Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Analysis of multi-server two-stage queueing network with split and blocking

Yang Woo Shin*. Dug Hee Moon**

**Changwon National University, Changwon, Gyeong-nam, KOREA*
(Tel:82-55-213-3462; e-mail: ywshin@ changwon.ac.kr).
*** Changwon National University, Changwon, Gyeong-nam, KOREA*
(e-mail:dhmoon@ changwon.ac.kr)

Abstract: Parallel lines are often used to increase production rate in manufacturing system or model the multiple-product manufacturing systems. In parallel systems, the main line splits into several lines in parallel, and after some operations, they merge into a main line again. This paper provides an approximation technique for multi-server two-stage networks with splits and blocking which will be a building block for analysis of the parallel system with splits.

Keywords: queueing network, multi-product manufacturing system, multi-server, split configuration

1. INTRODUCTION

Parallel lines are often used to increase production rate in manufacturing system or model the multiple-product manufacturing systems. In parallel systems, the main line splits into several lines in parallel, and after some operations, they merge into a main line again. For example, three types of lens module installed in smart phone are assembled in their own assembly line. After finishing assembly, all lens modules are sent to the resolution inspection machines which can inspect all types of lens modules automatically. Defective parts are separated from the main line to be either repaired or scraped.

There is an extensive literature on the analysis of flow lines for modelling single-product manufacturing system. For a detailed list of references for approximations and their application to manufacturing system, see review papers Dallery and Gershwin (1992), Papadopoulos and Heavey (1996), Li et al. (2009) and monographs Buzacott and Shantikumar (1993), Gershwin (1994), Altiok (1997), Li and Meerkov (2009). The study of the systems with split and/or merge configuration for multiple-product manufacturing system is relatively scarce, e.g. see Helber (2000) for the system with splits, Helber and Jusic (2004) for the system with merge and Colledani et al. (2005), Li (2005), Li and Huang (2005), Liu and Li (2009) for the system with merge and splits. Most of the approximation methods for flow lines and the system with merge and/or splits are either based on decomposition or on aggregation. The principal procedure of decomposition methods is to decompose the long line into subsystems with two-station that are mathematically tractable, and derives a set of equations that determine the unknown parameters of each subsystem, and finally develops an algorithm to solve these equations. The basic idea of aggregation method is to replace a two-station system by a single equivalent station and reduce the number of stations in the network. Thus an analysis of queueing network with two-station and one buffer is a building block in analysis of long line using decomposition or aggregation method. A few papers deal with a three-station merge system as a building block for an analysis of the system with merge configuration (see Tan (2001), Helber and Mehrtens (2003), Diamantidis et al. (2004), Diamantidis and Papadopoulos (2006)).

There is a tremendous literature on the analysis of flow lines with single server at each service stage. However, there is a little literature for analysis of the system with multiple servers (van Vuuren et al., 2005, Diamantidis et al., 2007). Recently, Shin and Moon (unpublished) presents a decomposition method for multi-server flow line with exponential service times and finite buffers. They use the subsystems with two service stations and two buffers to reflect the dependence between consecutive stages and improve the quality of approximation. To the best of our knowledge, there is little literature for analysis of the multiple-product manufacturing systems with multiple servers at each service station. As mentioned above, it is necessary to analyse two-stage networks for an approximation of such queueing networks using decomposition method. The purpose of this paper is to provide an approximation method for multi-server two-stage networks with split and blocking which will be a building block for analysis of multi-server-finite-buffer queueing networks with parallel lines and split configuration.

This paper is organized as follows. In Section 2, the model is described in detail. Approximation procedure is presented in Section 3. Numerical results and concluding remarks are presented in Sections 4 and 5, respectively.

2. MODEL DESCRIPTION

We consider the two-stage open queueing network in which the first stage is one queue with multiple servers and finite buffer and the second stage consists of J multi-server queues with finite buffer. Two stages are linked in split configuration as depicted in Fig. 1. The queue (split queue) at the first stage is denoted by S_0 and is called node 0 and the queues at the second stage will be denoted by S_i called node i, i=1,2,…,J. The node i consists of a service station M_i with m_i parallel identical servers and a buffer B_i of size K_i and let N_i=m_i+K_i, i=0,1,2,…,J. The index i can be viewed as the sequential number of the nodes at the second stage runs from 1 to J unless otherwise stated. The service time of each server at M_i is exponentially distributed with rate μ_i, i=0,1,2,…,J. Customers arrive to node 0 from outside according to a Poisson process with rate λ. When an arrival finds no free position in the buffer B_0 (buffer is full), it is lost. After completion of the service at node 0, the request proceeds to node i with probability p_i and is taken immediately into service if one of the servers at M_i is free, otherwise is placed into the buffer B_i and waits there until a server is available. If buffer B_i is full at the instant of service completion at node 0, the customer is forced to stay at node 0 until a space becomes available at node i. During this period, the server just completed its service is blocked. This type of blocking mechanism is called blocking-after-service (BAS). A customer who has completed its service at node i leaves the system.

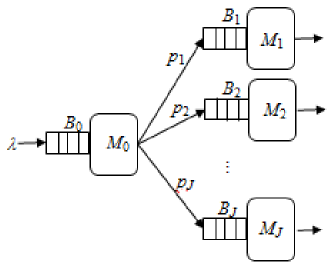

Fig. 1. A two-stage queueing network with split

Let X_0 be the number of customers at node 0 which includes the customers being served at M_0 and waiting requests at B_0 but does not include the customers blocked to enter the downstream node i, i =1,2,…,J, in stationary state. By X_i denote the number of customers at node i that include the customers blocked at node 0 and Y_i the number of customers who are blocked to enter into node i, i =1,2,…,J. Note that the state space of X_i is {0,1,…, H_i}, where H_i=N_i+m_0.

3. APPROXIMATION

For an investigation of the behaviour of node i and node 0, we consider the two-node tandem queue denoted by L_i where departures from the first stage to outside of the system are allowed as depicted in Fig.2. We denote the first stage and the second stage by S_0 and S_i, respectively. The stage S_0 consists of a service station M_0 with m_0 exponential servers with rate μ_0 in parallel and a buffer B_0 of size K_0. The stage S_i consists of a service station with m_i exponential servers with rate μ_i in parallel and a finite buffer of size K_i. The customers at S_0 may be blocked. The BAS blocking rule is assumed in the system. We assume that each customer at S_0 is classified by type j customer with probability p_j, j=1,2,…,J immediately after its service. Type i customer joins S_i and the customer of type j, $j\neq i$ leaves the system or is blocked after its service at S_0. We assume that the blocking probability of type j, $j\neq i$ customer depends not only on the state h of S_0 but also on the state $\boldsymbol{y}$=(y_1,…,y_J) of the number of blocked customers $\boldsymbol{Y}$=(Y_1,…,Y_J) and denote it by $\beta_j(h,\boldsymbol{y})$.

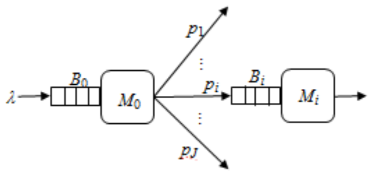

Fig. 2. Two-stage tandem queue L_i

Let $X_0^*(t)$ be the number of customers excluding the blocked one at S_0 and $X_i^*(t)$ the number of customers at S_i and the customers blocked to enter S_i at time t. Denote by $Y_j^*(t)$ the number of type j customers blocked at S_0. Let

$$\boldsymbol{Z}_i^*(t) = (X_0^*(t), X_i^*(t), \boldsymbol{Y}_i^*(t)),$$
$$\boldsymbol{Y}_i^*(t) = (Y_1^*(t), \dots, Y_{i-1}^*(t), Y_{i+1}^*(t), \dots, Y_J^*(t)).$$

The stochastic process $\boldsymbol{Z}_i^* = \{\boldsymbol{Z}_i^*(t), t \geq 0\}$ is a Markov chain. By X_0^*, X_i^* and Y_j^* denote the stationary versions of $X_0^*(t)$, $X_i^*(t)$ and $Y_j^*(t)$, respectively. Note that $0\leq X_0^* \leq N_0$ and for a given $X_0^* = h$, we have that

$$0\leq X_i^*\leq N_i+m_0-\max(h-K_0, 0).$$

Note that $Y_j^*=\max(X_i^*-N_i, 0)$ and for a given (X_0^*, X_i^*),

$$0\leq \boldsymbol{Y}_i^*\mathbf{e}= \sum_{k\neq i} Y_k^* \leq m_0-\max(X_i^*-N_i, 0)-\max(X_0^*-K_0, 0),$$

where $\mathbf{e}=(1,\dots,1)^T$ is column vector of appropriate size whose components are all 1. The state space of $\boldsymbol{Z}_i^*$ is $\boldsymbol{\mathcal{S}}_i=\bigcup_{h=0}^{N_0} \boldsymbol{h}$, where $\boldsymbol{h}$={(h,n,$\boldsymbol{y}$), $0\leq n\leq N_i+m_0-\max(h-K_0, 0)$, $0\leq \boldsymbol{y}\mathbf{e}\leq l(h,n)$} and $l(h,n)=m_0-\max(n-N_i, 0)-\max(h-K_0, 0)$.

The transition rates $q_i\big((h,n,\boldsymbol{y}),(h',n',\boldsymbol{y}')\big)$ of $\boldsymbol{Z}_i^*$ are as follows: for h=0,1,...,N_0 -1,

$$q_i\big((h,n,\boldsymbol{y}),(h',n',\boldsymbol{y}')\big) = \begin{cases} \lambda, & (h',n',\boldsymbol{y}') = (h+1,n,\boldsymbol{y}), \\ \mu_0(h,n,\boldsymbol{y})p_i, & (h',n',\boldsymbol{y}') = (h-1,n+1,\boldsymbol{y}), \\ \mu_0(h,n,\boldsymbol{y})\sum_{j\neq i} p_j\bar{\beta}_j(h,\boldsymbol{y}), & (h',n',\boldsymbol{y}') = (h-1,n,\boldsymbol{y}) \\ \mu_0(h,n,\boldsymbol{y})p_j\beta_j(h,\boldsymbol{y}), & (h',n',\boldsymbol{y}') = \big(h-1,n,\boldsymbol{y}+\mathbf{e}_j\big), j\neq i \\ \mu_i(n), & (h',n',\boldsymbol{y}') = (h,n-1,\boldsymbol{y}) \\ \mu_j(m_j), & (h',n',\boldsymbol{y}') = \big(h,n,\boldsymbol{y}-\mathbf{e}_j\big), j\neq i \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{e}_k$ is the (J-1)-dimensional vector whose kth component is 1 and others are all 0, $\bar{\beta}_j(h,\boldsymbol{y}) = 1-\beta_j(h,\boldsymbol{y})$,

$$\mu_0(h,n,\boldsymbol{y})=\min(h,\ m_0\text{-}\max(n\text{-}N_i,\ 0)\text{-}\boldsymbol{y}\mathbf{e})\mu_0,$$

$$\mu_j(n)=\min(n,\ m_j)\mu_j,\ j=1,2,\ldots,J.$$

Remark. One can consider the system whose routing probability p_j may depend on the number $\boldsymbol{Y}=\boldsymbol{y}$ of customers blocked at node j, j=1,2,…,J by replacing p_j with probability $p_j(\boldsymbol{y})$.

It can be easily seen that the Markov chain $\boldsymbol{Z}_i^*$ is a level dependent quasi-birth-and-death (LDQBD) process whose generator Q_i is of the form

$$\mathrm{Q_i} = \begin{pmatrix} \mathrm{B}_i^{(0)} & \mathrm{A}_i^{(0)} & & & \\ \mathrm{C}_i^{(1)} & \mathrm{B}_i^{(1)} & \mathrm{A}_i^{(1)} & & \\ & \ddots & \ddots & \ddots & \\ & & \mathrm{C}_i^{(N_0-1)} & \mathrm{B}_i^{(N_0-1)} & \mathrm{A}_i^{(N_0-1)} \\ & & & \mathrm{C}_i^{(N_0)} & \mathrm{B}_i^{(N_0)} \end{pmatrix},$$

where the block matrix $B_i^{(h)}$ is the square matrix of size $|\boldsymbol{h}|$, the number of states of level $\boldsymbol{h}$, h=0,1,2,…,N_0 and the components of matrices $A_i^{(h)}$, $B_i^{(h)}$ and $C_i^{(h)}$ are given in terms of $q_i\big((h,n,\boldsymbol{y}),(h',n',\boldsymbol{y}')\big)$. Let

$$\pi_i(h,n,\boldsymbol{y})=\mathrm{P}\big((X_0^*,\ X_i^*,\boldsymbol{Y}_i^*)=(h,n,\boldsymbol{y})\big),\ (h,n,\boldsymbol{y})\in\boldsymbol{\mathcal{S}}_i$$

$$\boldsymbol{\pi}_i(h)=(\pi_i(h,n,\boldsymbol{y}),\ (h,n,\boldsymbol{y})\in\boldsymbol{h}),\ h=0,1,..,N_0.$$

The stationary distribution $\boldsymbol{\pi}_i=(\boldsymbol{\pi}_i(0),\ \boldsymbol{\pi}_i(1),\ldots,\ \boldsymbol{\pi}_i(N_0))$ satisfies the following equations

$$\boldsymbol{\pi}_i(n)=\boldsymbol{\pi}_i(n\text{-}1)\ R_i(n),\ n=1,2,\ldots,N_0,$$

where $\boldsymbol{R}_i(n)$ is obtained recursively as follows

$$R_i(N_0) = A_i^{(N_0-1)}\left(-B_i^{(N_0)}\right)^{-1},$$

$$R_i(n) = A_i^{(n-1)}\left[-\left(B_i^{(n)}+R_i(n+1)C_i^{(n+1)}\right)\right]^{-1},$$

$$n=N_0\text{-}1,\ N_0-2,\ldots,1.$$

The vector $\boldsymbol{\pi}_i(0)$ is the unique solution of the equations

$$\boldsymbol{\pi}_i(0)\ (B_i^{(0)}+R_i(1)\ C_i^{(1)})=0$$

$$\boldsymbol{\pi}_i(0)\ [\mathbf{e}+\ \textstyle\sum_{1\le n\le N_0} R_i(1)R_i(2)\cdots R_i(n)\mathbf{e}]=1.$$

Let Ψ be the number of busy servers at M_0. The mean arrival rate $\gamma_i(n)=\mu_0\mathrm{E}[\Psi|X_i=n]$ from node 0 to node i given X_i=n. Once $\boldsymbol{\pi}_i$ is given, we can calculate the approximation formulae for $\gamma_i(n)$ and $\beta_i(h,\boldsymbol{y})$ as follows:

$$\begin{aligned}\gamma_i(n) &= \mu_0 p_i \mathrm{E}[\Psi\ |X_i^* = n] \\ &= \mu_0 p_i \sum_{\{(h,\boldsymbol{y}):(h,n,\boldsymbol{y})\in S_i^*(n)\}} E[\Psi_i\ |\boldsymbol{Z}_i^*=(n,h,\boldsymbol{y})] \\ &\qquad \times \mathrm{P}(X_0^* = h,\ \boldsymbol{Y}_i^* = \boldsymbol{y}\ |X_i^* = n) \qquad (1) \\ &= \sum_{\{(h,\boldsymbol{y}):(h,n,\boldsymbol{y})\in S_i^*(n)\}} \mu_0(h,n,\boldsymbol{y})p_i\frac{\pi_i(h,n,\boldsymbol{y})}{P(X_i^*=n)},\end{aligned}$$

where $S_i^*(n)=\{(h,k,\boldsymbol{y})\in\boldsymbol{\mathcal{S}}_i : k=n\}$,

$$\begin{aligned}\beta_i(h,\boldsymbol{y}) &= \mathrm{P}(X_i^*\ge N_i|X_0^*=h,\boldsymbol{Y}=\boldsymbol{y}) \\ &= \begin{cases}\mathrm{P}(X_i^*=N_i\ |X_0^*=h,\boldsymbol{Y}=\boldsymbol{y}), & y_i=0 \\ 1, & y_i>0\end{cases} \qquad (2) \\ &= \begin{cases}\pi_i(h,N_i,\boldsymbol{y}^*)/\sum_{n=0}^{N_i}\pi_i(h,n,\boldsymbol{y}^*), & y_i=0, \\ 1, & y_i>0,\end{cases}\end{aligned}$$

where $\boldsymbol{y}^*=(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_J)$.

Throughput from node i is given by

$$TP_i = \sum_{(h,n,\boldsymbol{y})\in\, \boldsymbol{S_i}} \mu_i(n)\pi_i(h,n,\boldsymbol{y}). \qquad (3)$$

We use the iterative algorithm to determine the parameters and calculate the throughput. Let $L_i^{(k)}$, k = 0, 1, 2, … be the system L_i in the kth iteration and by $\beta_i^{(k)}(h,\boldsymbol{y})$ and $\gamma_i^{(k)}(n)$ denote $\beta_i(h,\boldsymbol{y})$ and $\gamma_i(n)$ corresponding to the subsystem $L_i^{(k)}$.

Algorithm

Step 0 [Initial Step]

Stage 1. Initially we assume that no type j customers (j=2,…,J) at node S_0 are blocked and consider the system $L_1^{(0)}$ as two node tandem queue where the type j customers (j=2,…,J) leave the system immediately after completing their service. Calculate $\pi_1^{(0)}(h,n,\mathbf{0})$ with $\beta_j^{(0)}(h,\mathbf{0})$=0, j=2,…,J and then calculate $\beta_1^{(0)}(h,y_1,0,...,0))$ and $\gamma_1^{(0)}(n)$ using (1) and (2), respectively.

Stage *i*. For i=2,3,…,J, calculate $\pi_i^{(0)}(h,n,\boldsymbol{y}_i^*)$ with $\beta_j^{(0)}(h,y_1,...,y_j,0,...,0)$, j=1,2,…,i-1 and $\beta_j^{(0)}(h,\mathbf{0})$=0, j=i+1,…,J and then calculate $\beta_i^{(0)}(h,y_1,...,y_i,0,...,0)$ and $\gamma_i^{(0)}(n)$ using (1) and (2), respectively.

Step 1 For i = 1, 2, … , J, calculate the stationary distribution $\boldsymbol{\pi}_i^{(k)}$ of $L_i^{(k)}$ with $\beta_j^{(k)}(h,\boldsymbol{y})$, j<i and $\beta_j^{(k-1)}(h,\boldsymbol{y})$, j > i. Update $\beta_i^{(k)}(h,\boldsymbol{y})$ and $\gamma_i^{(k)}(n)$ using (1) and (2), respectively.

Step 2 [Stopping Criterion] Check the stopping criterion

$$TOL = \max_{i,h,\boldsymbol{y}} |\gamma_i^{(k)}(n)-\gamma_i^{(k-1)}(n)| < \varepsilon,$$

where ε>0 is a given tolerance. If the stopping condition is satisfied, then stop the iteration and calculate the throughput from node i using (3). Otherwise, go to Step 1.

No analytical proof of convergence or accuracy can be given for this algorithm. However extensive numerical experiments show the convergence of the algorithm. The complexity of this method is as follows. Within the iterative algorithm, calculating $\boldsymbol{\pi}_i(n)$ consumes most of the time. For solving the subsystem L_i, inversion of N_0+1 matrices of size $|\boldsymbol{h}|$, $0\le h\le N_0$ is required. The number of iterations needed is difficult to predict because it depends on the tolerance ε and the length of the line and system parameters.

4. NUMERICAL RESULTS

In this section, some numerical results are presented for the performance of approximations. We consider the system with parameters m_0=3, m_1=1, m_2=2, m_3=3 and the service rates μ_0=1.0, μ_i=1.0/m_i, i=1,2,3. The approximation results (App) for throughput, mean E[X_i] and standard deviation Std[X_i] of the number of customers at node i are numerically compared with those of simulation (Sim) in Tables 1-3 for K_0= 5 and arrival rates λ =1.0 and λ =3.0. Table 1 represents the results for the system with K_i=0, i=1,2,3. Similarly, Tables 2-3 present the results for the system with K_i= 2, i=1,2,3 and K_i=4-i, i=1,2,3, respectively. The maximal number of iterations of the algorithm to obtain each result for the tolerance $\varepsilon=10^{-5}$ in the examples is 5 for λ =3.0.

Simulation models are developed with ARENA (Kelton et al., 1998). Simulation run time is set to 11,000 unit times including 1,000 unit times of warm-up period. Each simulation is repeated sufficiently many times such that the half length of 95% confidence interval for throughput and E[X_i] are smaller than 1%. The relative percentage error is calculated by Err(%)=(App-Sim)×100/Sim. Numerical results show the approximation works very well.

Table 1. Comparisons with Simulation I (K_i=0, i=1,2,3)

λ	node	Throughput			E[X_i]			Std[X_i]		
		App	Sim	Err(%)	App	Sim	Err(%)	App	Sim	Err(%)
1.0	0	0.995	0.996	0.1	1.175	1.182	0.6	1.237	1.244	-0.6
	1	0.199	0.200	0.7	0.246	0.247	0.4	0.545	0.547	-0.3
	2	0.298	0.297	-0.4	0.650	0.648	-0.2	0.861	0.860	0.2
	3	0.497	0.498	0.2	1.661	1.675	0.8	1.389	1.396	-0.5
3.0	0	1.681	1.675	-0.4	5.488	5.489	0.0	1.500	1.494	0.4
	1	0.336	0.336	-0.1	0.489	0.495	1.1	0.795	0.802	-0.8
	2	0.504	0.502	-0.5	1.257	1.244	-1.0	1.241	1.240	0.1
	3	0.841	0.837	-0.4	3.410	3.409	0.0	1.598	1.605	-0.4

Table 2. Comparisons with Simulation II (K_i=2, i=1,2,3)

λ	node	Throughput			E[X_i]			Std[X_i]		
		App	Sim	Err(%)	App	Sim	Err(%)	App	Sim	Err(%)
1.0	0	0.999	1.000	0.1	1.069	1.078	0.9	1.114	1.124	-0.9
	1	0.200	0.201	0.4	0.250	0.249	-0.2	0.558	0.557	0.1
	2	0.300	0.299	-0.3	0.658	0.660	0.4	0.885	0.891	-0.7
	3	0.499	0.501	0.2	1.715	1.724	0.5	1.512	1.511	0.1
3.0	0	1.872	1.868	-0.2	5.303	5.304	0.0	1.731	1.732	0.0
	1	0.374	0.376	0.3	0.625	0.634	1.3	1.004	1.017	-1.3
	2	0.562	0.560	-0.4	1.632	1.653	1.3	1.631	1.664	-1.9
	3	0.936	0.933	-0.4	5.156	5.109	-0.9	1.991	2.025	-1.7

Table 3. Comparisons with Simulation III (K_1=3, K_2=2, K_3=1)

λ	node	Throughput			E[X_i]			Std[X_i]		
		App	Sim	Err(%)	App	Sim	Err(%)	App	Sim	Err(%)
1.0	0	0.998	0.998	0.0	1.092	1.089	0.2	1.142	1.136	0.6
	1	0.200	0.200	-0.1	0.250	0.250	-0.2	0.559	0.557	0.2
	2	0.299	0.299	0.0	0.657	0.663	-0.9	0.885	0.891	-0.7
	3	0.499	0.499	0.0	1.697	1.688	0.5	1.465	1.451	1.0
3.0	0	1.837	1.824	0.7	5.330	5.340	-0.2	1.697	1.688	0.5
	1	0.367	0.366	0.3	0.619	0.627	-1.4	1.023	1.045	-2.1
	2	0.551	0.544	1.2	1.586	1.585	0.1	1.610	1.626	-1.0
	3	0.918	0.913	0.6	4.440	4.433	0.2	1.749	1.766	-1.0

5. CONCLUSION

In this paper, we developed an approximation method for a multi-server two-stage queueing network with finite buffers and split. The system is decomposed into subsystems that have two service stations with multiple servers and two buffers and departures from the first stage to outside of the system are allowed. An iterative algorithm is presented for calculating the parameters. The approximation method proposed can be used to analyse the complex queueing networks with split configuration. Furthermore, the analysis of multi-server system can be used for allocation of the servers and buffers in design of manufacturing systems.

The analysis of the system with more complex splitting rule, the server failure, more complex arrival process and/or more general service distribution than exponential and the long system with split configuration are left as future research area.

Acknowledgement. The first and second authors were supported by Basic Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Grant numbers NRF-2012R1A1B3004158 and NRF-2013R1A1A2058943, respectively.

REFERENCES

Altiok, T. (1997). *Performance Analysis of Manufacturing Systems*. Springer, New York.

Buzacott, J.A. and Shantikumar, J.G. (1993). *Stochastic Models of Manufacturing Systems*. Prentice Hall, Engelwood Cliffs, NJ.

Colledani, M., Matta, A., and Tolio, T. (2005). Performance evaluation of production lines with finite buffer capacity producing two different products. *OR Spectrum*, 27, 243–263.

Dallery, Y. and Gershwin, S.B. (1992). Manufacturing flow line systems: A review of models and analytical results. *Queueing Systems*, 12, 3–94.

Diamantidis, A.C. and Papadopoulos, C.T. (2006). Markovian analysis of a discrete material manufacturing system with merge operations, operation-dependent and idleness failures. *Computers & Industrial Engineering*, 50, 466–487.

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

Diamantidis, A.C., Papadopoulos, C.T. and Heavey, C. (2007), Approximate analysis of serial flow lines with multiple parallel-machine stations, *IIE Transaction*, 39, 361-375.

Gershwin, S.B. (1994). *Manufacturing Systems Engineering*. Prentice Hall, Engelwood Cliffs, NJ.

Helber, S. (2000). Approximate analysis of unreliable transfer lines with splits in the flow of material. *Annals of Operations Research*, 93, 217–243.

Helber, S. and Jusic, H. (2004). A new decomposition approach for non-cyclic continuous material flow lines with a merging flow of material. *Annals of Operations Research*, 125, 117–139.

Helber, S. and Mehrtens, N. (2003). Exact analysis of continuous material merge system with limited buffer capacity and three stations. In: S. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and Modeling of Manufacturing Systems,* 85–85, Kluwer Academic, Dordrecht.

Kelton, W.D., Sadowski, R. P. and Sadowski, D. A. (1998). *Simulation with ARENA*, 2nd Ed. McGraw-Hill, New York.

Li, J., (2005) Overlapping decomposition: A system-theoretic method for modeling and analysis of complex manufacturing. *IEEE Transactions on Automation Sciences and Engineering*, 2, 40–53.

Li, J. and Meerkov, S. M. (2009). *Production Systems Engineering*, Springer.

Li, J., Blumenfeld, D.E., Huang, N., and Alden, J.M. (2009). Throughput analysis in production systems: recent advances and future topics. *International Journal of Production Research*, 47(14), 3823-3851.

Li, J., Huang, N. (2005). Modelling and analysis of a multiple product manufacturing system with split and merge. *International Journal of Production Research*, 43 (19), 4049-4066.

Liu, Y., Li, J. (2009). Modelling and analysis of split and merge production systems with Bernoulli reliability machines, *International Journal of Production Research*, 47(16), 4373-4397.

Papadopoulos, H.T. and Heavey, C. (1996). Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines, *European Journal of Operational Research*, 92, 1–27.

Shin, Y.W. and Moon, D. H. Approximation of throughput in tandem queues with multiple servers and blocking. submitted for publication.

Tan, B. (2001). A three-station merge system with unreliable stations and a shared buffer. *Mathematical and Computer Modeling*, 33, 1011–1026.

van Vuuren, M., Adan, I.J.B.F. and Resing-Sassen, S.A.E. (2005). Performance analysis of multi-server tandem queues with finite buffers and blocking. *OR Spectrum*, 26, 315-338.