OPERATIONS RESEARCH CENTER

Working Paper

Modeling and Analysis of Markovian Continuous Flow Production Systems with a Finite Buffer: A General Methodology and Applications

by

B. Tan
S. B. Gershwin

OR 381-07 **March 2007**

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Modeling and Analysis of Markovian Continuous Flow Production Systems with a Finite Buffer: Methodology and Applications

Barış Tan
Graduate School of Business
Koç University
Rumeli Feneri Yolu, Sarıyer
Istanbul, Turkey
btan@ku.edu.tr

Stanley B. Gershwin
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307
USA
gershwin@mit.edu

March 14, 2007

Abstract

Fluid flow models have been used in performance evaluation of production, computer, and telecommunication systems. Currently, a general methodology to model and analyze a given continuous flow production system is not available. In order to develop a general methodology to analyze Markovian continuous material flow production systems with a finite buffer, we modelled the fluid flow system as a continuous time, continuous-discrete state space stochastic process and determined the steady state distributions by using a level crossing analysis. Various performance measures such as the production rate and the expected buffer level are also determined directly from the steady-state distributions. The flexibility of our methodology allows us to analyze a wide range of models including models with machines that have multiple up and down states, models with multiple unreliable machines in series or parallel in each stage, models with merge-type structures, and models with phase-type failure and repair-time distributions by using the same methodology. Therefore the method is proposed as a general tool for performance evaluation of continuous material flow production systems with a single buffer.

1 Introduction

In this study, we consider a two-stage continuous flow system separated by a finite capacity buffer (Figure 1). The dynamics of each stage is described by a continuous-time Markov chain where each state is associated with a different flow rate. In our setting, the definition of a stage is very general:

it can be a single machine with an arbitrary number of states, a number of machines in series or parallel, etc.

There is a vast literature on continuous material flow models of unreliable production lines e.g. (Wijngaard 1979), (Gershwin and Schick 1980), (Dubois and Forestier 1982), (Yeralan, Franck, and Quasem 1986), (Yeralan and Tan 1997) among others. In all of these models, each machine has two states: a single up state that represents the condition of a fully productive machine and a single down state that represent the condition where the machine is not productive due to a failure and the failure and repair times are exponential random variables.

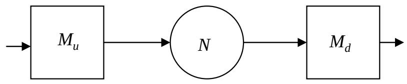

Figure 1: A Single Buffer Fluid Flow System with Two Stages

Models of unreliable production lines have also been extended to analyze various production systems. More detailed models of production systems where each stage is modelled by using more than two states have been used to approximate general processing, failure, and repair time distributions by using phase-type distributions (Altıok 1985), (Dallery 1994); to study quality-quantity interactions (Tempelmeier and Burger 2001), (Inman, Blumenfeld, Huang, and Li 2003), (Kim and Gershwin 2005), (Poffe and Gershwin 2005); or to develop new approximation methods with multiple up and down states (Levantesi, Matta, and Tolio 2003). Similarly, analysis of production lines with series or parallel structures (Mitra 1988), (Patchong and Willaeys 2001) or merge structures (Tan 2001), (Helber and Jusic 2004), (Diamantidis, Papadopoulos, and Vidalis 2004) also received attention.

Although, a variety of models are used to evaluate the performance of continuous flow production systems, currently there exists no unified methodology to analyze these systems. For discrete production lines with single buffer and identical processing rates, (Gershwin and Fallah-Fini 2007) recently proposed a general method. In the analysis of continuous flow models, once the state space is determined based on the underlying assumptions, the steady-state distribution is determined by analyzing the continuous time-continuous and discrete state space Markov process. In order to analyze this process, a set of differential equations that describe the behavior of the system is derived and then solved subject to boundary and normalization conditions. Without a general methodology, considerable effort is required to model and to analyze a given system. This study is motivated by the need to develop a unified methodology to analyze all Markovian single-buffer continuous-flow production systems.

Fluid flow models with a single buffer are also used to evaluate the performance of computer and telecommunication systems, e.g. (Anick, Mitra, and Sondhi 1982) and (Elwalid and Mitra 1991). There are general methodologies to analyze fluid flow models that appear in computer and telecommunication systems, e.g. (Serucola 2001), (Ahn and Ramaswami 2003) (Ahn, Jeon, and Ramaswami 2005), (Soares and Latouche 2006). Although the fluid flow models developed for production and computer/telecommunication systems are similar, the methods developed for telecommunication and computer systems cannot be used to analyze production systems directly due to operation dependent failures that are observed in production systems. When the failures are operation dependent, an idle machine due to being blocked or starved cannot fail. If it is partially blocked or partially starved and operating at a reduced rate, its failure rate will be lower than its rate when the buffer is partially full. As a result, the boundary processes are not the same as the partially-full buffer process and all three processes must be analyzed accordingly.

In this paper, we present a general methodology to analyze continuous flow production systems with a finite buffer. The model utilizes a level crossing analysis and links probabilities of entering boundary processes in specific states when the buffer is empty or full with the probabilities of exiting these states in specific states. The boundary processes give the conditional probabilities that link enter and exit probabilities. The inputs of the model are the transition rates of each stage and the processing rates associated with the discrete states of each stage. Therefore our model is quite general and allows one to analyze a wide range of models by determining the required inputs. We show examples of how different models analyzed can be analyzed directly by using our methodology.

The organization of the remaining part of the manuscript is as follows: In Section 2, we describe the model, its assumptions, and introduce the variables used in the model. In Section 3, we present our general methodology to analyze the Markov process. In Section 4, a number of performance measures are derived by using the solution given in Section 3. We explain the methodology in detail by modelling and analyzing a specific system in Section 5. Application of the proposed methodology in analysis of various production lines is illustrated with three examples in Section 6. Finally, conclusions are given in Section 7.

2 Model

2.1 Model Description

We consider a continuous material flow system with two stages separated by a buffer with capacity N (Figure 1). The state of the system at time t is $s(t) = (X, \alpha_u, \alpha_d, t)$ where X is the buffer level, $\alpha_u \in \{1, ..., I_u\}$ is the state of the upstream stage M_u and $\alpha_d \in \{1, ..., I_d\}$ is the state of the downstream stage M_d. There are $I_u \times I_d$ discrete states in the discrete state space S_M. We do not classify states as up or down states as most of the other studies in the literature. A state with a processing rate equal to zero can be considered as a down state.

The processing rate of M_u in state i is $\mu_i^u \geq 0$ and the processing rate of M_d in state j is $\mu_j^d \geq 0$. Vectors $\mu^u = \{\mu_i^u\}$ and $\mu^d = \{\mu_j^d\}$ contain these processing rates. Row vectors $\mathbf{m}^u = \{\mu_i^u | (i,j) \in S_M\}$ and $\mathbf{m}^d = \{\mu_j^d | (i,j) \in S_M\}$ to determine the processing rate of each stage at a given state. The rates of change in the buffer level in each state (α_u, α_d) is given in row vector $\mathbf{m}_S = \{|\mu_i^u - \mu_j^d| (i,j) \in S\}$ for $S = \Upsilon, \Delta$.

We partition the discrete states of the system into three sets depending on whether the buffer level goes up (Υ), down (Δ), or stays the same (Z) in that state:

- $\alpha_\Upsilon(x) = (\alpha_u, \alpha_d) \in \Upsilon$ if $\mu_i^u > \mu_j^d$
- $\alpha_\Delta(x) = (\alpha_u, \alpha_d) \in \Delta$ if $\mu_i^u < \mu_j^d$

- $\alpha_Z(x) = (\alpha_u, \alpha_d) \in Z$ if $\mu_i^u = \mu_j^d$.

The number of states in each of these sets are $I_\Upsilon = |\Upsilon|$, $I_\Delta = |\Delta|$, and $I_Z = |Z|$ respectively.

When $0 < X < N$, the transition time from state i to state j is an exponential random variable with rate λ_{ij}^u for M_u and λ_{ij}^d for M_d. Matrices $\lambda^u = \{\lambda_{ij}^u\}$ and $\lambda^d = \{\lambda_{ij}^d\}$ contain the transition rates for M_u and M_d respectively.

We assume that α_u cannot leave its state when M_u is completely blocked.Similarly, we assume that α_d cannot leave its state if M_d is completely starved. When M_u is partially blocked and continues production at a reduced rate while the buffer is full, the transition time from state i to state j is also an exponential random variable with rate ψ_{ij}^u. Similarly, when M_d is partially starved and continues production at a reduced rate while the buffer is empty, the transition rate from state i to state j is ψ_{ij}^d.

Although our methodology is developed to work with arbitrary values of ψ_{ij}^u and ψ_{ij}^d, in the examples we discuss in Section 5 and 6, we consider a specific case where the reduction in the transition rates at these boundaries is proportional to the reduction in the processing rate. That is when the buffer is empty and M_d is producing at a reduced rate of μ_i^u, $\psi_{jj'}^d = \frac{\mu_i^u}{\mu_j^d}\lambda_{jj'}^d$. This setting implies that when $\mu_i^u = 0$, $\psi_{jj'}^d = 0$ and therefore it is not possible to make a transition when M_d is completely starved. Similarly, when the buffer is full and M_u is producing at a reduced rate of μ_j^d, $\psi_{ii'}^u = \frac{\mu_j^d}{\mu_i^u}\lambda_{ii'}^d$. Similar to the previous case, when $\mu_j^d = 0$, $\psi_{ii'}^u = 0$ and therefore a transition is not possible when M_u is completely blocked.

The time-dependent probability density while the buffer is partially full is

$$f(x, i, j, t) = \frac{\partial}{\partial x}\text{prob}[X(t) \le x, \alpha_u(t) = i, \alpha_d(t) = j].$$

The process is assumed to be ergodic and the steady-state probabilities exist. The steady-state density functions are defined as

$$f(x, i, j) = \lim_{t\to\infty} f(x, i, j, t) \tag{1}$$

and arranged in column vectors as

$$\mathbf{f}_S(x) = \{f(x, i, j)|(i, j) \in S\}, \quad S = \Upsilon, \Delta, Z. \tag{2}$$

The probability density of the state $(0, i, j, t)$ when the buffer is empty is denoted with $p(0, i, j, t)$. Similarly, the probability density of the state (N, i, j, t) when the buffer is full is denoted with $p(N, i, j, t)$. The steady-state probabilities that the buffer is empty and full are
$P_0 = \lim_{t\to\infty} \text{prob}[X(t) = 0]$ and $P_N = \lim_{t\to\infty} \text{prob}[X(t) = N]$.

3 Analysis of Partially-Full and Boundary Processes

In this section, the steady-state distribution is determined by analyzing the continuous time-continuous and discrete state space Markov process. First, the differential equations that describe the dynamics of the system when the buffer is partially full and when the buffer is empty or full are derived. Then a number of equations are derived to solve these equations.

3.1 Partially-Full Buffer Process

State Transition Equations Conditioning the probability density of the state at time $t+h$ on the state of the system at time t yields

$$\begin{aligned} f(x,i,j,t+h) &= f(x-(\mu_i^u-\mu_j^d)h,i,j,t)\left(1-\sum_{\substack{i'=1\\ i'\neq i}}^{I_u}\lambda_{ii'}^u h\right)\left(1-\sum_{\substack{j'=1\\ j'\neq j}}^{I_d}\lambda_{jj'}^d h\right)\\ &+\sum_{\substack{i'=1\\ i'\neq i}}^{I_u} f(x-(\mu_{i'}^u-\mu_j^d)h,i',j,t)\lambda_{i'i}^u h\left(1-\sum_{\substack{j'=1\\ j'\neq j}}^{I_d}\lambda_{jj'}^d h\right)\\ &+\sum_{\substack{j'=1\\ j'\neq j}}^{I_d} f(x-(\mu_i^u-\mu_{j'}^d)h,i,j',t)\lambda_{j'j}^d h\left(1-\sum_{\substack{i'=1\\ i'\neq i}}^{I_u}\lambda_{ii'}^u h\right). \end{aligned} \tag{3}$$

The above equation can also be written in differential form by setting $h \to 0$ as

$$\begin{aligned} \frac{\partial f(x,i,j,t)}{\partial t}+(\mu_i^u-\mu_j^d)\frac{\partial f(x,i,j,t)}{\partial x} &= -f(x,i,j,t)\left(\sum_{\substack{i'=1\\ i'\neq i}}^{I_u}\lambda_{ii'}^u+\sum_{\substack{j'=1\\ j'\neq j}}^{I_d}\lambda_{jj'}^d\right)\\ &+\sum_{\substack{i'=1\\ i'\neq i}}^{I_u} f(x,i',j,t)\lambda_{i'i}^u+\sum_{\substack{j'=1\\ j'\neq j}}^{I_d} f(x,i,j',t)\lambda_{j'j}^d. \end{aligned} \tag{4}$$

In the steady-state, the above equation yields $I_u I_d$ equations given below:

$$(\mu_i^u-\mu_j^d)\frac{\partial f(x,i,j)}{\partial x} = -f(x,i,j)\left(\sum_{\substack{i'=1\\ i'\neq i}}^{I_u}\lambda_{ii'}^u+\sum_{\substack{j'=1\\ j'\neq j}}^{I_d}\lambda_{jj'}^d\right)+\sum_{\substack{i'=1\\ i'\neq i}}^{I_u} f(x,i',j)\lambda_{i'i}^u+\sum_{\substack{j'=1\\ j'\neq j}}^{I_d} f(x,i,j')\lambda_{j'j}^d. \tag{5}$$

Solution of the Internal Equations Note that the coefficient of $\frac{\partial f(x,i,j)}{\partial x}$ in the above equation can be positive, negative, or zero. We order the discrete states in the order Υ, Δ, and Z. Then the internal equation given in Equation (5) can be written in matrix form as

$$\begin{bmatrix} \frac{\partial \mathbf{f}_\Upsilon(x)}{\partial x} \\ \frac{\partial \mathbf{f}_\Delta(x)}{\partial x} \\ 0 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} \mathbf{f}_\Upsilon(x) \\ \mathbf{f}_\Delta(x) \\ \mathbf{f}_Z(x) \end{bmatrix} \tag{6}$$

where A_1 is a square matrix of size $(I_\Upsilon + I_\Delta) \times (I_\Upsilon + I_\Delta)$, A_4 is a square matrix of size $I_Z \times I_Z$, A_2 is a matrix of size $(I_\Upsilon + I_\Delta) \times I_Z$, and A_3 is a matrix of size $I_Z \times (I_\Upsilon + I_\Delta)$. These matrices are determined by the parameters of a given system.

Accordingly, the solution of this system of equations is

$$\begin{bmatrix} \mathbf{f}_\Upsilon(x) \\ \mathbf{f}_\Delta(x) \end{bmatrix} = e^{\Lambda x}\mathbf{w} \tag{7}$$

$$\mathbf{f}_Z(x) = \Omega \begin{bmatrix} \mathbf{f}_\Upsilon(x) \\ \mathbf{f}_\Delta(x) \end{bmatrix} = \Omega e^{\Lambda x}\mathbf{w} \tag{8}$$

where $\Lambda = A_1 - A_2 A_4^{-1} A_3$, $\Omega = -A_4^{-1} A_3$ and $\mathbf{w}$ is a column vector of length $I_\Upsilon + I_\Delta$. Therefore $I_\Upsilon + I_\Delta$ equations are needed to determine the weights uniquely.

3.2 Level Crossing Equivalence

In order to relate the densities of the partially-full buffer process and the boundary buffer processes when the buffer is empty or full, we use a level crossing analysis. With this approach, the entry and exit probabilities into the empty- and full-buffer processes are determined by using the density functions. Let $L(x, i, j, T)$ denote the number of level crossings in state (x, i, j, t) in the time interval $[t, t + T]$ for large T. It can be shown that

$$\lim_{T\to\infty} \frac{L(x, i, j, T)}{T} = |\mu_i^u - \mu_j^d| f(x, i, j). \tag{9}$$

In partially-full buffer states with $\mu_i^u > \mu_j^d$, $(\mu_i^u - \mu_j^d)f(x, i, j)$ is the expected number of downward crossings at buffer level x per unit time. Similarly in states with $\mu_i^u < \mu_j^d$, $(\mu_j^d - \mu_i^d)f(x, i, j)$ is the expected number of upward crossings per unit time. Since at any given buffer level, the expected number of upward and downward crossings are equal in the long run, we can also write

$$\sum_{i=1}^{I_u}\sum_{j=1}^{I_d}(\mu_i^u - \mu_j^d)f(x, i, j) = 0. \tag{10}$$

Using the solution of the densities given in Equations (7) and (8), Equation (10) can be written as

$$\begin{bmatrix} \mathbf{m}_\Upsilon & -\mathbf{m}_\Delta \end{bmatrix} e^{\Lambda x}\mathbf{w} = 0. \tag{11}$$

3.3 Empty Buffer Process

Now we will derive the equations that describe the dynamics of the system when the buffer is empty. As the buffer level decreases in states $(i,j) \in \Delta$, the buffer eventually becomes empty if no other transition occurs first. Once the buffer becomes empty, it stays empty until the system makes a transition to a state $(i,j) \in \Upsilon$.

When the buffer is empty and if $\mu_i^u = 0$ and $\mu_j^d > 0$ in state (i,j) then M_d is completely starved and cannot make a transition to any other state. However, if $\mu_j^d \geq \mu_i^u > 0$, M_d can continue its production at a reduced rate of μ_i^u. In this case, M_d can make a transition to another state with rate $\frac{\mu_i^u}{\mu_j^d}\lambda_{jj'}^d$. With these dynamics, it is not possible to reach state (i,j) with $\mu_i^u = 0$ and $\mu_j^d = 0$. Let S_0 be the set of reachable states when the buffer is empty: $S_0 = \Delta \cup Z \setminus \{(i,j) | \mu_i^u = 0, \mu_j^d = 0\}$ and $I_{S_0} = |S_0|$.

State Transition Equations The dynamics of the system in reachable states when the buffer is empty is governed by the following equations

$$\begin{aligned} \frac{dp(0,i,j,t)}{dt} &= -p(0,i,j,t)\left(\sum_{\substack{i'=1 \\ i' \neq i}}^{I_u} \lambda_{ii'}^u + \sum_{\substack{j'=1 \\ j' \neq j}}^{I_d} \frac{\mu_i^u}{\mu_j^d}\lambda_{jj'}^d \right) \\ &+ \sum_{\substack{i'=1 \\ i' \neq i \\ (i',j) \in S_0}}^{I_u} p(0,i',j,t)\lambda_{i'i}^u + \sum_{\substack{j'=1 \\ j' \neq j \\ (i,j') \in S_0}}^{I_d} p(0,i,j',t)\frac{\mu_i^u}{\mu_{j'}^d}\lambda_{j'j}^d, \quad (i,j) \in S_0. \end{aligned} \tag{12}$$

The empty buffer process ends with a transition into a state where the buffer level starts increasing. The dynamics of the transitions are

$$\frac{dp(0,i,j,t)}{dt} = \sum_{\substack{i'=1 \\ i' \neq i \\ (i',j) \in S_0}}^{I_u} p(0,i',j,t)\lambda_{i'i}^u + \sum_{\substack{j'=1 \\ j' \neq j \\ (i,j') \in S_0}}^{I_d} p(0,i,j',t)\frac{\mu_i^u}{\mu_{j'}^d}\lambda_{j'j}^d, \quad (i,j) \in \Upsilon. \tag{13}$$

Entry and Exit Probabilities The probability that the buffer becomes empty while the process has been in a specific state is the ratio of the number of downward crossings in this particular state and the all possible downward crossings at $X = 0^+$:

$$\text{prob}[\alpha_\Delta(0^+) = (i,j)] = \lim_{T \to \infty} \frac{L(0^+, i, j, T)/T}{\sum_{(i',j') \in \Delta} L(0^+, i', j', T)/T} \tag{14}$$

$$= \frac{(\mu_i^u - \mu_j^d) f(0^+, i, j)}{\sum_{(i',j') \in \Delta} (\mu_{i'}^u - \mu_{j'}^d) f(0^+, i', j')}, \quad (i,j) \in \Delta. \tag{15}$$

Similarly, the process exits the empty buffer state with a transition into a specific state is given as

$$\text{prob}[\alpha_\Upsilon(0^+) = (i,j)] = \frac{(\mu_i^u - \mu_j^d) f(0^+, i, j)}{\sum\limits_{(i',j')\in\Upsilon} (\mu_{i'}^u - \mu_{j'}^d) f(0^+, i', j')}, \quad (i,j) \in \Upsilon. \tag{16}$$

The empty buffer process relates the probabilities given in Equations (14) and (16). More specifically,

$$\text{prob}[\alpha_\Upsilon(0^+) = (i,j)] = \sum_{(i',j')\in\Delta} \text{prob}[\alpha_\Upsilon(0^+) = (i,j) \mid \alpha_\Delta(0^+) = (i',j')]\text{prob}[\alpha_\Delta(0^+) = (i',j')], \quad (i,j) \in \Upsilon. \tag{17}$$

By using the equivalence of the upward and downward crossings, Equation (17) simplifies to

$$(\mu_i^u - \mu_j^d) f(0^+, i, j) = \sum_{(i',j')\in\Delta} \text{prob}[\alpha_\Delta(0^+) = (i,j) \mid \alpha_\Upsilon(0^-) = (i',j')](\mu_{j'}^d - \mu_{i'}^u) f(0^+, i', j'), \quad (i,j) \in \Upsilon. \tag{18}$$

where the conditional probabilities $\text{prob}[\alpha_\Upsilon(0^+) = (i,j) \mid \alpha_\Delta(0^+) = (i',j')]$ are determined from Equations (12) and (13).

Determining Conditional Probabilities Let $\mathbf{p}_S^0(t) = \{p(0,i,j,t) \mid (i,j) \in S\}$ for $S = S_0, \Upsilon$. Then Equations (12) and (13) can be written in matrix form as

$$\frac{d\mathbf{p}_{S_0}^0(t)}{dt} = A_0 \mathbf{p}_{S_0}^0(t) \tag{19}$$

and

$$\frac{d\mathbf{p}_\Upsilon^0(t)}{dt} = B_0 \mathbf{p}_{S_0}^0(t) \tag{20}$$

where A_0 is a $I_{S_0} \times I_{S_0}$ square matrix and B_0 is a $I_\Upsilon \times I_{S_0}$ matrix. Then, the matrix $-B_0 A_0^{-1}$ gives the conditional probability that the empty buffer process exits in a particular state $(i,j) \in \Upsilon$ given that it starts in one of the transient states $(i',j') \in S_0$. Since the empty buffer process can start only in states $(i,j) \in \Delta$, let G_0 be a $I_\Upsilon \times I_\Delta$ matrix that is obtained by eliminating the columns of $-B_0 A_0^{-1}$ corresponding to states $S_0 \setminus \Delta$. Accordingly, Equation (18) can be written in matrix notation as

$$\begin{bmatrix} \text{diag}(\mathbf{m}_\Upsilon) & 0_{I_\Upsilon \times I_\Delta} \end{bmatrix} \mathbf{w} = G_0 \begin{bmatrix} 0_{I_\Delta \times I_\Upsilon} & \text{diag}(\mathbf{m}_\Delta) \end{bmatrix} \mathbf{w} \tag{21}$$

where $\text{diag}(\mathbf{a})$ is a diagonal matrix of vector $\mathbf{a}$ and $0_{k\times l}$ is a $k \times l$ matrix of zeros.

Since $\sum\limits_{(i,j)\in\Upsilon} \text{prob}[\alpha_\Upsilon(0^+) = (i,j)] = 1$, Equation (18) gives $I_\Upsilon - 1$ linearly independent equations that will be used to determine $\mathbf{w}$.

Empty Buffer Probability The probability that the buffer is empty in the long run is the ratio of the total time the buffer stays empty in a given time period. The expected time buffer stays in a transient state (i', j') before exiting into a partially full buffer state conditioned on the state (i, j) the process enters into the empty buffer process, $E[T^0_{(i,j),(i',j')}]$ is also determined from Equations (12) and (13). Namely, the matrix $-A_0^{-1}$ gives the expected sojourn time in a particular transient state $(i, j) \in S_0$ given that it starts in one of the transient states $(i', j') \in S_0$. Since the empty buffer process can start only in states $(i, j) \in \Delta$, let $E[T^0]$ be a $I_{S_0} \times I_\Delta$ matrix that is obtained by eliminating the columns of $-A_0^{-1}$ corresponding to states $S_0 \setminus \Delta$.

Since the expected total time the buffer stays empty in a given time period can be determined by multiplying the number of times the process enters into the empty buffer with the expected time it stays empty before it exits, the probability that the buffer is empty in the long run is

$$\begin{aligned} P_0 &= \sum_{(i,j)\in\Delta} \lim_{T\to\infty} \frac{L(0^+, i, j, T)}{T} \sum_{(i',j')\in S_0} E[T^0_{(i,j),(i',j')}] \\ &= \sum_{(i,j)\in\Delta} (\mu_j^d - \mu_i^u) f(0^+, i, j) \sum_{(i',j')\in S_0} E[T^0_{(i,j),(i',j')}] \end{aligned} \tag{22}$$

or equivalently in matrix form

$$P_0 = u_{I_{S_0}} E[T^0] \left[\begin{array}{cc} 0_{I_\Delta \times I_\Upsilon} & \text{diag}(\mathbf{m}_\Delta) \end{array} \right] \mathbf{w} \tag{23}$$

where u_k is a row vector of ones with length k.

3.4 Full Buffer Process

The last step of the analysis of the partially-full and boundary processes is the analysis of the full buffer process. As the buffer level increases in states $(i, j) \in \Upsilon$, the buffer eventually becomes full if no other transition occurs first. Once the buffer becomes full, it stays full until the system makes a transition to a state $(i, j) \in \Delta$.

When the buffer is full, if $\mu_i^u > 0$ and $\mu_j^d = 0$ in state (i, j) then M_u is completely starved and cannot make a transition to any other state. However, if $\mu_i^u \geq \mu_j^d > 0$, M_u can continue its production at a reduced rate of μ_j^d. In this case, M_u can make a transition to another state with rate $\frac{\mu_j^d}{\mu_i^u} \lambda_{ii'}^d$. With these dynamics, it is not possible to reach state (i, j) with $\mu_i^u = 0$ and $\mu_j^d = 0$. Let S_N be the set of reachable states when the buffer is full: $S_N = \Upsilon \cup Z \setminus \{(i, j) | \mu_i^u = 0, \mu_j^d = 0\}$ and $I_{S_N} = |S_N|$.

State Transition Equations The dynamics of the system in reachable states when the buffer is full is governed by the following equations

$$\frac{dp(N,i,j,t)}{dt} = -p(N,i,j,t)\left(\sum_{\substack{j'=1\\ j'\neq j}}^{I_d} \lambda^d_{jj'} + \sum_{\substack{i'=1\\ i'\neq i}}^{I_u} \frac{\mu^d_j}{\mu^u_i}\lambda^u_{ii'}\right) + \sum_{\substack{j'=1\\ j'\neq j\\ (i,j')\in S_N}}^{I_d} p(N,i,j',t)\lambda^d_{j'j} + \sum_{\substack{i'=1\\ i'\neq i\\ (i',j)\in S_N}}^{I_u} p(N,i',j,t)\frac{\mu^d_j}{\mu^u_{i'}}\lambda^u_{i'i}, \quad (i,j)\in S_N. \tag{24}$$

The full buffer process ends with a transition into a state where the buffer level starts decreasing. The dynamics of the transitions are

$$\frac{dp(N,i,j,t)}{dt} = \sum_{\substack{j'=1\\ j'\neq j\\ (i,j')\in S_N}}^{I_d} p(N,i,j',t)\lambda^d_{jj'} + \sum_{\substack{i'=1\\ i'\neq i\\ (i',j)\in S_N}}^{I_u} p(N,i',j,t)\frac{\mu^d_j}{\mu^u_{i'}}\lambda^u_{i'i}, \quad (i,j)\in\Delta. \tag{25}$$

Entry and Exit Probabilities The probability that the buffer becomes full while the process has been in a specific state is the ratio of the number of upward crossings in this particular state and the all possible upward crossings at $X = N^-$:

$$\text{prob}[\alpha_\Upsilon(N^-) = (i,j)] = \frac{(\mu^u_i - \mu^d_j)f(N^-,i,j)}{\sum_{(i',j')\in\Upsilon}(\mu^u_{i'} - \mu^d_{j'})f(N^-,i',j')}, \quad (i,j)\in\Upsilon. \tag{26}$$

Similarly, the probability that the process exits the full buffer state with a transition into a specific state is given as

$$\text{prob}[\alpha_\Delta(N^-) = (i,j)] = \frac{(\mu^u_i - \mu^d_j)f(N^-,i,j)}{\sum_{(i',j')\in\Delta}(\mu^u_{i'} - \mu^d_{j'})f(N^-,i',j')}, \quad (i,j)\in\Delta. \tag{27}$$

Determining Conditional Probabilities The full buffer process relates the probabilities given in Equations (26) and (27). More specifically,

$$\text{prob}[\alpha_\Delta(N^-) = (i,j)] = \sum_{(i',j')\in\Upsilon}\text{prob}[\alpha_\Delta(N^-) = (i,j) \mid \alpha_\Upsilon(N^-) = (i',j')]\text{prob}[\alpha_\Upsilon(N^-) = (i',j')], \quad (i,j)\in\Delta. \tag{28}$$

By using the equivalence of the upward and downward crossings, Equation (28) simplifies to

$$(\mu_j^d - \mu_i^u) f(N^-, i, j) = \sum_{(i',j') \in \Upsilon} \text{prob}[\alpha_\Delta(N^-) = (i,j) \mid \alpha_\Upsilon(N^-) = (i',j')](\mu_{i'}^u - \mu_{j'}^d) f(N^-, i', j'),$$
$$(i,j) \in \Delta \quad (29)$$

where the conditional probabilities $\text{prob}[\alpha_\Delta(N^-) = (i,j) \mid \alpha_\Upsilon(N^-) = (i',j')]$ are determined from Equations (24) and (25).

Let $\mathbf{p}_S^N(t) = \{p(N, i, j, t) \mid (i,j) \in S\}$ for $S = S_N, \Delta$. Then Equations (24) and (25) can be written in matrix form as

$$\frac{d\mathbf{p}_{S_N}^N(t)}{dt} = A_N \mathbf{p}_{S_N}^N(t) \quad (30)$$

and

$$\frac{d\mathbf{p}_\Delta^N(t)}{dt} = B_N \mathbf{p}_{S_N}^N(t) \quad (31)$$

where A_N is a $I_{S_N} \times I_{S_N}$ square matrix and B_N is a $I_\Delta \times I_{S_N}$ matrix. Then, the matrix $-B_N A_N^{-1}$ gives the conditional probability that the full buffer process exits in a particular state $(i,j) \in \Delta$ given that it starts in one of the transient states $(i',j') \in S_N$. Since the full buffer process can start only in states $(i,j) \in \Upsilon$, let G_N be a $I_\Delta \times I_\Upsilon$ matrix that is obtained by eliminating the columns of $-B_N A_N^{-1}$ corresponding to states $S_N \setminus \Upsilon$.

In matrix notation, Equation (28) can be written as

$$\left[\begin{array}{cc} 0_{I_\Delta \times I_\Upsilon} & \text{diag}(\mathbf{m}_\Delta) \end{array} \right] e^{\Lambda N} \mathbf{w} = G_N \left[\begin{array}{cc} \text{diag}(\mathbf{m}_\Upsilon) & 0_{I_\Upsilon \times I_\Delta} \end{array} \right] e^{\Lambda N} \mathbf{w}. \quad (32)$$

Since $\sum_{(i,j) \in \Delta} \text{prob}[\alpha_\Delta(N^-) = (i,j)] = 1$, Equation (29) gives $I_\Delta - 1$ linearly independent equations that will be used to determine $\mathbf{w}$.

Full Buffer Probability The probability that the buffer is full in the long run is the ratio of the total time the buffer stays full in a given time period. The expected time buffer stays full in a transient state (i',j') before exiting into a partially full buffer state conditioned on the state (i,j) the process enters into the full buffer process, $E[T_{(i,j),(i',j')}^N]$ is determined from Equations (24), and (25). More specifically, the matrix $-A_N^{-1}$ gives the expected sojourn time in a particular transient state $(i,j) \in S_N$ given that it starts in one of the transient states $(i',j') \in S_N$. Since the full buffer process can start only in states $(i,j) \in \Upsilon$, let $E[T^N]$ be a $I_{S_N} \times I_\Upsilon$ matrix that is obtained by eliminating the columns of $-A_N^{-1}$ corresponding to states $S_N \setminus \Upsilon$.

By multiplying the number of times the process enters into the full buffer with the expected time it stays full before it exits, we can determine total time the buffer stays full in a given time period. Then the probability that the buffer is full in the long run can be determined as

$$P_N = \sum_{(i,j) \in \Upsilon} \lim_{T \to \infty} \frac{L(N^-, i, j, T)}{T} \sum_{(i',j') \in S_N} E[T_{(i,j),(i',j')}^N]$$

$$= \sum_{(i,j)\in\Upsilon} (\mu_i^u - \mu_j^d) f(N^-, i, j) \sum_{(i',j')\in S_N} E[T^N_{(i,j),(i',j')}] \tag{33}$$

Equation (33) can be written in matrix form as

$$P_N = u_{I_{S_N}} E[T^N] \begin{bmatrix} \text{diag}(\mathbf{m}_\Upsilon) & 0_{I_\Upsilon \times I_\Delta} \end{bmatrix} e^{\Lambda N} \mathbf{w}. \tag{34}$$

3.5 Solution of the Probability Densities

As explained above, Equations (21) and (32) give a total of $I_\Upsilon + I_\Delta - 2$ equations. Since there are $I_\Upsilon + I_\Delta$ weights, two additional equations are required to uniquely determine $\mathbf{w}$.

The first additional equation is the equivalence of flow through the buffer. In the long run, the amount of material brought into the buffer must be equal to the amount of material taken from the buffer. Then

$$\int_0^N \sum_{i=1}^{I_u} \sum_{j=1}^{I_d} \mu_i^u f(x, i, j) dx = \int_0^N \sum_{i=1}^{I_u} \sum_{j=1}^{I_d} \mu_j^d f(x, i, j) dx \tag{35}$$

Note that Equation (35) can also be derived from the equivalence of the upward and downward crossings at any given buffer level by integrating Equation (11) from 0 to N that yields

$$\begin{bmatrix} \mathbf{m}_\Upsilon & -\mathbf{m}_\Delta \end{bmatrix} \left(\int_0^N e^{\Lambda x} dx \right) \mathbf{w} = 0. \tag{36}$$

The last equation is the normalization equation:

$$P_0 + \int_0^N \sum_{i=1}^{I_u} \sum_{j=1}^{I_d} f(x, i, j) dx + P_N = 1. \tag{37}$$

By using Equations (7), (8), (23) and (34), the normalization equation can be written in matrix form as

$$\Big(u_{I_{S_0}} E[T^0] \begin{bmatrix} 0_{I_\Delta \times I_\Upsilon} & \text{diag}(\mathbf{m}_\Delta) \end{bmatrix} + (u_{I_\Upsilon + I_\Delta} + u_{I_Z} \Omega) \left(\int_0^N e^{\Lambda x} dx \right) + u_{I_{S_N}} E[T^N] \begin{bmatrix} \text{diag}(\mathbf{m}_\Upsilon) & 0_{I_\Upsilon \times I_\Delta} \end{bmatrix} e^{\Lambda N} \Big) \mathbf{w} = 1 \tag{38}$$

Now Equations (21) and (32) with Equations (36) and (38) give $I_\Upsilon + I_\Delta$ linearly independent equations that uniquely determine w. Therefore all the steady-state probability distributions that describe the dynamics of the system are determined by these equations.

4 Performance Measures

When the probability densities are determined, all performance measures of interest can be determined. In a production setting, the main performance measures of interest are the production rate and the expected buffer level.

The production rate is the amount of material processed per unit time in the long run. The production rate of the first stage and the second stage are the same due to the conservation of material. Therefore we give the production rate of the first stage without loss of generality. The production rate in the internal states can be determined in a straight-forward way. Since the first stage can be forced to produce at a reduced rate due to partial blocking and the second stage cane be forced to produce at a reduced rate due to partial starvation, the production in these states must be determined based on how long the process stays in these boundary states and at what rate it produces the products. The following equation gives the production rate of the first stage:

$$\begin{aligned}\pi &= \sum_{(i,j)\in\Delta} \lim_{T\to\infty} \frac{L(0^+,i,j,T)}{T} \sum_{(i',j')\in S_0} \mu^u_{i'} E[T^0_{(i,j),(i',j')}] \\ &+ \int_0^N \sum_{i=1}^{I_u}\sum_{j=1}^{I_d} \mu^u_i f(x,i,j)dx \\ &+ \sum_{(i,j)\in\Upsilon} \lim_{T\to\infty} \frac{L(N^-,i,j,T)}{T} \sum_{(i',j')\in S_N} \mu^d_{j'} E[T^N_{(i,j),(i',j')}] \qquad (39)\end{aligned}$$

By using Equation (11) we can determine the production rate in terms of the densities as

$$\begin{aligned}\pi &= \sum_{(i,j)\in\Delta} (\mu^d_j - \mu^u_i) f(0^+,i,j) \sum_{(i',j')\in S_0} \mu^u_{i'} E[T^0_{(i,j),(i',j')}] \\ &+ \int_0^N \sum_{i=1}^{I_u}\sum_{j=1}^{I_d} \mu^u_i f(x,i,j)dx \\ &+ \sum_{(i,j)\in\Upsilon} (\mu^u_i - \mu^d_j) f(N^-,i,j) \sum_{(i',j')\in S_N} \mu^d_{j'} E[T^N_{(i,j),(i',j')}] \qquad (40)\end{aligned}$$

or in matrix form

$$\begin{aligned}\pi &= \Bigg(\mathbf{m}^u_{S_0} E[T^0] \left[\ 0_{I_\Delta\times I_\Upsilon}\ \ \mathrm{diag}(\mathbf{m}_\Delta)\ \right] + \left(\mathbf{m}^u_{\Upsilon+\Delta} + \mathbf{m}^u_Z\Omega\right) \int_0^N e^{\Lambda x}dx \\ &+ \mathbf{m}^d_{S_N} E[T^N] \left[\ \mathrm{diag}(\mathbf{m}_\Upsilon)\ \ 0_{I_\Upsilon\times I_\Delta}\ \right] e^{\Lambda N}\Bigg) \mathbf{w} \qquad (41)\end{aligned}$$

where row vector $\mathbf{m}^k_D$ is the vector obtained from $\mathbf{m}^k$ by taking the elements in its subset D for $k \in \{u,d\}$.

Finally, the expected buffer level is

$$E[X] = \int_0^N \sum_{i=1}^{I_u} \sum_{j=1}^{I_d} x f(x,i,j) dx + N P_N. \tag{42}$$

or in matrix form

$$E[X] = \left((u_{I_\Upsilon + I_\Delta} + u_{I_Z} \Omega) \left(\int_0^N x e^{\Lambda x} dx \right) + u_{I_{S_N}} E[T^N] \left[\begin{array}{cc} \mathrm{diag}(\mathbf{m}_\Upsilon) & 0_{I_\Upsilon \times (I_\Delta + I_Z)} \end{array} \right] e^{\Lambda N} N \right) \mathbf{w}.$$

5 An Example: A Model with Machines that have Multiple Up and Multiple Down States

In this section, a specific system with multiple up and down states is modelled and analyzed in detail by using the methodology given above. In the following section, four additional examples are given.

5.1 Model Description

The system we consider is a two-stage system where the first stage has two up (State 1 and State -1) and three down states (State D_1, D_{-1}, and D_Q and the second stage has one up (State 1') and one down state (State 0'). The processing rates of the upstream stage in both of the up states are equal to μ_u and the processing rate of the downstream stage in its up state is μ_d and the processing rates of the down states for both stages are equal to zero. This system is analyzed in detail in (Poffe and Gershwin 2005). Figure 2 depicts the state transitions for M_u and M_d for this specific case. Figure 3 shows a sample realization of this system.

5.2 Model Inputs

Our solution methodology requires only matrices λ^u and λ^d matrices and vectors μ^u and μ^d as its inputs. The transition rate for M_u is given as

$$\lambda^u = \begin{bmatrix} -g-p & g & p & 0 & 0 \\ 0 & -p-h & 0 & p & h \\ r & 0 & -r & 0 & 0 \\ 0 & r & 0 & -r & 0 \\ r_Q & 0 & 0 & 0 & -r_Q \end{bmatrix} \tag{43}$$

where the states are ordered as $\{1, -1, D_1, D_{-1}, D_Q\}$. The processing rates in these states are

$$\mu^u = \left[\begin{array}{cccc} \mu_u & \mu_u & 0 & 0 \end{array} \right].$$

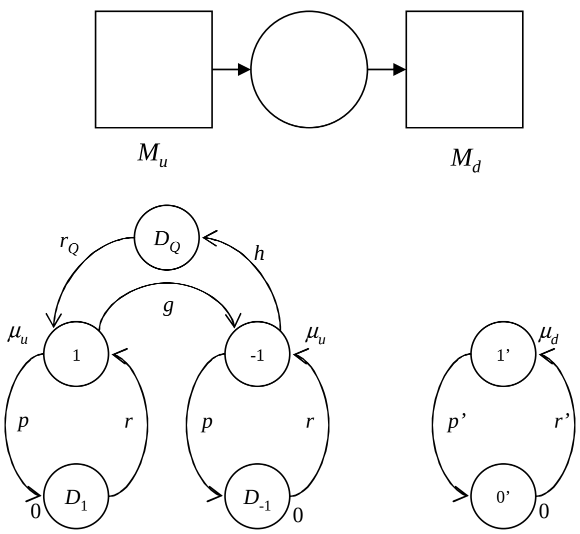

Figure 2: Modelling of a system with multiple up and multiple down states for analysis with the general methodology

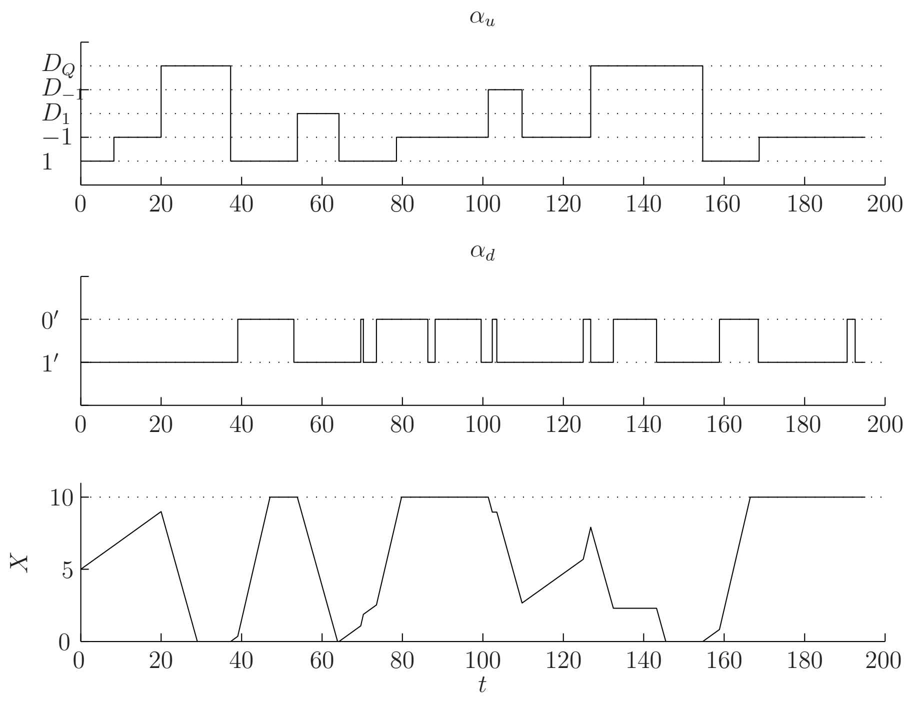

Figure 3: Sample path for a system with multiple up and down states ($\mu_u = 1.2$, $\mu_d = 1$, $p = 0.01$, $r = 0.1$, $p' = 0.05$, $r' = 0.10$, $g = 0.05$, $h = 0.10$, $r_Q = 0.10$)

Similarly,

$$\lambda^d = \begin{bmatrix} -p' & p' \\ r & -r' \end{bmatrix} \tag{44}$$

where the states are ordered as $\{1', 0'\}$. In these states the processing rates of M_d are given as

$$\mu^d = \begin{bmatrix} \mu_d & 0 \end{bmatrix}.$$

5.3 Analysis of the Model

Once these inputs are given, we can specify matrices A_1, A_2, A_3, A_4, A_0, B_0, A_N, B_N and vectors $\mathbf{m}_\Upsilon$, $\mathbf{m}_\Delta$, $\mathbf{m}_Z$, $\mathbf{m}^u_{S_0}$, $\mathbf{m}^d_{S_0}$, $\mathbf{m}^u_{S_N}$, and $\mathbf{m}^d_{S_N}$ directly. Once these matrices and vectors are specified, the methodology outlined in the preceding sections yield the desired performance measures directly.

Table given in (45) lists the states, the corresponding processing rates, and the classification of each state in sets Υ, Δ, and Z depending on μ_u and μ_d.

					S		
α_u	α_d	$\mathbf{m}^u$	$\mathbf{m}^d$	$\mathbf{m}_S$	$\mu_1 > \mu_2$	$\mu_1 = \mu_2$	$\mu_1 < \mu_2$
1	$1'$	μ_u	μ_d	$\mu_u - \mu_d$	Υ	Z	Δ
-1	$1'$	μ_u	μ_d	$\mu_u - \mu_d$	Υ	Z	Δ
1	$0'$	μ_u	0	μ_d	Υ	Υ	Υ
-1	$0'$	μ_u	0	μ_u	Υ	Υ	Υ
D_1	$1'$	0	μ_d	μ_d	Δ	Δ	Δ
D_{-1}	$1'$	0	μ_d	μ_d	Δ	Δ	Δ
D_Q	$1'$	0	μ_d	μ_d	Δ	Δ	Δ
D_1	$0'$	0	0	0	Z	Z	Z
D_{-1}	$0'$	0	0	0	Z	Z	Z
D_Q	$0'$	0	0	0	Z	Z	Z

(45)

There are 10 discrete states in the state space. For example, when $\mu_u > \mu_d$, $I_\Upsilon = 4$, $I_\Delta = 3$, and $I_Z = 3$. In this case,

$$\mathbf{m}_\Upsilon = \begin{bmatrix} \mu_u - \mu_d & \mu_u - \mu_d & \mu_u & \mu_u \end{bmatrix},$$

$$\mathbf{m}_\Delta = \begin{bmatrix} \mu_d & \mu_d & \mu_d \end{bmatrix},$$

$$\mathbf{m}_Z = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

For this specific case, the submatrices A_1, A_2, A_3, and A_4 are written as

$$A_1 = \begin{bmatrix} \frac{-p-g-p'}{\mu_u-\mu_d} & 0 & \frac{r'}{\mu_u-\mu_d} & 0 & \frac{r}{\mu_u-\mu_d} & 0 & \frac{r_Q}{\mu_u-\mu_d} \\ \frac{g}{\mu_u-\mu_d} & \frac{-p-h-p'}{\mu_u-\mu_d} & 0 & \frac{r'}{\mu_u-\mu_d} & 0 & \frac{r}{\mu_u-\mu_d} & 0 \\ \frac{p'}{\mu_u} & 0 & \frac{-p-g-r'}{\mu_u} & 0 & 0 & 0 & 0 \\ 0 & \frac{p'}{\mu_u} & \frac{g}{\mu_u} & \frac{-p-h-r'}{\mu_u} & 0 & 0 & 0 \\ -\frac{p}{\mu_d} & 0 & 0 & 0 & \frac{r+p'}{\mu_d} & 0 & 0 \\ 0 & -\frac{p}{\mu_d} & 0 & 0 & 0 & \frac{r+p'}{\mu_d} & 0 \\ 0 & -\frac{h}{\mu_d} & 0 & 0 & 0 & 0 & \frac{r_Q+r'}{\mu_d} \end{bmatrix}, \tag{46}$$

$$A_2 = \begin{bmatrix} 0 & 0 & \frac{r}{\mu_u} & 0 & -\frac{r'}{\mu_d} & 0 & 0 \\ 0 & 0 & 0 & \frac{r}{\mu_u} & 0 & -\frac{r'}{\mu_d} & 0 \\ 0 & 0 & \frac{r_Q}{\mu_u} & 0 & 0 & 0 & -\frac{r'}{\mu_d} \end{bmatrix}^{\mathrm{T}}, \tag{47}$$

$$A_3 = \begin{bmatrix} 0 & 0 & p & 0 & p' & 0 & 0 \\ 0 & 0 & 0 & p & 0 & p' & 0 \\ 0 & 0 & 0 & h & 0 & 0 & p' \end{bmatrix}, \tag{48}$$

$$A_4 = \begin{bmatrix} -r-r' & 0 & 0 \\ 0 & -r-r' & 0 \\ 0 & 0 & -r_Q-r' \end{bmatrix}. \tag{49}$$

The submatrices for the cases $\mu_u = \mu_d$ and $\mu_u < \mu_d$ can be written similarly. We will only discuss the case $\mu_u > \mu_d$ in detail.

When $\mu_u \neq \mu_d$, the buffer level does not change when both stages are in down states. Since these states cannot be reached when the buffer is empty or full, $S_0 = \Delta$ and $S_N = \Upsilon$. Therefore $I_{S_0} = I_\Delta = 3$ and $I_{S_N} = I_\Upsilon = 4$.

For the empty buffer process, since M_d is completely starved in all transient states. Therefore

$$\mathbf{m}^u_{S_0} = \mathbf{m}^d_{S_0} = \begin{bmatrix} 0 & 0 \end{bmatrix}.$$

The matrices A_0 and B_0 for the empty buffer process are

$$A_0 = \begin{bmatrix} -r & 0 & 0 \\ 0 & -r & 0 \\ 0 & 0 & -r_Q-r' \end{bmatrix} \tag{50}$$

and

$$B_0 = \begin{bmatrix} r & 0 & r_Q \\ 0 & r & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{51}$$

Since $S_0 = \Delta$, $G_0 = -B_0 A_0^{-1}$.

For the full buffer process, M_u is partially blocked in states $(1,1')$ and $(-1,1')$ and completely blocked in states $(1,1')$ and $(-1,1')$. Accordingly,

$$\mathbf{m}^u_{S_N} = \mathbf{m}^d_{S_N} = \begin{bmatrix} \mu_d & \mu_d & 0 & 0 \end{bmatrix}.$$

Then the matrices A_N and B_N are

$$A_N = \begin{bmatrix} -p\frac{\mu_d}{\mu_u} - g\frac{\mu_d}{\mu_u} - p' & 0 & r' & 0 \\ g\frac{\mu_d}{\mu_u} & -p\frac{\mu_d}{\mu_u} - h\frac{\mu_d}{\mu_u} - p' & 0 & r' \\ p' & 0 & -r' & 0 \\ 0 & p' & g\frac{\mu_d}{\mu_u} & -r' \end{bmatrix} \tag{52}$$

$$B_N = \begin{bmatrix} p\frac{\mu_d}{\mu_u} & 0 & 0 & 0 \\ 0 & p\frac{\mu_d}{\mu_u} & 0 & 0 \\ 0 & h\frac{\mu_d}{\mu_u} & 0 & 0 \end{bmatrix}. \tag{53}$$

Since $S_N = \Upsilon$, $G_N = -B_N A_N^{-1}$.

5.4 Performance Evaluation

Now since all the input matrices and vectors are determined, the solution methodology outlined in the preceding sections yields the probability densities and the performance measures directly. Namely, inserting these matrices and vectors into Equations (21) and (32) with Equations (36) and (38) yields a system of equations that determine the weight vector $\mathbf{w}$. Then Equations (41) and (43) yield production rate and the expected buffer level.

All the results in this section and in Section 6 are validated by simulation. Each model is simulated by using both a continuous flow and also a discrete event simulation model. When the continuous simulation is run for 10^6 events, the percentage error between the analytical production rate and the simulated production rate is less than 10^{-5}. The time required to determine the performance measures by using the general methodology is very short and not affected by the buffer level.

Figures 4 and 5 show that increasing the processing rate of each stage increases the production rate until it reaches its limit. However, the expected buffer level increases with the processing rate of the first stage and it reaches its capacity and decreases with the processing rate of the second stage and it approaches zero. Figure 6 shows that increasing the buffer level increases the production rate and the expected buffer level as expected.

6 Modelling of Various Systems

In this section, we will model different systems to illustrate the application of our methodology in the analysis of different production lines. The first model is a system where each stage has a number of identical machines in parallel. The second model is a system where the up- and down-times of each station are Erlang random variables with different number of stages. Then a model of a three

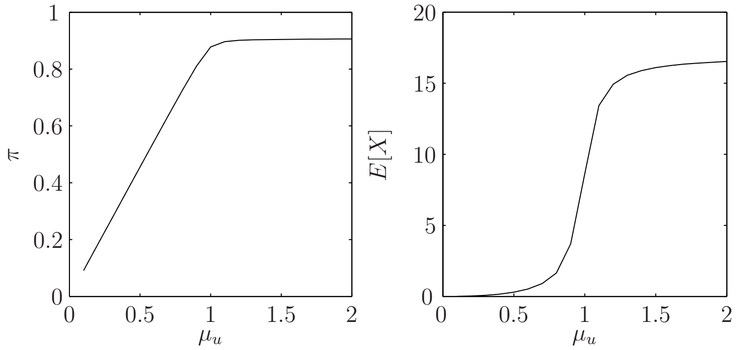

Figure 4: Effect of the processing rate of the upstream station in the model with multiple up and down states ($\mu_d = 1$, $p = 0.005$, $r = 0.15$, $p' = 0.015$, $r' = 0.15$, $g = 0.01$, $h = 0.20$, $r_Q = 0.15$, $N = 17$)

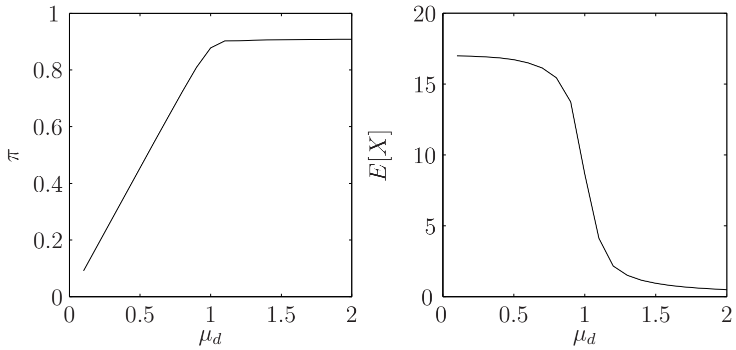

Figure 5: Effect of the processing rate of the downstream station in the model with multiple up and down states ($\mu_u = 1$, $p = 0.005$, $r = 0.15$, $p' = 0.015$, $r' = 0.15$, $g = 0.01$, $h = 0.20$, $r_Q = 0.15$, $N = 17$)

station merge system with a shared buffer is discussed. Finally, a model of a system where each stage has a number of machines in series is given. The way these systems are modelled is shown and the inputs are given explicitly for each model.

6.1 A Model with Parallel Machines

We now model a system where M_u has m_u and M_d has m_d identical machines in parallel similar to the one analyzed in Mitra (1988). Each machine is unreliable and has one up and one down state. In the upstream stage, the processing rate of each machine is μ_u and the failure and repair rates are p_u and r_u. In the downstream stage, the processing rate of each machine is μ_d and the failure and repair rates are p_d and r_d.

In this model M_k has $m_k + 1$ states where state i indicates that i machines are operational in

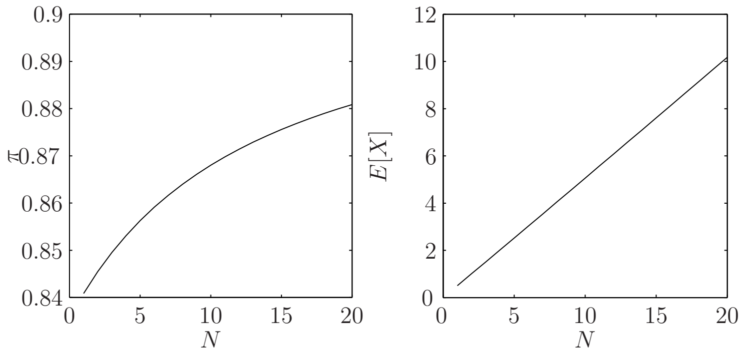

Figure 6: Effect of the buffer capacity in the model with multiple up and down states ($\mu_1 = 1$, $\mu_2 = 1$, $p = 0.005$, $r = 0.15$, $p' = 0.015$, $r' = 0.15$, $g = 0.01$, $h = 0.20$, $r_Q = 0.15$)

that state, $i = 0, ..., m_k$. Accordingly, the effective processing rate in state i for stage k is $i\mu_k$, $k = u, d$.

Possible transitions for stage k are

- from state i to state $i-1$ with rate ip_k, $i = 1, ..., m_k$ and
- from state i to state $i+1$ with rate $(m_k - i)r$, $i = 0, ..., m_k - 1$.

Figure 7 depicts the state transitions for M_u and M_d for a specific case where M_u has $m_u = 3$ machines and M_d has $m_d = 2$ machines in parallel.

The matrices λ^u and λ^d and the vectors μ^u and μ^d for this specific case are given below:

$$\lambda^u = \begin{bmatrix} -3r_u & 3r_u & 0 & 0 \\ p_u & -p_u - 2r_u & 2r_u & 0 \\ 0 & 2p_u & -2p_u - r_u & r_u \\ 0 & 0 & 3p_u & -3p_u \end{bmatrix} \tag{54}$$

where the states are ordered as $\{0, 1, 2, 3\}$. The processing rates in these states are

$$\mu^u = \begin{bmatrix} 0 & \mu_u & 2\mu_u & 3\mu_u \end{bmatrix}.$$

Similarly,

$$\lambda^d = \begin{bmatrix} -2r_d & 2r_d & 0 \\ p_d & -p_d - r_d & r_d \\ 0 & 2p_d & -2p_d \end{bmatrix} \tag{55}$$

where the states are ordered as $\{0', 1', 2'\}$. In these states the processing rates of M_d are given as

$$\mu^d = \begin{bmatrix} 0 & \mu_d & 2\mu_d \end{bmatrix}.$$

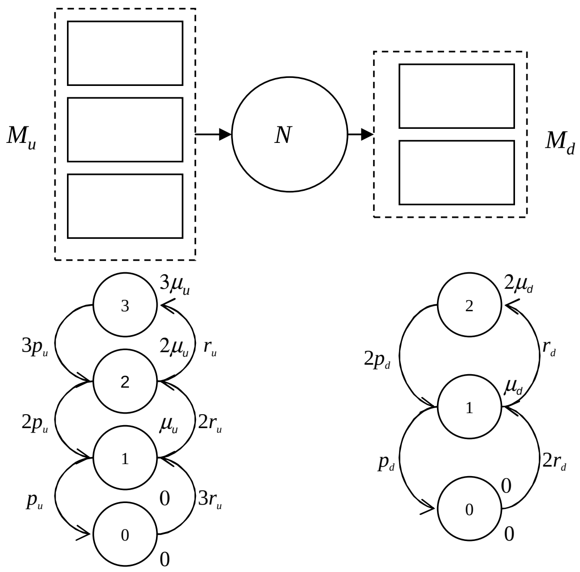

Figure 7: Modelling of a system with parallel machines for the analysis by using the general methodology

There are a total of twelve states in the state space. Once these inputs are given, the methodology described above yields the desired performance measures directly. Figure 8 shows the effect of the number of parallel stations on the production rate and the expected buffer level. In this specific case, the production rate of the second stage is kept equal to the production rate of the first stage as the number of parallel stations in the second stage increases. The figures shows that as the number of parallel stations increase both the production rate and the expected buffer level increases.

6.2 A Model with a Shared Buffer

We now consider a three station merge system with a shared buffer. This system was analyzed in detail in (Tan 2001). Helber and Jusic (2004) also analyzes a similar system. In the upstream stage, there are two unreliable machines with processing rates μ_1 and μ_2. In the downstream stage, there is only one machine with processing rate μ_3. The failure and repair rates for each machine are p_i and r_i for $i = 1, 2, 3$. Figure 9 depicts the state transitions for M_u and M_d for this specific case.

Similar to the first example, we will specify the matrices λ^u and λ^d and the vectors μ^u and μ^d as the inputs of the solution methodology. The transition rates for M_u are given as

$$\lambda^u = \begin{bmatrix} -p_1 - p_2 & p_2 & p_1 & 0 \\ r_2 & -p_1 - r_2 & 0 & p_1 \\ r_1 & 0 & -p_2 - r_1 & p_2 \\ 0 & r_1 & r_2 & -r_1 - r_2 \end{bmatrix} \tag{56}$$

where the states are ordered as $\{11, 10, 01, 00\}$. The processing rates in these states are

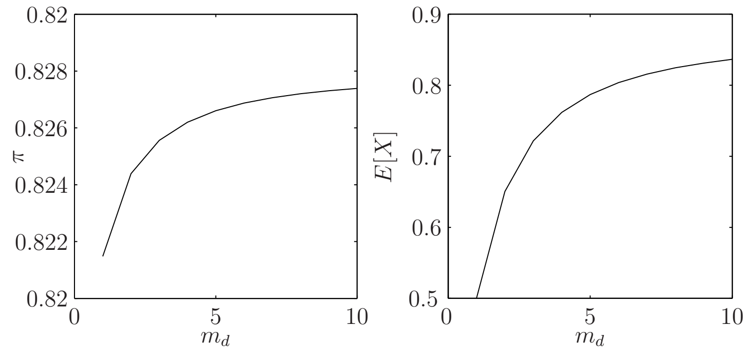

Figure 8: Effect of the number of parallel machines ($\mu_u = 1$, $p_u = 0.01$, $r_u = 0.09$, $m_u = 1$, $\mu_d = \mu_u \frac{m_u}{m_d}$, $p_d = 0.01$, $r_d = 0.09$, $N = 1$)

$$\mu^u = \left[\begin{array}{cccc} \mu_1 + \mu_2 & \mu_1 & \mu_2 & 0 \end{array} \right].$$

Similarly,

$$\lambda^d = \left[\begin{array}{cc} -p_3 & p_3 \\ r_3 & -r_3 \end{array} \right] \tag{57}$$

where the states are ordered as $\{1, 0\}$. In these states the processing rates of M_d are given as

$$\mu^d = \left[\begin{array}{cc} \mu_3 & 0 \end{array} \right].$$

There are eight discrete states in the state space. Once these inputs are given, the methodology described above yields the desired performance measures directly. We compare this case with the results given in (Tan 2001). Since a specific case with hot standby is analyzed in (Tan 2001), the method described above is modified accordingly. Figure 10 shows the effect of μ_3 on the production rate and the expected buffer level obtained by using the methodology given here and the results in (Tan 2001) that are equal to each other.

6.3 A Model with Erlang Up and Down Times

We now model a production system where the failure and repair times are Erlang-type random variables. We assume that the failure time of M_u is an Erlang random variable with κ_f^u stages. The expected failure time is $MTTF_u$ and the squared coefficient of variation of the failure time is $scv_f^u = 1/\kappa_f^u$. The repair time of M_u is also an Erlang random variable with κ_r^u stages. The expected failure time is $MTTR_u$ and the squared coefficient of variation of the failure time is $scv_r^u = 1/\kappa_r^u$.

Similarly, failure time of M_d is an Erlang random variable with κ_f^d stages. The expected failure time is $MTTF_d$ and the squared coefficient of variation of the failure time is $scv_f^d = 1/\kappa_f^d$. The

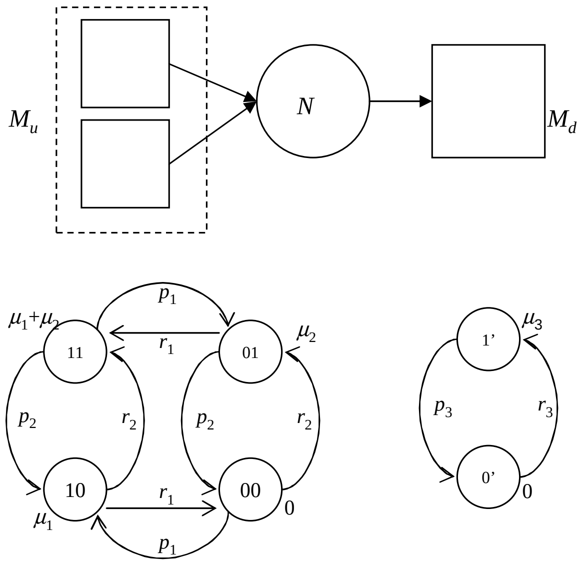

Figure 9: Modelling of a system with a shared buffer for analysis by using the general methodology

repair time of M_d is also an Erlang random variable with κ_r^d stages. The expected failure time is $MTTR_d$ and the squared coefficient of variation of the failure time is $scv_r^d = 1/\kappa_r^d$.

The processing rates of M_u and M_d are μ_u and μ_d respectively. In this model M_u has $\kappa_f^u + \kappa_r^u$ states and M_d has $\kappa_f^d + \kappa_r^d$ states. The states are indexed from 1 to $\kappa_f^k + \kappa_r^k$ and ordered such that states $i = 1, ..., \kappa_f^k$ are for the up states and states $i = \kappa_f^k + 1, ..., \kappa_f^k + \kappa_r^k$ are for the down states of M_k, $k = u, d$.

Possible transitions for M_k are

- from up state i to up state $i+1$ with rate $p_k = \kappa_f^k / MTTF_k$, $i = 1, ..., \kappa_f^k$,
- from down state i to down state $i+1$ with rate $r_k = \kappa_r^k / MTTR_k$, $i = \kappa_f^k + 1, ..., \kappa_f^k + \kappa_r^k - 1$,
- from state κ_f^k to state $\kappa_f^k + 1$ with rate p_k, and
- from state state $\kappa_f^k + \kappa_r^k$ to state 1 with rate r_k.

For example, let us consider a specific case with $\kappa_f^u = 2$, $\kappa_r^u = 2$, $\kappa_f^d = 1$, and $\kappa_r^u = 3$. For this specific system, Figure (11) depicts the state transition diagram.

The matrices λ^u and λ^d and the vectors μ^u and μ^d for this specific case are given below:

$$\lambda^u = \begin{bmatrix} -p_u & p_u & 0 & 0 \\ 0 & -p_u & p_u & 0 \\ 0 & 0 & -r_u & r_u \\ r^u & 0 & 0 & -r_u \end{bmatrix}, \tag{58}$$

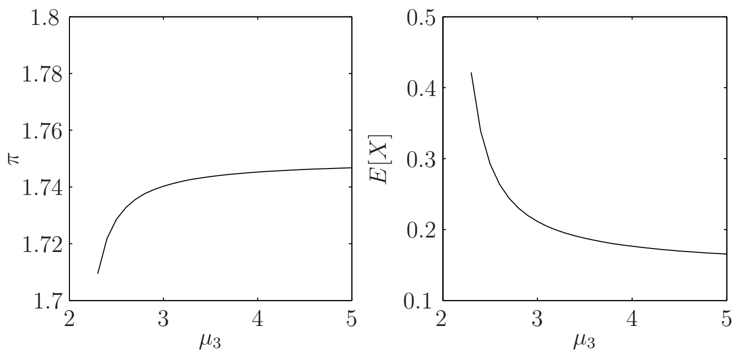

Figure 10: Effect of the processing rate ($\mu_1 = 1.2$, $\mu_2 = 1$, $p_1 = 0.1$, $p_2 = 0.1$, $p_3 = 0.2$, $r_1 = 0.9$, $r_2 = 0.9$, $r_3 = 0.9$, $N = 1$)

$$\mu^u = \left[\begin{array}{cccc} \mu_u & \mu_u & 0 & 0 \end{array} \right],$$

$$\lambda^d = \left[\begin{array}{cccc} -p_d & p_d & 0 & 0 \\ 0 & -r_d & r_d & 0 \\ 0 & 0 & -r_d & r_d \\ r^d & 0 & 0 & -r_d \end{array} \right], \tag{59}$$

$$\mu^d = \left[\begin{array}{cccc} \mu_d & 0 & 0 & 0 \end{array} \right].$$

Figures 12 and 13 show the effects of the failure and repair time variabilities of each stage on the production rate and the expected buffer level. Figure 12 shows that as the coefficient of variation of the failure times of first and the second stages increase, the production rate decreases. On the other hand, a decrease in the variability of the failure time of the upstream machine results in an increase in the expected buffer level. Similarly, Figure 13 shows the effect of the repair time variability of the firs and the second stage on the production rate and the expected buffer level. A decrease in repair time variability of either stage increases the production rate. On the other hand, a decrease of the repair time variability of only the first stage increases the expected buffer level.

6.4 A Model with Series Machines

We now consider a production line where M_u has m_u and M_d has m_d machines in series. The machines are indexed from 1 to $m_u + m_d$. Each machine is unreliable and has one up and one down state. The processing rate of machine i is μ_i and the failure and repair times are exponential random variables with rates p_i and r_i, $i = 1, ..., m_u + m_d$.

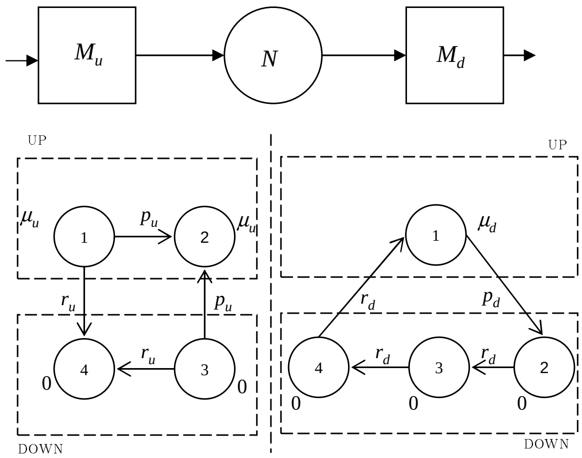

Figure 11: Modelling of a system with Erlang Up and Down times for analysis by using the general methodology

The state of the upstream stage is a tuple with its ith element is 1 if machine i is operational and 0 otherwise. Similarly, the state of the downstream stage is also a tuple with its ith element is 1 if machine $i - m_u$ is operational and 0 otherwise. Accordingly, M_k has 2^{m_k} states, $k = u, d$.

Since each stage is operational only when all the machines are up, the first stage produces at the rate of $\mu_u = \min\{\mu_1, ..., \mu_{m_u}\}$ when all the stations are up and it does not produce in all the other states. Similarly, the second stage produces at the rate of $\mu_d = \min\{\mu_{m_u+1}, ..., \mu_{m_u+m_d}\}$ when all the stations are up and it does not produce in all the other states.

From a given state, there are m_u possible transitions in the upstream stage and m_d possible transitions in the downstream stage. Each transition corresponds to a failure with rate $p_i \frac{\mu_k}{\mu_i}$ or a repair with rate r_i in one of its machines, $k = u, d$. Since, a machine can fail only when it is operational, it cannot fail if one of the other machines is down. As a result, although there are 2^{m_k} states for stage k, only $m_k + 1$ of them will be non-transient.

Figure 14 depicts the state transitions for M_u and M_d for a specific case where M_u has 3 machines and M_d has 2 machines in series.

The matrices λ^u and λ^d and the vectors μ^u and μ^d for this specific case are given below:

$$\lambda^u = \begin{bmatrix} -\mu_u(\frac{p_1}{\mu_1} + \frac{p_2}{\mu_2} + \frac{p_3}{\mu_3}) & p_1\frac{\mu_u}{\mu_1} & p_2\frac{\mu_u}{\mu_2} & p_3\frac{\mu_u}{\mu_3} \\ r_1 & -r_1 & 0 & 0 \\ r_2 & 0 & -r_2 & 0 \\ r_3 & 0 & 0 & -r_3 \end{bmatrix} \tag{60}$$

where the states are ordered as $\{(1,1,1), (0,1,1), (1,0,1), (1,1,0)\}$. The processing rates in these states are

$$\mu^u = \begin{bmatrix} \mu_u & 0 & 0 & 0 \end{bmatrix}$$

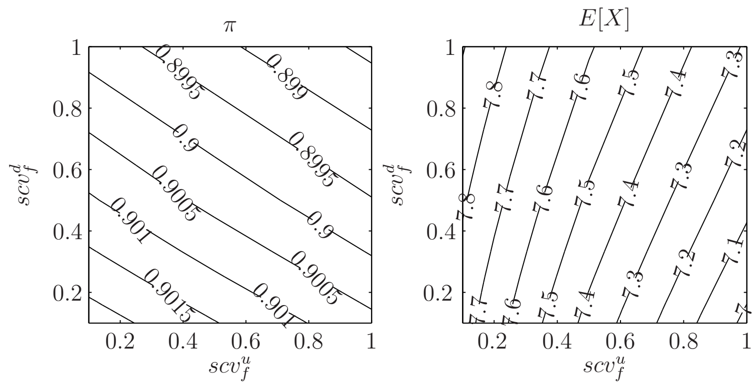

Figure 12: Effect of the failure time variability ($\mu_u = 1$, $\mu_d = 1$, $MTTF_u = 200$, $MTTF_d = 100$, $MTTR_u = 6.67$, $MTTR_d = 10$, $N = 10$)

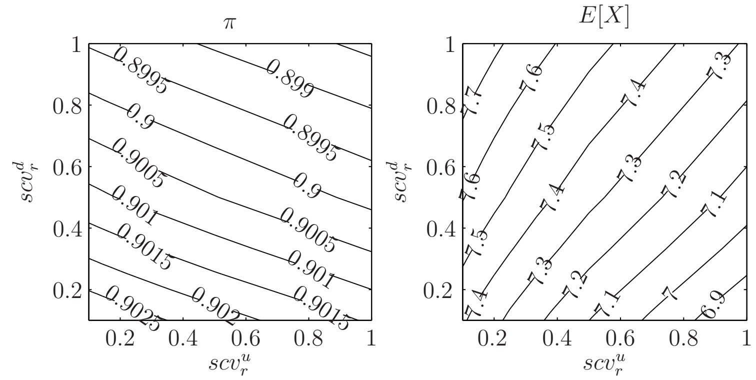

Figure 13: Effect of the repair time variability ($\mu_u = 1$, $\mu_d = 1$, $MTTF_u = 200$, $MTTF_d = 100$, $MTTR_u = 6.67$, $MTTR_d = 10$, $N = 10$)

where $\mu_u = \min\{\mu_1, \mu_2, \mu_3\}$. Similarly

$$\lambda^d = \begin{bmatrix} -\mu_d(\frac{p_4}{\mu_4} + \frac{p_5}{\mu_5}) & p_4\frac{\mu_d}{\mu_4} & p_5\frac{\mu_d}{\mu_5} \\ r_4 & -r_4 & 0 \\ r_5 & 0 & -r_5 \end{bmatrix} \tag{61}$$

where the states are ordered as $\{(1,1),(1,0),(0,1)\}$. In these states the processing rates of M_d are given as

$$\mu^d = \begin{bmatrix} \mu_d & 0 & 0 \end{bmatrix}$$

where $\mu_d = \min\{\mu_4, \mu_5\}$. There are a total of twelve states in the state space. Once these inputs are given, the methodology described above yields the desired performance measures directly.

Consider the problem of locating a finite buffer in a continuous material flow production line with no interstation buffers. Once the buffer is located between machine k and $k+1$, the line is

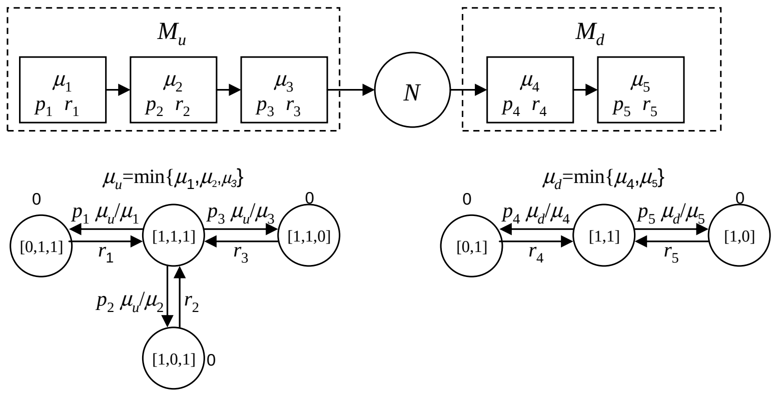

Figure 14: Modelling of a system with series machines for analysis by using the general methodology

divided into two stages. The resulting two-stage system can be analyzed by using the methodology outlined above. Figure 15 shows the effect of the buffer placement on the production rate for a production line with ten identical stations. As expected, for this homogeneous system placing the buffer in the middle, between Machine 5 and 6 maximizes the production rate.

However, when the machines are not identical, the buffer location that maximizes the production rate can be different. Figure 16 shows the effect of the buffer placement on the production rate for a production line with ten non-identical stations. In this case, placing the buffer between Machine 5 and 6 maximizes the production rate.

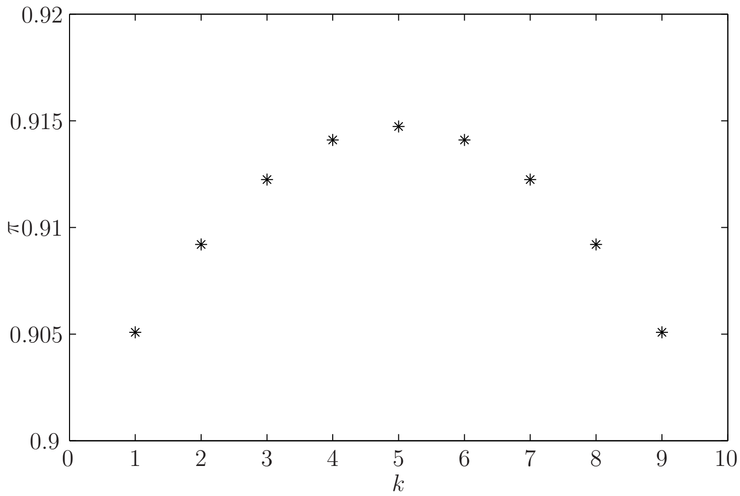

Figure 15: Effect of the buffer placement on the production rate ($\mu_i = 1$, $p_i = 0.01$, $r_i = 0.9$, $i = 1, ..., 10$, $N = 1$)

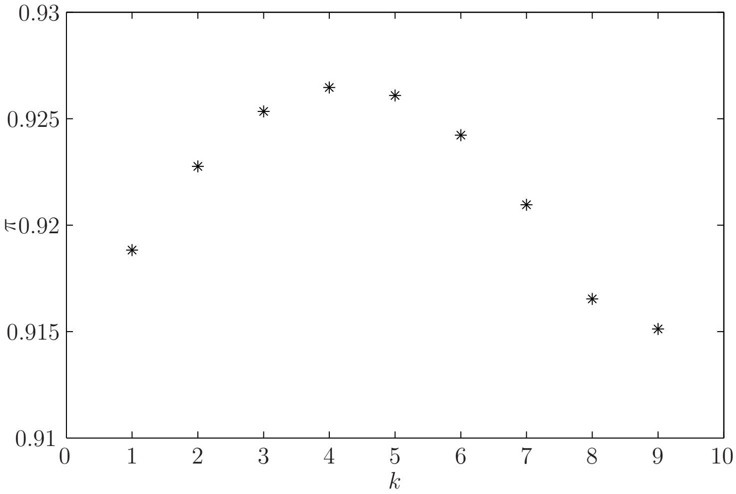

Figure 16: Effect of the buffer placement on the production rate ($p_i = 0.01$, $r_i = 0.9$, $i = 1 : 10$, $\mu_j = 1$, $j = 1 : 8$, $\mu_k = 4$, $k = 9, 10$, $N = 1$)

7 Conclusion

We presented a general methodology to analyze continuous-flow material flow two stage-single buffer production systems. The method handles general Markovian transitions and different processing rates associated with each state for both stages.

A wide range of models can be analyzed by our methodology directly by determining the transition rates of each stage and the flow rates associated with the discrete states of each stage. We illustrated the generality of our method by showing how a number of different models analyzed in the literature can be handled by using our general methodology. We validated all the results with simulation and observed that the percentage error between the simulated and analytical results is less than 10^{-5} when the system is simulated for 10^6 events. The run time of the methodology is very fast and not affected by the buffer size.

Our methodology can also be used in performance evaluation of computer and telecommunication systems. Since the operation-dependent failure mechanism differentiates the models of production and computer/telecommunication models, setting the operation dependent failure rates equal to the original rates in our methodology allows us to use the same tool in the performance evaluation of computer and telecommunication systems.

Therefore we propose our model as a general tool to model and analyze single buffer fluid flow systems.

Acknowledgement

The first author acknowledges the support from TÜBA and TÜBİTAK. The second author acknowledges the support from the Singapore-MIT Alliance and the General Motors Research and

Development Center.

References

Ahn, S., J. Jeon, and V. Ramaswami (2005). Steady state analysis for finite fluid flow models using finite QBDs. *Queueing Systems 49*, 223259.

Ahn, S. and V. Ramaswami (2003). Fluid flow models and queues: a connection by stochastic coupling. *Stochastic Models 19*(3), 325–348.

Altıok, T. (1985). On the phase-type approximation of general distributions. *AIIE Transactions 17*, 110–116.

Anick, D., D. Mitra, and M. Sondhi (1982). Stochastic theory of a data handling system with multiple sources. *Bell System Technology Journal 61*, 1871–1894.

Dallery, Y. (1994). On modeling failure and repair times in stochastic models of manufacturing systems using generalized exponential distributions. *Queueing Systems Theory Applications 15*(1-4), 199–209.

Diamantidis, A., C. Papadopoulos, and M. Vidalis (2004). Exact analysis of a discrete material three-station one-buffer merge system with unreliable machines. *International Journal of Production Research 42*(4), 651–675.

Dubois, D. and J. P. Forestier (1982). Productivité et en-cours moyens d'un ensemble de deux machines séparées par une zone de stockage. *RAIRO Automatique 16*, 105–132.

Elwalid, A. and D. Mitra (1991). Analysis and design of rate-based congestion control of high speed networks, i: stochastic fluid models, access regulation. *Queueing Systems Theory Applications 9*, 19–64.

Gershwin, S. and S. Fallah-Fini (2007). A general model and analysis of a discrete two-machine production line. In *Analysis of Manufacturing Systems 2007*.

Gershwin, S. B. and I. C. Schick (1980). Continuous model of an unreliable two-machine material flow system with a finite interstage buffer. Report LIDS-R-1039, MIT Laboratory for Information and Decision Systems.

Helber, S. and H. Jusic (2004). A new decomposition approach for non-cyclic continuous material flow lines with a merging flow of material. *Annals of Operations Research 125*, 117139.

Inman, R. R., D. E. Blumenfeld, N. Huang, and J. Li (2003). Designing production systems for quality: research opportunities from an automotive industry perspective. *International Journal of Production Research 41*(9), 1953–1971.

Kim, J. and S. B. Gershwin (2005). Integrated quality and quantity modeling of a production line. *OR Spectrum 2*, 287–314.

Levantesi, R., A. Matta, and T. Tolio (2003). Performance evaluation of continuous production lines with machines having different processing times and multiple failure modes. *Performance Evaluation 51*, 247–268.

Mitra, D. (1988). Stochastic theory of a fluid model of multiple failure-susceptible producers and consumers coupled by a buffer. *Advances in Applied Probability 20*, 646–676.

Patchong, A. and D. Willaeys (2001). Modeling and analysis of an unreliable flow line composed of parallel-machine stages. *IIE Transactions 33*, 559–568.

Poffe, A. and S. Gershwin (2005). Integrating quality and quantity modelling in a production line. Technical Report 377-05, Operations Research Center Working Paper Series.

Serucola, B. (2001). A finite buffer fluid queue driven by a markovian queue. *Queueing Systems 38*, 213–220.

Soares, A. D. S. and G. Latouche (2006). Matrix-analytic methods for fluid queues with finite buffers. *Performance Evaluation 63*, 295314.

Tan, B. (2001). A three station continuous materials flow merge system with unreliable stations and a shared buffer. *Mathematical and Computer Modelling 33*(8-9), 1011–1026.

Tempelmeier, H. and M. Burger (2001). Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production. *IIE Transactions 33*, 293 – 302.

Wijngaard, J. (1979). The effect of interstage buffer storage on the output of two unreliable production units in series with different production rates. *AIIE Transactions 11*(1), 42–47.

Yeralan, S., W. Franck, and M. A. Quasem (1986). A continuous materials flow production line model with station breakdown. *European Journal of Operational Research 27*, 289–300.

Yeralan, S. and B. Tan (1997). A station model for continuous materials flow production. *International Journal of Production Research 35*(9), 2525–2541.