are the slowest when $p_1 = p_2$; larger drifts, generated by $|p_1 - p_2|$, result in faster transients.
- For all $p^*$ and $N$,
  - the effect of SLE on the transients of $PR$ becomes negligible when $p_2$ tends to 1; this conclusion is similar to that observed in the case of identical machines;
  - the transients of $WIP$ becomes faster as $p_2$ is increasing;
  - the effect of SLE on the transients of $WIP$ becomes negligible as $p_2$ tends to 1; this conclusion is different from that observed in the case of identical machines, where the PEF for $WIP$ is non-zero for any $p$.  

151
- Finally, since $\Psi_{11}\tilde{x}_1(0) < \Psi_{21}\tilde{x}_1(0)$, $\lambda_1$ has a larger effect on the transients of $PR$ than on $WIP$, which implies that, as in the identical machine case, the transients of $PR$ are faster than those of $WIP$.

![Figure 6.17: PEFs as functions of $p_2$ for $p^* = 0.49$ and $N = 2$](image)

To illustrate these conclusions, consider again the three serial lines defined in (6.58). The transients of their $PR$ and $WIP$ are shown in Figure 6.18. As one can see, similar to the $x(n)$ case, $L_1$ has the slowest transients in $PR(n)$ and $WIP(n)$, while $L_3$ has the fastest ones. In addition, in all three lines, $PR$ has faster transients than $WIP$.

![Figure 6.18: Transients of $PR(n)$ and $WIP(n)$ for lines $L_1$, $L_2$ and $L_3$](image)
6.3.3 Transients in the unconstrained case

**Transients of states and outputs:** For \( N \gg 1 \), the properties of SLE and PEFs remain qualitatively the same as in the constrained case. This is illustrated in Figure 6.19 for \( N = 10 \), according to which \( \lambda_1 \) and \( \Psi_{11}\tilde{x}_1(0) \) reach their maxima at \( p_1 = p_2 \), while \( \Psi_{21}\tilde{x}_1(0) \) is monotonically increasing and decreasing in \( p_1 \) and \( p_2 \), respectively; also both \( \Psi_{11}\tilde{x}_1(0) \) and \( \Psi_{21}\tilde{x}_1(0) \) tend to 0 as \( p_2 \) tends to 1. For small \( N \), however, SLE and PEFs may behave qualitatively different, and their general properties are very sensitive to the differences between \( p_1 \) and \( p_2 \). This is illustrated in Figure 6.20 for \( N = 2 \), where for some \( p_1 \)’s SLE is monotonically increasing, while for others, it is decreasing. We do not pursue these issues further but rather use the unconstrained case to analyze reversibility of the transients.

**Reversibility:** *Reversibility of states:* Consider a two-machine Bernoulli serial line \( L \) and its reverse \( L_r \) (see Figure 6.21). It is obvious from (6.56) and (6.57) that \( \lambda_1 \) is not affected by the order the two machines. For \( N \leq 30 \), using the symbolic derivation function of MATLAB®, we determine that the transition matrices of \( L \) and \( L_r \) have the same characteristic polynomials and, thus, the same eigenvalues. For arbitrary \( N \), we provide:

**Numerical Fact 6.6** *Lines \( L \) and \( L_r \) have the same set of eigenvalues.*

**Justification:** See Appendix D.

Thus, the transients of the states remain the same when the parts flow is reversed.

*Reversibility of outputs:* The reversibility properties of \( PR(n) \) and \( WIP(n) \) are characterized as follows:

**Numerical Fact 6.7** *Assume that the buffers are initially empty for both lines,*
Figure 6.19: SLE and PEFs as functions of machine efficiencies for the unconstrained case with $N = 10$
Figure 6.20: SLE and PEFs as functions of machine efficiencies for the unconstrained case with $N = 2$

\[
\begin{align*}
\lambda_i & \quad \Psi_{1i}^k(0) \\
\Psi_{2i}^k(0) &
\end{align*}
\]

Figure 6.21: Two-machine Bernoulli line and its reverse

(a) Line $L$ 
(b) Reversed line $L_r$
\[ x^L(n) = x^{L_r}(n) = [1 \ 0 \ \cdots \ 0]^T. \quad (6.64) \]

Then, lines \( L \) and \( L_r \) have identical pre-exponential factors for \( PR(n) \).

**Justification:** See Appendix D.

Since both \( \lambda_1 \) and \( \Psi_{1i}\bar{x}_i(0), \ i = 0, \ldots, N \), are reversible, we conclude that \( PR(n) \), under condition (6.64), is also reversible, i.e.,

\[ PR^L(n) = PR^{L_r}(n), \quad n = 0, 1, 2, \ldots \quad (6.65) \]

As for the reversibility of \( WIP(n) \), the following is obtained:

**Numerical Fact 6.8** Assume that buffer \( b \) is initially empty and buffer \( b^r \) is initially full, i.e.,

\[ x^L(n) = [1 \ 0 \ \cdots \ 0]^T, \quad x^{L_r}(n) = [0 \ \cdots \ 0 \ 1]^T. \quad (6.66) \]

Then,

\[ WIP^L(n) + WIP^{L_r}(n) = WIP^{L_s} + WIP^{L_r}_{ss}, \quad n = 1, 2, \ldots \quad (6.67) \]

**Justification:** See Appendix D.

As an example, consider a production line, \( L \), defined by assumptions (i)-(vi) with \( p_1 = 0.9, \ p_2 = 0.7, \ N = 5 \), and its reverse \( L_r \). The transients of \( PR(n) \) and \( WIP(n) \) for \( L \) and \( L_r \) are shown in Figure 6.22. Clearly, production rate has the same transients in \( L \) and \( L_r \) under zero initial conditions, while \( WIP^L(n) \) and \( WIP^{L_r}(n) \) has complementary transients under initial condition (6.66).
6.4 Geometric Lines with Identical Machines

6.4.1 Problem formulation

For machines with downtime relatively longer than the cycle time, the Bernoulli model is not applicable. To investigate transients in this case, we address in this section serial lines with geometric machines. Specifically, consider a serial line with two identical geometric machines (see Figure 6.23). This system at hand is described by an ergodic Markov chain. As discussed in Section 6.2, the transients of a production system are characterized by the second largest eigenvalue of its transition matrix. With this in mind, the problems addressed in this section are as follows:

- Analyze the second largest eigenvalue of an individual geometric machine as a function of $P$ and $R$. In particular, investigate the effect of $T_{up} = 1/P$ and $T_{down} = 1/R$ on SLE, under the assumption that the machine efficiency $e$ is fixed.
Carry out similar analyses for two-machine lines. In addition, investigate explicitly the transients of the production rate, \( PR(n) \), i.e., the probability that \( m_2 \) is up and the buffer is not empty at time slot \( n = 1, 2, \ldots \).

Note that the steady state production rate, \( PR(\infty) =: PR_{ss} \), of a production line with geometric machines can be evaluated using the method developed in [74]. Here, we are interested in how \( PR(n) \) approaches the steady state value \( PR_{ss} \).

The interest in the effect of \( T_{up} \) and \( T_{down} \) on the transients stems from the following: It has been shown in Subsection 3.2.3 that

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

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

### 6.4.2 Transients of individual machines

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

\[
x(n+1) = Ax(n), \quad x_0(n) + x_1(n) = 1,
\]

\[
A = \begin{bmatrix} 1 - R & P \\ R & 1 - P \end{bmatrix}.
\]

The eigenvalues of \( A \) are

\[
\lambda_0 = 1, \quad \lambda_1 = 1 - P - R.
\]
and, therefore, the dynamics of the machine states can be expressed as

\[ x_0(n) = (1 - e) + [x_0(0) - (1 - e)](1 - P - R)^n = (1 - e) \left( 1 - \frac{\Delta}{1 - e} \lambda^n \right), \quad (6.71) \]

\[ x_1(n) = e + [x_1(0) - e] (1 - P - R)^n = e \left( 1 + \frac{\Delta}{e} \lambda^n \right), \quad (6.72) \]

\[ \Delta = x_1(0) - e = (1 - e) - x_0(0). \quad (6.73) \]

To investigate the effects of up- and downtime on the transients, consider \( \lambda_1 \) as a function of \( R \) for a fixed \( e \), i.e., \( \lambda_1(R) = 1 - \frac{R}{e} \). The behavior of \( |\lambda_1| \) as a function of \( R \) is illustrated in Figure 6.24. From this figure, we conclude:

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

![Figure 6.24: Behavior of \(|\lambda_1|\) as a function of \( R \)](image)

Next, we address the issue of separate effects of uptime and of downtime on the transients. Recall that, as mentioned in Subsection 3.2.3, increasing the uptime by a factor \( (1 + \alpha) \), \( \alpha > 0 \), or decreasing the downtime by the same factor leads to the same steady state performance for an individual machine since the efficiency in both
cases is the same, i.e.,

\[ e' = \frac{1}{1 + \frac{T_{\text{down}}}{(1+\alpha)T_{\text{up}}}}. \]  

(6.74)

However, the transient properties resulting from both cases are different. Indeed, consider a geometric machine with breakdown and repair probabilities \( P \) and \( R \), respectively. Let \( \lambda^u_1 \) denote the SLE of the machine with the uptime increased by \( (1 + \alpha) \), \( \alpha > 0 \) and \( \lambda^d_1 \) denote the SLE for the same machine with the downtime decreased by the same factor. Then,

**Theorem 6.3** For an individual geometric machine,

\[ |\lambda^u_1| > |\lambda^d_1|, \]  

(6.75)

if

\[ e > 0.5, \quad \frac{T_{\text{down}}}{1 + \alpha} > 2. \]  

(6.76)

**Proof:** See Appendix D.

This theorem implies that if the machine efficiency is larger than 0.5 and the decreased downtime is larger than two cycle times (which typically occur in practice), decreasing the downtime leads to faster transients than increasing the uptime, while preserving the steady state production rate in both cases the same.

To conclude this subsection, we evaluate the eigenvalues of a system consisting of two geometric machines operating independently (without a buffer). In this case, the
transition matrix is:

\[ A = \begin{bmatrix}
(1 - R)^2 & P(1 - R) & P(1 - R) & P^2 \\
(1 - R)R & (1 - P)(1 - R) & RP & P(1 - P) \\
R(1 - R) & RP & (1 - P)(1 - R) & (1 - P)P \\
R^2 & (1 - P)R & (1 - P)R & (1 - P)^2
\end{bmatrix}, \quad (6.77) \]

which implies that the four eigenvalues are:

\[ 1, 1 - P - R, 1 - P - R, (1 - P - R)^2. \quad (6.78) \]

These eigenvalues are used in Subsections 6.4.3 and 6.4.4 for the analysis of transients in two-machine geometric serial lines with a buffer of capacity \( N \geq 1 \).

### 6.4.3 Transients of two-machine lines with \( N = 1 \)

For a serial line with two geometric machines, the state of the system can be denoted by a triple \((h, s_1, s_2)\), where \( h \in \{0, 1, \ldots, N\} \) is the state of the buffer and \( s_i \in \{0, 1\}, \ i = 1, 2 \), are the states of the first and the second machine, respectively. The behavior of the system is described by an ergodic Markov chain. For \( N = 1 \), the transition probability matrix is:

\[ A = \begin{bmatrix}
A_1 & 0 & A_2 & 0 \\
0 & A_3 & 0 & A_4
\end{bmatrix}, \quad (6.79) \]

where

\[ A_1 = \begin{bmatrix}
(1 - R)^2 & (1 - R)P \\
(1 - R)R & (1 - R)(1 - P) \\
R(1 - R) & RP \\
R^2 & R(1 - P)
\end{bmatrix}, \quad A_2 = \begin{bmatrix}
(1 - R)P \\
(1 - R)(1 - P) \\
RP \\
R(1 - P)
\end{bmatrix} \]
\[ A_3 = \begin{bmatrix} (1 - R) P & P^2 & (1 - R)^2 \\ R P & P (1 - P) & (1 - R) R \\ (1 - P) (1 - R) & (1 - P) P & R (1 - R) \end{bmatrix}, \]

\[ A_4 = \begin{bmatrix} (1 - R) P & P^2 \\ R P & P (1 - P) \\ (1 - P) (1 - R) & (1 - P) P \\ (1 - P) R & (1 - P)^2 \end{bmatrix} \]

and 0's are zero-matrices of appropriate dimensionalities. The eight eigenvalues of \( A \) are

\[ [1, 1 - P - R, 1 - P - R, (1 - P - R)^2, (1 - R)^2, 0, 0, 0], \quad (6.80) \]

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

Clearly, the two eigenvalues \( 1 - P - R \) represent, as it follows from Subsection 6.4.2, the dynamics of the individual machines; the eigenvalue \( (1 - P - R)^2 \) represents the transients of a pair of individual machines (note that the states of the machines are determined independently); therefore, the remaining non-zero eigenvalue \( (1 - R)^2 \) can be viewed as describing the transients of the buffer. The last statement is supported by the following two arguments:

First, using the notations

\[ \lambda_m = 1 - P - R, \quad \lambda_b = (1 - R)^2, \]
the transients of the states, i.e.,

\[ x_{h,i,j}(n) = P[h(n) = h, s_1(n) = i, s_2(n) = j], \ n = 0, 1, \ldots, \]

can be represented as

\[ x_{h,i,j}(n) = x_{h,i,j}(1 + B\lambda_b^n + C\lambda_m^n + D(\lambda_m^2)^n), \quad h \in \{0, 1\}, i, j \in \{0, 1\}, \ n = 0, 1, \ldots, \]

(6.81)

where

\[ x_{h,i,j} = \lim_{n \to \infty} x_{h,i,j}(n) \]

and \( B, C \) and \( D \) are constants defined by initial conditions.

**Theorem 6.4** Consider a serial line with two identical geometric machines and \( N = 1 \). Assume that initially the machines are in the steady states, i.e.,

\[ P[s_1(0) = 1] = P[s_2(0) = 1] = e. \]  

(6.82)

Then, in expression (6.81), \( C = D = 0, \forall i, j, h \in \{0, 1\} \).

**Proof:** See Appendix D.

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

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

From (6.80), it is not immediately clear which of the eigenvalues is the SLE. Obviously, the SLE can be either \( 1 - P - R \) or \( (1 - R)^2 \), i.e., either \( \lambda_m \) or \( \lambda_b \). Which one is, in fact, the SLE depends on the relationship between \( P \) and \( R \). To investigate
when $\lambda_m$ or $\lambda_b$ is SLE, consider the simplex $0 < P < R < 1$ in the $(P,R)$-plane (see Figure 6.25). Each point $(P,R)$ implies $e > 0.5$ and each line, $P = kR$, $k < 1$, represents a set of points $(P,R)$ with identical efficiency $e = \frac{1}{1+k}$. Let $\lambda_1$ denote the SLE, i.e., $|\lambda_1| = \max\{|\lambda_m|, \lambda_b\}$. Then, it can be shown that

$$|\lambda_1| = \begin{cases} 
\lambda_m, & \text{if } 0 < P < R(1 - R), \\
\lambda_b, & \text{if } R(1 - R) < P < (1 - R)(2 - R), \\
-\lambda_m, & \text{if } (2 - R)(1 - R) < P < 1.
\end{cases} \quad (6.83)$$

This leads to the partitioning of the simplex according to SLE as shown in Figure 6.25. Specifically, in area I, the transients of the system are defined mostly by an individual machine; in area II, the transients are defined mostly by the buffer; in area III, the transients are again defined mostly by the machine, however, since the eigenvalue in this area is negative, the transients in area III are oscillatory.

![Figure 6.25: Partitioning of the simplex 0 < P < R < 1 according to SLE](image)

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

**Theorem 6.5** Consider a serial line with two identical geometric machines and $N = 1$. Then, for any fixed $e > 0.5$, the SLE is a monotonically decreasing function of $R$ for $R \in (0,0.5)$.
Proof: See Appendix D.

Thus, for $T_{\text{down}} > 2$, shorter up- and downtimes lead to faster transients than longer ones, even if the efficiency $e > 0.5$ remains the same. This phenomenon is illustrated in Figure 6.26.

![Figure 6.26: Transients of $PR$ for $e = 0.9$](image)

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

**Theorem 6.6** Consider a serial line with two identical geometric machines and $N = 1$. Let $|\lambda^u_1|$ and $|\lambda^d_1|$ denote the SLEs resulting from increasing the uptime by $(1 + \alpha)$, $\alpha > 0$, or decreasing its downtime by the same factor, respectively. Then, under assumption (6.76),

$$|\lambda^u_1| > |\lambda^d_1|. \quad (6.84)$$

Proof: See Appendix D.

Thus, the qualitative effect of the uptime and the downtime on the transients in two-machine lines with $N = 1$ remains the same as that for individual machines: under (6.76), it is better to reduce the downtime than increase the uptime in order
to shorten the transients. This phenomenon is illustrated in Figure 6.27.

![Figure 6.27: Transients of $PR$ with increased uptime or decreased downtime for $e = 0.7$, $R = 0.1$ and $e' = 0.9$](image)

6.4.4 Transients of two-machine lines with $N \geq 2$

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

As in the previous subsection, we denote the state of the system by $(h, s)$, where $h \in H = \{0, 1, \ldots, N\}$ is the buffer state and $s = (s_1, s_2) \in S = \{0, 1\} \times \{0, 1\}$ denotes the states of the two machines. The transition probability matrix for this serial line is given by:

$$A = \{p(h', s'|h, s)\},$$

where all variables vary over their ranges. Furthermore, let

$$A_M = \{p(s'|s)\}$$

denote the transition matrix of the machine states given in (6.77). Since the machine
states are independent of the buffer state,
\[ p(h', s'|h, s) = p(h'|h, s)p(s'|s). \] (6.85)

**Theorem 6.7** The eigenvalues of $A_M$ constitute a subset of the eigenvalues of $A$.

**Proof:** See Appendix D.

The eigenvalues of $A_M$, as computed in Subsection 6.4.2, are $1, \lambda_m, \lambda_m, \lambda_m^2$. Thus, these eigenvalues are also eigenvalues of the serial line. Therefore, $\lambda_m$ provides a lower bound on the SLE of the transition matrix $A$.

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

\[ PR(n) = P[\text{buffer is not empty at } n|m_2 \text{ is up at } n]P[m_2 \text{ is up at } n]. \] (6.86)

The second term in the right hand side of (6.86), as it follows from Section 6.4.2, is given by $e(1 + \frac{\Delta}{e} \lambda_m^n)$, where $\Delta$ is defined in (6.73). We approximate the first term by reducing the geometric line to a Bernoulli one with the machines defined by $p^{Ber} = e = \frac{R}{P+R}$, the buffer capacity $N^{Ber} = \left\lfloor \frac{N}{T_{\text{down}}} + 1 \right\rfloor$, where $[x]$ denotes the nearest integer to $x$, and the cycle time $T_{\text{down}}$. For such a line, $P^{Ber}[\text{the buffer is not empty at slot } k], k = 0, 1, 2, \ldots$, can be easily calculated. Since

\[ eP^{Ber}[\text{the buffer is not empty at slot } k] = PR^{Ber}(k), \quad k = 0, 1, 2, \ldots, \] (6.87)

we obtain the following estimate of the production rate for the geometric line:

\[ \overline{PR}(kT_{\text{down}}) = PR^{Ber}(k) \left( 1 + \frac{\Delta}{e} \lambda_m^{kT_{\text{down}}} \right)^2, \quad k = 0, 1, 2, \ldots, \] (6.88)

where the additional multiplier $(1 + \frac{\Delta}{e} \lambda_m^{kT_{\text{down}}})$ is intended to account for the transients of the first machine.
The accuracy of (6.88) has been investigated numerically using 50,000 lines constructed by selecting the machine and buffer parameters randomly and equiprobably from the sets:

\[ e \in [0.6, 0.95], \quad R \in [0.05, 0.5], \quad N \in \{2, 3, \ldots, 40\}. \quad (6.89) \]

A typical example is shown in Figure 6.28, where the accuracy \( \epsilon(kT_{\text{down}}) \) is defined by

\[ \epsilon(kT_{\text{down}}) = \frac{\hat{PR}(kT_{\text{down}})}{PR(\infty)} - \frac{PR(kT_{\text{down}})}{PR(\infty)}, \quad k = 0, 1, 2, \ldots \quad (6.90) \]

and \( \hat{PR}(kT_{\text{down}}) \) is calculated using (6.88), while \( PR(kT_{\text{down}}) \) is obtained by simulations. The average of \( |\epsilon(kT_{\text{down}})| \) is given in Figure 6.29. As one can see, the accuracy is sufficiently high.

Figure 6.28: Illustration of the accuracy of expression (6.88) for \( e = 0.9 \) and \( R = 0.1 \)
Using approximation (6.88), the effects of up- and downtime on the transients can be evaluated. Since this is carried out numerically, we formulated the results as numerical facts.

**Numerical Fact 6.9** *Consider a geometric line with two identical machines having \( e > 0.5 \) and \( N \geq 2 \). Then, for any \( T_{\text{down}} > 2 \), shorter up- and downtimes practically always lead to faster transients than longer ones.*

**Justification:** See Appendix D.

**Numerical Fact 6.10** *Under condition (6.76), reducing downtime practically always leads to shorter transients than increasing uptime.*

**Justification:** See Appendix D.

## 6.5 Application: Float-based Production Systems

### 6.5.1 System description

Consider a production system consisting, for simplicity, of two departments (or operations, cells, machines, etc.). Assume that the average production rates of the first and
the second departments in isolation (i.e., when each is neither starved nor blocked) are \( p_1 \) and \( p_2 \), respectively, and that the second department is more efficient than the first, i.e.,

\[ p_1 < p_2. \]

In this case, in its steady state, the system cannot produce during a shift of duration \( T_{\text{shift}} \) more than \( p_1 T_{\text{shift}} \) parts, even if the buffer between the departments is infinite [75].

Assume that the production schedule calls for producing more than that but less than or equal to \( p_2 T_{\text{shift}} \). In this case, two modes of operation are possible. The first one is to have the whole system operate overtime. The second is to have only the first department operate overtime and build up a supply of parts, typically referred to as the “float”, between the two departments, so that during the next shift, the whole system is producing at the rate of the second, i.e., the best, department. Clearly, the second mode of operation is preferable since it calls for only a part of the system to operate overtime.

The block diagram of such a system is shown in Figure 6.30, where the storage between the two departments is referred to as the float, rather than the buffer, in order to indicate that the departments operate during a different periods of time: \( T_{\text{shift}} \) for the second department and \( T_{\text{shift}} + T_{\text{float}} \) for the first, where \( T_{\text{float}} \) is the additional time necessary to fill up the float of capacity \( F \).

![Figure 6.30: Float-based production system with two departments](image)

Many machining systems in large volume manufacturing operate using this approach. Clearly, this float-based system is different from the traditional buffer-based system in that it always operates in the transient regime, since at the beginning of
the shift the “buffer” is full and is gradually approaching its steady state occupancy during the whole system operation. This allows for the production rate of the system during the shift to be equal to that on the best, rather than the worst, department.

A question arises: What is the smallest float capacity, which is necessary and sufficient to ensure that during the span of the shift the production rate of the system is equal or arbitrarily close to that of the best department in isolation? This is the question addressed in this section. Specifically, three approaches to lean float design. The first two are based on the average part flow and the first passage time theory [76], respectively. The third one is based on the transient analysis of Bernoulli lines with non-identical machines developed in Section 6.3. We show that although all three methods provide somewhat similar results, the third one is preferable. In addition, since the calculations involved in the transient analysis are somewhat complicated, we indicate how to modify the outcomes of the simplest one, the first method, in order to match the performance of the third.

6.5.2 Terminal performance measures and problem addressed

Consider a float-based production system shown in Figure 6.30 defined by assumptions (a)-(e) of Subsection 2.4.1 (with machines and buffer replaced by departments and float, respectively). In addition, assume that the system operates on a shift-basis, with shift duration $T_{\text{shift}}$ time slots. At the end of each shift, $d_2$ is shut down, while $d_1$ will continue operating until the float is full. This assumption implies that the system operates over a finite period of time. Therefore, transient properties of the system must be used.

In the framework of the above model, along with the transient performance measures $PR(n)$ and $WIP(n)$, the following terminal performance measures are necessary to describe float-based production systems:

- **Production deficit, PD**: the average production losses during a shift compared
to the maximum production capability of the system:

$$PD = \frac{\sum_{n=1}^{T_{shift}} [p_2 - PR(n)]}{p_2 T_{shift}} \cdot 100\%.$$ \hspace{1cm} (6.91)

- **Service level, SL**: the ratio of production rate at the end of a shift to the efficiency of the best department in the system:

$$SL = \frac{PR(T_{shift})}{p_2}.$$ \hspace{1cm} (6.92)

- **Float build-up time, T_{float}**: the average time necessary for department $d_1$ to build up the float after each shift:

$$T_{float} = \frac{F - WIP(T_{shift})}{p_1}.$$ \hspace{1cm} (6.93)

In terms of the terminal performance measures, the problem addressed here is the following: Given $p_i$, $i = 1, 2$, and $T_{shift}$, select the smallest float capacity, $F$, which ensures that the total production during a shift is close to $p_2 T_{shift}$, i.e., $PD \approx 0$ and $SL \approx 1$, and evaluate $T_{float}$.

Note that the terminal performance measures, involved in the above problem, can be evaluated as follows:

$$PD = \frac{\sum_{n=1}^{T_{shift}} [p_2 - C_1 x(n)]}{p_2 T_{shift}} \cdot 100\%,$$ \hspace{1cm} (6.94)

$$SL = \frac{C_1 x(T_{shift})}{p_2} = \frac{C_1 A T_{shift} x(0)}{p_2},$$ \hspace{1cm} (6.95)

$$T_{float} = \frac{F - C_2 x(T_{shift})}{p_1} = \frac{F - C_2 A T_{shift} x(0)}{p_1},$$ \hspace{1cm} (6.96)

where $x(n)$ is calculated using (6.2) and $C_1$ and $C_2$ are the first and second row of the matrix $C$ given in (6.54).

To illustrate the behavior of the transient and terminal performance measures,
consider a float-based system with

\[ p_1 = 0.6, \quad p_2 = 0.8, \quad F = 50, \quad T_{\text{shift}} = 500. \tag{6.97} \]

The evolution of the system is calculated using (6.5)-(6.7), and the transients of \( PR(n) \) and \( WIP(n) \) are shown in Figure 6.31. Clearly, at the beginning of the shift, due to the availability of work-in-process in the float, the system operates with production rate \( p_2 \). As the work-in-process is depleted almost linearly with rate \( (p_2 - p_1) \), \( PR(n) \) starts to decrease after \( n = F \) cycle times and reaches its steady state (close to \( p_1 \)) before the end of the shift.

![Graphs of PR(n) and WIP(n)](image)

Figure 6.31: Transient performance measures

To illustrate the behavior of the terminal performance measures, we investigate them as functions of float capacity \( F \). The results are shown in Figure 6.32. Based on this figure, we conclude that the smallest float capacity that leads to \( PD \approx 0 \) and \( SL \approx 1 \) is \( F \approx 125 \), while the shortest overtime for the first department is \( T_{\text{float}} \approx 167 \) cycles.

### 6.5.3 Design of lean float

**Approaches:** To design the lean float of an unbalanced production system, the following three approaches are used:

- *Stationary Flow Time:* To maintain the production rate of a float-based pro-
Figure 6.32: Terminal performance measures
duction line at $p_2$, the float must be non-empty during the entire shift. As illustrated in Figure 6.31, the work-in-process of a production line with float is decreasing practically linearly with rate $(p_2 - p_1)$ before reaching the steady state. Therefore, to remain non-empty, the capacity of the float can be estimated based on the stationary flow time as follows:

$$\frac{F}{p_2 - p_1} \geq T_{\text{shift}},$$

(6.98)
i.e., the lean float capacity can be calculated as

$$F(p_1, p_2, T) = \lceil (p_2 - p_1)T_{\text{shift}} \rceil, \quad (6.99)$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to $x$.

Although (6.99) provides a simple estimate of lean float capacity, its shortcomings are obvious: First, this approach is based on the stationary flow and, thus, ignores random fluctuations. Second, although it is intuitively clear that the lean float should depend on $p_1$ and $p_2$, according to (4.2), $F$ remains the same for any $p_1$ and $p_2$ as long as the difference $p_2 - p_1$ is constant. In the following, we refer to this method as Approach I and denote the float capacity obtained as $F_I$.

- **Average first passage time:** Another way to estimate the lean float capacity is based on the Markov chain of the system with the states being the occupancy of the float. Since this Markov chain is ergodic, as mentioned in Subsection 6.2.1, the system reaches state 0, i.e. empty float, with probability 1.

Let $t_{ij}$ denote the average time for the system to reach state $j$ from state $i$, which is referred to as the *average first passage time*. Then $t_{F0}$ is the average time that the float is depleted from being full. It follows from the total probability formula that the average first passage times to state 0 from any other state can
be calculated as follows

\[
\begin{bmatrix}
t_{10} \\
t_{20} \\
\vdots \\
t_{F0}
\end{bmatrix} = \begin{bmatrix} 1 \\
1 \\
\vdots \\
1
\end{bmatrix} + \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1F} \\
P_{21} & P_{22} & \cdots & P_{2F} \\
\vdots & \vdots & \ddots & \vdots \\
P_{F1} & P_{F2} & \cdots & P_{FF}
\end{bmatrix} \begin{bmatrix} t_{10} \\
t_{20} \\
\vdots \\
t_{F0}
\end{bmatrix},
\]

(6.100)

where \( P_{ij} \) is the transition probability from state \( i \) to state \( j \). Since the float should not be empty during a shift, the problem of design amounts to finding the smallest \( F \) such that

\[ t_{F0} \geq T_{\text{shift}}. \]

(6.101)

To find this \( F \), we first calculate \( t_{F0} \) for \( F = F_I \), and then either increase or decrease it by one, depending on the relationship between \( t_{F,I0} \) and \( T_{\text{shift}} \). We refer to this method as Approach II and denote the float capacity obtained by this approach as \( F_{II} \).

- **Transients analysis:** The above two approaches estimate the desired float capacity using the stationary flow time and the average first passage time, respectively, both of which are based on the stationary behavior of the system. However, since the system under consideration operates during a finite period of time, the third method is based on transient analysis. Specifically, we calculate the production rate of the system, \( PR(n) \), \( n = 0, 1, \ldots, T_{\text{shift}} \), using (6.5)-(6.7) and select the smallest \( F \) so that

\[ PR(n) \geq p_2(1 - \epsilon), \quad \forall n \leq T_{\text{shift}}, \]

(6.102)

where \( 0 < \epsilon \ll 1 \). This \( F \) is calculated starting also from \( F_I \) and then increasing or decreasing it appropriately. We denote the resulting float capacity as \( F_{III} \) and refer to this method as Approach III.
Evaluation of the approaches: Next, we investigate the efficacy of the three approaches proposed above. To accomplish this, we construct 15 production lines as follows: First, we select $p_2$ from set $\{0.8, 0.9, 0.95\}$. Then, we select $p_1$ from set $\{p_2 - 0.4, p_2 - 0.3, p_2 - 0.2, p_2 - 0.1, p_2 - 0.05\}$. Also, we assume that $T_{\text{shift}} = 500$. Under these assumptions, we calculate the float capacities for all 15 pairs of $(p_1, p_2)$, using the three approaches, where $\epsilon$ for Approach III is selected as 0.01. The results are summarized in Tables 6.6–6.9.

<table>
<thead>
<tr>
<th>$(p_1, p_2)$</th>
<th>$F_I$</th>
<th>$F_{II}$</th>
<th>$F_{III}(\epsilon = 0.01)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.55, 0.95)</td>
<td>200</td>
<td>201</td>
<td>226</td>
</tr>
<tr>
<td>(0.65, 0.95)</td>
<td>150</td>
<td>151</td>
<td>174</td>
</tr>
<tr>
<td>(0.75, 0.95)</td>
<td>100</td>
<td>101</td>
<td>120</td>
</tr>
<tr>
<td>(0.85, 0.95)</td>
<td>50</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td>(0.90, 0.95)</td>
<td>25</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>(0.50, 0.90)</td>
<td>200</td>
<td>201</td>
<td>228</td>
</tr>
<tr>
<td>(0.60, 0.90)</td>
<td>150</td>
<td>151</td>
<td>176</td>
</tr>
<tr>
<td>(0.70, 0.90)</td>
<td>100</td>
<td>101</td>
<td>123</td>
</tr>
<tr>
<td>(0.80, 0.90)</td>
<td>50</td>
<td>51</td>
<td>69</td>
</tr>
<tr>
<td>(0.85, 0.90)</td>
<td>25</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>(0.40, 0.80)</td>
<td>200</td>
<td>201</td>
<td>231</td>
</tr>
<tr>
<td>(0.50, 0.80)</td>
<td>150</td>
<td>151</td>
<td>180</td>
</tr>
<tr>
<td>(0.60, 0.80)</td>
<td>100</td>
<td>101</td>
<td>128</td>
</tr>
<tr>
<td>(0.70, 0.80)</td>
<td>50</td>
<td>51</td>
<td>74</td>
</tr>
<tr>
<td>(0.75, 0.80)</td>
<td>25</td>
<td>27</td>
<td>46</td>
</tr>
</tbody>
</table>

It can be concluded from Table 6.6 that Approaches I and II result in very close float capacities, while Approach III usually results in 15–30% larger floats. In addition, in Approach III, more efficient departments lead to smaller float capacity, even if $(p_2 - p_1)$ remains the same. This can be explained by the fact that more efficient departments are more synchronized, i.e., both departments are up or down simultaneously with larger probability; such a “synchronization” slows down the depletion of the float and, as a result, requires less parts to maintain a given service level.

While the relative production deficit under float capacities obtained by all methods
Table 6.7: Production deficit

<table>
<thead>
<tr>
<th>$(p_1, p_2)$</th>
<th>$PD(F_I)$</th>
<th>$PD(F_{II})$</th>
<th>$PD(F_{III})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.55, 0.95)</td>
<td>1.14</td>
<td>1.03</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.65, 0.95)</td>
<td>1.11</td>
<td>1.00</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.75, 0.95)</td>
<td>1.05</td>
<td>0.94</td>
<td>0.04</td>
</tr>
<tr>
<td>(0.85, 0.95)</td>
<td>0.97</td>
<td>0.86</td>
<td>0.07</td>
</tr>
<tr>
<td>(0.90, 0.95)</td>
<td>0.97</td>
<td>0.84</td>
<td>0.13</td>
</tr>
<tr>
<td>(0.50, 0.90)</td>
<td>1.29</td>
<td>1.17</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.60, 0.90)</td>
<td>1.28</td>
<td>1.17</td>
<td>0.03</td>
</tr>
<tr>
<td>(0.70, 0.90)</td>
<td>1.26</td>
<td>1.14</td>
<td>0.04</td>
</tr>
<tr>
<td>(0.80, 0.90)</td>
<td>1.26</td>
<td>1.13</td>
<td>0.07</td>
</tr>
<tr>
<td>(0.85, 0.90)</td>
<td>1.36</td>
<td>1.09</td>
<td>0.14</td>
</tr>
<tr>
<td>(0.40, 0.80)</td>
<td>1.56</td>
<td>1.43</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.50, 0.80)</td>
<td>1.61</td>
<td>1.48</td>
<td>0.03</td>
</tr>
<tr>
<td>(0.60, 0.80)</td>
<td>1.65</td>
<td>1.52</td>
<td>0.04</td>
</tr>
<tr>
<td>(0.70, 0.80)</td>
<td>1.78</td>
<td>1.63</td>
<td>0.08</td>
</tr>
<tr>
<td>(0.75, 0.80)</td>
<td>2.06</td>
<td>1.73</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 6.8: Service level

<table>
<thead>
<tr>
<th>$(p_1, p_2)$</th>
<th>$SL(F_I)$</th>
<th>$SL(F_{II})$</th>
<th>$SL(F_{III})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.55, 0.95)</td>
<td>0.7720</td>
<td>0.7858</td>
<td>0.9914</td>
</tr>
<tr>
<td>(0.65, 0.95)</td>
<td>0.8264</td>
<td>0.8372</td>
<td>0.9914</td>
</tr>
<tr>
<td>(0.75, 0.95)</td>
<td>0.8810</td>
<td>0.8887</td>
<td>0.9903</td>
</tr>
<tr>
<td>(0.85, 0.95)</td>
<td>0.9358</td>
<td>0.9403</td>
<td>0.9907</td>
</tr>
<tr>
<td>(0.90, 0.95)</td>
<td>0.9633</td>
<td>0.9659</td>
<td>0.9905</td>
</tr>
<tr>
<td>(0.50, 0.90)</td>
<td>0.7586</td>
<td>0.7722</td>
<td>0.9912</td>
</tr>
<tr>
<td>(0.60, 0.90)</td>
<td>0.8156</td>
<td>0.8260</td>
<td>0.9905</td>
</tr>
<tr>
<td>(0.70, 0.90)</td>
<td>0.8726</td>
<td>0.8799</td>
<td>0.9903</td>
</tr>
<tr>
<td>(0.80, 0.90)</td>
<td>0.9298</td>
<td>0.9338</td>
<td>0.9913</td>
</tr>
<tr>
<td>(0.85, 0.90)</td>
<td>0.9584</td>
<td>0.9627</td>
<td>0.9905</td>
</tr>
<tr>
<td>(0.40, 0.80)</td>
<td>0.7280</td>
<td>0.7420</td>
<td>0.9913</td>
</tr>
<tr>
<td>(0.50, 0.80)</td>
<td>0.7912</td>
<td>0.8017</td>
<td>0.9910</td>
</tr>
<tr>
<td>(0.60, 0.80)</td>
<td>0.8544</td>
<td>0.8614</td>
<td>0.9911</td>
</tr>
<tr>
<td>(0.70, 0.80)</td>
<td>0.9176</td>
<td>0.9212</td>
<td>0.9908</td>
</tr>
<tr>
<td>(0.75, 0.80)</td>
<td>0.9490</td>
<td>0.9528</td>
<td>0.9906</td>
</tr>
</tbody>
</table>
Table 6.9: Float build-up time

<table>
<thead>
<tr>
<th>$(p_1, p_2)$</th>
<th>$T_{\text{float}}(F_I)$</th>
<th>$T_{\text{float}}(F_{II})$</th>
<th>$T_{\text{float}}(F_{III})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.55, 0.95)</td>
<td>354.3</td>
<td>355.3</td>
<td>363.6</td>
</tr>
<tr>
<td>(0.65, 0.95)</td>
<td>223.1</td>
<td>223.9</td>
<td>230.8</td>
</tr>
<tr>
<td>(0.75, 0.95)</td>
<td>127.1</td>
<td>127.8</td>
<td>133.3</td>
</tr>
<tr>
<td>(0.85, 0.95)</td>
<td>54.0</td>
<td>54.6</td>
<td>59.0</td>
</tr>
<tr>
<td>(0.90, 0.95)</td>
<td>23.7</td>
<td>24.4</td>
<td>28.1</td>
</tr>
<tr>
<td>(0.50, 0.90)</td>
<td>389.1</td>
<td>390.1</td>
<td>400.1</td>
</tr>
<tr>
<td>(0.60, 0.90)</td>
<td>241.0</td>
<td>241.8</td>
<td>250.1</td>
</tr>
<tr>
<td>(0.70, 0.90)</td>
<td>135.4</td>
<td>136.2</td>
<td>143.1</td>
</tr>
<tr>
<td>(0.80, 0.90)</td>
<td>56.5</td>
<td>57.2</td>
<td>63.1</td>
</tr>
<tr>
<td>(0.85, 0.90)</td>
<td>24.2</td>
<td>25.7</td>
<td>30.7</td>
</tr>
<tr>
<td>(0.40, 0.80)</td>
<td>485.4</td>
<td>486.7</td>
<td>500.3</td>
</tr>
<tr>
<td>(0.50, 0.80)</td>
<td>288.1</td>
<td>289.2</td>
<td>300.5</td>
</tr>
<tr>
<td>(0.60, 0.80)</td>
<td>156.8</td>
<td>157.7</td>
<td>167.4</td>
</tr>
<tr>
<td>(0.70, 0.80)</td>
<td>63.4</td>
<td>64.2</td>
<td>73.0</td>
</tr>
<tr>
<td>(0.75, 0.80)</td>
<td>26.3</td>
<td>28.1</td>
<td>36.5</td>
</tr>
</tbody>
</table>

are very small (less than 2%), the service levels resulted by $F_I$ and $F_{II}$ are significantly lower than that obtained by $F_{III}$. Specifically, Approaches I and II may result in $SL$ less than 0.75, whereas Approach III gives $SL \geq 1 - \epsilon = 0.99$. Of course, larger float capacities in Approach III lead to longer float build-up time, but the difference in comparison with Approaches I and II is insignificant.

Thus, the main difference in performance resulting from three approaches is the service level. To illustrate this phenomenon, consider again the production line analyzed in Subsection 6.5.2. The lean float capacities are calculated with $\epsilon = 0.01$ in Approach III. The performance measures resulted from the three approaches are shown in Table 6.10 and the corresponding behavior $PR(n)$ and $WIP(n)$ is illustrated in Figure 6.33. Clearly, in all three cases, $WIP(n)$ is linearly decreasing with rate $p_2 - p_1$. However, due to a larger float in Approach III, more jobs are available at the end of the shift, which leads the production rate close to $p_2$, while the production rate resulted from $F_I$ and $F_{II}$ drop significantly.

Finally, we discuss the computational effort involved in the three approaches.
Table 6.10: Comparison of Approaches I, II and III

<table>
<thead>
<tr>
<th></th>
<th>Approach I</th>
<th>Approach II</th>
<th>Approach III</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F )</td>
<td>100</td>
<td>101</td>
<td>128</td>
</tr>
<tr>
<td>( PD )</td>
<td>1.65</td>
<td>1.52</td>
<td>0.04</td>
</tr>
<tr>
<td>( SL )</td>
<td>0.8544</td>
<td>0.8614</td>
<td>0.9911</td>
</tr>
<tr>
<td>( T_{\text{float}} )</td>
<td>156.8</td>
<td>157.7</td>
<td>167.4</td>
</tr>
</tbody>
</table>

From this point of view, Approach I is, of course, the simplest, since Approaches II and III require iterations, which are time-consuming when \( T_{\text{shift}} > 1000 \). On the other hand, the desired end-of-shift service level can only be assured using Approach III. Therefore, it is recommended that to design a lean float-based production system with minimal computational effort, one should first calculate \( F_I \) and then increase it by 15–30% (depending on the desired service level, available storage space, operational expenses, etc.).

6.6 Summary

- The transients of production systems can be characterized in terms of the following metrics: the second largest eigenvalue (SLE) of the transition matrix; pre-exponential factors (PEF); settling time; and production losses due to transients.
- The transients become more sluggish in systems with many machines and long
buffers.

- The transients of production rate are faster than those of work-in-process.
- Production losses due to transients can be quite substantial (up to 12% in systems with low efficiency machines).
- In order to avoid production losses due to transients, the buffers at the beginning of the shift must be at least half full.
- The transients in systems with non-identical machines are, in general, faster than in systems with identical machines.
- Transients are reversible in the same manner as the steady state values of the performance measures are.
- Shorter up- and downtimes lead to faster transients than longer ones.
- A reduction in downtime leads to faster transients than a similar increase of the uptime.
- For production systems with floats, the lean float capacity should be selected as

\[
F \in \{1.15 \lceil (p_2 - p_1)T_{\text{shift}} \rceil, 1.3 \lceil (p_2 - p_1)T_{\text{shift}} \rceil\},
\]

where \(\lceil x \rceil\) denotes the smallest integer larger than or equal to \(x\), \(p_i, i = 1, 2\), are the efficiencies of the two departments, and \(T_{\text{shift}}\) is the shift duration in units of cycle time.
CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

In this dissertation, system-theoretic properties of production lines are described. Specifically, the follow results have been reported:

- **Reversibility:** The reversibility property is observed in serial lines and closed lines, regardless of the reliability models of the machines. Serial lines with quality issues also possess the reversibility if no inspection device is present. In lines with rework, the reversibility is violated due to the asymmetric blockage and starvation at the split and merge machine. In addition, it is shown that in a production line with rework more efficient machines should be placed towards the end of the line in order to have higher throughput.

- **Monotonicity:** In this work, the monotonicity property refers to the monotonic property of system throughput as a function of machine, buffer, and system parameters. For all system studied in this work, the throughput is always monotonic to all machine and buffer parameters except for lines with QQC machines. In addition, it has been shown that the throughput of a closed line is a non-monotonic concave function of the number of carriers in the system.

- **Effects of up- and downtime:** To characterize the effects of up- and downtime on system performance, the following results have been obtained in serial lines:
Shorter up- and downtime lead to larger throughput than longer ones, even if the machine efficiencies remain the same.

Decreasing $T_{down}$ by any factor leads to a larger throughput than increasing $T_{up}$ by the same factor.

**Improvability:** In this work, we investigate the improvability properties of serial lines and closed lines. As a result, the following practical indicators are obtained:

- A serial line is practically unimprovable with respect to workforce or cycle time re-allocation if each buffer is, on average, close to being half full.

- A serial line is practically unimprovable with respect to buffer capacity re-allocation if the average occupancy of each buffer is close to the average availability of its immediate downstream buffer.

- The throughput of a closed line can be increased by adding a carrier if the sum of the frequencies of machine starvations is larger than the sum of the frequencies of machine blockages; if the inequality is reversed, the throughput can be increased by removing a carrier.

**Bottlenecks:** The bottleneck studied in this work is defined as the machine (or buffer), which has the largest effect on the system performance. For serial lines, an arrow-based method is described to identify the bottleneck machine and bottleneck buffer. This method is extended to closed lines by using virtual, rather than real, blockages and starvations of the machines. In production lines with quality issues, local bottleneck is introduced based on segmentation or overlapping decomposition. Then, the arrow-based method is applied to identify each local bottleneck in order to facilitate determining the global one.

**Transient properties:** In serial production lines, the transients (i.e., the pro-
cess of reaching the steady states) of the throughput are typically orders of magnitude faster than the transients of work-in-process. Larger buffers, as well as longer lines, lead to longer transients. Larger difference of machine efficiencies leads to shorter transients. To ensure faster transients of throughput, machines with shorter up- and downtime should be used rather than longer one. Decreasing $T_{\text{down}}$ by any factor leads to a transients in throughput than increasing $T_{\text{up}}$ by the same factor. The throughput losses due to transients may be as high as 10%, if at the beginning of the shift all buffers were empty. To eliminate throughput losses due to transients, all buffers should be initially at least half full.

7.2 Future work

Future work on system-theoretic properties of production systems includes:

- Extensions of the results obtained to other quality models, e.g., geometric, exponential or general.

- Extension of the transient analysis to exponential, Weibull, gamma, log-normal, and other reliability models.

- Generalization of the results for production systems with different topologies, e.g., assembly systems, closed lines, re-entrant lines, etc.

- Generalization of the results for small volume job-shop production environment with high product variety (e.g., semiconductor fabrication).

- Investigation of active feedback control and related system-theoretic properties for production systems.
APPENDICES
APPENDIX A

PROOFS AND JUSTIFICATIONS FOR
CHAPTER III

To prove the theorems and justify the numerical facts in Chapter III, we need the formulas for performance evaluation of serial exponential lines. The formulas for $PR$, $ST_i$ and $BL_i$ evaluation in two-machine exponential lines have been derived in [77]. Based on these formulas, recursive aggregation procedures were developed in [21] and [23] to estimate these performance measures in $M > 2$-machine lines. Expressions for $WIP_i$, $i = 1, \ldots, M$, however, are not available in the literature. Since these quantities and the aggregation procedures are used in the investigations below, we derive formulas for evaluating the work-in-process, and review the recursive procedure for exponential lines:

Performance evaluation of synchronous exponential lines

Two-machine case: The system under consideration is described by a continuous time, mixed state Markov process. The state of the system is denoted by a triple $(h, s_1, s_2)$, where $h$ is the state of the buffer and $s_i$, $i = 1, 2$, are the states of the first and the second machine, respectively. The transition rates of this process from state $(s_1 = i, s_2 = j)$ to state $(s_1 = k, s_2 = l)$, i.e., $\nu_{kl,ij}$, are as follows:

\[
\begin{align*}
\nu_{11,00} &= 0, & \nu_{10,00} &= \mu_1, & \nu_{01,00} &= \mu_2, \\
\nu_{11,01} &= \mu_1, & \nu_{10,01} &= 0, & \nu_{00,01} &= \lambda_2, \\
\nu_{11,10} &= \mu_2, & \nu_{01,10} &= 0, & \nu_{00,10} &= \lambda_1.
\end{align*}
\]
\[ \nu_{10,11} = \lambda_2, \quad \nu_{01,11} = \lambda_1, \quad \nu_{00,11} = 0. \]

While the discrete part of the Markov process at hand is described by the above transition rates, the continuous part, i.e., buffer occupancy, is characterized by

\[
\left. \frac{dh(t)}{dt} \right|_{0 < h(t) < N} = \begin{cases} 
1, & \text{if } s_1(t) = 1, s_2(t) = 0, \\
0, & \text{if } s_1(t) = s_2(t), \\
-1, & \text{if } s_1(t) = 0, s_2(t) = 1,
\end{cases}
\]

\[
\left. \frac{dh(t)}{dt} \right|_{h(t)=0} = \begin{cases} 
1, & \text{if } s_1(t) = 1, s_2(t) = 0, \\
0, & \text{otherwise},
\end{cases}
\]

\[
\left. \frac{dh(t)}{dt} \right|_{h(t)=N} = \begin{cases} 
-1, & \text{if } s_1(t) = 0, s_2(t) = 1, \\
0, & \text{otherwise}.
\end{cases}
\]

The states of the system can be separated into two groups: boundary \((0, s_1, s_2), (N, s_1, s_2)\) and internal \((h, s_1, s_2), 0 < h < N\). It turns out that the boundary states are described by probability mass function

\[
P[h(t) = 0, s_1(t) = i, s_2(t) = j] = P_{0,ij}(t), \tag{A.1}
\]

\[
P[h(t) = N, s_1(t) = i, s_2(t) = j] = P_{N,ij}(t), \tag{A.2}
\]

while internal states are described by probability density function

\[
P \left[ h - \frac{\Delta h}{2} \leq h(t) \leq h + \frac{\Delta h}{2}, s_1(t) = i, s_2(t) = j \right] = f_{H,I,J}(h, i, j, t) \Delta h + o(\Delta h), \tag{A.3}
\]

where \(\Delta h \ll 1\).

Thus, the steady state pdf for the the internal states \(h \in (0, N)\) are characterized by the following equations:

\[
f_{H,I,J}(h, 1, 1) = f_{H,I,J}(h, 1, 0) \frac{\mu_2}{\lambda_1 + \lambda_2} + f_{H,I,J}(h, 0, 1) \frac{\mu_1}{\lambda_1 + \lambda_2},
\]
\[
\frac{\partial f_{H,I,J}(h, 1, 0)}{\partial h} = f_{H,I,J}(h, 1, 1)\lambda_2 - f_{H,I,J}(h, 1, 0)(\lambda_1 + \mu_2) + f_{H,I,J}(h, 0, 0)\mu_1,
\]
\[
\frac{\partial f_{H,I,J}(h, 0, 1)}{\partial h} = f_{H,I,J}(h, 0, 1)(\lambda_2 + \mu_1) - f_{H,I,J}(h, 1, 1)\lambda_1 - f_{H,I,J}(h, 0, 0)\mu_2,
\]
\[
f_{H,I,J}(h, 0, 0) = f_{H,I,J}(h, 1, 0) \frac{\lambda_1}{\mu_1 + \mu_2} + f_{H,I,J}(h, 0, 1) \frac{\lambda_2}{\mu_1 + \mu_2}. \tag{A.4}
\]

Since \(h(t)\) is continuous in \(t\), events

\[ A_1 = \{ h(t) = h \text{ and } h(t) \text{ is increasing in } t \} \]

and

\[ A_2 = \{ h(t) = h \text{ and } h(t) \text{ is decreasing in } t \} \]

occur alternately with the equal frequencies. Therefore,

\[ f_{H,I,J}(h, 1, 0) = f_{H,I,J}(h, 0, 1). \tag{A.5} \]

The stationary probabilities of the boundary states, \(h = 0\) and \(h = N\), are derived as follows: The evolution equations for these probabilities are:

\[
P_{0,11}(t + \delta t) = P_{0,11}(t)[1 - (\lambda_1 + \lambda_2)\delta t] + P_{0,01}(t)\mu_1\delta t + o(\delta t),
\]
\[
P_{0,10}(t + \delta t) = 0,
\]
\[
P_{0,01}(t + \delta t) = P_{0,01}(t)[1 - (\mu_1 + \lambda_2)\delta t] + P_{0,11}(t)\lambda_1\delta t + P_{0,00}(t)\mu_2\delta t + f_{H,I,J}(0, 0, 1, t)\delta t + o(\delta t),
\]
\[
P_{0,00}(t + \delta t) = P_{0,00}(t)[1 - (\mu_1 + \mu_2)\delta t] + P_{0,01}(t)\lambda_2\delta t + o(\delta t),
\]
\[
P_{N,11}(t + \delta t) = P_{N,11}(t)[1 - (\lambda_1 + \lambda_2)\delta t] + P_{N,10}(t)\mu_2\delta t + o(\delta t),
\]
\[
P_{N,10}(t + \delta t) = P_{N,10}(t)[1 - (\lambda_1 + \mu_2)\delta t] + P_{N,11}(t)\lambda_2\delta t + P_{N,00}(t)\mu_1\delta t + f_{H,I,J}(N, 1, 0, t)\delta t + o(\delta t),
\]
\[
P_{N,01}(t + \delta t) = 0,
\]
\[
P_{N,00}(t + \delta t) = P_{N,00}(t)[1 - (\mu_1 + \mu_2)\delta t] + P_{N,10}(t)\lambda_1\delta t + o(\delta t).\]
In the limit $\delta t \to 0$, these equations lead to the following steady state relationships:

\[
\begin{align*}
P_{0,11} &= \frac{\mu_1}{\lambda_1 + \lambda_2} P_{0,01}, \\
P_{0,10} &= 0, \\
P_{0,01} &= \frac{1}{\mu_1 + \lambda_2} \left[ f_{H,I,J}(0, 0, 1) + \lambda_1 P_{0,11} + \mu_2 P_{0,00} \right], \\
P_{0,00} &= \frac{\lambda_2}{\mu_1 + \mu_2} P_{0,01}, \\
P_{N,11} &= \frac{\mu_2}{\lambda_1 + \lambda_2} P_{N,10}, \\
P_{N,10} &= \frac{1}{\lambda_1 + \mu_2} \left[ f_{H,I,J}(N, 1, 0) + \lambda_2 P_{N,11} + \mu_1 P_{N,00} \right], \\
P_{N,01} &= 0, \\
P_{N,00} &= \frac{\lambda_1}{\mu_1 + \mu_2} P_{N,10}.
\end{align*}
\] (A.6)

Solving (A.4) and (A.5), we obtain

\[
\begin{align*}
f_{H,I,J}(h, 1, 0) &= f_{H,I,J}(h, 0, 1) = C_0 e^{Kh}, \\
f_{H,I,J}(h, 1, 1) &= C_0 D_1 e^{Kh}, \\
f_{H,I,J}(h, 0, 0) &= \frac{C_0}{D_1} e^{Kh},
\end{align*}
\] (A.7)

where

\[
K = \begin{cases} 
\frac{(\mu_1 + \mu_2 + \lambda_1 + \lambda_2)(\lambda_2 \mu_1 - \lambda_1 \mu_2)}{(\mu_1 + \mu_2)(\lambda_1 + \lambda_2)}, & \text{for } e_1 \neq e_2, \\
0, & \text{for } e_1 = e_2,
\end{cases}
\] (A.8)

\[
D_1 = \frac{\mu_1 + \mu_2}{\lambda_1 + \lambda_2},
\] (A.9)

and $C_0$ is a free constant. Substituting (A.7) into (A.6) results in

\[
\begin{align*}
P_{0,11} &= \frac{(\mu_1 + \mu_2)C_0}{\lambda_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)}, \\
P_{0,10} &= 0, \\
P_{0,01} &= \frac{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)C_0}{\lambda_2 \mu_1 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)}, \\
P_{0,00} &= \frac{(\lambda_1 + \lambda_2)C_0}{\mu_1 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)},
\end{align*}
\] (A.10)

189
\[ P_{N,11} = \frac{\mu_1 + \mu_2}{\lambda_1(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)} C_0 e^{KN}, \]
\[ P_{N,10} = \frac{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)}{\lambda_1\mu_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)} C_0 e^{KN}, \]
\[ P_{N,01} = 0, \]
\[ P_{N,00} = \frac{\lambda_1 + \lambda_2}{\mu_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)} C_0 e^{KN}, \]

To evaluate the free constant, substitute (A.7) and (A.10) into the total probability formula,
\[
\sum_{i=0}^{1} \sum_{j=0}^{1} \int_{0}^{-N} f_{H,I,J}(h,i,j) dh + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{0,ij} + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij} = 1, \quad (A.11)
\]
leading to
\[
C_0 = \frac{1}{D_2 + D_3 + D_4}, \quad (A.12)
\]
where
\[
D_2 = \begin{cases} 
(2 + D_1 + \frac{1}{D_1}) e^{KN} - 1, & \text{for } e_1 \neq e_2, \\
(2 + D_1 + \frac{1}{D_1}) N, & \text{for } e_1 = e_2,
\end{cases}
\]
\[
D_3 = \frac{(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)(\lambda_2 + \mu_1) + \lambda_1\mu_2 - \lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)}, \quad (A.13)
\]
\[
D_4 = \frac{(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)(\lambda_1 + \mu_2) + \lambda_2\mu_1 - \lambda_1\mu_2}{\lambda_1\mu_2(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)} e^{KN}.
\]

Thus, the steady state distributions of buffer occupancy are as follows:
\[
f_{H,I,J}(h,1,1) = \frac{D_1}{D_2 + D_3 + D_4} e^{Kh},
\]
\[
f_{H,I,J}(h,1,0) = \frac{1}{D_2 + D_3 + D_4} e^{Kh},
\]
\[
f_{H,I,J}(h,0,1) = \frac{1}{D_2 + D_3 + D_4} e^{Kh},
\]
\[
f_{H,I,J}(h,0,0) = \frac{1}{D_1(D_2 + D_3 + D_4)} e^{Kh},
\]
\[
P_{0,11} = \frac{\mu_1 + \mu_2}{\lambda_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)},
\]
190
Let

\[ P_{0,10} = 0, \]
\[ P_{0,01} = \frac{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)}{\lambda_2 \mu_1 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)}, \quad (A.14) \]
\[ P_{0,00} = \frac{\lambda_1 + \lambda_2}{\mu_1 + \mu_2} \]
\[ \frac{\lambda_1(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)}{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)} e^{KN}, \]
\[ P_{N,11} = \frac{\lambda_1 \mu_2}{\mu_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)} e^{KN}, \]
\[ P_{N,10} = \frac{\lambda_1 \mu_2}{\mu_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)} e^{KN}, \]
\[ P_{N,01} = \frac{\lambda_1 + \lambda_2}{\mu_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)(D_2 + D_3 + D_4)} e^{KN}. \]

The marginal steady state pdf of buffer occupancy is given by

\[ f_H(h) = \sum_{i=0}^{1} \sum_{j=0}^{1} f_{H,I,J}(h, i, j) + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{0,ij} \delta(h) + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij} \delta(h - N). \quad (A.15) \]

Using the above distributions, \( PR, WIP, BL_1 \) and \( ST_2 \) can be expressed as follows:

\[ PR = P[s_2 = 1](1 - P[h = 0, s_1 = 0|s_2 = 1]) = P[s_1 = 1](1 - P[h = N, s_2 = 0|s_1 = 1]) = e_2 \left( 1 - \frac{P_{0,01}}{e_2} \right) = e_1 \left( 1 - \frac{P_{N,10}}{e_1} \right), \]
\[ WIP = \int_0^N f_H(h) dh = \sum_{i=0}^{1} \sum_{j=0}^{1} \int_0^{N^+} f_{H,I,J}(h, i, j) dh + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij} N, \]
\[ BL_1 = P[h = N, s_1 = 1, s_2 = 0] = P_{N,10}, \]
\[ ST_2 = P[h = 0, s_1 = 0, s_2 = 1] = P_{0,01}. \]

Let

\[ Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N) = \begin{cases} 
\frac{(1-\phi_1)(1-\phi_2)}{1-\phi_1 \phi_2 - \phi_1 \phi_2}, & \text{if } \frac{\lambda_1}{\mu_1} \neq \frac{\lambda_2}{\mu_2}, \\
\frac{\lambda_1(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)}{\lambda_1 + \mu_1 + \mu_2 + \mu_2(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)N}, & \text{if } \frac{\lambda_1}{\mu_1} = \frac{\lambda_2}{\mu_2}, \end{cases} \quad (A.16) \]
\[ \phi = \frac{e_1(1 - e_2)}{e_2(1 - e_1)}, \quad (A.17) \]

\[ \beta = \frac{(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)(\lambda_1 \mu_2 - \lambda_2 \mu_1)}{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)}. \quad (A.18) \]

It can be shown that

\[ \frac{P_{0.01}}{e_2} = Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N), \]

\[ \frac{P_{N.10}}{e_1} = Q(\lambda_2, \mu_2, \lambda_1, \mu_1, N). \]

Therefore,

\[ PR = e_2[1 - Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N)] \]

\[ = e_1[1 - Q(\lambda_2, \mu_2, \lambda_1, \mu_1, N)], \quad (A.19) \]

\[ WIP = \begin{cases} \frac{D_5}{D_2 + D_3 + D_4}, & \text{for } e_1 \neq e_2, \\
\frac{(D_2 + D_4)N}{D_2 + D_3 + D_4}, & \text{for } e_1 = e_2, \end{cases} \quad (A.20) \]

\[ BL_1 = e_1 Q(\lambda_2, \mu_2, \lambda_1, \mu_1, N), \quad (A.21) \]

\[ ST_2 = e_2 Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N), \quad (A.22) \]

where

\[ D_5 = \frac{D_2[1 + (KN - 1)e^{KN}]}{K(e^{KN} - 1)} + D_4N. \quad (A.23) \]

**M > 2-machine case:** For longer lines, an aggregation procedure was developed in [21] to estimate the performance measures:

**Recursive Aggregation Procedure A.1:**

\[ \mu_i^b(s + 1) = \mu_i(1 - Q(\lambda_{i+1}^b(s + 1), \mu_{i+1}^b(s + 1), \lambda_i^f(s), \mu_i^f(s), N_i)), \]

\[ i = 1, \ldots, M - 1, \]

\[ \lambda_i^b(s + 1) = \lambda_i + \mu_i Q(\lambda_{i+1}^b(s + 1), \mu_{i+1}^b(s + 1), \lambda_i^f(s), \mu_i^f(s), N_i), \quad (A.24) \]
\[ i = 1, \ldots, M - 1, \]
\[ \mu^f_i(s + 1) = \mu_i(1 - Q(\lambda^f_{i-1}(s + 1), \mu^f_{i-1}(s + 1), \lambda^b_i(s + 1), \mu^b_i(s + 1), N_{i-1})), \]
\[ i = 2, \ldots, M, \]
\[ \lambda^f_i(s + 1) = \lambda_i + \mu_iQ(\lambda^f_{i-1}(s + 1), \mu^f_{i-1}(s + 1), \lambda^b_i(s + 1), \mu^b_i(s + 1), N_{i-1}), \]
\[ i = 2, \ldots, M, \]
\[ s = 1, 2, \ldots, \]

with initial conditions
\[ \lambda^f_i(0) = \lambda_i, \quad \mu^f_i(0) = \mu_i, \quad i = 2, \ldots, M - 1, \]

and boundary conditions
\[ \lambda^f_1(s) = \lambda_1, \quad \mu^f_1(s) = \mu_1, \quad s = 1, 2, \ldots, \]
\[ \lambda^b_M(s) = \lambda_M, \quad \mu^b_M(s) = \mu_M, \quad s = 1, 2, \ldots, \]

where function \( Q \) is defined by (A.16).

Clearly, this procedure iterates \( \lambda_i \)'s, \( \mu_i \)'s and \( N_i \)'s, resulting in \( 4(M - 1) \) sequences of numbers
\[ \lambda^f_2(s), \ldots, \lambda^f_M(s), \quad \mu^f_2(s), \ldots, \mu^f_M(s), \]
\[ \lambda^b_2(s), \ldots, \lambda^b_{M-1}(s), \quad \mu^b_2(s), \ldots, \mu^b_{M-1}(s), \]
\[ s = 1, 2, \ldots. \]

It was proved that these sequences are convergent and the follow limits exist [22]:
\[ \lim_{s \to \infty} \lambda^f_i(s) =: \lambda^f_i, \quad \lim_{s \to \infty} \mu^f_i(s) =: \mu^f_i, \]
\[ \lim_{s \to \infty} \lambda^b_i(s) =: \lambda^b_i, \quad \lim_{s \to \infty} \mu^b_i(s) =: \mu^b_i. \]  \hspace{1cm} (A.25)

Based on these limits, the production rate of synchronous exponential \( M > 2 \)-
machine lines is given as follows:

\[
\hat{PR} = e_M^f = e_M^b \\
= e_i^b [1 - Q(\lambda_i^f, \mu_i^f, \lambda_{i+1}^b, \mu_{i+1}^b, N_i)] \\
= e_i^f [1 - Q(\lambda_i^b, \mu_i^b, \lambda_i^f, \mu_i^f, N_i)], \quad i = 1, \ldots, M - 1, \quad (A.26)
\]

\[
\hat{WIP}_i = \begin{cases} 
\frac{D_i}{D_2 + D_3 + D_4}, & \text{for } e_i^f \neq e_{i+1}^b, \\
\frac{(D_2 + D_4)N_i}{D_2 + D_3 + D_4}, & \text{for } e_i^f = e_{i+1}^b,
\end{cases} \quad (A.27)
\]

\[
\hat{BL}_i = e_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, N_i), \quad i = 1, \ldots, M - 1, \quad (A.28)
\]

\[
\hat{ST}_i = e_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, N_{i-1}), \quad i = 2, \ldots, M, \quad (A.29)
\]

where

\[
e_i^f = \frac{\mu_i^f}{\lambda_i^f + \mu_i^f}, \quad e_i^b = \frac{\mu_i^b}{\lambda_i^b + \mu_i^b}, \quad i = 1, \ldots, M.
\]

and \(K, D_i, i = 2, \ldots, 5\), are defined in (A.8), (A.13) and (A.23) with \(\lambda_1, \mu_1, \lambda_2, \mu_2, \)
\(N\) substituted by \(\lambda_i^f, \mu_i^f, \lambda_i^b + 1, \mu_i^b + 1, N_i\), respectively. The accuracy of (A.26)-(A.29)
has been investigated in [75] and it was shown that Recursive Aggregation Procedure
A.1 provides a sufficiently accurate tool for estimating the performance measures in
synchronous exponential lines.

\[\square\]

**Performance evaluation of asynchronous exponential lines**

**Two-machine case:** In the asynchronous case, the transition rates, \(\nu_{kl,ij}\), remain
the same as in synchronous lines. The dynamics of the buffer occupancy become

\[
\frac{dh(t)}{dt} \bigg|_{0 < h(t) < N} = \begin{cases} 
c_1 - c_2, & \text{if } s_1(t) = 1, s_2(t) = 1, \\
c_1, & \text{if } s_1(t) = 1, s_2(t) = 0, \\
c_2, & \text{if } s_1(t) = 0, s_2(t) = 1, \\
0, & \text{if } s_1(t) = 0, s_2(t) = 0,
\end{cases}
\]

194
\[
\frac{dh(t)}{dt} \bigg|_{h(t)=0} = \begin{cases} 
\max(c_1 - c_2, 0), & \text{if } s_1(t) = 1, s_2(t) = 1, \\
\ c_1, & \text{if } s_1(t) = 1, s_2(t) = 0, \\
0, & \text{otherwise},
\end{cases}
\]
\[
\frac{dh(t)}{dt} \bigg|_{h(t)=N} = \begin{cases} 
\min(c_1 - c_2, 0), & \text{if } s_1(t) = 1, s_2(t) = 1, \\
\ -c_2, & \text{if } s_1(t) = 0, s_2(t) = 1, \\
0, & \text{otherwise}.
\end{cases}
\]

Thus, the steady state pdf for the the internal states are characterized by the following equations:

\[
\frac{\partial f_{H,I,J}(h,1,1)}{\partial h} = -\frac{\lambda_1 + \lambda_2}{c_1 - c_2} f_{H,I,J}(h,1,1) + \frac{\mu_2}{c_1 - c_2} f_{H,I,J}(h,1,0) + \frac{\mu_1}{c_1 - c_2} f_{H,I,J}(h,0,1)
\]
\[
\frac{\partial f_{H,I,J}(h,1,0)}{\partial h} = -\frac{\lambda_1 + \mu_2}{c_1} f_{H,I,J}(h,1,0) + \frac{\mu_2}{c_1} f_{H,I,J}(h,1,1) + \frac{\mu_1}{c_1} f_{H,I,J}(h,0,0),
\]
\[
\frac{\partial f_{H,I,J}(h,0,1)}{\partial h} = \frac{\lambda_2 + \mu_1}{c_2} f_{H,I,J}(h,0,1) - \frac{\lambda_1}{c_2} f_{H,I,J}(h,1,1) - \frac{\mu_2}{c_2} f_{H,I,J}(h,0,0),
\]
\[
f_{H,I,J}(h,0,0) = \frac{\lambda_1}{\mu_1 + \mu_2} f_{H,I,H}(h,1,0) + \frac{\lambda_2}{\mu_1 + \mu_2} f_{H,I,J}(h,0,1).
\]

In addition, since as in the synchronous case, events $A_1$ and $A_2$ occur with equal frequencies,

\[
(c_1 - c_2) f_{H,I,J}(h,1,1) + c_1 f_{H,I,J}(h,1,0) = c_2 f_{H,I,J}(h,0,1).
\]

Using (A.31), (A.30) can be simplified to

\[
f_{H,I,J}(h,1,1) = \frac{c_1}{c_2 - c_1} f_{H,I,J}(h,1,0) - \frac{c_2}{c_2 - c_1} f_{H,I,J}(h,0,1),
\]
\[
\frac{\partial f_{H,I,J}(h,1,0)}{\partial h} = E_1 f_{H,I,J}(h,1,0) + E_2 f_{H,I,J}(h,0,1),
\]
\[
\frac{\partial f_{H,I,J}(h,0,1)}{\partial h} = E_3 f_{H,I,H}(h,1,0) + E_4 f_{H,I,J}(h,0,1),
\]
\[
f_{H,I,J}(h,0,0) = \frac{\lambda_1}{\mu_1 + \mu_2} f_{H,I,H}(h,1,0) + \frac{\lambda_2}{\mu_1 + \mu_2} f_{H,I,J}(h,0,1),
\]
where

\[
E_1 = \frac{\lambda_2(c_1\mu_1+c_2\mu_2)-\mu_2(c_2-c_1)(\mu_1+\mu_2+\lambda_1+\lambda_2)}{c_1(c_2-c_1)(\mu_1+\mu_2)},
\]
\[
E_2 = -\frac{\lambda_2(c_1\mu_1+c_2\mu_2)}{c_1(c_2-c_1)(\mu_1+\mu_2)},
\]
\[
E_3 = -\frac{\lambda_1(c_1\mu_1+c_2\mu_2)}{c_2(c_2-c_1)(\mu_1+\mu_2)},
\]
\[
E_4 = \frac{\lambda_1(c_1\mu_1+c_2\mu_2)+\mu_1(c_2-c_1)(\mu_1+\mu_2+\lambda_1+\lambda_2)}{c_2(c_2-c_1)(\mu_1+\mu_2)}.
\]

The probabilities of the boundary states for asynchronous lines can be characterized as follows:

- If \( c_1 < c_2 \),

\[
P_{0,11} = \frac{1}{\lambda_1 + \lambda_2} \left[ \mu_1 P_{0,01} + (c_2 - c_1) f_{H,I,J}(0,1,1) \right],
\]
\[
P_{0,10} = 0,
\]
\[
P_{0,01} = \frac{1}{\mu_1 + \lambda_2} \left[ \lambda_1 P_{0,11} + \mu_2 P_{0,00} + c_2 f_{H,I,J}(0,0,1) \right],
\]
\[
P_{0,00} = \frac{\lambda_2}{\mu_1 + \mu_2} P_{0,01},
\]
\[
P_{N,11} = 0,
\]
\[
P_{N,10} = \frac{c_2 - c_1}{\mu_2} f_{H,I,J}(N,1,1),
\]
\[
P_{N,01} = 0,
\]
\[
P_{N,00} = \frac{\lambda_1}{\mu_1 + \mu_2} P_{N,10},
\]
\[
P_{N,00} = \frac{c_2}{\mu_2} f_{H,I,J}(N,0,1).
\]

- If \( c_1 > c_2 \),

\[
P_{0,11} = 0,
\]
\[
P_{0,10} = 0,
\]
\[
P_{0,01} = \frac{c_1 - c_2}{\mu_1} f_{H,I,J}(0,1,1),
\]
\[
P_{0,00} = \frac{c_1}{\mu_1} f_{H,I,J}(0,1,0),
\]
\[
P_{0,00} = \frac{\lambda_2}{\mu_1 + \mu_2} P_{0,01}.
\]
\[ P_{N,11} = \frac{1}{\lambda_1 + \lambda_2}[\mu_2 P_{N,10} + (c_1 - c_2)f_{H,I,J}(N, 1, 1)], \quad \text{(A.34)} \]
\[ P_{N,10} = \frac{1}{\lambda_1 + \mu_2}[\lambda_2 P_{N,11} + \mu_1 P_{N,00} + c_1 f_{H,I,J}(N, 1, 0)], \]
\[ P_{N,01} = 0, \]
\[ P_{N,00} = \frac{\lambda_1}{\mu_1 + \mu_2} f_{H,I,J}(N, 0, 1). \]

Finally, the total probability formula takes the form:
\[ \sum_{i=0}^{1} \sum_{j=0}^{1} \int_{0}^{N^-} f_{H,I,J}(h, i, j) dh + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{0,ij} + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij} = 1. \]

Using the same procedure as in the synchronous case, the solutions to the above equations can be obtained as follows:

- If \( c_1 < c_2, \)

\[ f_{H,I,J}(h, 1, 0) = C_0 \left[ e^{K_1 h} - F_1 e^{(K_1 - K_2)N} \cdot e^{K_2 h} \right], \]
\[ f_{H,I,J}(h, 0, 1) = C_0 \left[ -\frac{F_2}{2E_2} e^{K_1 h} + \frac{F_1 F_3}{2E_2} e^{(K_1 - K_2)N} \cdot e^{K_2 h} \right], \]
\[ f_{H,I,J}(h, 1, 1) = \frac{c_1}{c_2 - c_1} f_{H,I,J}(h, 1, 0) - \frac{c_2}{c_2 - c_1} f_{H,I,J}(h, 0, 1), \]
\[ f_{H,I,J}(h, 0, 0) = \frac{\lambda_1}{\mu_1 + \mu_2} f_{H,I,H}(h, 1, 0) + \frac{\lambda_2}{\mu_1 + \mu_2} f_{H,I,J}(h, 0, 1), \]
\[ P_{0,11} = \frac{c_1(\lambda_2 + \mu_1 + \mu_2)f_{H,I,J}(0, 1, 0) - c_2 \lambda_2 f_{H,I,J}(0, 0, 1)}{\lambda_2(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)}, \]
\[ P_{0,10} = 0, \]
\[ P_{0,01} = \frac{(\mu_1 + \mu_2)[c_1 \lambda_1 f_{H,I,J}(0, 1, 0) + c_2 \lambda_2 f_{H,I,J}(0, 0, 1)]}{\lambda_2 \mu_1(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)}, \]
\[ P_{0,00} = \frac{c_1 \lambda_1 f_{H,I,J}(0, 1, 0) + c_2 \lambda_2 f_{H,I,J}(0, 0, 1)}{\mu_1(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)}, \]
\[ P_{N,11} = 0, \quad \text{(A.35)} \]
\[ P_{N,10} = \frac{c_1 f_{H,I,J}(N, 1, 0) - c_2 f_{H,I,J}(N, 0, 1)}{\mu_2}, \]
\[ P_{N,01} = 0, \]
\[ P_{N,00} = \frac{\lambda_1 [c_1 f_{H,I,J}(N, 1, 0) - c_2 f_{H,I,J}(N, 0, 1)]}{\mu_2(\mu_1 + \mu_2)}. \]

197
If $c_1 > c_2$,

$$f_{H,I,J}(h,1,0) = C_0 \left[ e^{K_1 h} - F_1 e^{(K_1 - K_2)N} \cdot e^{K_2 h} \right],$$

$$f_{H,I,J}(h,0,1) = C_0 \left[ -\frac{F_2}{2E_2} e^{K_1 h} + \frac{F_1 F_3}{2E_2} e^{(K_1 - K_2)N} \cdot e^{K_2 h} \right],$$

$$f_{H,I,J}(h,1,1) = \frac{c_1}{c_2 - c_1} f_{H,I,J}(h,1,0) - \frac{c_2}{c_2 - c_1} f_{H,I,J}(h,0,1),$$

$$f_{H,I,J}(h,0,0) = \frac{\lambda_1}{\mu_1 + \mu_2} f_{H,I,H}(h,1,0) + \frac{\lambda_2}{\mu_1 + \mu_2} f_{H,I,J}(h,0,1),$$

$$P_{0,11} = 0,$$

$$P_{0,10} = 0,$$

$$P_{0,01} = \frac{-c_1 f_{H,I,J}(0,1,0) + c_2 f_{H,I,J}(0,0,1)}{\mu_1},$$

$$P_{0,00} = \frac{c_1}{\mu_1} f_{H,I,J}(0,1,0),$$

$$P_{N,11} = \frac{c_2(\lambda_1 + \mu_1 + \mu_2) f_{H,I,J}(N,0,1) - c_1 \lambda_1 f_{H,I,J}(N,1,0)}{\mu_1 \lambda_1 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)},$$

$$P_{N,10} = \frac{(\mu_1 + \mu_2)c_1 \lambda_1 f_{H,I,J}(N,1,0) + c_2 \lambda_2 f_{H,I,J}(N,0,1)}{\mu_2 \lambda_2 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)},$$

$$P_{N,01} = 0,$$

$$P_{N,00} = \frac{c_1 \lambda_1 f_{H,I,J}(N,1,0) + c_2 \lambda_2 f_{H,I,J}(N,0,1)}{\mu_2 (\lambda_1 + \mu_1 + \lambda_2 + \mu_2)},$$

where

$$K_1 = \frac{(E_1 + E_4)}{2} + \frac{\sqrt{(E_1 - E_4)^2 + 4E_2 E_3}}{2},$$

$$K_2 = \frac{(E_1 + E_4)}{2} - \frac{\sqrt{(E_1 - E_4)^2 + 4E_2 E_3}}{2},$$

$$C_0 = \frac{1}{F_4 F_5 + F_6 F_9 + F_6 + F_7},$$

$$F_1 = \begin{cases} 
\frac{c_2(\mu_1 + \mu_2 + \lambda_1)(E_1 - E_4 - K_1 + K_2) + 2\lambda_1 c_1 E_2}{c_2(\mu_1 + \mu_2 + \lambda_1)(E_1 - E_4 + K_1 - K_2) + 2\lambda_1 c_1 E_2}, & \text{for } c_1 < c_2, \\
\frac{c_2 c_1 c_2(\mu_1 + \mu_2 + \lambda_2)(E_1 - E_4 - K_1 + K_2) + 2\lambda_1 c_1 E_2}{c_2 c_1 c_2(\mu_1 + \mu_2 + \lambda_2)(E_1 - E_4 + K_1 - K_2) + 2\lambda_1 c_1 E_2}, & \text{for } c_1 > c_2,
\end{cases}$$

$$F_2 = E_1 - E_4 - (K_1 - K_2),$$

$$F_3 = E_1 - E_4 + (K_1 - K_2).$$
Using the above distribution, the performance measures can be expressed as follows:

\[ F_4 = \frac{\lambda_1}{\mu_1 + \mu_2} + \frac{c_2}{c_2 - c_1}, \]
\[ F_5 = \frac{\lambda_2}{\mu_1 + \mu_2} - \frac{c_1}{c_2 - c_1}, \]
\[ F_6 = \begin{cases} 
2E_2c_1(\lambda_1 + \mu_1)(\lambda_2 + \mu_1 + \mu_2)(1 - F_1e^{(K_1 - K_2)N}) \\
+ c_2\lambda_2(\lambda_2 + \mu_2)(-F_2 + F_1F_3e^{(K_1 - K_2)N}) \\
\end{cases}, \text{ for } c_1 < c_2, \]  
\[ F_7 = \begin{cases} 
e^{K_1N}\left[c_2(\lambda_2 + \mu_2)(\lambda_1 + \mu_1 + \mu_2)(-F_2 + F_1F_3e^{(K_2 - K_1)N}) \\
+ 2E_2c_1\lambda_1(\lambda_1 + \mu_1)(1 - F_1e^{(K_2 - K_1)N}) \right] \\
\end{cases}, \text{ for } c_1 < c_2, \]
\[ F_8 = \begin{cases} 
e^{K_1N-1}_{K_1} - \frac{e^{K_2N-1}_{K_2}}{F_1} \cdot F_1e^{(K_1 - K_2)N}, \text{ for } c_1 > c_2 \\
\end{cases} \]
\[ F_9 = \begin{cases} 
\frac{1-e^{K_1N}_{K_1}}{K_1} \cdot \frac{F_2}{2E_2} + \frac{e^{K_2N-1}_{K_2}}{F_1} \cdot \frac{F_1F_3e^{(K_2 - K_1)N}}{2E_2}, \text{ for } c_1 > c_2 \\
\frac{1-e^{K_1N}_{K_1}}{K_1} \cdot \frac{F_2}{2E_2} + \frac{e^{K_2N-1}_{K_2}}{F_1} \cdot \frac{F_1F_3}{2E_2}, \text{ for } c_1 < c_2. \end{cases} \]

The marginal pdf of buffer occupancy can be expressed as

\[ f_H(h) = \sum_{i=0}^{1} \sum_{j=0}^{1} f_{H,I,J}(h, i, j) + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{0,ij}\delta(h) + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij}\delta(h - N) \] (A.39)

Using the above distribution, the performance measures can be expressed as follows:

\[ TP = c_2 \sum_{i=0}^{1} \left\{ \int_{0^+}^{N^-} f_{H,I,J}(h, i, 1)dh + P_{N,i1} \right\} + \min(c_1, c_2)P_{0,11}, \]  
(A.40)
\[ WIP = \sum_{i=0}^{1} \sum_{j=0}^{1} \int_{0^+}^{N^-} f_{H,I,J}(h, i, j)dh + \sum_{i=0}^{1} \sum_{j=0}^{1} P_{N,ij}N, \]  
(A.41)
\[ ST_2 = e_2 - \frac{TP}{c_2}, \]  
(A.42)
\[ BL_1 = e_1 - \frac{TP}{c_1}. \]  

(A.43)

Substituting (A.35) and (A.36) into (A.40) and (A.41) leads to

\[ TP = \frac{G_4 + G_5 e^{K_1 N} + G_6 e^{K_2 N}}{G_1 + G_2 e^{K_1 N} + G_3 e^{K_2 N}}, \]  

(A.44)

\[ WIP = \frac{F_4 F_8 + F_5 F_9 + F_6 + F_7}{F_4 F_{10} + F_5 F_{11} + F_7 N}, \]  

(A.45)

where

\[
G_1 = \mu_1 G_7^2 + \mu_1 G_7 [c_1 (\mu_1 + \mu_2 + \lambda_2) - c_2 (\mu_1 + \mu_2 + \lambda_1)], \\
G_2 = \begin{cases} 
\mu_2 \lambda_1 c_2 [(c_1 - c_2)(\mu_1 - \mu_2) - (c_2 \lambda_1 + c_1 \lambda_2) - G_7], & \text{for } c_1 < c_2, \\
\mu_1 \lambda_2 c_1 [(c_1 - c_2)(\mu_1 - \mu_2) - (c_2 \lambda_1 + c_1 \lambda_2) + G_7], & \text{for } c_1 > c_2, \\
\frac{e_2 (c_2 - c_1 \epsilon_2) G_1 + c_1 \epsilon_1 (1 - \epsilon_2) G_2}{c_1 \epsilon_1 (\epsilon_2 - 1)}, & \text{for } c_1 < c_2, \\
\frac{e_1 (c_1 - c_2 \epsilon_1) G_1 + c_2 \epsilon_2 (1 - \epsilon_1) G_2}{c_2 \epsilon_2 (\epsilon_1 - 1)}, & \text{for } c_1 > c_2, \\
c_2 \epsilon_2 G_1, & \text{for } c_1 < c_2, \\
c_1 \epsilon_1 G_1, & \text{for } c_1 > c_2, \\
G_4 = c_1 \epsilon_1 G_2, & \text{for } c_1 < c_2, \\
c_2 \epsilon_2 G_2, & \text{for } c_1 > c_2, \\
G_5 = c_1 \epsilon_1 G_3, & \text{for } c_1 < c_2, \\
c_2 \epsilon_2 G_3, & \text{for } c_1 > c_2, \\
G_6 = \sqrt{[c_1 (\mu_1 + \mu_2 + \lambda_2) - c_2 (\mu_1 + \mu_2 + \lambda_1)]^2 + 4c_1 c_2 \lambda_1 \lambda_2},
\]
\[ F_{10} = \begin{cases} 
\frac{1+(K_1 N - 1) e^{K_1 N}}{K_1^2} + \frac{1+(K_2 N - 1) e^{K_2 N}}{K_2^2} \cdot F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 < c_2, \\
\frac{1+(K_1 N - 1) e^{K_1 N}}{K_1^2} - \frac{1+(K_2 N - 1) e^{K_2 N}}{K_2^2} \cdot F_1, & \text{for } c_1 > c_2, \\
F_2 [\frac{1+(K_1 N - 1) e^{K_1 N}}{2E_2 K_1^2}] + \frac{1+(K_2 N - 1) e^{K_2 N}}{2E_2 K_2^2} F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 < c_2, \\
- \frac{1+(K_1 N - 1) e^{K_1 N}}{2E_2 K_1^2} F_2 + \frac{1+(K_2 N - 1) e^{K_2 N}}{2E_2 K_2^2} F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 > c_2. 
\end{cases} \]

\[ F_{11} = \begin{cases} 
\frac{1+(K_1 N - 1) e^{K_1 N}}{K_1^2} + \frac{1+(K_2 N - 1) e^{K_2 N}}{K_2^2} \cdot F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 < c_2, \\
\frac{1+(K_1 N - 1) e^{K_1 N}}{K_1^2} - \frac{1+(K_2 N - 1) e^{K_2 N}}{K_2^2} \cdot F_1, & \text{for } c_1 > c_2, \\
F_2 [\frac{1+(K_1 N - 1) e^{K_1 N}}{2E_2 K_1^2}] + \frac{1+(K_2 N - 1) e^{K_2 N}}{2E_2 K_2^2} F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 < c_2, \\
- \frac{1+(K_1 N - 1) e^{K_1 N}}{2E_2 K_1^2} F_2 + \frac{1+(K_2 N - 1) e^{K_2 N}}{2E_2 K_2^2} F_1 e^{(K_1 - K_2) N}, & \text{for } c_1 > c_2. 
\end{cases} \]

\( M > 2 \)-machine lines: To evaluate performance measures of \( M > 2 \)-machine asynchronous exponential lines. The following recursive procedure is developed [23]:

**Recursive Aggregation Procedure A.2:**

\[
bl_i(s+1) = \frac{e_i c^f_i(s) - TP(\lambda_i, \mu_i, c^f_i(s), \lambda_{i+1}, \mu_{i+1}, c_{i+1}(s+1), N_i)}{e_i c^f_i(s)}, \quad 1 \leq i \leq M - 1,
\]

\[
c^b_i(s+1) = c_i [1 - bl_i(s+1)], \quad 1 \leq i \leq M - 1,
\]

\[
st_i(s+1) = \frac{e_i c^b_i(s+1) - TP(\lambda_{i-1}, \mu_{i-1}, c^f_{i-1}(s+1), \lambda_i, \mu_i, c^b_i(s+1), N_{i-1})}{e_i c^b_i(s+1)}, \quad 2 \leq i \leq M,
\]

\[
c^f_i(s+1) = c_i [1 - st_i(s+1)], \quad 2 \leq i \leq M, \quad \text{(A.46)}
\]

with initial conditions

\[ c^f_i(0) = c_i, \quad i = 2, \ldots, M - 1, \]

and boundary conditions

\[ c^f_1(s) = c_1, \quad c^b_M(s) = c_M, \quad s = 0, 1, 2, \ldots, \]

where \( TP(\lambda_1, \mu_1, c_1, \lambda_2, \mu_2, c_2, N) \) is calculated according to (A.19) or (A.44), whichever is applicable.

It can be shown that aggregation procedure (A.46) is convergent and, therefore,
the following limits exist:

\[
\lim_{s \to \infty} b_i(s) =: b_i, \quad i = 1, \ldots, M - 1, \quad \lim_{s \to \infty} s_i(s) =: s_i, \quad i = 2, \ldots, M,
\]

\[
\lim_{s \to \infty} c_i^f(s) =: c_i^f, \quad i = 1, \ldots, M, \quad \lim_{s \to \infty} c_i^b(s) =: c_i^b, \quad i = 1, \ldots, M.
\]

Then, using (A.42), (A.43), and (A.44), the performance measure estimates can be given in terms of the limits above:

\[
\hat{T}P = e_M c_M^f = e_1 c_1^b,
\]

\[
\hat{W}IP = \sum_{i=1}^{M-1} \hat{W}IP_i,
\]

\[
\hat{B}L_i = e_i b_i, \quad i = 1, \ldots, M - 1,
\]

\[
\hat{S}T_i = e_i s_i, \quad i = 2, \ldots, M,
\]

where \( \hat{W}IP_i \) are defined by expressions (A.45) with \( c_1 \) and \( c_2 \) substituted by \( c_i^f \) and \( c_i^b + 1 \), respectively. Similarly, \( \hat{T}P \) can be calculated using expressions (A.44) with \( c_i^f \) and \( c_i^b + 1 \) substituted for \( c_1 \) and \( c_2 \), respectively, and \( N_i \) for \( N \).

**Proof of Theorem 3.4**

**Two-machine lines:** For two-machine lines, the theorem follows immediately from (A.19)-(A.22) and (A.42)-(A.45) for synchronous and asynchronous lines, respectively.

**M > 2-machine lines:** Consider a synchronous exponential line. Let \( \lambda_j^f, \mu_j^f, \lambda_j^b \) and \( \mu_j^b, 1 \leq j \leq M \), denote the steady states of recursive procedure (A.24) applied to the line. Introduce the notations \( \overline{\lambda}_j^f = \lambda_{M-j}^f, \overline{\mu}_j^f = \mu_{M-j}^f, \overline{\lambda}_j^b = \lambda_{M-j}^b, \) and \( \overline{\mu}_j^b = \mu_{M-j}^b \). Observe that \( \overline{\lambda}_j^f, \overline{\mu}_j^f, \overline{\lambda}_j^b \) and \( \overline{\mu}_j^b \) solve the equilibrium equations of recursive procedure (A.24) for the reversed line. Since the equilibrium equations possess a unique solution [21], \( \overline{\lambda}_j^f, \overline{\mu}_j^f, \overline{\lambda}_j^b \) and \( \overline{\mu}_j^b \) must be the limiting values obtained by recursive procedure
(A.24) for the reversed line. Therefore,

\[ \hat{PR}_{Lr} = \frac{P_M}{\lambda_M + P_M} = \frac{\mu^b_1}{\lambda^b_1 + \mu^b_1} = \hat{PR}_{L}. \tag{A.51} \]

In addition, since the performance measures are calculated by two-machine line formulas with machine having parameters \((\lambda^f_i, \mu^f_i)\) and \((\lambda^b_{i+1}, \mu^b_{i+1})\), using the reversibility of two-machine exponential line leads to

\[ \hat{WIP}_{Lr} = N_{M-i} - \hat{WIP}_{M-i}^L, \quad i = 1, \ldots, M - 1, \]
\[ \hat{B}_{Lr}^i = \hat{S}_{(M-i+1)r}^L, \quad i = 1, \ldots, M - 1. \]

For an asynchronous exponential line, using the same argument for recursive procedure (A.46), the reversibility can be proved.

\[ \blacksquare \]

**Justification of Numerical Fact 3.1**

This justification is carried out by simulations. Since similar numerical investigations are used later in this work, we define them below as a standard procedure:

**Simulation Procedure A.1:**

1. Select initial status of each machine up with probability \(e_i\) (\(p_i\) in the case of Bernoulli machines) and down with probability \(1 - e_i\), \(i = 1, \ldots, M\).

2. For each line under consideration, carry out 20 runs of the simulation code.

3. In each run, use the first 20,000 time units as a warm-up period and the subsequent 400,000 time units to statistically evaluate \(PR\) (\(TP\)), \(WIP_i\), \(ST_i\) and \(BL_i\); this results in 95\% confidence intervals of less than 0.001 for \(PR\); 0.02 for \(WIP_i\); and 0.002 for \(ST_i\) and \(BL_i\).

A total of 100,000 lines were constructed by selecting the parameters of the ma-
chines and buffers randomly and equiprobably from the following sets

\[ M \in \{2, \ldots, 5\}, \]  
(A.52)

\[ T_{\text{down},i} \in [5, 15], \quad i = 1, \ldots, M, \]  
(A.53)

\[ e_i \in [0.60, 0.95], \quad i = 1, \ldots, M, \]  
(A.54)

\[ N_i \in \{8, 9, 10, 11, 12\}, \quad i = 1, \ldots, M - 1, \]  
(A.55)

\[ f_{\text{up},i}, f_{\text{down},i} \in \{W, ga, LN, exp\}, \quad i = 1, \ldots, M, \]  
(A.56)

\[ CV_{\text{up},i}, CV_{\text{down},i} \in [0.2, 1], \quad i = 1, \ldots, M. \]  
(A.57)

We evaluated the \( TP, WIP_i, ST_i \) and \( BL_i \) of each line, thus generated, and those of its reverse. As a result, the difference between the left-hand sides and right-hand sides of (3.15)-(3.17) is less than 0.001 for all systems studied. Thus, we conclude that Numerical Fact 3.1 indeed holds.

\[ \blacksquare \]

**Justification of Numerical Fact 3.2**

To justify this numerical fact, we used the 100,000 lines generated in the justification of Numerical Fact 3.1. For each line generated, we evaluated the throughput using Recursive Aggregation Procedures A.1 and A.2 for exponential lines and Simulation Procedure A.1 for non-exponential lines. Then, we investigated \( TP \) as a function of \( T_{\text{up},i}, T_{\text{down},i}, CV_{\text{up},i}, CV_{\text{down},i}, c_i, i = 1, \ldots, M, \) and \( N_i, i = 1, \ldots, M - 1. \) Specifically, we increased each parameter of a line by 50%, one at a time, evaluated the resulting throughput and examined whether the corresponding statement of Numerical Fact 3.2 holds. Among the 100,000 lines studied, no counterexamples of Numerical Fact 3.2 were found. Thus, we conclude that Numerical Fact 3.2 indeed takes place.

\[ \blacksquare \]
Proof of Theorem 3.5

Let \( \lambda_i, \mu_i, \lambda_i^2, \mu_i^2, i = 1, 2 \), be the failure and repair rates of machine \( m_i \) of lines \( l_1 \) and \( l_2 \), respectively. Assume

\[
\mu_i^1 = a_1 \mu_i^2, \quad \mu_i^2 = a_2 \mu_i^2.
\]

Then, from (3.19),

\[
a_1 > 1, \quad a_2 > 1.
\]

In addition, from \( e_i^1 = e_i^2, i = 1, 2 \), it follows that

\[
\lambda_i^1 = a_1 \lambda_i^2, \quad \lambda_i^2 = a_2 \lambda_i^2.
\]

When \( \frac{\lambda_i^1}{\mu_i^1} = \frac{\lambda_i^2}{\mu_i^2}, i = 1, 2 \), we obtain:

\[
Q_i^1 = \frac{\lambda_i^1 (\lambda_i^1 + \lambda_i^2)(\mu_i^1 + \mu_i^2)}{(\lambda_i^1 + \mu_i^1)[(\lambda_i^1 + \lambda_i^2)(\mu_i^1 + \mu_i^2) + \lambda_i^2 \mu_i^1 (\lambda_i^1 + \lambda_i^2 + \mu_i^1 + \mu_i^2) N]}
\]

\[
= \frac{[(1 - e_i^1)(a_1 \lambda_i^2 + a_2 \lambda_i^2)(a_1 \mu_i^2 + a_2 \mu_i^2)]/(a_1 \lambda_i^2 + a_2 \lambda_i^2)(a_1 \mu_i^2 + a_2 \mu_i^2) + a_2 \mu_i^2) + a_1 a_2 \lambda_i^2 \mu_i^1 (a_1 \lambda_i^2 + a_2 \lambda_i^2 + a_1 \mu_i^2 + a_2 \mu_i^2) N]}
\]

\[
Q_i^2 = \frac{(1 - e_i^2)(\lambda_i^2 + \lambda_i^2)(\mu_i^2 + \mu_i^2) + \lambda_i^2 \mu_i^1 (\lambda_i^1 + \lambda_i^2 + \mu_i^1 + \mu_i^2) N]}
\]

Thus,

\[
\frac{1}{Q_i^1} = \frac{[(a_1 \lambda_i^1 + a_2 \lambda_i^2)(a_1 \mu_i^1 + a_2 \mu_i^2) + a_1 a_2 \lambda_i^2 \mu_i^1 (a_1 \lambda_i^1 + a_2 \lambda_i^2 + a_1 \mu_i^2 + a_2 \mu_i^2) N)]}{{(1 - e_i^1)(a_1 \lambda_i^2 + a_2 \lambda_i^2)(a_1 \mu_i^2 + a_2 \mu_i^2) N]}}
\]

\[
= \frac{1}{1 - e_i^1} \left[ 1 + \lambda_i^2 \mu_i^1 \left( \frac{a_1 a_2}{a_1 \mu_i^1 + a_2 \mu_i^2} + \frac{a_1 a_2}{a_1 \lambda_i^1 + a_2 \lambda_i^2} \right) N \right]
\]

\[
= \frac{1}{1 - e_i^1} \left[ 1 + \lambda_i^2 \mu_i^1 \left( \frac{1}{a_2 \mu_i^2} + \frac{1}{a_1 \mu_i^2} + \frac{1}{a_2 \lambda_i^2} + \frac{1}{a_1 \lambda_i^2} \right) N \right]
\]

\[
= \frac{1}{Q_i^2}.
\]
Therefore, $Q'^1 < Q'^2$ and

$$PR'^1 = e'^1_2 (1 - Q'^1) > e'^2_2 (1 - Q'^2) = PR'^2.$$ 

When $\frac{\lambda'^i_1}{\mu'_1} \neq \frac{\lambda'^i_2}{\mu'_2}$, since $e'^i_1 = e'^i_2$, $i = 1, 2$, it follows that $\phi'^1 = \phi'^2$. Then,

$$\beta'^1 = \frac{(\lambda'^1_1 + \lambda'^2_1 + \mu'_1 + \mu'_2)(\lambda'^1_1 \mu'_2 - \lambda'^2_1 \mu'_1)}{(\lambda'^1_1 + \lambda'^2_1)(\mu'_1 + \mu'_2)}$$

$$= \frac{(a_1 \lambda'^1_1 + a_2 \lambda'^2_1 + a_1 \mu'_1 + a_2 \mu'_2)(a_1 \lambda'^1_1 \mu'_2 - a_2 \lambda'^2_1 \mu'_1)}{(a_1 \lambda'^1_1 + a_2 \lambda'^2_1)(a_1 \mu'_1 + a_2 \mu'_2)}$$

$$= \frac{a_1 a_2 (\lambda'^1_1 \mu'_2 - \lambda'^2_1 \mu'_1) + a_1 a_2 (\lambda'^1_1 \mu'_2 - \lambda'^2_1 \mu'_1)}{(a_1 \lambda'^1_1 + a_2 \lambda'^2_1)(a_1 \mu'_1 + a_2 \mu'_2)}$$

$$= \frac{\lambda'^1_1 \mu'^1_2 - \lambda'^2_1 \mu'^2_1}{a_2 \mu'^1_2 + a_1 \mu'^2_2} + \frac{\lambda'^1_1 \mu'^2_2 - \lambda'^2_1 \mu'^1_2}{a_2 \lambda'^1_2 + a_1 \lambda'^2_2}.$$

If $\lambda'^1_1 \mu'^2_1 > \lambda'^2_1 \mu'^1_1$,

$$\beta'^1 > \frac{\lambda'^1_1 \mu'^2_1 - \lambda'^2_1 \mu'^1_1}{\mu'^2_1 + \mu'^1_2} + \frac{\lambda'^1_1 \mu'^2_1 - \lambda'^2_1 \mu'^1_1}{\lambda'^1_1 + \lambda'^2_2} = \beta'^2 > 0.$$

If $\lambda'^1_1 \mu'^2_1 < \lambda'^2_1 \mu'^1_1$, then

$$0 > \beta'^1 > \beta'^2.$$

Hence, in both cases,

$$e^{-\beta'^1} < e^{-\beta'^2},$$

Thus, $Q'^1 < Q'^2$, i.e.,

$$PR'^1 > PR'^2.$$ 

Proof of Theorem 3.6

Consider a two-machine synchronous exponential line defined by parameters $\lambda_i$, $\mu_i$, $i = 1, 2$, and $N$. Let $l_1$ denote the line with $T_{down,1}$ decreased by a factor $1 + \alpha$,
\[\alpha > 0, \text{ and } l_2 \text{ denote the line with } T_{up,1} \text{ increased by the same factor. Then,}\]

\[Q^{l_1} = Q(\lambda_1, \mu_1(1 + \alpha), \lambda_2, \mu_2, N),\]

\[Q^{l_2} = Q(\frac{\lambda_1}{1 + \alpha}, \mu_1, \lambda_2, \mu_2, N).\]

When \(\frac{\lambda_1}{\mu_1(1 + \alpha)} \neq \frac{\lambda_2}{\mu_2}\),

\[\beta^{l_1} = \frac{(\lambda_1 + \lambda_2 + \mu_1(1 + \alpha) + \mu_2)(\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha))}{(\lambda_1 + \lambda_2)(\mu_1(1 + \alpha) + \mu_2)}\]

\[= \frac{\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha)}{\mu_1(1 + \alpha) + \mu_2} + \frac{\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha)}{\lambda_1 + \lambda_2},\]

\[\beta^{l_2} = \frac{(\frac{\lambda_1}{1 + \alpha} + \lambda_2 + \mu_1 + \mu_2)(\frac{\lambda_1}{1 + \alpha}\mu_2 - \lambda_2\mu_1)}{(\frac{\lambda_1}{1 + \alpha} + \lambda_2)(\mu_1 + \mu_2)}\]

\[= \frac{(\lambda_1 + (1 + \alpha)\lambda_2 + (1 + \alpha)(\mu_1 + \mu_2))(\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha))}{(\lambda_1 + (1 + \alpha)\lambda_2)(\mu_1 + \mu_2)}\]

\[= \frac{\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha)}{(1 + \alpha)(\mu_1 + \mu_2)} + \frac{\lambda_1\mu_2 - \lambda_2\mu_1(1 + \alpha)}{(1 + \alpha)(\lambda_1 + (1 + \alpha)\lambda_2)}.\]

If \(\lambda_1\mu_2 < \lambda_2\mu_1(1 + \alpha)\), it follows that \(\beta^{l_1} > \beta^{l_2} > 0\). If \(\lambda_1\mu_2 > \lambda_2\mu_1(1 + \alpha)\), it implies that \(0 > \beta^{l_1} > \beta^{l_2}\). In both cases, \(e^{-\beta^{l_1}N} < e^{-\beta^{l_2}N}\). Therefore, \(Q^{l_1} < Q^{l_2}\) and

\[PR^{l_1} > PR^{l_2}.\]

When \(\frac{\lambda_1}{\mu_1(1 + \alpha)} = \frac{\lambda_2}{\mu_2}\),

\[Q^{l_1} = \left(\lambda_1(\lambda_1 + \lambda_2)(\mu_1(1 + \alpha) + \mu_2)/\left((\lambda_1 + \mu_1(1 + \alpha))(\lambda_1 + \lambda_2)(\mu_1(1 + \alpha) + \mu_2))\right)\right)

\[+ \lambda_2\mu_1(\lambda_1 + \lambda_2 + \mu_1(1 + \alpha) + \mu_2))\right)\}

\[Q^{l_2} = \frac{\lambda_1}{1 + \alpha}(\frac{\lambda_1}{1 + \alpha} + \lambda_2)(\mu_1 + \mu_2)

\[= \left(\lambda_1(\lambda_1 + (1 + \alpha)\lambda_2)(\mu_1 + \mu_2)/\left((\lambda_1 + (1 + \alpha)\mu_1)(\lambda_1

\[+(1 + \alpha)\lambda_2)(\mu_1 + \mu_2) + \lambda_2\mu_1(\lambda_1 + (1 + \alpha)\lambda_2 + (1 + \alpha)(\mu_1 + \mu_2))\right)\right)\].\]
Then,
\[
\frac{1}{Q^1} = \frac{1}{\lambda_1} + \frac{\lambda_2 \mu_1}{\mu_1 + \frac{\mu_2}{1+\alpha}} + \frac{\lambda_2 \mu_1 (1 + \alpha)}{\lambda_1 \lambda_1 + \lambda_2},
\]
\[
\frac{1}{Q^2} = (1 + \alpha) \left[ \frac{1}{\lambda_1} + \frac{\lambda_2 \mu_1}{\mu_1 + \mu_2} + \frac{\lambda_2 \mu_1}{\lambda_1} \cdot \frac{1 + \alpha}{\lambda_1 + (1 + \alpha) \lambda_2} \right]
\]
and
\[
\frac{1}{Q^1} > \frac{1}{Q^2},
\]
which implies that \( Q^1 < Q^2 \). Therefore, again
\[
PR^1 > PR^2.
\]

Justification of Numerical Fact 3.3

For exponential lines, this numerical fact was justified by calculations (using Recursive Aggregation Procedure A.1 and A.2) and for non-exponential lines by simulations (using Simulation Procedure A.1).

A total of 100,000 lines were constructed by randomly and equiprobably selecting machine and buffer parameters from sets (A.52)-(A.55). For non-exponential lines, additional parameters were selected from (A.56)-(A.57).

Justification of statement (1): For each of these lines, we constructed 10 additional lines with \( N_i \) and \( e_i \) being the same but with new breakdown and repair rates, \( a_i \lambda_i \) and \( a_i \mu_i, \ i = 1, \ldots, M \), respectively, where \( a_i \) is randomly and equiprobably selected from
\[
a_i \in (1, 3), \quad i = 1, \ldots, M. \tag{A.58}
\]
It turns out that the throughput of every additional line was higher than the original line for all cases studied. Thus, we conclude that statement (1) of Numerical Fact 3.3 holds.
Justification of statement (2): For each of the 100,000 lines constructed above, we increased the uptime of one machine by $1 + \alpha$, where $\alpha$ is randomly and equiprobably selected from

$$\alpha \in (0, 2),$$

(A.59)

and evaluated the throughput of the resulting system. Then, we kept the uptime of this machine at its original value and decreased its downtime by the same factor $1 + \alpha$. The throughput of this system was evaluated and compared to the uptime increasing case. We repeated the procedure for every machine in this line. As a result, decreasing downtime led to a higher throughput than increasing uptime of the same machine in all cases studied. Therefore, we conclude that statement (2) of Numerical Fact 3.3 is justified.

\[\blacksquare\]

Justification of CT-Improvability Indicator 3.1: To carry out the justification, we introduce

**CT-Continuous Improvement Procedure A.1:**

1. Evaluate the buffer occupancy, $WIP_i$, $i = 1, \ldots, M - 1$.

2. Determine the buffer for which $|WIP_i - N_i/2|$ is the largest. Assume this is buffer $k$.

3. If $WIP_k - N_k/2$ is positive, re-allocate a small fraction of cycle time, $\Delta \tau$, from $m_{k+1}$ to $m_k$; if negative, re-allocate from $m_k$ to $m_{k+1}$.

4. Return to step (1).

5. Continue this process until $\max_{i=1,\ldots,M-1} |WIP_i - N_i/2|$ is sufficiently small.
For exponential lines, this indicator was justified by calculations (using Recursive Aggregation Procedure A.1 and A.2) and for non-exponential lines by simulations (using Simulation Procedure A.1). In either case, the parameters of the machines and buffers were selected randomly and equiprobably from (A.52)-(A.55) and $\tau^*$ is selected as

$$\tau^* = M.$$  \hspace{1cm} (A.60)

For non-exponential lines, additional parameters were selected from (A.56)-(A.57). A total of 50,000 lines are generated. Some examples, thus formed, are shown in Tables A.1-A.3.

<table>
<thead>
<tr>
<th>line</th>
<th>$T_{down,1}$</th>
<th>$T_{down,2}$</th>
<th>$\epsilon_1$</th>
<th>$\epsilon_2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.0571</td>
<td>14.3547</td>
<td>0.9209</td>
<td>0.7436</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>6.2105</td>
<td>9.5075</td>
<td>0.8506</td>
<td>0.9125</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>7.5477</td>
<td>13.6560</td>
<td>0.6813</td>
<td>0.8817</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>5.1527</td>
<td>12.4679</td>
<td>0.7558</td>
<td>0.9261</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>9.1865</td>
<td>13.4622</td>
<td>0.7838</td>
<td>0.6709</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>10.9153</td>
<td>6.1975</td>
<td>0.6133</td>
<td>0.7605</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>14.3424</td>
<td>7.6445</td>
<td>0.6561</td>
<td>0.9055</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>11.4583</td>
<td>14.6689</td>
<td>0.8327</td>
<td>0.9046</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9.5735</td>
<td>9.5069</td>
<td>0.7443</td>
<td>0.9156</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>11.2731</td>
<td>11.9908</td>
<td>0.7390</td>
<td>0.7448</td>
<td>11</td>
</tr>
</tbody>
</table>

For each of the lines considered, CT-Continuous Improvement Procedure A.1 was carried out and $\tau_i^{unimp}$ along with $TP^{unimp}$ were obtained.

To evaluate the accuracy of these allocations, a full search in $\tau^*$ (with $\Delta \tau = 0.001$) was carried out to determine $\tau_i^{opt}$ and $TP^{opt}$. As measures of accuracy, the following were used:

$$\Delta_\tau = \max_{i=1,\ldots,M} \left| \frac{\tau_i^{unimp} - \tau_i^{opt}}{\tau_i^{opt}} \right| \cdot 100\%,$$  \hspace{1cm} (A.61)

$$\Delta_{TP} = \left| \frac{TP^{unimp} - TP^{opt}}{TP^{opt}} \right| \cdot 100\%.$$  \hspace{1cm} (A.62)

As it turns out, in all cases analyzed the unimprovable allocations were within
Table A.2: Examples of non-exponential five-machine lines

<table>
<thead>
<tr>
<th>Line</th>
<th>i</th>
<th>( f_{up,i} )</th>
<th>( CV_{up,i} )</th>
<th>( f_{down,i} )</th>
<th>( CV_{down,i} )</th>
<th>( T_{down,i} )</th>
<th>( e_i )</th>
<th>( N_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>W</td>
<td>0.4825</td>
<td>7.1518</td>
<td>0.7558</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>LN</td>
<td>0.3093</td>
<td>W</td>
<td>0.2819</td>
<td>13.5381</td>
<td>0.6606</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ga</td>
<td>0.6549</td>
<td>exp</td>
<td>1</td>
<td>9.8212</td>
<td>0.8251</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LN</td>
<td>0.7256</td>
<td>LN</td>
<td>0.6059</td>
<td>14.9183</td>
<td>0.7732</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.3797</td>
<td>14.4976</td>
<td>0.6018</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>LN</td>
<td>0.5815</td>
<td>LN</td>
<td>0.6455</td>
<td>7.3337</td>
<td>0.7088</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>W</td>
<td>0.9049</td>
<td>W</td>
<td>0.8062</td>
<td>11.6122</td>
<td>0.7192</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>W</td>
<td>0.5657</td>
<td>LN</td>
<td>0.8767</td>
<td>9.9696</td>
<td>0.6021</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LN</td>
<td>0.6877</td>
<td>W</td>
<td>0.3263</td>
<td>12.0850</td>
<td>0.6295</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ga</td>
<td>0.4799</td>
<td>W</td>
<td>0.6014</td>
<td>6.7101</td>
<td>0.7475</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>W</td>
<td>0.9008</td>
<td>ga</td>
<td>0.6693</td>
<td>5.8872</td>
<td>0.6963</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>W</td>
<td>0.9912</td>
<td>exp</td>
<td>1</td>
<td>7.0249</td>
<td>0.8819</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>W</td>
<td>0.6037</td>
<td>exp</td>
<td>1</td>
<td>11.2444</td>
<td>0.7414</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ga</td>
<td>0.4175</td>
<td>exp</td>
<td>1</td>
<td>8.2783</td>
<td>0.7662</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>7.4815</td>
<td>0.7695</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>LN</td>
<td>0.5304</td>
<td>LN</td>
<td>0.7327</td>
<td>5.9019</td>
<td>0.9040</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>ga</td>
<td>0.6717</td>
<td>10.0097</td>
<td>0.7870</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>exp</td>
<td>1</td>
<td>ga</td>
<td>0.8811</td>
<td>6.2931</td>
<td>0.6391</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>W</td>
<td>0.8826</td>
<td>W</td>
<td>0.6796</td>
<td>11.6755</td>
<td>0.6385</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.2921</td>
<td>14.8140</td>
<td>0.9062</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>LN</td>
<td>0.8901</td>
<td>exp</td>
<td>1</td>
<td>9.7771</td>
<td>0.9013</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.9212</td>
<td>11.6542</td>
<td>0.7576</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>LN</td>
<td>0.7500</td>
<td>ga</td>
<td>0.6525</td>
<td>13.1047</td>
<td>0.8582</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LN</td>
<td>0.7803</td>
<td>exp</td>
<td>1</td>
<td>11.8892</td>
<td>0.9168</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.4667</td>
<td>10.2316</td>
<td>0.9146</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table A.3: Examples of non-exponential five-machine lines (cont.)

<table>
<thead>
<tr>
<th>Line</th>
<th>i</th>
<th>$f_{up,i}$</th>
<th>$CV_{up,i}$</th>
<th>$f_{down,i}$</th>
<th>$CV_{down,i}$</th>
<th>$T_{down,i}$</th>
<th>$e_i$</th>
<th>$N_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
<td>ga</td>
<td>0.9627</td>
<td>ga</td>
<td>0.5137</td>
<td>8.0893</td>
<td>0.7525</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>W</td>
<td>0.8345</td>
<td>LN</td>
<td>0.7744</td>
<td>11.7826</td>
<td>0.8013</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>W</td>
<td>0.2034</td>
<td>W</td>
<td>0.4894</td>
<td>8.4346</td>
<td>0.7184</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.7935</td>
<td>exp</td>
<td>1</td>
<td>12.0255</td>
<td>0.7683</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>LN</td>
<td>0.7688</td>
<td>LN</td>
<td>0.3289</td>
<td>11.9677</td>
<td>0.7499</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>LN</td>
<td>0.3430</td>
<td>exp</td>
<td>1</td>
<td>12.1406</td>
<td>0.7576</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>LN</td>
<td>0.5469</td>
<td>exp</td>
<td>1</td>
<td>8.1310</td>
<td>0.7540</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>LN</td>
<td>0.5438</td>
<td>ga</td>
<td>0.3047</td>
<td>5.5005</td>
<td>0.7481</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.9525</td>
<td>7.9897</td>
<td>0.6998</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>ga</td>
<td>0.6858</td>
<td>ga</td>
<td>0.8286</td>
<td>11.8815</td>
<td>0.7705</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>LN</td>
<td>0.2739</td>
<td>exp</td>
<td>1</td>
<td>13.7011</td>
<td>0.6370</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>LN</td>
<td>0.2686</td>
<td>ga</td>
<td>0.3830</td>
<td>14.7001</td>
<td>0.6986</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ga</td>
<td>0.7471</td>
<td>exp</td>
<td>1</td>
<td>8.4040</td>
<td>0.8633</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.8232</td>
<td>exp</td>
<td>1</td>
<td>13.6312</td>
<td>0.6883</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>W</td>
<td>0.8045</td>
<td>ga</td>
<td>0.6328</td>
<td>10.7755</td>
<td>0.9015</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>W</td>
<td>0.6782</td>
<td>LN</td>
<td>0.4681</td>
<td>7.4056</td>
<td>0.8688</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>ga</td>
<td>0.2417</td>
<td>W</td>
<td>0.2227</td>
<td>5.0504</td>
<td>0.8649</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ga</td>
<td>0.7362</td>
<td>W</td>
<td>0.8836</td>
<td>5.0268</td>
<td>0.7387</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.6096</td>
<td>exp</td>
<td>1</td>
<td>13.2138</td>
<td>0.8867</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>6.8641</td>
<td>0.7909</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>W</td>
<td>0.2522</td>
<td>LN</td>
<td>0.2597</td>
<td>13.3459</td>
<td>0.6406</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>W</td>
<td>0.3015</td>
<td>LN</td>
<td>0.7171</td>
<td>8.2771</td>
<td>0.6389</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>LN</td>
<td>0.8226</td>
<td>exp</td>
<td>1</td>
<td>13.2628</td>
<td>0.6520</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.3720</td>
<td>ga</td>
<td>0.9374</td>
<td>8.2221</td>
<td>0.9192</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>12.8257</td>
<td>0.7916</td>
<td>-</td>
</tr>
</tbody>
</table>
5% of the optimal ones (see Tables A.4 and A.5 for illustration). Thus, we consider CT-Improvability Indicator 3.1 justified.

Table A.4: Accuracy of CT-Improvability Indicator 3.1 for two-machine exponential lines

<table>
<thead>
<tr>
<th>Line</th>
<th>$TP_{opt}$</th>
<th>$\tau_{1, opt}$</th>
<th>$\tau_{2, opt}$</th>
<th>$TP_{unimp}$</th>
<th>$\tau_{1, unimp}$</th>
<th>$\tau_{2, unimp}$</th>
<th>$\Delta TP$</th>
<th>$\Delta \tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7260</td>
<td>1.0390</td>
<td>0.9610</td>
<td>0.7241</td>
<td>1.0593</td>
<td>0.9407</td>
<td>0.2586</td>
<td>2.1124</td>
</tr>
<tr>
<td>2</td>
<td>0.8170</td>
<td>0.9742</td>
<td>1.0258</td>
<td>0.8169</td>
<td>0.9694</td>
<td>1.0306</td>
<td>0.0196</td>
<td>0.4927</td>
</tr>
<tr>
<td>3</td>
<td>0.6811</td>
<td>0.8865</td>
<td>1.1135</td>
<td>0.6810</td>
<td>0.8815</td>
<td>1.1185</td>
<td>0.0091</td>
<td>0.5640</td>
</tr>
<tr>
<td>4</td>
<td>0.7568</td>
<td>0.9177</td>
<td>1.0823</td>
<td>0.7567</td>
<td>0.9128</td>
<td>1.0872</td>
<td>0.0136</td>
<td>0.5339</td>
</tr>
<tr>
<td>5</td>
<td>0.5932</td>
<td>1.0080</td>
<td>0.9920</td>
<td>0.5902</td>
<td>1.0497</td>
<td>0.9503</td>
<td>0.0565</td>
<td>4.2036</td>
</tr>
<tr>
<td>6</td>
<td>0.5711</td>
<td>0.9466</td>
<td>1.0534</td>
<td>0.5694</td>
<td>0.9092</td>
<td>1.0908</td>
<td>0.3117</td>
<td>3.9510</td>
</tr>
<tr>
<td>7</td>
<td>0.6448</td>
<td>0.9385</td>
<td>1.0615</td>
<td>0.6402</td>
<td>0.8958</td>
<td>1.1042</td>
<td>0.7099</td>
<td>4.5498</td>
</tr>
<tr>
<td>8</td>
<td>0.7814</td>
<td>0.9838</td>
<td>1.0162</td>
<td>0.7809</td>
<td>0.9757</td>
<td>1.0243</td>
<td>0.0620</td>
<td>0.8233</td>
</tr>
<tr>
<td>9</td>
<td>0.7257</td>
<td>0.9499</td>
<td>1.0501</td>
<td>0.7245</td>
<td>0.9336</td>
<td>1.0664</td>
<td>0.1598</td>
<td>1.7160</td>
</tr>
<tr>
<td>10</td>
<td>0.6249</td>
<td>1</td>
<td>1</td>
<td>0.6245</td>
<td>0.9951</td>
<td>1.0049</td>
<td>0.0663</td>
<td>0.4900</td>
</tr>
</tbody>
</table>

Table A.5: Accuracy of CT-Improvability Indicator 3.1 for five-machine non-exponential lines

<table>
<thead>
<tr>
<th>Line</th>
<th>$TP_{opt}$</th>
<th>$TP_{unimp}$</th>
<th>$\Delta TP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5075</td>
<td>0.4989</td>
<td>1.6946</td>
</tr>
<tr>
<td>2</td>
<td>0.5003</td>
<td>0.4962</td>
<td>0.8195</td>
</tr>
<tr>
<td>3</td>
<td>0.5782</td>
<td>0.5630</td>
<td>2.6288</td>
</tr>
<tr>
<td>4</td>
<td>0.5409</td>
<td>0.5507</td>
<td>1.8118</td>
</tr>
<tr>
<td>5</td>
<td>0.6702</td>
<td>0.6570</td>
<td>1.9696</td>
</tr>
<tr>
<td>6</td>
<td>0.5750</td>
<td>0.5610</td>
<td>2.4348</td>
</tr>
<tr>
<td>7</td>
<td>0.5702</td>
<td>0.5438</td>
<td>4.6300</td>
</tr>
<tr>
<td>8</td>
<td>0.5376</td>
<td>0.5178</td>
<td>3.6830</td>
</tr>
<tr>
<td>9</td>
<td>0.6942</td>
<td>0.6910</td>
<td>0.4610</td>
</tr>
<tr>
<td>10</td>
<td>0.5199</td>
<td>0.5149</td>
<td>0.9617</td>
</tr>
</tbody>
</table>

An illustration of this indicator and CT-Continuous Improvement Procedure A.1 is given in Tables A.6 and A.7 for a five-machine exponential line with $N = [10, 10, 10, 11]$ and $\tau^* = 5$. Initially, the cycle time of the machines was allocated according to a decreasing pattern. Using CT-Continuous Improvement Procedure A.1, the unimprovable allocation was obtained as a bowl, resulting in $TP = 0.5149$, whereas the initial $TP$ was 0.4381, i.e., a 17% improvement.
<table>
<thead>
<tr>
<th>$WIP_1$</th>
<th>$WIP_2$</th>
<th>$WIP_3$</th>
<th>$WIP_4$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
<th>$\tau_4$</th>
<th>$\tau_5$</th>
<th>$TP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.40</td>
<td>4.04</td>
<td>0.94</td>
<td>1.35</td>
<td>1.249</td>
<td>1.160</td>
<td>1.122</td>
<td>0.926</td>
<td>0.543</td>
<td>0.4381</td>
</tr>
<tr>
<td>5.41</td>
<td>4.04</td>
<td>0.87</td>
<td>1.52</td>
<td>1.249</td>
<td>1.160</td>
<td>1.122</td>
<td>0.826</td>
<td>0.643</td>
<td>0.4377</td>
</tr>
<tr>
<td>5.28</td>
<td>3.57</td>
<td>1.01</td>
<td>1.49</td>
<td>1.249</td>
<td>1.160</td>
<td>1.022</td>
<td>0.926</td>
<td>0.643</td>
<td>0.4425</td>
</tr>
<tr>
<td>5.30</td>
<td>3.58</td>
<td>0.93</td>
<td>1.71</td>
<td>1.249</td>
<td>1.160</td>
<td>1.022</td>
<td>0.826</td>
<td>0.743</td>
<td>0.4424</td>
</tr>
<tr>
<td>5.21</td>
<td>3.21</td>
<td>1.06</td>
<td>1.68</td>
<td>1.249</td>
<td>1.160</td>
<td>0.922</td>
<td>0.926</td>
<td>0.743</td>
<td>0.4460</td>
</tr>
<tr>
<td>5.17</td>
<td>2.98</td>
<td>1.55</td>
<td>1.66</td>
<td>1.249</td>
<td>1.160</td>
<td>0.822</td>
<td>1.026</td>
<td>0.743</td>
<td>0.4476</td>
</tr>
<tr>
<td>5.15</td>
<td>2.97</td>
<td>1.27</td>
<td>1.91</td>
<td>1.249</td>
<td>1.160</td>
<td>0.822</td>
<td>0.926</td>
<td>0.843</td>
<td>0.4481</td>
</tr>
<tr>
<td>5.14</td>
<td>2.79</td>
<td>1.74</td>
<td>1.87</td>
<td>1.249</td>
<td>1.160</td>
<td>0.722</td>
<td>1.026</td>
<td>0.843</td>
<td>0.4481</td>
</tr>
<tr>
<td>5.13</td>
<td>2.76</td>
<td>1.48</td>
<td>2.21</td>
<td>1.249</td>
<td>1.160</td>
<td>0.722</td>
<td>0.926</td>
<td>0.943</td>
<td>0.4491</td>
</tr>
<tr>
<td>5.11</td>
<td>2.63</td>
<td>1.92</td>
<td>2.16</td>
<td>1.249</td>
<td>1.160</td>
<td>0.622</td>
<td>1.026</td>
<td>0.943</td>
<td>0.4497</td>
</tr>
<tr>
<td>5.13</td>
<td>2.63</td>
<td>1.72</td>
<td>2.72</td>
<td>1.249</td>
<td>1.160</td>
<td>0.622</td>
<td>0.926</td>
<td>1.043</td>
<td>0.4495</td>
</tr>
<tr>
<td>5.13</td>
<td>2.52</td>
<td>2.09</td>
<td>2.50</td>
<td>1.249</td>
<td>1.160</td>
<td>0.522</td>
<td>1.026</td>
<td>1.043</td>
<td>0.4494</td>
</tr>
<tr>
<td>5.16</td>
<td>2.56</td>
<td>2.00</td>
<td>3.30</td>
<td>1.249</td>
<td>1.160</td>
<td>0.522</td>
<td>0.926</td>
<td>1.143</td>
<td>0.4484</td>
</tr>
<tr>
<td>5.15</td>
<td>2.46</td>
<td>2.32</td>
<td>3.09</td>
<td>1.249</td>
<td>1.160</td>
<td>0.422</td>
<td>1.026</td>
<td>1.143</td>
<td>0.4484</td>
</tr>
<tr>
<td>5.15</td>
<td>2.40</td>
<td>2.69</td>
<td>2.82</td>
<td>1.249</td>
<td>1.160</td>
<td>0.322</td>
<td>1.126</td>
<td>1.143</td>
<td>0.4487</td>
</tr>
<tr>
<td>5.17</td>
<td>2.48</td>
<td>2.63</td>
<td>3.80</td>
<td>1.249</td>
<td>1.160</td>
<td>0.322</td>
<td>1.026</td>
<td>1.243</td>
<td>0.4471</td>
</tr>
<tr>
<td>4.52</td>
<td>2.68</td>
<td>2.74</td>
<td>4.00</td>
<td>1.249</td>
<td>1.060</td>
<td>0.422</td>
<td>1.026</td>
<td>1.243</td>
<td>0.4555</td>
</tr>
<tr>
<td>3.96</td>
<td>2.89</td>
<td>2.84</td>
<td>4.14</td>
<td>1.249</td>
<td>0.960</td>
<td>0.522</td>
<td>1.026</td>
<td>1.243</td>
<td>0.4615</td>
</tr>
<tr>
<td>3.97</td>
<td>2.82</td>
<td>3.30</td>
<td>3.82</td>
<td>1.249</td>
<td>0.960</td>
<td>0.422</td>
<td>1.126</td>
<td>1.243</td>
<td>0.4613</td>
</tr>
<tr>
<td>3.54</td>
<td>3.02</td>
<td>3.35</td>
<td>3.87</td>
<td>1.249</td>
<td>0.860</td>
<td>0.522</td>
<td>1.126</td>
<td>1.243</td>
<td>0.4644</td>
</tr>
<tr>
<td>3.22</td>
<td>3.28</td>
<td>3.34</td>
<td>3.90</td>
<td>1.249</td>
<td>0.760</td>
<td>0.622</td>
<td>1.126</td>
<td>1.243</td>
<td>0.4665</td>
</tr>
<tr>
<td>4.21</td>
<td>3.54</td>
<td>3.69</td>
<td>4.33</td>
<td>1.149</td>
<td>0.860</td>
<td>0.622</td>
<td>1.126</td>
<td>1.243</td>
<td>0.4862</td>
</tr>
<tr>
<td>3.82</td>
<td>3.94</td>
<td>3.67</td>
<td>4.38</td>
<td>1.149</td>
<td>0.760</td>
<td>0.722</td>
<td>1.126</td>
<td>1.243</td>
<td>0.4886</td>
</tr>
<tr>
<td>3.81</td>
<td>3.80</td>
<td>4.31</td>
<td>3.93</td>
<td>1.149</td>
<td>0.760</td>
<td>0.622</td>
<td>1.226</td>
<td>1.243</td>
<td>0.4895</td>
</tr>
<tr>
<td>3.92</td>
<td>3.95</td>
<td>4.23</td>
<td>5.43</td>
<td>1.149</td>
<td>0.760</td>
<td>0.622</td>
<td>1.126</td>
<td>1.343</td>
<td>0.4846</td>
</tr>
<tr>
<td>5.11</td>
<td>4.19</td>
<td>4.57</td>
<td>5.86</td>
<td>1.049</td>
<td>0.860</td>
<td>0.622</td>
<td>1.126</td>
<td>1.343</td>
<td>0.4998</td>
</tr>
<tr>
<td>5.06</td>
<td>4.23</td>
<td>4.59</td>
<td>5.88</td>
<td>1.049</td>
<td>0.850</td>
<td>0.632</td>
<td>1.126</td>
<td>1.343</td>
<td>0.4998</td>
</tr>
<tr>
<td>5.02</td>
<td>4.30</td>
<td>4.62</td>
<td>5.92</td>
<td>1.049</td>
<td>0.840</td>
<td>0.642</td>
<td>1.126</td>
<td>1.343</td>
<td>0.4999</td>
</tr>
<tr>
<td>4.96</td>
<td>4.35</td>
<td>4.64</td>
<td>5.96</td>
<td>1.049</td>
<td>0.830</td>
<td>0.652</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5008</td>
</tr>
<tr>
<td>4.93</td>
<td>4.41</td>
<td>4.63</td>
<td>5.94</td>
<td>1.049</td>
<td>0.820</td>
<td>0.662</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5009</td>
</tr>
<tr>
<td>4.88</td>
<td>4.44</td>
<td>4.63</td>
<td>5.96</td>
<td>1.049</td>
<td>0.810</td>
<td>0.672</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5018</td>
</tr>
<tr>
<td>4.84</td>
<td>4.50</td>
<td>4.64</td>
<td>5.96</td>
<td>1.049</td>
<td>0.800</td>
<td>0.682</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5015</td>
</tr>
<tr>
<td>4.79</td>
<td>4.54</td>
<td>4.62</td>
<td>5.96</td>
<td>1.049</td>
<td>0.790</td>
<td>0.692</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5021</td>
</tr>
<tr>
<td>4.77</td>
<td>4.62</td>
<td>4.64</td>
<td>6.00</td>
<td>1.049</td>
<td>0.780</td>
<td>0.702</td>
<td>1.126</td>
<td>1.343</td>
<td>0.5016</td>
</tr>
<tr>
<td>4.75</td>
<td>4.58</td>
<td>4.63</td>
<td>5.84</td>
<td>1.049</td>
<td>0.780</td>
<td>0.702</td>
<td>1.136</td>
<td>1.333</td>
<td>0.5028</td>
</tr>
<tr>
<td>4.70</td>
<td>4.63</td>
<td>4.60</td>
<td>5.82</td>
<td>1.049</td>
<td>0.770</td>
<td>0.712</td>
<td>1.136</td>
<td>1.333</td>
<td>0.5035</td>
</tr>
<tr>
<td>4.70</td>
<td>4.61</td>
<td>4.69</td>
<td>5.80</td>
<td>1.049</td>
<td>0.770</td>
<td>0.702</td>
<td>1.146</td>
<td>1.333</td>
<td>0.5033</td>
</tr>
</tbody>
</table>

Table A.6: Illustration of CT-Continuous Improvement Procedure A.1
Table A.7: Illustration of CT-Continuous Improvement Procedure A.1 (cont.)

<table>
<thead>
<tr>
<th>$WIP_1$</th>
<th>$WIP_2$</th>
<th>$WIP_3$</th>
<th>$WIP_4$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
<th>$\tau_4$</th>
<th>$\tau_5$</th>
<th>$TP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.65</td>
<td>4.65</td>
<td>4.66</td>
<td>5.76</td>
<td>1.049</td>
<td>0.760</td>
<td>0.712</td>
<td>1.146</td>
<td>1.333</td>
<td>0.5039</td>
</tr>
<tr>
<td>4.62</td>
<td>4.70</td>
<td>4.66</td>
<td>5.78</td>
<td>1.049</td>
<td>0.750</td>
<td>0.722</td>
<td>1.146</td>
<td>1.333</td>
<td>0.5039</td>
</tr>
<tr>
<td>4.75</td>
<td>4.74</td>
<td>4.71</td>
<td>5.85</td>
<td>1.039</td>
<td>0.760</td>
<td>0.722</td>
<td>1.146</td>
<td>1.333</td>
<td>0.5047</td>
</tr>
<tr>
<td>4.73</td>
<td>4.73</td>
<td>4.69</td>
<td>5.66</td>
<td>1.039</td>
<td>0.760</td>
<td>0.722</td>
<td>1.156</td>
<td>1.323</td>
<td>0.5058</td>
</tr>
<tr>
<td>4.74</td>
<td>4.73</td>
<td>4.77</td>
<td>5.62</td>
<td>1.039</td>
<td>0.760</td>
<td>0.712</td>
<td>1.166</td>
<td>1.323</td>
<td>0.5055</td>
</tr>
<tr>
<td>4.69</td>
<td>4.77</td>
<td>4.75</td>
<td>5.62</td>
<td>1.039</td>
<td>0.750</td>
<td>0.722</td>
<td>1.166</td>
<td>1.323</td>
<td>0.5062</td>
</tr>
<tr>
<td>4.82</td>
<td>4.80</td>
<td>4.80</td>
<td>5.69</td>
<td>1.029</td>
<td>0.760</td>
<td>0.722</td>
<td>1.166</td>
<td>1.323</td>
<td>0.5074</td>
</tr>
<tr>
<td>4.78</td>
<td>4.86</td>
<td>4.78</td>
<td>5.65</td>
<td>1.029</td>
<td>0.750</td>
<td>0.732</td>
<td>1.166</td>
<td>1.323</td>
<td>0.5069</td>
</tr>
<tr>
<td>4.78</td>
<td>4.84</td>
<td>4.86</td>
<td>5.62</td>
<td>1.029</td>
<td>0.750</td>
<td>0.722</td>
<td>1.176</td>
<td>1.323</td>
<td>0.5074</td>
</tr>
<tr>
<td>4.91</td>
<td>4.86</td>
<td>4.91</td>
<td>5.67</td>
<td>1.019</td>
<td>0.760</td>
<td>0.722</td>
<td>1.176</td>
<td>1.323</td>
<td>0.5089</td>
</tr>
<tr>
<td>4.89</td>
<td>4.83</td>
<td>4.89</td>
<td>5.52</td>
<td>1.019</td>
<td>0.760</td>
<td>0.722</td>
<td>1.186</td>
<td>1.313</td>
<td>0.5095</td>
</tr>
<tr>
<td>4.85</td>
<td>4.89</td>
<td>4.88</td>
<td>5.51</td>
<td>1.019</td>
<td>0.750</td>
<td>0.732</td>
<td>1.186</td>
<td>1.313</td>
<td>0.5096</td>
</tr>
<tr>
<td>4.97</td>
<td>4.93</td>
<td>4.93</td>
<td>5.56</td>
<td>1.009</td>
<td>0.760</td>
<td>0.732</td>
<td>1.186</td>
<td>1.313</td>
<td>0.5107</td>
</tr>
<tr>
<td>4.94</td>
<td>5.00</td>
<td>4.91</td>
<td>5.55</td>
<td>1.009</td>
<td>0.750</td>
<td>0.742</td>
<td>1.186</td>
<td>1.313</td>
<td>0.5107</td>
</tr>
<tr>
<td>4.96</td>
<td>4.99</td>
<td>4.98</td>
<td>5.49</td>
<td>1.009</td>
<td>0.750</td>
<td>0.732</td>
<td>1.196</td>
<td>1.313</td>
<td>0.5149</td>
</tr>
</tbody>
</table>

**Justification of BC-Improvability Indicator 3.2**

The approach of CT-Improvability Indicator 3.1 justification was used again except that the full search was carried out with respect to $N_i$’s, rather than $\tau_i$’s. The buffer capacity allocations, obtained using BC-Improvability Indicator 3.2 and the full search, were denoted as $N_i^{unimp}$ and $N_i^{opt}$, respectively, while the corresponding production rates were $TP^{unimp}$ and $TP^{opt}$. The accuracy of the indicator was quantified in terms of

$$\Delta_{TP} = \frac{|TP^{unimp} - TP^{opt}|}{TP^{opt}} \cdot 100\%.$$  

(A.63)

The machine characteristics were chosen from the sets

- $f_{t_{up},i}, f_{t_{down},i} \in \{exp, W, ga, LN\}, \quad i = 1, \ldots, 5,$
- $CV_{up,i}, CV_{down,i} \in [0.2, 1], \quad i = 1, \ldots, 5,$
- $T_{down,i} \in [5, 15], \quad i = 1, \ldots, 5,$
- $e_i \in [0.55, 0.9], \quad i = 1, \ldots, 5.$

215
A total of 50,000 lines are generated. Some examples of five-machine lines, thus formed, are shown in Tables A.8 and A.9. For each of these lines, the total buffer capacity was selected as $N^* = 40$.

### Table A.8: Examples of non-exponential five-machine lines

<table>
<thead>
<tr>
<th>Line</th>
<th>$i$</th>
<th>$f_{up,i}$</th>
<th>$CV_{up,i}$</th>
<th>$f_{down,i}$</th>
<th>$CV_{down,i}$</th>
<th>$T_{down,i}$</th>
<th>$e_i$</th>
<th>$N_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>W</td>
<td>0.6039</td>
<td>LN</td>
<td>0.8776</td>
<td>5.5421</td>
<td>0.7187</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>LN</td>
<td>0.4710</td>
<td>ga</td>
<td>0.8442</td>
<td>7.2775</td>
<td>0.8406</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>LN</td>
<td>0.5337</td>
<td>W</td>
<td>0.8306</td>
<td>10.9987</td>
<td>0.7913</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>LN</td>
<td>0.7313</td>
<td>LN</td>
<td>0.5625</td>
<td>8.1183</td>
<td>0.7419</td>
<td>37</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>ga</td>
<td>0.9756</td>
<td>W</td>
<td>0.2070</td>
<td>5.2817</td>
<td>0.5747</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>W</td>
<td>0.2757</td>
<td>ga</td>
<td>0.8722</td>
<td>6.3129</td>
<td>0.6700</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>W</td>
<td>0.7803</td>
<td>exp</td>
<td>1</td>
<td>6.5742</td>
<td>0.9294</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>ga</td>
<td>0.2000</td>
<td>ga</td>
<td>0.4855</td>
<td>13.7374</td>
<td>0.7316</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>exp</td>
<td>1</td>
<td>W</td>
<td>0.3779</td>
<td>9.9440</td>
<td>0.5501</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>W</td>
<td>0.4197</td>
<td>ga</td>
<td>0.5051</td>
<td>5.7966</td>
<td>0.9293</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>LN</td>
<td>0.5160</td>
<td>ga</td>
<td>0.6645</td>
<td>9.0213</td>
<td>0.8148</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>W</td>
<td>0.4301</td>
<td>exp</td>
<td>1</td>
<td>13.3912</td>
<td>0.6511</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.3658</td>
<td>14.9227</td>
<td>0.8979</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>W</td>
<td>0.3231</td>
<td>W</td>
<td>0.9833</td>
<td>13.2212</td>
<td>0.6001</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>LN</td>
<td>0.6939</td>
<td>LN</td>
<td>0.7332</td>
<td>14.0609</td>
<td>0.7638</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>LN</td>
<td>0.4501</td>
<td>exp</td>
<td>1</td>
<td>7.9909</td>
<td>0.8214</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>W</td>
<td>0.2281</td>
<td>5.1004</td>
<td>0.7080</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>ga</td>
<td>0.2325</td>
<td>LN</td>
<td>0.3670</td>
<td>13.1779</td>
<td>0.7916</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>exp</td>
<td>1</td>
<td>W</td>
<td>0.3368</td>
<td>11.2724</td>
<td>0.6076</td>
<td>37</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>8.9908</td>
<td>0.6075</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>ga</td>
<td>0.3038</td>
<td>exp</td>
<td>1</td>
<td>8.2323</td>
<td>0.5980</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>ga</td>
<td>0.2776</td>
<td>exp</td>
<td>1</td>
<td>12.9370</td>
<td>0.6775</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>exp</td>
<td>1</td>
<td>ga</td>
<td>0.8797</td>
<td>10.5801</td>
<td>0.8697</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>9.5964</td>
<td>0.7087</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>LN</td>
<td>0.7272</td>
<td>ga</td>
<td>0.5039</td>
<td>11.4333</td>
<td>0.7193</td>
<td>-</td>
</tr>
</tbody>
</table>

To carry out BC-Improvability Indicator 3.2: we introduce

**BC-Continuous Improvement Procedure A.1:**

1. Evaluate the average occupancy of each buffer, $WIP_i$, $i = 1, \ldots, M - 1$.

2. Determine the buffer for which $|WIP_{i-1} - (N_i - WIP_i)|$ is the largest. Assume this is buffer $k$. 

216
Table A.9: Examples of non-exponential five-machine lines (cont.)

<table>
<thead>
<tr>
<th>Line</th>
<th>i</th>
<th>$f_{up,i}$</th>
<th>$CV_{up,i}$</th>
<th>$f_{down,i}$</th>
<th>$CV_{down,i}$</th>
<th>$T_{down,i}$</th>
<th>$e_i$</th>
<th>$N_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
<td>LN</td>
<td>0.9261</td>
<td>W</td>
<td>0.7585</td>
<td>11.9306</td>
<td>0.8868</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>ga</td>
<td>0.3701</td>
<td>12.1931</td>
<td>0.8921</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>W</td>
<td>0.8824</td>
<td>exp</td>
<td>1</td>
<td>10.7019</td>
<td>0.8922</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>W</td>
<td>0.6162</td>
<td>LN</td>
<td>0.9431</td>
<td>5.5913</td>
<td>0.8698</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>LN</td>
<td>0.3233</td>
<td>exp</td>
<td>1</td>
<td>10.4924</td>
<td>0.7000</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>ga</td>
<td>0.8113</td>
<td>W</td>
<td>0.2739</td>
<td>12.0217</td>
<td>0.6670</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>W</td>
<td>0.8571</td>
<td>Ga</td>
<td>0.2524</td>
<td>12.8855</td>
<td>0.6443</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.7246</td>
<td>5.5566</td>
<td>0.9269</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.3233</td>
<td>exp</td>
<td>1</td>
<td>10.4924</td>
<td>0.7000</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>ga</td>
<td>0.2811</td>
<td>exp</td>
<td>1</td>
<td>8.8016</td>
<td>0.5500</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>ga</td>
<td>0.5989</td>
<td>LN</td>
<td>0.6521</td>
<td>13.1424</td>
<td>0.7613</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>W</td>
<td>0.2322</td>
<td>W</td>
<td>0.9889</td>
<td>7.5795</td>
<td>0.5942</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>ga</td>
<td>0.9558</td>
<td>exp</td>
<td>1</td>
<td>11.9192</td>
<td>0.6543</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>ga</td>
<td>0.8231</td>
<td>LN</td>
<td>0.3146</td>
<td>12.8043</td>
<td>0.8907</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>LN</td>
<td>0.6844</td>
<td>exp</td>
<td>1</td>
<td>11.4923</td>
<td>0.8900</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>Ga</td>
<td>0.5187</td>
<td>13.2304</td>
<td>0.5529</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>14.7083</td>
<td>0.8941</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LN</td>
<td>0.2234</td>
<td>exp</td>
<td>1</td>
<td>11.7865</td>
<td>0.8665</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>W</td>
<td>0.5345</td>
<td>W</td>
<td>0.7578</td>
<td>12.2501</td>
<td>0.8188</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>LN</td>
<td>0.2634</td>
<td>exp</td>
<td>1</td>
<td>13.6322</td>
<td>0.7782</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>exp</td>
<td>1</td>
<td>exp</td>
<td>1</td>
<td>13.4672</td>
<td>0.5604</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>W</td>
<td>0.6076</td>
<td>LN</td>
<td>0.5718</td>
<td>8.5633</td>
<td>0.7672</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>exp</td>
<td>1</td>
<td>LN</td>
<td>0.4914</td>
<td>10.6609</td>
<td>0.5709</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>W</td>
<td>0.8402</td>
<td>exp</td>
<td>1</td>
<td>7.8958</td>
<td>0.9324</td>
<td>-</td>
</tr>
</tbody>
</table>
(3) If $WIP_{i-1} - (N_i - WIP_i)$ is positive, transfer a unit of buffer capacity from $b_k$ to $b_{k+1}$; if $WIP_{i-1} - (N_i - WIP_i)$ is negative, re-allocate a unit of buffer capacity from $b_{k+1}$ to $b_k$.

(4) Return to step (1).

(5) Continue this process until arriving at a limit cycle (i.e., when $N_i$’s repeat themselves) and choose the buffer capacity allocation on the limit cycle, which maximizes $TP$.

As a result, BC-Continuous Improvement Procedure A.1 leads to near-optimal buffer capacity allocations (see Table A.10 for illustration). Thus, we consider BC-Improvability Indicator 3.2 justified.

Table A.10: Accuracy of BC-Improvability Indicator 3.2

<table>
<thead>
<tr>
<th>Line</th>
<th>$N^{unimp}$</th>
<th>$TP^{unimp}$</th>
<th>$N^{opt}$</th>
<th>$TP^{opt}$</th>
<th>$\Delta TP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[5, 7, 11, 17]</td>
<td>0.5511</td>
<td>[5, 7, 11, 17]</td>
<td>0.5511</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>[5, 7, 17, 11]</td>
<td>0.5218</td>
<td>[9, 6, 18, 7]</td>
<td>0.5300</td>
<td>1.55</td>
</tr>
<tr>
<td>3</td>
<td>[7, 9, 13, 11]</td>
<td>0.4946</td>
<td>[6, 8, 16, 8]</td>
<td>0.4961</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>[4, 8, 13, 15]</td>
<td>0.4930</td>
<td>[4, 8, 13, 15]</td>
<td>0.4930</td>
<td>1.34</td>
</tr>
<tr>
<td>5</td>
<td>[12, 11, 9, 8]</td>
<td>0.5141</td>
<td>[12, 11, 8, 9]</td>
<td>0.5142</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>[4, 7, 16, 13]</td>
<td>0.6706</td>
<td>[4, 7, 16, 13]</td>
<td>0.6706</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>[10, 15, 11, 4]</td>
<td>0.4780</td>
<td>[10, 16, 9, 5]</td>
<td>0.4786</td>
<td>0.13</td>
</tr>
<tr>
<td>8</td>
<td>[11, 12, 12, 5]</td>
<td>0.4548</td>
<td>[11, 11, 17, 1]</td>
<td>0.4569</td>
<td>0.46</td>
</tr>
<tr>
<td>9</td>
<td>[14, 16, 7, 3]</td>
<td>0.5104</td>
<td>[12, 20, 3, 5]</td>
<td>0.5122</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>[9, 14, 11, 6]</td>
<td>0.4403</td>
<td>[10, 14, 14, 2]</td>
<td>0.4463</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Justification of Numerical Fact 3.4:

A total of 50,000 systems were generated with parameters selected randomly and equiprobably from sets (A.52)-(A.57). For each system, the throughput and machines’ blockages and starvations were evaluated using Simulation Procedure A.1 and the $c$-BN was identified by $c$-BN Indicator 3.1. The partial derivatives involved in (3.13)
were estimated as

\[
\frac{\partial T P}{\partial c_i} \approx \frac{\Delta T P}{\Delta c_i},
\]

(A.64)

with \( \Delta c_i = 0.02 \). The results showed that c-BN Indicator 3.1 identified the c-BN and BN-b with roughly the same accuracy as in the case of exponential lines. Thus, we conclude that Numerical Fact 3.4 holds.
APPENDIX B

PROOFS AND JUSTIFICATIONS FOR CHAPTER IV

To prove Theorems 4.1, 4.2, 4.3 and 4.5, we need the formulas to evaluate the performance measures of a two-machine closed Bernoulli line, which are derived below.

Performance analysis of two-machine closed Bernoulli lines: Under assumptions (a)-(h) of Subsection 2.4.1, a closed Bernoulli line is described by an ergodic Markov chain with states being the probability of occupancy of buffer $b_1$ (since, given the probability of occupancy of $b_1$, the probability of occupancy of $b_0$ can be immediately calculated). Let $P_j$ be the stationary probability that $b_1$ contains $j$ parts, i.e.,

$$P_j = P\{h_1 = j\}, \quad j = 0, 1, \ldots, N_1.$$  \hfill (B.1)

Then, the performance analysis of the system at hand amounts to evaluating $P_j$’s, and then evaluating $PR$, $ST_i$ and $BL_i$, $i = 1, 2$. It turns out that it is convenient to calculate $P_j$ separately for three cases of relationships among $S$, $N_0$, and $N_1$. This is carried out below.

**Case 1:** $S < \min(N_1, N_0)$. The balance equations for this case are:

$$P_0 = (1 - p_1)P_0 + (1 - p_1)p_2 P_1,$$

$$P_1 = p_1 P_0 + [p_1 p_2 + (1 - p_1)(1 - p_2)] P_1 + (1 - p_1)p_2 P_2,$$

$$P_i = p_1 (1 - p_2) P_{i-1} + [p_1 p_2 + (1 - p_1)(1 - p_2)] P_i +$$
\( (1 - p_1)p_2 P_{i+1}, \quad i = 2, \ldots, S - 1, \)

\[
P_S = p_1(1 - p_2)P_{S-1} + (1 - p_2)P_S,
\]

\[
P_j = 0, \quad j = S + 1, \ldots, N_1,
\]

\[
\sum_{i=0}^{N_1} P_i = 1.
\]

Their solution is:

\[
P_i = \frac{\alpha^i}{1 - p_2} P_0, \quad i = 1, \ldots, S - 1,
\]

\[
P_S = \frac{\alpha^S}{1 - p_2} (1 - p_1)P_0,
\]

\[
P_j = 0, \quad j = S + 1, \ldots, N_1,
\]

where

\[
P_0 = Q_1^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} \frac{1-p}{S+1-2p}, & \text{if } p_1 = p_2 = p, \\ \frac{(1-p)(1-\alpha(p_1, p_2))}{1-p_1(1-p_2)\alpha(p_1, p_2)}, & \text{if } p_1 \neq p_2, \end{cases}
\]

\[
\alpha(p_1, p_2) = \frac{p_1(1 - p_2)}{1- \frac{1}{p_1(1 - p_2)}}, \quad B.4
\]

**Case 2a:** \( N_1 < S \leq N_0 \). In this case, \( m_1 \) is never starved and \( m_2 \) is never blocked. In other words, the closed loop does not impede the open system performance. Thus, the stationary probability mass function is the same as the corresponding open line. Therefore, it follows that, for \( N_1 < S \leq N_0 \),

\[
P_0 = Q_2^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} \frac{1-p}{\min(N_1, N_0)+1-p}, & \text{if } p_1 = p_2 = p, \\ \frac{(1-p)(1-\alpha(p_1, p_2))}{1-p_1(1-p_2)\alpha(p_1, p_2)}, & \text{if } p_1 \neq p_2. \end{cases}
\]

**Case 2b:** \( N_0 < S \leq N_1 \). Here, in the reversed flow scenario, the first machine, \( m_2 \), is never starved and the second machine, \( m_1 \), is never blocked. Thus, the line again is equivalent to an open line with the same machines but with the in-process
buffer of capacity $N_0$. Therefore, in this case,

$$P_0 = Q_{2}^c(p_2, p_1, N_0, N_1, S).$$  \hspace{1cm} \text{(B.6)}$$

**Case 3:** $S > \max(N_1, N_0)$. The balance equations in this case are:

$$P_j = 0, \quad j = 0, 1, \ldots, S - N_0 - 1,$$

$$P_{S-N_0} = [(1 - p_1) + p_1p_2] P_{S-N_0} + p_2(1 - p_1) P_{S-N_0+1},$$

$$P_i = P_{i-1}p_1(1 - p_2) + [p_1p_2 + (1 - p_1)(1 - p_2)] P_i + (1 - p_1)p_2 P_{i+1}, \quad i = S - N_0 + 1, \ldots, N_1 - 1,$$

$$P_{N_1} = P_{N_1-1}p_1(1 - p_2) + (p_1p_2 + 1 - p_2) P_{N_1},$$

$$\sum_{i=0}^{N_1} P_i = 1.$$

Their solution is:

$$P_i = \alpha(p_1, p_2)^{i-(S-N_0)} P_{S-N_0}, \quad i = S - N_0, \ldots, N_1, \hspace{1cm} \text{(B.7)}$$

where $\alpha(p_1, p_2)$ is given in (B.4) and

$$P_{S-N_0} = \frac{1}{\sum_{i=0}^{N_1+N_0-S} \alpha(p_1, p_2)^i}. \hspace{1cm} \text{(B.8)}$$

Given the above, the probability that $b_0$ is full and $m_1$ is down can be expressed as

$$Q_{3}^c(p_1, p_2, N_1, N_0, S) = \begin{cases} 
\frac{1-p}{N_1+N_0-S+1}, & \text{if } p_1 = p_2 = p, \\
\frac{(1-p_1)(1-\alpha(p_1,p_2))}{1-\alpha(p_1,p_2)^{N_1+N_0-S+1}}, & \text{if } p_1 \neq p_2. 
\end{cases} \hspace{1cm} \text{(B.9)}$$

Using the three probability mass functions derived above and following the same arguments as in open Bernoulli line [14], we obtain the following expressions for the performance measures of a two-machine closed Bernoulli line:

$$PR_{cl} = p_2[1 - Q_{3}^c(p_1, p_2, N_1, N_0, S)] = p_1[1 - Q_{3}^c(p_2, p_1, N_1, N_0, S)], \hspace{1cm} \text{(B.10)}$$
\[ WIP_1^{cl} = \begin{cases} 
\left( \sum_{i=1}^{S} \frac{i\alpha(p_1, p_2)^i}{1-p_2} - \frac{p_1S\alpha(p_1, p_2)^S}{1-p_2} \right) 
\cdot Q_1^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S \leq \min(N_1, N_0), \\
\sum_{i=1}^{N_1} \frac{i\alpha(p_1, p_2)^i}{1-p_2} Q_2^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } N_1 < S \leq N_0, \\
S - \sum_{i=1}^{N_0} \frac{i\alpha(p_1, p_2)^i}{1-p_2} Q_2^{cl}(p_2, p_1, N_1, N_0, S), & \text{if } N_0 < S \leq N_1, \\
S - N_2 + \sum_{i=1}^{N_1+N_2-S} i\alpha(p_1, p_2)^i 
\cdot Q_3^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S > \max(N_1, N_0), 
\end{cases} \]  

\[ (B.11) \]

\[ \alpha(p_1, p_2) = \frac{p_1(1-p_2)}{p_2(1-p_1)}, \]  

\[ (B.16) \]

\[ Q^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} 
Q_1^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S \leq \min(N_0, N_1), \\
Q_3^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S > \max(N_0, N_1), \\
Q_2^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S > \min(N_1, N_0) \text{ and } S \leq \max(N_1, N_0), 
\end{cases} \]  

\[ (B.17) \]
\[ Q_1^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} \frac{1-p}{S+1-2p}, & \text{if } p_1 = p_2 = p, \\ \frac{(1-p_1)(1-\alpha)}{1-p_1(1-p_1)} S, & \text{if } p_1 \neq p_2, \end{cases} \quad (B.18) \]

\[ Q_2^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} \frac{1-p}{\min(N_1, N_0)+1-p}, & \text{if } p_1 = p_2 = p, \\ \frac{(1-p_1)(1-\alpha)}{1-p_2 \alpha \min(N_1, N_0)} S, & \text{if } p_1 \neq p_2, \end{cases} \quad (B.19) \]

\[ Q_3^{cl}(p_1, p_2, N_1, N_0, S) = \begin{cases} \frac{1-p}{N_1+N_0-S+1}, & \text{if } p_1 = p_2 = p, \\ \frac{(1-p_1)(1-\alpha)}{1-\alpha N_1+N_0-S+1}, & \text{if } p_1 \neq p_2, \end{cases} \quad (B.20) \]

\[ \blacksquare \]

**Proof of Theorem 4.1**

Follows immediately from (B.10).

\[ \blacksquare \]

**Justification of Numerical Fact 4.1**

This numerical fact was justified using simulation. Specifically, We constructed 30,000 closed lines by selecting \( M \), \( p_i \)'s, and \( N_i \)'s randomly and equiprobably from the following sets:

\[ M \in \{3, 5, 10\}, \quad (B.21) \]

\[ p_i \in [0.7, 0.95], \quad i = 1, \ldots, M, \quad (B.22) \]

\[ N_i \in \{1, 2, 3, 4, 5\}, \quad i = 1, \ldots, M-1 \quad (B.23) \]

\[ N_0 \in \{1, 2, \ldots, 15\}, \quad (B.24) \]

\[ S \in \{1, 2, \ldots, \sum_{i=0}^{M-1} N_i - 1\}, \quad (B.25) \]

Using Simulation Procedure A.1, we evaluated the throughput of each line thus constructed and that of its reverse. In these simulations, the carriers were initially placed in the empty carrier buffer, with the excess carriers (if any) placed randomly.
and equiprobably in the in-process buffers. As a result, $PR_{cl}^L$ and $PR_{cl}^{Lr}$ were with 0.001 in all case studied. Thus, we conclude that Numerical Fact 4.1 is justified.


\[ \begin{align*}
\text{Proof of Theorem 4.2} \\
\text{In this proof, we again consider three cases:} \\
\textbf{Case 1: } S < \min(N_1, N_0).
\end{align*} \]

For $p_1 \neq p_2$, function $Q_1^{cl}(p_1, p_2, N_1, N_0, S)$ given in (B.3) can be rewritten as

\[
Q_1^{cl}(p_1, p_2, N_1, N_0, S) = \frac{(1 - p_1)}{1 + \alpha + \cdots + \alpha^{S-3} + \left[ 1 + \frac{p_1(1-p_2)(p_1+p_2)}{p_2^2} \right] \alpha^{S-2}}.
\]

Clearly, it is strictly decreasing in $p_1$ and $S$. Similarly, $Q_1^{cl}(p_2, p_1, N_0, N_1, S)$ is strictly decreasing in $p_2$ and $S$. Thus,

\[
PR_{cl} = p_2[1 - Q_1^{cl}(p_1, p_2, N_1, N_0, S)]
\]

\[
= p_1[1 - Q_1^{cl}(p_2, p_1, N_0, N_1, S)],
\]

is strictly increasing in $p_1$, $p_2$, and $S$, and is independent of, i.e. constant in, $N_1$ and $N_0$.

For $p_1 = p_2 = p$, it is easy to show that

\[
PR_{cl} = \frac{p(S - p)}{S + 1 - 2p} = p - \frac{p(1 - p)}{S + 1 - 2p}, \quad S \geq 2,
\]

which again implies that $PR_{cl}$ is strictly increasing in $p$ and $S$, and is independent of, i.e. constant in, $N_1$ and $N_0$.

\[
\textbf{Case 2: } \min(N_1, N_0) < S \leq \max(N_1, N_0).
\]

In this situation, the closed line is exactly equivalent to an open line. Thus, $PR_{cl}$ is strictly increasing in $p_1$ and $p_2$, monotonically increasing in $N_1$ and $N_0$ and inde-
pendent of, i.e. constant in, $S$.

**Case 3**: $S > \max(N_1, N_0)$.

For $p_1 \neq p_2$,

$$Q_3^d(p_1, p_2, N_1, N_0, S) = \frac{(1 - p_1)}{1 + \alpha + \cdots + \alpha^{N_1+N_0-S}},$$

which implies that this function is strictly decreasing in $p_1$, strictly decreasing in $N_1$ and $N_0$, and strictly increasing in $S$. Therefore,

$$PR_{cl} = p_2[1 - Q_3^d(p_1, p_2, N_1, N_0, S)]$$

$$= p_1[1 - Q_3^d(p_2, p_1, N_0, N_1, S)],$$

is strictly increasing in $p_1$ and $p_2$, strictly decreasing in $S$, and strictly increasing in $N_1$ and $N_0$.

For $p_1 = p_2 = p$,

$$PR_{cl} = \frac{p(N_1 + N_0 - S - p)}{N_1 + N_0 - S + 1}, \quad S \geq 2,$$

and the same conclusions hold.

Thus, $PR_{cl}$ is strictly increasing in $p_1$ and $p_2$, non-strictly increasing in $N_1$ and $N_0$, and non-monotonic concave in $S$.

---

**Justification of Numerical Fact 4.2**

To investigate the monotonicity property of $PR_{cl}$, the same approach used in the justification of Numerical Fact 3.2 was applied to the 30,000 lines generated in the justification of Numerical Fact 4.1. Among all cases studied, no counterexample was found. Thus, we conclude that Numerical Fact 4.2 indeed takes place.
Proof of Theorem 4.3

It has been shown that in open lines \[14\]

\[
PR_o(p_1, p_2, N_1) = p_2[1 - Q_o(p_1, p_2, N_1)],
\]  

(B.26)

where

\[
Q_o(p_1, p_2, N_1) = \begin{cases} 
\frac{1-p}{N_1+1-p}, & \text{if } p_1 = p_2 = p, \\
\frac{(1-p)(1-\alpha)}{1-\frac{p}{p_2}^\alpha}, & \text{if } p_1 \neq p_2.
\end{cases}
\]  

(B.27)

It can be shown that

\[
Q_o(p_1, p_2, N_1) < Q_c^1(p_1, p_2, N_1, N_0, S), \quad S \leq \min(N_1, N_0),
\]

\[
Q_o(p_1, p_2, N_1) < Q_c^2(p_1, p_2, N_1, N_0, S), \quad N_0 < S \leq N_1,
\]

\[
Q_o(p_1, p_2, N_1) = Q_c^2(p_1, p_2, N_1, N_0, S), \quad N_1 < S \leq N_0,
\]

\[
Q_o(p_1, p_2, N_1) < Q_c^3(p_1, p_2, N_1, N_0, S), \quad S > \max(N_1, N_0).
\]

Therefore, the pair \((N_0, S)\) is unimpeding, i.e.

\[
PR_c(p_1, p_2, N_1, N_0, S) = PR_o(p_1, p_2, N_1),
\]

if and only if \(N_1 < S \leq N_0\).

\[
\Box
\]

Proof of Theorem 4.4

When (4.4) occurs, \(m_1\) is never starved and \(m_M\) is never blocked. Hence, the closed nature of the line does not impact the open line performance.

\[
\Box
\]

Justification of Numerical Fact 4.3

The justification was carried out using Simulation Procedure A.1. In these simulations, the carriers were initially placed in the empty carrier buffer, with the excess
carriers (if any) placed randomly and equiprobably in the in-process buffers. As a result, \( PR_{cl} \), \( ST_{i}^{cl} \), \( BL_{i}^{cl} \), \( PR_{o} \), \( ST_{i}^{o} \), \( BL_{i}^{o} \) have been evaluated with 95% confidence intervals 0.001 for production rates and 0.002 for blockages and starvations.

We constructed 30,000 closed lines by selecting \( M \), \( p_{i} \)'s, and \( N_{i} \)'s randomly and equiprobably from the following sets:

\[
M \in \{3, 5, 10\},
\]

\[
p_{i} \in [0.7, 0.95], \quad i = 1, \ldots, M,
\]

\[
N_{i} \in \{1, 2, 3, 4, 5\}, \quad i = 1, \ldots, M - 1
\]

\[
N_{0} \in \{3, 4, \ldots, \sum_{i=1}^{M-1} N_{i} - 1\}.
\]

Note that on the set (B.31), condition (4.4) does not hold. In addition, \( \delta \) was selected as 0.01.

For each of these lines, using Simulation Procedure A.1, we evaluated \( PR_{cl} \) and \( PR_{o} \) and selected randomly and equiprobably a pair \( (N_{0}^{*}, S^{*}(N_{0}^{*})) \) that was practically unimpeding in the sense of Definition 4.2. To verify whether this pair remained practically unimpeding with other values of \( p_{i} \)'s, for each of the 30,000 lines mentioned above, we constructed 10 additional lines with \( N_{i} \), \( i = 0, 1, \ldots, M - 1 \), being the same but with new \( p_{i} \)'s selected randomly and equiprobably from the set (B.29). Again, using Simulation Procedure A.1, we evaluated \( PR_{cl} \) and \( PR_{o} \) for each of the new lines, thus constructed, and checked whether Definition 4.2 holds. The results are shown in Figure B.1. As one can see, \( (N_{0}^{*}, S^{*}(N_{0}^{*})) \) remains unimpeding in almost 100% of the cases analyzed. Thus, we conclude that Numerical Fact 4.3 indeed takes place.

\[\blacksquare\]

**Proof of Theorem 4.5**

Two cases are considered:
Figure B.1: Accuracy of Numerical Fact 4.3

Case 1: \( N_1 \neq N_0 \). Under this condition, it follows from Theorem 4.2 that a closed line with two machines is \( S^+ \)-improvable if \( S < \min(N_1, N_0) + 1 \), i.e. \( S^*_{\text{min}} = \min(N_1, N_0) + 1 \), and

\[
ST_{1}^{cl} = p_1 Q^{cl}(p_2, p_1, N_0, N_1, S),
\]
\[
BL_{1}^{cl} = \begin{cases} 
    p_1 (1 - p_2) Q^{cl}(p_2, p_1, N_0, N_1, S), & \text{if } S = N_1 < N_0, \\
    0, & \text{if } S < N_1,
\end{cases}
\]
\[
ST_{2}^{cl} = p_2 Q^{cl}(p_1, p_2, N_1, N_0, S),
\]
\[
BL_{2}^{cl} = \begin{cases} 
    (1 - p_1) p_2 Q^{cl}(p_1, p_2, N_1, N_0, S), & \text{if } S = N_0 < N_1, \\
    0, & \text{if } S < N_0.
\end{cases}
\]

Clearly, \( ST_{1}^{cl} > BL_{1}^{cl} \) and \( ST_{2}^{cl} > BL_{2}^{cl} \). Therefore, for \( S^+ \)-improvable situation,

\[
ST_{1}^{cl} + ST_{2}^{cl} > BL_{1}^{cl} + BL_{2}^{cl}.
\]  
(B.36)

Similarly, a closed line with two machines is \( S^- \)-improvable if \( S > \max(N_1, N_0) \), i.e. \( S^*_{\text{max}} = \max(N_1, N_0) \), and

\[
ST_{1}^{cl} = 0,
\]
\[
BL_{1}^{cl} = p_1 (1 - p_2) Q^{cl}(p_2, p_1, N_0, N_1, S),
\]
\[
(B.37)
\]
\[
(B.38)
\]
\[ ST_2^{cl} = 0, \quad (B.39) \]
\[ BL_2^{cl} = (1 - p_1)p_2 Q^{cl}(p_1, p_2, N_1, N_0, S). \quad (B.40) \]

Therefore, for \( S^- \)-improvable situation,

\[ ST_1^{cl} + ST_2^{cl} < BL_1^{cl} + BL_2^{cl}. \quad (B.41) \]

**Case 2:** \( N_1 = N_0 = N \). Under this condition,

\[
Q_1^{cl}(p_1, p_2, N_1 = N, N_0 = N, S = N) < Q_3^{cl}(p_1, p_2, N_1 = N, N_0 = N, S = N + 1),
\]

and therefore,

\[
PR_{cl}(p_1, p_2, N_1 = N, N_0 = N, S = N) > PR_{cl}(p_1, p_2, N_1 = N, N_0 = N, S = N + 1).
\]

In other words, the line is \( S^+ \)-improvable if \( S < \min(N_1, N_0) = N \) and \( S^- \)-improvable if \( S > \min(N_1, N_0) = N \), i.e. \( S_{\min}^* = S_{\max}^* = N \). Then, using again (B.34) and (B.39), we obtain that

\[
ST_1^{cl} + ST_2^{cl} > BL_1^{cl} + BL_2^{cl}, \quad \text{if} \quad S < N, \quad (B.42)
\]
\[
ST_1^{cl} + ST_2^{cl} < BL_1^{cl} + BL_2^{cl}, \quad \text{if} \quad S > N, \quad (B.43)
\]

which completes the proof.

**Proof of Theorem 4.6**

Follows directly from Theorem 4.5 and the definition of \( N_0 \)-improvability.
Justification of Numerical Fact 4.4

We constructed 300,000 closed lines by selecting $M$, $p_i$’s, and $N_i$’s randomly and equiprobably from the following sets:

$$M \in \{3, 5, 10\}, \quad (B.44)$$

$$p_i \in [0.7, 0.95], \quad i = 1, \ldots, M, \quad (B.45)$$

$$N_i \in \{1, 2, 3, 4, 5\}, \quad i = 1, \ldots, M - 1 \quad (B.46)$$

$$N_0 \in \{1, 2, \ldots, 15\}. \quad (B.47)$$

For each of these lines, using Simulation Procedure A.1, we evaluated $PR_{cl}(S)$ for all $S \in \{1, 2, \ldots, \sum_{i=0}^{M-1} N_i - 1\}$ and determined $S_{opt}$ for which $PR_{cl}(S)$ is maximized. Also, for each of these lines, using Simulation Procedure A.1 and Numerical Fact 4.4, we obtained $S_{unimp}$, i.e., the value of $S$ at which the improvability indicator $I$ changes its sign. Then we compared the values of $PR_{cl}(S_{opt})$ and $PR_{cl}(S_{unimp})$. The results are shown in Figure B.2. As one can see, the two production rates are within 1% of each other in almost 100% of the cases analyzed. Thus we conclude that Numerical Fact 4.4 indeed defines the conditions of $S$-improvability.

![Figure B.2: Accuracy of Numerical Fact 4.4](image)

Proof of Theorem 4.7

Follows the proof of the BN theorem obtained in [20] for open lines.
Justification of Bottleneck Indicator 4.1

This justification has been carried out as follows: A total of 1,000,000 closed lines were generated with parameters selected randomly and equiprobably from sets (B.44)-(B.47) and $S \in \{M, M+1, \ldots, \sum_{i=0}^{M-1} N_i\}$. Each of these lines has been analyzed using Simulation Procedure 3.1. Specifically, the probabilities of blockages and starvations of all machines were estimated and, in addition, partial derivatives of the production rate with respect to $p_i$’s were evaluated (with $\Delta p_i = 0.01$). The probabilities of blockages and starvations were used to identify the BN by Bottleneck Indicator 4.1, and the partial derivatives were used to identify the BN by (4.10). If the BNs identified by both method were the same, we conclude that Bottleneck Indicator 4.1 holds; otherwise, we conclude that it does not.

The results obtained using this approach are summarized in Figure B.3. Among the 1,000,000 lines analyzed, 87.59% had a single machine with no emanating arrows, and the BN machine was identified by Bottleneck Indicator 4.1 correctly in 92.76% of these cases. For the 12.41% of systems with more than one machine having no emanating arrows, Bottleneck Indicator 4.1 identified correctly the PBN in 71.1% of cases, while the PBN was indeed in the set of local BN’s in 97.20% of cases. These results are similar to those for BN identification in open lines [20]. Thus, we conclude that Bottleneck Indicator 4.1 provides a sufficiently accurate tool for bottleneck identification in closed lines.

Justification of Numerical Fact 4.5

To carry out the justification, a total of 100,000 closed lines has been generated with the parameters selected randomly and equiprobably from the sets

\[
M \in \{3, 5, 10\}, \quad \text{(B.48)}
\]
\[
T_{\text{down},i} \in [1, 5], \quad i = 1, \ldots, M, \quad \text{(B.49)}
\]
Figure B.3: Accuracy of Bottleneck Indicator 4.1

\[
\begin{align*}
    e_i &\in [0.75, 0.95], \quad i = 1, \ldots, M, \\
    c_i &\in [0.8, 1.2], \quad i = 1, \ldots, M - 1, \\
    f_{t_{up},i}, f_{t_{down},i} &\in \{exp, W, ga, LN\}, \quad i = 1, \ldots, M, \\
    CV_{up,i}, CV_{down,i} &\in [0.1, 1], \quad i = 1, \ldots, M, \\
    N_i &\in \{2, 3, \ldots, 10\}, \quad i = 1, \ldots, M - 1, \\
    N_0 &\in \{10, 11, \ldots, 50\}, \\
    S &\in \{1, 2, \ldots, 50\}.
\end{align*}
\] (B.50) (B.51) (B.52) (B.53) (B.54) (B.55) (B.56)

For each of the 100,000 closed line, thus constructed, we evaluated its throughput, \(TP_L\) and that of its reverse, \(TP_{L^r}\), by simulations using Simulation Procedure A.1. As a result, it has been determined that the maximal difference between \(TP_L\) and \(TP_{L^r}\) is less than 0.001. Thus, we conclude that Numerical Fact 4.5 holds.
Justification of Numerical Fact 4.6

This justification has been carried out using the 100,000 closed lines generated above and Simulation Procedure A.1. The approach used in the justification of Numerical Fact 3.2 are applied again. In all cases analyzed, no counterexamples were found.

Proof of Theorem 4.8

Under the above condition, $m_1$ is not starved, $m_M$ is not blocked, and, therefore, $TP_{cl} = TP_o$.

Justification of Numerical Fact 4.7

A set of 10,000 closed lines was generated with parameters selected randomly and equiprobably from (B.48)-(B.54), (B.52), (B.53) and $\delta = 0.01$ has been used. The empty carrier buffer capacity has been selected randomly and equiprobably from

$$N_0 \in \left\{5, 6, \ldots, \sum_{i=1}^{M-1} N_i - 1 \right\}, \quad (B.57)$$

provided that for the selected $N_0$ an unimpeding $S$ exists. Then, 20 more lines have been constructed with $N_i, i = 0, \ldots, M-1$, being the same but with new machine parameters randomly and equiprobably selected from (B.49)-(B.51), (B.52) and (B.53). For each of the new lines, we evaluated $TP_o$ and $TP_{cl}(S), S = 1, \ldots, \sum_{i=0}^{M-1} N_i$. As a result, we determine that a practically unimpeding pair $(N_0, S)$ remains practically unimpeding in 86.9% of the cases analyzed.

This number is unusually low compared to the Bernoulli case. The reason for this is that buffers with the same capacity exhibit significantly different disturbance rejection capabilities when the machine up- and downtimes are modified, while it is
not the case for the Bernoulli lines. An example of this phenomenon is illustrated in Figure B.4.

Despite of the above, since among all the systems studied, the average $\Delta_{imp}$ is 0.0083 and the maximum is 0.0515, we conclude that Numerical Fact 4.7 holds.

To prove Theorem 4.9, we use the following technique: Consider a two-machine closed line and a two-machine open line (referred to as the equivalent open line) shown in Figure B.5(a) and (b), respectively. Clearly, the machines in both lines are identical.
but the buffers are different.

\[ S = N_0 \]

\[ N_0 < S < \max(N_1, N_0) \]

\[ b_1 \]

\[ b_2 \]

\[ m_1 \]

\[ m_2 \]

\[ f_{up,1}, f_{down,1} \]

\[ f_{up,2}, f_{down,2} \]

(a) Closed line  
(b) Equivalent open line

Figure B.5: Closed and equivalent open two-machine lines

Let \( TP_{eo}, WIP_{1eo}, BL_{1eo} \) and \( ST_{2eo} \) denote the throughput, work-in-process, probability of blockage of \( m_1 \) and probability of starvation of \( m_2 \) in the equivalent open line, respectively. Then

**Lemma B.1** If

\[ N_{eo} = \begin{cases} 
S, & S \leq \min(N_1, N_0), \\
\min(N_1, N_0), & \min(N_1, N_0) < S \leq \max(N_1, N_0), \\
N_1 + N_0 - S, & \max(N_1, N_0) < S,
\end{cases} \tag{B.58} \]

the performance characteristics of the closed two-machine line defined by assumptions (a)-(h) of Subsection 2.4.2 and the equivalent open two-machine line defined by assumptions (a)-(e) of Subsection 2.4.2 are related as follows:

\[ TP_{cl} = TP_{eo}, \tag{B.59} \]

\[ WIP_{1cl} = \max(0, S - N_0) + WIP_{1eo}, \tag{B.60} \]

\[ BL_{1cl} = \begin{cases} 
BL_{1eo}, & S \geq N_1, \\
0, & \text{otherwise},
\end{cases} \tag{B.61} \]

\[ ST_{1cl} = \begin{cases} 
BL_{1eo}, & S \leq N_1, \\
0, & \text{otherwise},
\end{cases} \tag{B.62} \]
\[
\begin{align*}
BL_{2}^{cl} &= \begin{cases} 
ST_{2}^{eo}, & S \geq N_{0}, \\
0, & \text{otherwise},
\end{cases} \\
ST_{2}^{cl} &= \begin{cases} 
ST_{2}^{eo}, & S \leq N_{0}, \\
0, & \text{otherwise}.
\end{cases}
\end{align*}
\] (B.63) (B.64)

**Proof:** Consider the closed and equivalent open lines of Figure B.5 and assume that the machines in both lines follow the same sample path, i.e., \(m_{i}, \ i = 1, 2\), in the closed line is up (respectively, down) if and only if this machine is the open line is up (respectively, down). Under this assumption, we show that if \(N_{eo}\) is selected as in (B.58), the probabilities of blockages and starvations of the machines in both lines are related so that the throughputs of the two lines are identical.

**Case 1:** \(S \leq \min(N_{1}, N_{0})\). In this case, \(N_{eo} = S\) and

\[
H_{1}^{cl}(t) = h, \ 0 \leq h \leq S \iff H_{eo}(t) = h, \ 0 \leq h \leq S,
\]

where \(H_{1}^{cl}(t)\) and \(H_{eo}(t)\) are the occupancies of \(b_{1}\) and \(b\) at time \(t\) in the closed and open lines, respectively. Thus, denoting the occupancy of \(b_{0}\) at time \(t\) as \(H_{0}^{cl}(t)\), we have:

\[
ST_{1}^{cl} = P_{cl}[H_{0}^{cl} = 0, m_{1} \text{ is up, } m_{2} \text{ is down}] + \\
+ P_{cl}[H_{0}^{cl} = 0, m_{1} \text{ is up, } m_{2} \text{ is up}] \max(0, c_{1} - c_{2})
\] = \[P_{cl}[H_{1}^{cl} = S, m_{1} \text{ is up, } m_{2} \text{ is down}] + \\
+ P_{cl}[H_{1}^{cl} = S, m_{1} \text{ is up, } m_{2} \text{ is up}] \max(0, c_{1} - c_{2})
\] = \[P_{eo}[H_{eo} = N_{eo}, m_{1} \text{ is up, } m_{2} \text{ is down}] + \\
+ P_{eo}[H_{eo} = N_{eo}, m_{1} \text{ is up, } m_{2} \text{ is up}] \max(0, c_{1} - c_{2})
\] = \[BL_{1}^{eo},
\]

\[
BL_{1}^{cl} = P_{cl}[H_{1}^{cl} = N_{1}, m_{1} \text{ is up, } m_{2} \text{ is down}]
\]
\begin{align*}
+P_{cl}[H_1^{cl} = N_1, m_1 \text{ is up}, m_2 \text{ is up}] \max(0,c_1 - c_2)
&= P_{eo}[H_{eo} = N_1, m_1 \text{ is up}, m_2 \text{ is down}] \\
&+P_{eo}[H_{eo} = N_1, m_1 \text{ is up}, m_2 \text{ is up}] \max(0,c_1 - c_2)
&= \begin{cases} 
BL_1^{eo}, & \text{if } S = N_1 \leq N_0, \\
0, & \text{otherwise}
\end{cases}
\tag{B.65}
\end{align*}

and

\begin{align*}
ST_2^{cl} &= P_{cl}[H_1^{cl} = 0, m_1 \text{ is up}] \\
&= P_{eo}[H_{eo} = 0, m_1 \text{ is up}] \\
&= ST_2^{eo}, \\
BL_2^{cl} &= P_{cl}[H_0^{cl} = N_0, m_1 \text{ is down}, m_2 \text{ is up}] \\
&+P_{cl}[H_0^{cl} = N_0, m_1 \text{ is up}, m_2 \text{ is up}] \max(0,c_2 - c_1)
&= P_{eo}[H_{eo} = S - N_0, m_1 \text{ is down}, m_2 \text{ is up}] \\
&+P_{eo}[H_{eo} = S - N_0, m_1 \text{ is up}, m_2 \text{ is up}] \max(0,c_2 - c_1)
&= \begin{cases} 
ST_2^{eo}, & \text{if } S = N_0 \leq N_1, \\
0, & \text{otherwise}.
\end{cases}
\end{align*}

In addition,

\begin{align*}
TP_{cl} &= c_2[e_2 - ST_2^{cl}] = c_2[e_2 - ST_2^{eo}] = TP_{eo}, \\
WIP_1^{cl} &= E[H_1^{cl}] = E[H_{eo}] = WIP_{eo}.
\end{align*}

**Case 2a:** \(N_1 < S \leq N_0\). In this case, \(N_{eo} = N_1\) and

\[H_1^{cl}(t) = h, \ 0 \leq h \leq N_1 \iff H_{eo}(t) = h, \ 0 \leq h \leq N_1.\]

Using the same argument as in Case 1, it can be shown that

\[TP_{cl} = TP_{eo},\]
\[ WIP_{cl}^1 = WIP_{eo}, \]
\[ ST_{cl}^1 = 0, \]
\[ BL_{cl}^1 = BL_{eo}^1, \]
\[ ST_{cl}^2 = ST_{eo}^2, \]
\[ BL_{cl}^2 = 0. \]

**Case 2b:** \( N_0 < S \leq N_1 \). In this case, \( N_{eo} = N_0 \) and

\[ H_{cl}^1(t) = h, \ S - N_0 \leq h \leq S \iff H_{eo}(t) = h - (S - N_0), \ S - N_0 \leq h \leq S. \]

Again, using the same argument in Case 1, it can be obtained that

\[ TP_{cl} = TP_{eo}, \]
\[ WIP_{cl}^1 = (S - N_0) + WIP_{eo}, \]
\[ ST_{cl}^1 = BL_{eo}, \]
\[ BL_{cl}^1 = 0, \]
\[ ST_{cl}^2 = 0, \]
\[ BL_{cl}^2 = ST_{eo}^2. \]

**Case 3:** \( S \geq \max(N_1, N_0) \). In this case, \( N_{eo} = N_1 + N_0 - S \)

\[ H_{cl}^1(t) = h, \ S - N_0 \leq h \leq N_1 \iff H_{eo}(t) = h - (S - N_0), \ S - N_0 \leq h \leq N_1. \]

Finally, it can be shown that

\[ TP_{cl} = TP_{eo}, \]
\[ WIP_{cl}^1 = (S - N_0) + WIP_{eo}, \]
\[ ST_{cl}^1 = \begin{cases} 
BL_{eo}, & \text{if } S = N_1 \geq N_0, \\
0, & \text{otherwise}, 
\end{cases} \]
\[ BL_{cl}^1 = BL_{eo}. \]
Proof of Theorem 4.9

Using Lemma B.1, it can be shown that

\[ S_{\min}^* = \min(N_1, N_0), \quad S_{\max}^* = \max(N_1, N_0). \]  \hspace{1cm} (B.66)

Therefore,

\[ S < S_{\min}^* \iff BL_{1}^{co} = BL_{2}^{co} = 0 \iff I = \sum_{i=1}^{2} (ST_{i}^{cl} - BL_{i}^{cl}) > 0, \]

\[ S > S_{\max}^* \iff ST_{1}^{co} = ST_{2}^{co} = 0 \iff I = \sum_{i=1}^{2} (ST_{i}^{cl} - BL_{i}^{cl}) < 0. \]

Justification of Numerical Fact 4.8

A set of 50,000 closed lines has been generated with parameters randomly and equiprobably selected from (B.48)-(B.55) and (B.52), (B.53). For each of these lines, using Simulation Procedure A.1, we evaluated \( TP_{cl}(S) \) for all \( S \in \{1, 2, \ldots, \sum_{i=0}^{M-1} N_{i} - 1\} \) and determined \( S_{opt} \) for which \( TP_{cl}(S) \) is maximized. Also, for each of these lines, using Numerical Fact 4.8, we obtained \( S_{unimp} \), i.e., the value of \( S \) at which the improvability indicator \( I \) changes its sign. Then, we compared the values of \( TP_{cl}(S_{opt}) \) and \( TP_{cl}(S_{unimp}) \). As a result, we determined that the two throughputs are within 1\% of each other in 99.12\% of the cases analyzed. Thus, we conclude that Numerical Fact 4.8 holds.
APPENDIX C

PROOFS AND JUSTIFICATIONS FOR
CHAPTER V

To prove the theorems and justify the numerical facts in Sections 5.2 and 5.3, we need the formulas to evaluate the performance measures of Bernoulli lines with non-perfect quality machines, which are derived below.

Production lines with a single inspection machine \( m_M \): When the only inspection machine is the last one, \( m_M \), the performance analysis can be carried out using the recursive aggregation procedure developed in [14] for serial lines producing no defectives, i.e., defined by assumptions (a)-(e) of Subsection 2.4.1:

Recursive Procedure C.1 [14]:

\[
\begin{align*}
    p_i^b(s+1) &= p_i[1-Q(p_{i+1}^b(s+1), p_i^f(s), N_i)], \quad i = 1, \ldots, M-1, \quad (C.1) \\
    p_i^f(s+1) &= p_i[1-Q(p_{i-1}^f(s+1), p_i^b(s+1), N_{i-1})], \quad i = 2, \ldots, M, \quad (C.2) \\
    &\quad s = 0, 1, 2, \ldots,
\end{align*}
\]

with initial conditions

\[
    p_i^f(0) = p_i, \quad i = 1, \ldots, M, \quad (C.3)
\]

and boundary conditions

\[
    p_1^f(s) = p_1, \quad p_M^b(s) = p_M, \quad s = 0, 1, \ldots, \quad (C.4)
\]
where

\[
Q(p_1, p_2; N_1) = \begin{cases} 
\frac{(1-p_1)(1-\alpha(p_1, p_2))}{1-p_2}, & \text{if } p_1 \neq p_2, \\
\frac{1-p}{N_1+1-p}, & \text{if } p_1 = p_2 = p,
\end{cases}
\]

\( (C.5) \)

\[
\alpha(p_1, p_2) = \frac{p_1(1-p_2)}{p_2(1-p_1)}. \quad (C.6)
\]

It has been shown in [14] that this procedure is convergent and the following limits exist:

\[
p^f_i = \lim_{s \to \infty} p^f_i(s), \quad p^b_i = \lim_{s \to \infty} p^b_i(s). \quad (C.7)
\]

Then, the performance measures of a line defined by assumptions (a)-(h) with one inspection machine \( m_M \) are given by:

\[
\widehat{PR} = p^f_M q_M, \quad (C.8)
\]

\[
\widehat{SR}_M = p^f_M (1-q_M), \quad (C.9)
\]

\[
\widehat{CR} = p^b_i = p^f_i (1 - Q(p^f_{i-1}, p^b_i, N_{i-1})) = p^f_i (1 - Q(p^b_{i+1}, p^f_i, N_i)), \quad (C.10)
\]

\[
\widehat{WIP}_i = \begin{cases} 
\frac{p^f_i}{p^f_{i+1} - p^f_i \alpha N_i(p^f_{i+1})} \left( \frac{1-\alpha N_i(p^f_{i+1})}{1-\alpha(p^f_{i+1})} - N_i \alpha N_i(p^f_{i+1}) \right), & \text{if } p^f_i \neq p^b_{i+1}, \\
\frac{N_i(N_i+1)}{2(N_i+1-p)}, & \text{if } p^f_i = p^b_{i+1} = p,
\end{cases}
\]

\( (C.11) \)

\[
\widehat{ST}_i = p_i Q(p^f_{i-1}, p^b_i, N_{i-1}), \quad i = 2, \ldots, M, \quad (C.12)
\]

\[
\widehat{BL}_i = p_i Q(p^b_{i+1}, p^f_i, N_i), \quad i = 1, \ldots, M-1, \quad (C.13)
\]

where \( q_M \), referred to as the quality buy rate of the system, is given by

\[
q_M = \prod_{i \in I_{np}} g_i. \quad (C.14)
\]

The accuracy of these estimates is the same as that of serial Bernoulli line reported in [14].

Clearly, expressions (C.8)-(C.14) can be used for performance evaluation of Bernoulli
lines without inspection machines as well. In this case, (C.8) and (C.9) represent the production rate of non-defective and defective parts, respectively.

**Production lines with multiple inspection machines or a single inspection machine other than** $m_M$: **Aggregation procedure:** For the case of multiple inspection machines or a single inspection machine $m_i, i \neq M$, Recursive Procedure C.1 is modified as follows:

**Recursive Procedure C.2:**

\[
\begin{align*}
    p_i^b(s + 1) &= p_i[1 - Q(p_{i+1}^b(s + 1), p_i^f(s), N_i)], & i &= 1, \ldots, M - 1, \\
p_i^f(s + 1) &= p_i[1 - Q(p_{i-1}^f(s + 1), p_i^b(s + 1), N_{i-1})], & i \notin I_{\text{insp}}, & (C.15) \\
p_i^f(s + 1) &= p_i[1 - Q(p_{i-1}^f(s + 1), p_i^b(s + 1), N_{i-1})]q_i, & i \in I_{\text{insp}}, \\
    s &= 0, 1, 2, \ldots,
\end{align*}
\]

with initial conditions $p_i^f(0), i = 1, \ldots, M$, boundary conditions $p_1^f(s)$ and $p_M^b(s)$, $s = 0, 1, \ldots$, and function $Q$ defined by (C.3), (C.4) and (C.5), respectively, and with $q_i$ given by

\[
q_i = \prod_{i-k<i<i} g_i,
\]

(C.16)

where $i-k$ is the index of the upstream inspection machine, $m_{i-k}$, which is the closest to the inspection machine $m_i$. The quantity $q_i, i \in I_{\text{insp}}$, is referred to as the quality buy rate of inspection machine $m_i, i \in I_{\text{insp}}$.

**Theorem C.1** Sequences $p_2^f(s), \ldots, p_M^f(s)$ and $p_1^b(s), \ldots, p_{M-1}^b(s)$, $s = 1, 2, \ldots$, defined by Recursive Procedure C.2, are convergent and the following limits exist:

\[
\begin{align*}
    p_i^f &= \lim_{s \to \infty} p_i^f(s), & p_i^b &= \lim_{s \to \infty} p_i^b(s). & (C.17)
\end{align*}
\]

**Proof:** Similar to the proof of Lemma 1 in [14].
Performance measure estimates: Using $p_i^f$ and $p_i^b$, the estimates of the performance measures for the production line at hand are defined as follows:

$$\hat{PR} = p_M^f,$$  \hspace{1cm} (C.18)

$$\hat{SR}_i = p_i^b [1 - Q(p_{i-1}^f, p_i^b, N_{i-1})](1 - q_i)$$

$$\quad = p_{i-1}^f [1 - Q(p_i^b, p_{i-1}^f, N_{i-1})](1 - q_i), \quad i \in I_{insp}, \hspace{1cm} (C.19)$$

$$\hat{CR} = p_1^b,$$  \hspace{1cm} (C.20)

and the expressions for $\hat{WIP}_i$, $\hat{ST}_i$ and $\hat{BL}_i$ remain the same as in the previous case, i.e., are defined by (C.11)-(C.13). Also, as in the previous case, it can be shown that

$$\hat{CR} = \hat{PR} + \hat{SR},$$  \hspace{1cm} (C.21)

where

$$\hat{SR} = \sum_{i \in I_{insp}} \hat{SR}_i.$$  \hspace{1cm} (C.22)

Accuracy of the estimates: The accuracy of these estimates has been investigated numerically using a C++ code, which simulates production lines defined by assumptions (a)-(h) of Subsection 5.2.1 and Simulation Procedure A.1 (evaluating not only $PR$, $BL_i$ and $ST_i$ but also $SR_i$ and $CR$).

Using this procedure, a total of 100,000 lines with $M = 10$ were investigated with the parameters $p_i$’s, $g_i$’s, $N_i$’s and the number of inspection machines, $|I_{insp}|$, selected randomly and equiprobably from the sets

$$p_i \in [0.7, 0.95],$$  \hspace{1cm} (C.23)

$$g_i \in [0.7, 1),$$  \hspace{1cm} (C.24)

$$N_i \in \{1, 2, 3, 4, 5\},$$  \hspace{1cm} (C.25)

$$|I_{insp}| \in \{1, 2, 3\}.$$  \hspace{1cm} (C.26)
The positions of the inspection machines were selected randomly and equiprobably from the sets \{2, \ldots, 10\} (without replacement). Each remaining (producing) machine was chosen to be perfect or non-perfect with probability 1/2. For each non-perfect machine, the quality parameter was selected randomly and equiprobably from the set (C.24), and the efficiency of all machines has been selected randomly and equiprobably from set (C.23).

The performance measures of the systems, thus constructed, have been calculated analytically using expressions (C.18)-(C.20), (C.11)-(C.13) and numerically using Simulation Procedure A.1. The accuracy of the analytical estimates, as compared with those obtained by simulations, has been evaluated by

\[
\epsilon_{PR} = \frac{|PR - \hat{PR}|}{PR} \cdot 100\%, \\
\epsilon_{SR} = \max_{i \in I_{insp}} \frac{|SR_i - \hat{SR}_i|}{SR_i} \cdot 100\%, \\
\epsilon_{CR} = \frac{|CR - \hat{CR}|}{CR} \cdot 100\%, \\
\epsilon_{WIP} = \frac{1}{M-1} \sum_{i=1}^{M-1} \frac{|WIP_i - \hat{WIP}_i|}{WIP_i} \cdot 100\%, \\
\epsilon_{ST} = \frac{1}{M-1} \sum_{i=2}^{M} |ST_i - \hat{ST}_i|, \\
\epsilon_{BL} = \frac{1}{M-1} \sum_{i=1}^{M-1} |BL_i - \hat{BL}_i|.
\]

The results are as follows: Among the 100,000 lines analyzed, the averages of \(\epsilon_{PR}, \epsilon_{SR}\) and \(\epsilon_{CR}\) are less than 1%, the average of \(\epsilon_{WIP}\) is less than 3%, and the averages of \(\epsilon_{ST}\) and \(\epsilon_{BL}\) are less than 0.005. Therefore, we conclude that recursive procedure (C.15) results in an acceptable precision of performance estimates for serial lines under consideration.
Proof of Theorem 5.1

Follows directly from Theorem 3.1.

Proof of Theorem 5.2

The proof of the monotonicity of $\hat{PR}, \hat{SR}$ and $\hat{CR}$ with respect to $p_i$ and $N_i$ is similar to the proof of Theorem 2.3 in [14]. As for the monotonicity with respect to $g_i$, it follows from (C.15) and the properties of function $Q$ that $p^f_i(s)$ and $p^b_i(s)$, $s = 1, 2, \ldots$, are monotonically increasing and decreasing in $g_j$, $j \in I_{np}$, respectively. Since $\hat{PR} = p^f_M = \lim_{s \to \infty} p^f_M(s)$ and $\hat{CR} = p^b_i = \lim_{s \to \infty} p^b_i(s)$, the above implies that $\hat{CR}$ and $\hat{PR}$ are monotonically decreasing and increasing in $g_i$, $i \in I_{np}$, respectively. Also, since $\hat{SR} = \hat{CR} - \hat{PR}$, the scrap rate is monotonically decreasing in $g_i$. Finally, if the only inspection machine is $m_M$, $\hat{CR}$ is equal to the production rate of the corresponding line without scrap and, therefore, $\hat{CR}$ is independent of $g_i$.

Proof of Theorem 5.3

As it follows from (C.20), $\hat{CR}(1)$, is equal to $\hat{PR}(2)$. Therefore,

$$\frac{\partial \hat{CR}(1)}{\partial p_i} > \frac{\partial \hat{CR}(1)}{\partial p_j}, \quad \forall j \neq i \iff \frac{\partial \hat{PR}(2)}{\partial p_i} > \frac{\partial \hat{PR}(2)}{\partial p_j}, \quad \forall j \neq i.$$  \hfill (C.27)

Due to (C.18),

$$\frac{\partial \hat{PR}(1)}{\partial p_l} = \frac{\partial \hat{CR}(1)}{\partial p_l} \prod_{i \in I_{np}} g_i, \quad l = 1, \ldots, M.$$  \hfill (C.28)

Hence,

$$\frac{\partial \hat{CR}(1)}{\partial p_i} > \frac{\partial \hat{CR}(1)}{\partial p_j} \iff \frac{\partial \hat{PR}(1)}{\partial p_i} > \frac{\partial \hat{PR}(1)}{\partial p_j} \iff \frac{\partial \hat{PR}(2)}{\partial p_i} > \frac{\partial \hat{PR}(2)}{\partial p_j}, \quad \forall j \neq i.$$
Proof of Theorem 5.4

Follows immediately from (C.8), which implies that

$$\frac{\partial \hat{PR}}{\partial g_i} = \frac{\hat{PR}}{g_i}, \quad i \in I_{np}. \quad (C.29)$$

Justification of Numerical Fact 5.1

The justification has been carried out using Simulation Procedure A.1 and the 100,000 production lines constructed as mentioned in the performance analysis of Bernoulli lines with non-perfect quality machines. As a result, it has been determined that (5.14) implies and is implied by (5.7) in 96.2% and (5.15) implies and is implied by (5.8) in 95.8% of cases considered. Thus, we conclude that $\Delta \hat{PR}^l / \Delta p^l$ and $\Delta \hat{PR}^l / \Delta g^l$ can be used in stead of $\Delta PR^l / \Delta p^l$ and $\Delta PR^l / \Delta g^l$ for PR- and Q-BNs identification.

Proof of Theorem 5.5

Based on recursive procedure (C.15) and using the chain rule, we write

$$\frac{\partial PR}{\partial g^l_i} = \frac{\partial PR}{\partial q^l} \cdot \frac{\partial q^l}{\partial g^l_i} = \frac{\partial PR}{\partial q^l} \cdot \hat{q}^l_i,$$

where $q^l$ is the quality buy rate of the $l$-th inspection machine and $g^l_i$ is the quality parameter of the machines in the $l$-th Q-segment. Therefore, the LQ-BN of each segment is the machine with the smallest $g_i$.

Further, we prove by contradiction that one of the LQ-BNs is the Q-BN. Indeed, assume that none of the LQ-BNs is the Q-BN. Then, a non-perfect quality machine, other than a LQ-BN, is the Q-BN. This implies that in one of the Q-segments, this machine has a larger effect on the $\hat{PR}$ than its LQ-BN, which is a contradiction.
Justification of Numerical Fact 5.2

The justification has been carried out using the 100,000 production lines constructed as mentioned in the performance analysis of Bernoulli lines with non-perfect quality machines. Both simulation and calculation approaches have been used. In the simulation approach, \( ST_i \) and \( BL_i \) were estimated using Simulation Procedure A.1, LPR-BNs have been identified using Bottleneck Indicator 3.1 and PR-BN has been identified using (5.7). In the calculation approach, \( \hat{ST}_i \) and \( \hat{BL}_i \) were calculated using recursive procedure (C.15), \( \hat{LPR} \)-BNs have been identified using Bottleneck Indicator 3.1 and \( \hat{PR} \)-BN has been identified using (5.14). The results are as follows:

- In the simulation approach, LPR-BNs are identified correctly in 91.1% of all segments, and the PR-BN is one of the LPR-BNs in 93.9% of the cases.

- In the calculation approach, \( \hat{LPR} \)-BNs are identified correctly in 90.3% of all segments, and the \( \hat{PR} \)-BN is one of the \( \hat{LPR} \)-BNs in 94.5% of the cases.

Based on these data, we conclude that Numerical Fact 5.2 can be used for PR-BN identification.

Justification of Numerical Fact 5.3

The justification is carried out by calculating \( \hat{PR} \) for each possible position of the inspection module and determining whether there exists a \( p^* \) such that for all \( p_{insp} \geq p^* \), the stated fact takes place. Specifically, 100,000 of five- and 100,000 of ten-machine lines defined by assumptions (a)-(h) with \( |I_{insp}| = 0 \) have been investigated. The parameters \( p_i \), \( g_i \) and \( N_i \) have been selected randomly and equiprobably from sets (C.23)-(C.25). The position of the last non-perfect quality machine, \( m_{last} \), was selected randomly and equiprobably from the set \( \{1, 2, \ldots, M - 1\} \). Each of the machines upstream of \( m_{last} \) was chosen to be perfect and non-perfect with probability 1/2. For each of these lines, BNs have been identified using Bottleneck Indicator 3.1.
It turned out that all BNs were upstream of \( m_{\text{last}} \) in 42.7% of cases; in the remaining 57.3% of cases, one or more BNs were downstream of \( m_{\text{last}} \).

The inspection module consisted of a buffer and an inspection machine. The buffer capacity was selected as 2 or 3 with probability 1/2.

The production rate of each line, thus constructed, with the inspection module placed in all possible positions, was evaluated using recursive procedure (C.15) for \( p_{\text{insp}} \in [0.6, 1) \).

The results of this analysis are as follows: Numerical Fact 5.3 holds in 99.7% of cases when all BNs of the original line are upstream of \( m_{\text{last}} \) and in 97.2% of cases when one or more BNs are downstream of \( m_{\text{last}} \). Thus, we conclude that Numerical Fact 5.3 practically always takes place.

\[ \blacksquare \]

**Proof of Theorem 5.6**

Based on the results of [14], \( PR \) in a two-machine line defined by assumptions (a)-(h) can be expressed in the closed form as

\[
PR = p_2[1 - Q(p_1, p_2, N_1)]g_1(p_1)g_2(p_2)
= p_1[1 - Q(p_2, p_1, N_1)]g_1(p_1)g_2(p_2),
\]

where \( Q(\cdot) \) is defined in (C.5). Therefore,

\[
\frac{\partial PR}{\partial p_1} = p_2 \left( \frac{dg_1(p_1)}{dp_1} (1 - Q(p_1, p_2, N_1)) - \frac{\partial Q(p_1, p_2, N_1)}{\partial p_1} g_1(p_1) \right) g_2(p_2),
\]

\[
\frac{\partial PR}{\partial p_2} = p_1 \left( \frac{dg_2(p_2)}{dp_2} (1 - Q(p_2, p_1, N_1)) - \frac{\partial Q(p_2, p_1, N_1)}{\partial p_2} g_2(p_2) \right) g_1(p_1),
\]

where

\[
\frac{\partial Q(x, y, N)}{\partial x} = \begin{cases} 
\frac{y(1-x)(N(x,y)-1)+(y-x)Nd^N(x,y)}{(y-xN(x,y))^{2-N}(1-x)}, & \text{if } x \neq y, \\
-\frac{N(N+1)}{2p(N+1-p)^2}, & \text{if } x = y = p.
\end{cases}
\]

Clearly, \( \frac{\partial PR}{\partial p_1} \) and \( \frac{\partial PR}{\partial p_2} \) are continuous with respect to \( p_1 \) and \( p_2 \), respectively.
Assume that \( \{1\} \subseteq I_{QQC} \), then

\[
\lim_{p_1 \to 0} \frac{\partial PR}{\partial p_1} = g_1(0)g_2(p_2) > 0. \tag{C.34}
\]

For \( N_1 \geq 2 \), it is possible to show that

\[
\lim_{p_1 \to 1} \frac{\partial PR}{\partial p_1} = p_2 g_2(p_2) \left. \frac{dg_1(p_1)}{dp_1} \right|_{p_1=1} < 0. \tag{C.35}
\]

Therefore, there exist \( p_1^*, p_1^{**} \in (0, 1) \), such that

\[
\frac{\partial PR}{\partial p_1} > 0, \quad \forall 0 < p_1 < p_1^*;
\]
\[
\frac{\partial PR}{\partial p_1} < 0, \quad \forall p_1^{**} < p_1 < 1.
\]

Similar arguments in the case of \( \{2\} \subseteq I_{QQC} \) lead to

\[
\frac{\partial PR}{\partial p_2} > 0, \quad \forall 0 < p_2 < p_2^*;
\]
\[
\frac{\partial PR}{\partial p_2} < 0, \quad \forall p_2^{**} < p_2 < 1.
\]

\[\blacksquare\]

**Justification of Numerical Fact 5.4**

We constructed 500,000 lines by selecting \( M, p_i \)'s and \( N_i \)'s randomly and equiprobably from the following sets, respectively:

\[
M \in \{3, 5, 10\}, \tag{C.36}
\]
\[
p_i \in [0.7, 0.95], \tag{C.37}
\]
\[
N_i \in \{2, 3, 4, 5\}. \tag{C.38}
\]

Each machine was chosen to be of perfect quality, non-perfect without QQC or non-perfect with QQC with probability 1/3. For non-perfect machines without QQC, the
machine quality parameter $g_i$ was selected from

$$g_i \in [0.7, 1). \quad (C.39)$$

For each QQC machine, $g_i(p_i)$ was selected randomly and equiprobably from the three sets of functions shown in Figure C.1 (i.e., linear, concave and convex) and then randomly and equiprobably within each set.

![Figure C.1: QQC functions for the justification of Numerical Fact 5.4](image)

The production rate was evaluated using Recursive Procedure C.1. Among all the lines, thus constructed, $\hat{PR}(p_i)$, $i \in I_{QQC}$, was always concave and no counterexamples of Numerical Fact 5.4 were found.

**Justification of Numerical Fact 5.5**

The justification has been carried out using the 500,000 production lines constructed as mentioned in Subsection 5.3.2. For each line, $\hat{PR}$ was calculated using Recursive Procedure C.1 and $PR$ was evaluated using Simulation Procedure A.1.

As a result, it has been determined that (5.24) implies and is implied by (5.26), and (5.25) implies and is implied by (5.27) in 93.7% and 95.8% of cases considered, respectively. Thus, we conclude that (5.26) and (5.27) can be used instead of (5.24) and (5.25) for BN identification.
To investigate system-theoretic properties of Bernoulli lines with rework, we develop a method to evaluate the performance measures of such systems based on overlapping decomposition:

**Performance evaluation of Bernoulli lines with rework**

As mentioned in Subsection 5.4.1, a Bernoulli line with rework can be reduced to four serial lines using overlapping decomposition. To calculate the efficiencies of the virtual machines in these serial lines, we use the Recursive Procedure C.1, the performance measures $\hat{PR}$, given by (C.8) with $q_M = 1$, and $\hat{BL}$ and $\hat{ST}$, given by (C.12) and (C.13). In addition, we need the probability that in a serial line with $M$ machines the first buffer is full while the second machine is either blocked or down and the probability that the last buffer is empty; we denote these probabilities as $\hat{bl}_1$ and $\hat{st}_M$, respectively. As it follows from (C.12) and (C.12), they can be evaluated as

\[
\hat{bl}_1 = \frac{\hat{BL}_1}{p_1} = Q(p_2, p_1, N_1), \tag{C.40}
\]

and

\[
\hat{st}_M = \frac{\hat{ST}_M}{p_M} = Q(p_{M-1}, p_M, N_{M-1}). \tag{C.41}
\]

**Recursive aggregation procedure for overlapping decomposition:** Consider the virtual lines of Figure 5.20(b) and introduce the following notations:

\[
P R_l = \text{Production rate of virtual line } l, l = 1, 2, 3, 4,
\]

\[
BL_l^i = P[\text{machine } m_i \text{ in virtual line } l \text{ is blocked}],
\]

\[
l = 1, 2, 3, 4,
\]

\[
ST_l^i = P[\text{machine } m_i \text{ in virtual line } l \text{ is starved}],
\]

\[
l = 1, 2, 3, 4,
\]

\[
bl_j = P[b_j \text{ is full and } m_{j+1} \text{ is either down or blocked}],
\]

\[
st_{j_1} = P[b_{j-1} \text{ is empty}]\quad \text{and} \quad st_{j_2} = P[b_{r_M} \text{ is empty}],
\]
\[ \text{blk}_1 = P[b_k \text{ is full and } m_{k+1} \text{ is either down or blocked}], \]
\[ \text{blk}_2 = P[b_{r0} \text{ is full and } m_{r1} \text{ is either down or blocked}], \]
\[ \text{st}_k = P[b_{k-1} \text{ is empty}]. \]

The estimates of these probabilities, which allow us to evaluate the parameters \( p^1_j, p^2_j, p^3_j \) and \( p^2_k, p^3_k, p^4_k \) of the virtual machines \( m^1_j, m^2_j, m^4_j \) and \( m^2_k, m^3_k, m^4_k \), can be calculated using a recursive procedure described below.

**Recursive Procedure C.3:**

**Step 0:** Select the initial conditions \( \hat{\text{st}}_k(0), \hat{\text{bl}}_j(0), \) and \( \hat{\text{bl}}_{k1}(0) \) randomly and equiprobably from the interval \((0,1)\).

**Step 1:** Consider virtual line 4 of Figure 5.20(b) and update the efficiencies of machines \( m^4_j \) and \( m^4_k \) as follows:

\[ p^4_k(n+1) = p_k(1-q)[1-\hat{\text{st}}_k(n)][1-\hat{\text{bl}}_{k1}(n)], \]
\[ p^4_j(n+1) = p_j[1-\hat{\text{bl}}_j(n)], \]
\[ n = 0, 1, \ldots. \]

Then perform Recursive Procedure C.1 on virtual line 4 and, using (C.8), (C.12), (C.13), (C.40), and (C.41), calculate \( \hat{\text{PR}}_4(n+1), \hat{\text{ST}}_4^4(n+1), \hat{\text{BL}}_4^4(n+1), \hat{\text{bl}}_{k2}(n+1), \) and \( \hat{\text{st}}_{j2}(n+1) \).

**Step 2:** Consider virtual line 3 of Figure 5.20(b) and update the efficiency of machine \( m^3_k \) as follows:

\[ p^3_k(n+1) = p_k q[1-\hat{\text{st}}_k(n)][1-\hat{\text{bl}}_{k2}(n+1)], \]
\[ n = 0, 1, \ldots. \]

Then perform Recursive Procedure C.1 on virtual line 3 and, using (C.8), (C.12), (C.13), and (C.40), calculate \( \hat{\text{PR}}_3(n+1), \hat{\text{ST}}_3^3(n+1), \hat{\text{BL}}_3^3(n+1), \) and \( \hat{\text{bl}}_{k1}(n+1) \).
Step 3: Consider virtual line 1 of Figure 5.20(b) and update the efficiency of machine $m_j^1$ as follows:

$$p_j^1(n+1) = p_j[1 - \hat{bl}_j(n)]\hat{st}_{j2}(n+1),$$

$$n = 0, 1, \ldots.$$  

Then perform Recursive Procedure C.1 on virtual line 1 and, using (C.8), (C.12), (C.13), and (C.41), calculate $\hat{PR}_1(n+1)$, $\hat{ST}_i^1(n+1)$, $\hat{BL}_i^1(n+1)$, and $\hat{st}_{j1}(n+1)$.

Step 4: Consider virtual line 2 of Figure 5.20(b) and update the efficiencies of machines $m_j^2$ and $m_k^2$ as follows:

$$p_j^2(n+1) = p_j[1 - \hat{st}_{j1}(n+1)\hat{st}_{j2}(n+1)],$$

$$p_k^2(n+1) = p_k[1 - \hat{bl}_{k1}(n+1)][1 - \hat{bl}_{k2}(n+1)],$$

$$n = 0, 1, \ldots.$$  

Then perform Recursive Procedure C.1 on virtual line 2 and using (C.8), (C.12), (C.13), (C.40), and (C.41), calculate $\hat{PR}_2(n+1)$, $\hat{ST}_i^2(n+1)$, $\hat{BL}_i^2(n+1)$, $\hat{bl}_j(n+1)$, and $\hat{st}_k(n+1)$.

Step 5: If the stopping rule

$$\sum_{l=1}^{4} |\hat{PR}_l(n+1) - \hat{PR}_l(n)| < \varepsilon, \quad \varepsilon \ll 1,$$

is satisfied, the procedure is terminated; otherwise return to Step 1.

Denote the limits of this procedure as

$$\hat{PR}_l := \lim_{n \to \infty} \hat{PR}_l(n), \quad l = 1, 2, 3, 4,$$

$$\hat{bl}_j := \lim_{n \to \infty} \hat{bl}_j(n),$$

$$\hat{st}_{j1} := \lim_{n \to \infty} \hat{st}_{j1}(n), \quad \hat{st}_{j2} := \lim_{n \to \infty} \hat{st}_{j2}(n),$$

$$\hat{st}_k := \lim_{n \to \infty} \hat{st}_k(n),$$

(C.43)
\[
\hat{b}_{k_1} := \lim_{n \to \infty} \hat{b}_{k_1}(n), \quad \hat{b}_{k_2} := \lim_{n \to \infty} \hat{b}_{k_2}(n).
\]

Unfortunately, the existence of, and convergence to, these limits cannot be proved analytically (due to a non-monotonic behavior of \(\hat{b}_{k_2}(n), n = 0, 1, \ldots\)). Therefore, it has been investigated numerically: A total of 5,000,000 lines have been generated by selecting \(j, k, M, M_r, p_i\)'s, \(N_i\)'s, and \(q\) randomly and equiprobably from the sets:

\[
j \in \{2, 3, 4, 5\}, \quad (C.44)
\]

\[
k - j \in \{2, 3, 4, 5\}, \quad (C.45)
\]

\[
M - k \in \{1, 2, 3, 4\}, \quad (C.46)
\]

\[
M_r \in \{1, 2, 3, 4\}, \quad (C.47)
\]

\[
p_i \in [0.7, 0.95], \quad i = 1, \ldots, M, \quad (C.48)
\]

\[
p_i \in [0.1, 0.5], \quad i = r1, \ldots, rM_r, \quad (C.49)
\]

\[
N_i \in \{1, 2, 3, 4, 5\}, \quad i \in I_b, \quad (C.50)
\]

\[
q \in [0.7, 0.95]. \quad (C.51)
\]

Note that the efficiencies of the machines in the rework loop are selected lower than those of the main line because, in practice, the capacity of the repair part of the system is typically smaller than that of the main line.

For each of the lines, we ran Recursive Procedure C.1 (with \(\varepsilon = 10^{-10}\) in (C.42)) and observed the convergence taking place within a second using a standard laptop. Thus, we conclude that this procedure can be used for analysis of production lines with rework defined by assumptions (a)-(k).

Performance measure estimates and their accuracy: Based on the limits (C.43) of Recursive Procedure C.1, introduce:

\[
\hat{P}_{R_{lwr}} = \hat{P}_{R_3}, \quad (C.52)
\]

\[
\hat{S}_{T_{i\downarrow}}^{lwr} = \hat{S}_{T_{i\downarrow}}, \quad \hat{B}_{L_{i\downarrow}}^{lwr} = \hat{B}_{L_{i\downarrow}}, \quad i \neq j, k; \ l = 1, 2, 3, 4, \quad (C.53)
\]
\[
\begin{align*}
\hat{ST}_{j_1} &= \hat{s}_{j_1}p_j, \quad \hat{ST}_{j_2} = \hat{s}_{j_2}p_j, \quad \hat{BL}_{j} = \hat{b}_{j}p_j, \quad (C.54) \\
\hat{ST}_{k} &= \hat{s}_{k}p_k, \quad \hat{BL}_{k_1} = \hat{b}_{k_1}p_k, \quad \hat{BL}_{k_2} = \hat{b}_{k_2}p_k. \quad (C.55)
\end{align*}
\]

The accuracy of these estimates has been investigated using a C++ code that simulates the system at hand and Simulation Procedure A.1. Specifically, for the system parameters \(M = 10, j = 4, k = 7, \) and \(M_r = 2,\) we constructed 100,000 lines with \(p_i's, \) \(N_i's\) and \(q's\) selected randomly and equiprobably from sets \((C.48)-(C.51).\)

The following metrics were used to evaluate the accuracy of the estimates:

\[\epsilon_{PR} = \frac{|PR_{lw} - \hat{PR}_{lw}|}{PR_{lw}} \cdot 100\%, \quad (C.56)\]

\[\epsilon_{ST} = \frac{1}{M + M_r} \sum_{i \in S_1} |ST_{lw}^i - \hat{ST}_{lw}^i|, \quad (C.57)\]
\[S_1 = \{1, \ldots, j - 1, j_1, j_2, j + 1, \ldots, rM_r\},\]

\[\epsilon_{BL} = \frac{1}{M + M_r} \sum_{i \in S_2} |BL_{lw}^i - \hat{BL}_{lw}^i|, \quad (C.58)\]
\[S_2 = \{1, \ldots, k - 1, k_1, k_2, k + 1, \ldots, rM_r\}.\]

Among the 100,000 lines studied, the average of \(\epsilon_{PR}\) was 3.97%, with very few extreme cases resulting in \(\epsilon_{PR}\) up to 20.4%. The averages of \(\epsilon_{ST}\) and \(\epsilon_{BL}\) were both less than 0.01. Therefore, we conclude that Recursive Procedure 5.1 and \((C.52)-(C.55)\) provide an effective tool for performance evaluation of serial lines with rework defined by assumptions (a)-(k).

\[\blacksquare\]

**Justification of Numerical Fact 5.6**

To investigate the monotonicity property of \(PR_{lw},\) the same approach used in the justification of Numerical Fact 3.2 was applied to the 100,000 lines generated above. The production rate was evaluated using Recursive Procedure C.3. Among all cases, no counterexamples were found. Thus, we conclude that Numerical Fact 5.6 holds.

\[\blacksquare\]

256
Justification of Numerical Facts 5.7-5.9

The justification of Numerical Facts 5.7-5.9 has been carried out as follows: A total of 100,000 lines have been generated with $M = 10$, $j = 4$, $k = 7$, and $M_r = 2$ and parameters of machines and buffers selected randomly and equiprobably from sets (C.48)-(C.51). For the calculation-based approach, each of these lines has been analyzed using Recursive Procedure C.3 and the LBNs have been identified by calculating blockages and starvations using (C.53)-(C.55). Then, the partial derivatives have been estimated using

$$\frac{\partial \hat{PR}_{lwr}}{\partial p_i} \approx \frac{\Delta \hat{PR}_{lwr}}{\Delta p_i},$$

with $\Delta p_i = 0.001$. For the measurement-based approach, each of these lines has been analyzed using Simulation Procedure A.1 and the partial derivatives have been estimated via simulation using

$$\frac{\partial PR_{lwr}}{\partial p_i} \approx \frac{\Delta PR_{lwr}}{\Delta p_i},$$

with $\Delta p_i = 0.01$.

Based on this procedure, the following results, quantifying the term “practically always,” have been obtained:

- Numerical Fact 5.7 holds in 97.7% of cases using calculations and 93.3% of cases using measurements.

- Numerical Fact 5.8:
  - Claim ($\alpha$) holds in 88.1% of cases using calculations and 86.9% of cases using measurements.
  - Claim ($\beta$) holds in 95.3% of cases using calculations and 94.9% of cases using measurements.

- Numerical Fact 5.9:
– Claim (α) holds in 91.3% of cases using calculations and 92.2% of cases using measurements.

– Claim (β) holds in 94.6% of cases using calculations and 94.0% of cases using measurements.

– Claim (γ) holds in 96.7% of cases using calculations and 95.4% of cases using measurements.

– Claim (δ) holds in 97.0% of cases using calculations and 97.2% of cases using measurements.

Based on these data, we conclude that Numerical Facts 5.7-5.9 indeed take place.

■
APPENDIX D

PROOFS AND JUSTIFICATIONS FOR CHAPTER VI

Justification of Numerical Fact 6.1

The justification was carried out by enumerating \((p, N)\) from the set

\[
p \in \{0.001, 0.002, \ldots, 0.999\}, \quad (D.1)
\]

\[
N \in \{2, 3, \ldots, 10\}. \quad (D.2)
\]

For each \((p, N)\) pair, \(\lambda_1(p, N)\) was evaluated numerically using MATLAB. For all cases studied, no counterexamples were found. Thus, we conclude that Numerical Fact 6.1 holds.

\[\blacksquare\]

Proof of Theorem 6.1

The similarity transformation for (6.1) with \(N = 1\) is

\[
Q = \begin{bmatrix}
1 & 1 \\
-\frac{1}{\sqrt{2+p^2-2p}} & \frac{1-p}{\sqrt{2+p^2-2p}}
\end{bmatrix}. \quad (D.3)
\]

This leads to

\[
\tilde{C} = CQ^{-1} = \frac{1}{2-p} \begin{bmatrix}
p & p \sqrt{2+p^2-2p} \\
1 & \sqrt{2+p^2-2p}
\end{bmatrix}. \quad (D.4)
\]
\[ PR(n) = \frac{p}{2-p} + \frac{p \sqrt{2 + p^2 - 2p}}{2-p} \bar{x}_1(0) \lambda_1^n, \]  
\[ WIP(n) = \frac{1}{2-p} + \frac{\sqrt{2 + p^2 - 2p}}{2-p} \bar{x}_1(0) \lambda_1^n \]  

and, therefore, to (6.30).

---

**Justification of Numerical Fact 6.2**

The justification was carried out using the same approach as the justification of Numerical Fact 6.1. For each \((p, N)\) pair, \(\Psi_{11}\) and \(\Psi_{21}\) were calculated using the symbolic manipulation function of MATLAB. For all lines studied, no counterexamples were found. Thus, we conclude that Numerical Fact 6.2 holds.

---

**Proof of Theorem 6.2**

Expression (6.45), (6.46) follow directly from (6.35), (6.41) by evaluating the time necessary for \(\widehat{PR}(n)\) and \(\widehat{WIP}(n)\) to reach 95% of their steady state values. To prove (6.47) and (6.48), we observe that

\[ \hat{t}_{sWIP} - \hat{t}_{sPR} = 1 - \frac{\ln(20\beta)}{\ln(\lambda_1)} - 1 + \frac{\ln(20\gamma)}{\ln(\lambda_1)} = \frac{\ln(\gamma/\beta)}{\ln(\lambda_1)}, \]  

and

\[ \frac{\gamma}{\beta} = \frac{1 - \frac{p(N+1-p)}{N}}{1 - \frac{p(N+1-p)}{N}} \cdot \frac{2}{N+1} \geq 1, \quad 0 < N < \infty, \quad 0 < p < 1. \]  

---

**Justification of Numerical Fact 6.3**

It can be seen from Figure 6.14 that \(\Lambda_T^{[N/2]}(p, N, M) \leq 0\) for all cases with \(N = 5\) and \(N = 10\). Although \(\Lambda_T^{[N/2]}(p, N, M) > 0\) for \(N = 2\), the maximum loss is less than 1%. Thus, we conclude that Numerical Fact 6.3 holds.
Justification of Numerical Fact 6.4: To justify this numerical fact, a total of 10,000 pairs of \((p^*, N)\) were generated by randomly and equiprobably selecting these parameters from

\[
\begin{align*}
p^* & \in (0.25, 1) \quad \text{(D.9)} \\
N & \in \{3, 4, \ldots, 50\}. \quad \text{(D.10)}
\end{align*}
\]

For each \((p^*, N)\), \(\lambda_1\) was evaluated as a function of \(p_2\). Among all systems considered, no counterexamples were found. Thus, we conclude that Numerical Fact 6.4 holds.

Justification of Numerical Fact 6.5: The same 10,000 pairs of \((p^*, N)\) generated in the justification of Numerical Fact 6.4 were used. PEFs \(\Psi_{11}(0)\) and \(\Psi_{21}(0)\) were evaluated as functions of machine efficiency. Among all systems studied, no counterexamples were found. Thus, we conclude that Numerical Fact 6.5 holds.

Justification of Numerical Fact 6.6: For \(N \leq 30\), the justification was carried out analytically by using the symbolic derivation function of MATLAB® and it was shown that the transition matrices of \(L\) and \(L_r\) have the same characteristic polynomial and, thus, the same set of eigenvalues.

For \(N > 30\), the justification was carried out numerically by calculating and comparing the eigenvalues of the transition probability matrices for both lines. Specifically, we generated a total of 50,000 production lines by selecting \(p_i\) and \(N\) randomly and equiprobably from the following sets:

\[
\begin{align*}
p_i & \in (0.5, 1), \quad i = 1, 2, \quad \text{(D.11)} \\
N & \in \{31, 32, \ldots, 50\}. \quad \text{(D.12)}
\end{align*}
\]
As a result, for all lines investigated, the maximum difference between the eigenvalues of $L$ and $L_r$ was less than $10^{-6}$. Therefore, we conclude that the reversed line has the same eigenvalues as the original line.

**Justification of Numerical Fact 6.7:** Since both lines have the same set of eigenvalues, $\{1, \lambda_1, \ldots, \lambda_N\}$, the production rate of the lines can be expressed as follows:

$$PR^L(n) = \tilde{C}_{10} + \tilde{C}_{11}\tilde{x}_1(0)\lambda_1^n + \cdots + \tilde{C}_{1N}\tilde{x}_N(0)\lambda_N^n,$$

(D.13)

$$PR^{L_r}(n) = \tilde{C}_{r10} + \tilde{C}_{r11}\tilde{x}_1^r(0)\lambda_1^n + \cdots + \tilde{C}_{r1N}\tilde{x}_N^r(0)\lambda_N^n.$$

(D.14)

For $N \leq 5$, using the symbolic derivation function of MATLAB®, it was shown that

$$\tilde{C}_{1i}\tilde{x}_i^r(0) = \tilde{C}_{r1i}\tilde{x}_i(0), \quad i = 1, \ldots, N, \quad \forall p_1, p_2 \in (0, 1).$$

(D.15)

For $N > 5$, we generated 50,000 production lines by randomly and equiprobably selecting $p_i$ and $N$ from

$$p_i \in (0.5, 1), \quad i = 1, 2,$$

(D.16)

$$N \in \{6, 7, \ldots, 50\}.$$

(D.17)

Then, $\tilde{C}_{1i}\tilde{x}_i(0)$ and $\tilde{C}_{r1i}\tilde{x}_i^r(0)$, $i = 0, \ldots, N$, were evaluated by using (6.12)-(6.14). As a result, among all lines considered, the maximum difference between $\tilde{C}_{1i}\tilde{x}_i(0)$ and $\tilde{C}_{r1i}\tilde{x}_i^r(0)$ was less than $10^{-4}$. Therefore, we conclude that $PR$ has the same pre-exponential factors in $L$ and $L_r$ under initial condition (6.64).

**Justification of Numerical Fact 6.8:** We used the 50,000 lines generated in the justification of Numerical Fact 6.7. For each line, $WIP^L(n)$ and $WIP^{L_r}(n)$ were
calculated using (6.4). To evaluate the accuracy of (6.67), define
\[ \epsilon = \max_{n=1,2,...} \frac{|WIP_L(n) + WIP_{Lr}(n) - (WIP_{ss}^L + WIP_{ss}^{Lr})|}{N}, \]
where \( WIP_{ss}^L \) and \( WIP_{ss}^{Lr} \), calculated from [17], denote the steady state values of the work-in-process for \( L \) and \( L_r \), respectively. As a result, among the 50,000 lines studied, the average \( \epsilon \) was 0.008 with maximum 0.029. Thus, we conclude that Numerical Fact 6.8 indeed takes place.

\[ \blacksquare \]

**Proof of Theorem 6.3:** It follows from (6.70) that
\[ \lambda_1^u = 1 - \frac{P}{1 + \alpha} - R, \quad \lambda_1^d = 1 - P - (1 + \alpha)R. \]  
(D.19)
Solving inequalities \(|\lambda_1^u| - |\lambda_1^d| > 0\) and \(|\lambda_1^u| - |\lambda_1^d| < 0\) results in
\[ |\lambda_1^u| - |\lambda_1^d| > 0, \text{ if } (1 + \frac{\alpha}{2})R < e', \quad |\lambda_1^u| - |\lambda_1^d| < 0, \text{ if } (1 + \frac{\alpha}{2})R > e'. \]

It follows immediately from (6.76) that
\[ (1 + \frac{\alpha}{2})R < (1 + \alpha)R < 0.5 < e < e'. \]
Thus, under condition (6.76), \(|\lambda_1^u| > |\lambda_1^d|\).
\[ \blacksquare \]

**Proof of Theorem 6.4:** Since the algebraic and geometric multiplicities of every eigenvalue of matrix \( A \), defined in (6.81), are equal to each other, there exists a nonsingular matrix \( Q \) such that
\[ A = Q^{-1} \tilde{A} Q, \quad \tilde{A} = \text{diag}[1 \quad \lambda_b \quad \lambda_m \quad \lambda_m^2 \quad 0 \quad 0 \quad 0]. \]
Thus,

\[ x(n) = Ax(n - 1) = Q^{-1} \tilde{A}Qx(n - 1) = Q^{-1} \tilde{A}_n Qx(0), \]

where

\[ \tilde{A}_n = \text{diag}[1 \quad \lambda_b^n \quad \lambda_m^n \quad (\lambda_m^2)^n \quad 0 \quad 0 \quad 0]. \]

Hence, the evolution of the states can be expressed as

\[ x_{h,i,j}(n) = x_{h,i,j}[1 + \tilde{B}\tilde{x}_2(0)\lambda_b^n + (\tilde{C}_1\tilde{x}_3(0) + \tilde{C}_2\tilde{x}_4(0))\lambda_m^n + \tilde{D}\tilde{x}_5(0)(\lambda_m^2)^n]. \]

\[ h \in \{0, 1\}, \quad i, j \in \{0, 1\}, \quad n = 1, 2, \ldots, \]  

(D.20)

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

Then, it follows from (6.81) that

\[ C = \tilde{C}_1\tilde{x}_3(0) + \tilde{C}_2\tilde{x}_4(0), \quad D = \tilde{D}\tilde{x}_5(0). \]  

(D.21)

For matrix \( Q \), it can be shown that

\[
\begin{bmatrix}
q_3 \\
q_4 \\
q_5 \\
\end{bmatrix}
= \frac{P^2}{(-R + P + R^2)(R + P)^2}.
\]

\[
\begin{bmatrix}
R^2 & -RP & R^2 & -RP & R^2 & -RP & R^2 & -RP \\
R & R & -P & -P & R & R & -P & -P \\
-R^3 & -R^3 & R^2 & R^2 & -R & -R & R^2 & R^2 \\
\end{bmatrix}
\]

Moreover, initial condition (6.82) implies that

\[
\sum_{h,j} x_{h,1,j}(0) = \sum_{h,i} x_{h,i,1}(0) = e, \quad \sum_{h,j} x_{h,0,j}(0) = \sum_{h,i} x_{h,i,0}(0) = 1 - e.
\]
In addition, since \( m_1 \) and \( m_2 \) are independent,

\[
\sum_{h,i \neq j} x_{h,i,j}(0) = 2e(1 - e), \quad \sum_{h} x_{h,0,0}(0) = (1 - e)^2, \quad \sum_{h} x_{h,1,1}(0) = e^2.
\]

Thus, under (6.82),

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

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

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

Therefore, due to (D.21), \( C = D = 0 \).

\[\square\]

**Proof of Theorem 6.5:** Since \(|1 - P - R|\) and \((1 - R)^2\) are both monotonically decreasing functions of \( R \) on \((0, 0.5)\) for a fixed \( e \), the SLE of the system is a monotonically decreasing function of \( R \) on \((0, 0.5)\).

\[\square\]

**Proof of Theorem 6.6:** It follows from Theorem 6.3 that \(|\lambda_m^u| > |\lambda_m^d|\). In addition, \(\lambda_b^u = (1 - R)^2 > [1 - (1 + \alpha)R]^2 = \lambda_b^d\). Thus, \(|\lambda_1^u| = \max(|\lambda_m^u|, \lambda_b^u) > \max(|\lambda_m^d|, \lambda_b^d) = |\lambda_1^d|\).

\[\square\]

**Proof of Theorem 6.7:** Let \( \lambda \) be an eigenvalue of \( A_M \), with left eigenvector \( v = (v(s))_{s \in S} \). Thus,

\[
(vA_M)(s) = \sum_{s' \in S} v(s')p(s', s) = \lambda v(s), \quad s \in S.
\]

Consider now the vector \( w = (w(h, s))_{h \in H, s \in S} \) with \( w(h, s) = v(s) \) for all \( h \) and \( s \); that is, \( w \) is a row vector indexed by the elements of the set \( H \times S \), and the value of
its \((h, s)\) component equals \(v(s)\), independently of \(h\). Then, using (6.85), we obtain that

\[
(wA)(s, h) = \sum_{h', s'} w(h', s')p(h', s'|h, s) = \sum_{h', s'} v(s')p(s'|s)p(h'|h, s) \\
= \lambda v(s) \sum_{h'} p(h'|h, s) = \lambda v(s) \equiv \lambda w(s, h)
\]

for every \(s\) and \(h\). Hence \(w\) is a left eigenvector of \(A\) with eigenvalue \(\lambda\). This establishes that any eigenvalue of the matrix \(A_M\) is an eigenvalue of the matrix \(A\) as well, as claimed.

\[
\text{Justification of Numerical Fact 6.9:} \quad \text{This justification is carried out by evaluating the settling time of production rate, } t_{sPR}, \text{ which is the time necessary for } PR \text{ to reach and remain within } \pm 5\% \text{ of its steady state value, provided that the buffer is initially empty. A total of 10,000 lines were generated with } e \text{ and } N \text{ randomly and equiprobably selected from the sets (6.89), respectively. For each line, thus constructed, } t_{sPR} \text{ was evaluated using approximation (6.88) as a function of } R. \text{ As a result, we obtained that } t_{sPR} \text{ is a monotonically decreasing function of } R \text{ on } R \in (0, 0.5) \text{ in 99\% of all cases studied. Thus, we conclude that shorter up- and downtimes lead, practically always, to faster transients, i.e., Numerical Fact 6.9 holds.}
\]

\[
\text{Justification of Numerical Fact 6.10:} \quad \text{To justify this numerical fact, the 50,000 lines generated as mentioned in Subsection 6.4.4 were used to investigate the effects of increasing uptime or decreasing downtime on } t_{sPR}. \text{ To accomplish this, we selected } \alpha \text{ randomly and equiprobably from the set } \alpha \in [0.05, 1] \text{ and evaluated the settling times } t_{sPR}^u \text{ and } t_{sPR}^d, \text{ resulting from increasing uptime by } (1 + \alpha) \text{ and decreasing downtime by } (1 + \alpha), \text{ respectively. It turned out that } t_{sPR}^u \text{ was longer than } t_{sPR}^d \text{ in 96.12\% of}
\]
all cases studied. For the remaining 3.88% of cases, $t_{sPR}^u$ was shorter than $t_{sPR}^d$ by at most 1 cycle time. Therefore, we conclude that Numerical Fact 6.10 takes place.
BIBLIOGRAPHY
BIBLIOGRAPHY


274