International Journal of Production Research
Vol. 47, No. 14, 15 July 2009, 3823–3851

Throughput analysis of production systems: recent advances and future topics

Jingshan Li[a]*, Dennis E. Blumenfeld[b], Ningjian Huang[b] and Jeffrey M. Alden[b]

[a]*Department of Electrical and Computer Engineering, and Center for Manufacturing, University of Kentucky, Lexington, USA;* [b]*Manufacturing Systems Research Laboratory, General Motors Research & Development Center, Warren, USA*

(*Received 15 July 2007; final version received 6 November 2007*)

Throughput analysis is important for the design, operation and management of production systems. A substantial amount of research has been devoted to developing analytical methods to estimate the throughput of production systems with unreliable machines and finite buffers. In this paper we summarise the recent studies in this area. In addition to the performance evaluation of serial lines, approximation methods for more complex systems, such as assembly/disassembly systems, parallel lines, split and merge, closed-loop systems, etc., are discussed. Moreover, we propose future research topics from the automotive manufacturing systems perspective.

Keywords: manufacturing systems; production modelling

1. Introduction

1.1 *Motivation*

Throughput analysis is important for the design, operation and management of production systems. The purpose of this paper is to summarise the recent advances in analytical models for the throughput analysis of production systems with unreliable machines and finite buffers in a large-volume manufacturing environment and propose future research topics from the automotive industry perspective.

Due to the randomness in production (machine breakdowns, random processing times, etc.), the number of parts produced by a production system is a random variable. Its expectation, i.e. throughput, characterises the production volume. Often, this measure is normalised as the production rate (PR, i.e. the average number of parts produced by the last machine in the production system per unit of time) or line efficiency. Analysis of the throughput of production systems has attracted much attention in the last 50 years (see reviews by Koenigsberg (1959), Buxey *et al.* (1973), Buzacott and Hanifin (1978), Dallery and Gershwin (1992), Govil and Fu (1996), and Papadopoulos and Heavey (1996), monographs by Viswanadham and Narahari (1992), Buzacott and Shanthikumar (1993), Papadopoulos *et al.* (1993), Gershwin (1994), Altiok (1997) and Li and Meerkov (2007), the edited volume of Gershwin *et al.* (2003) and the Proceeding Series of Conferences on Analysis of Manufacturing Systems (1997, 1999, 2001, 2003, 2005, 2007)). However, exact analytical results only exist for the two-machine–one-buffer system and

---

*Corresponding author. Email: jingshan@engr.uky.edu

ISSN 0020–7543 print/ISSN 1366–588X online

DOI: 10.1080/00207540701829752
http://www.informaworld.com

systems with infinite buffer capacity or without buffers. Li *et al.* (2006) summarise and compare eight different two-machine line models and show that a similar performance in throughput can be observed for all models. For longer lines and assembly systems, the analytical results for the two-machine line have been used as a building block for the approximation of system performance. A review of serial line models can be found in Dallery and Gershwin (1992), Papadopoulos and Heavey (1996) and Govil and Fu (1996). To extend these studies and to outline the recent advances in the throughput analysis of more complex systems, this paper concentrates on investigating and summarising recent analytical work on complex systems (for example, assembly systems, parallel lines, closed-loop systems, etc.). Moreover, we present future research topics from the automotive manufacturing system perspective. Clearly, many of these are also applicable to other manufacturing industries, such as appliance, semiconductor, electronics, etc.

1.2 *Problem addressed*

The analysis of production systems in large-volume manufacturing can be categorised into the study of systems with reliable machines and with unreliable machines, both with finite and infinite buffer capacities (see Figure 1). Since most large-volume manufacturing systems consist of automated machines with breakdowns and finite inventory, we limit our discussion to the analysis of production systems with unreliable machines and finite buffer capacities. Random processing times are applicable to manual operations, whereas constant cycle times are more suitable for automated production lines. The latter is the main focus of this paper. In addition, systems with fixed processing times can be divided into homogeneous (where all machines have identical speeds) and inhomogeneous (where machines have different speeds) lines. Both cases are discussed in this paper. (Note that, in Figure 1, research issues within the categories of reliable machines and random processing times are omitted.)

From the model point of view, we can categorise the production system models into discrete and continuous models. In terms of machine reliability, the discrete models consist of the Bernoulli machine reliability model and geometric machines, whereas the continuous models include exponential machine models and non-exponential lines (see Figure 2), both including homogenous and inhomogeneous cases.

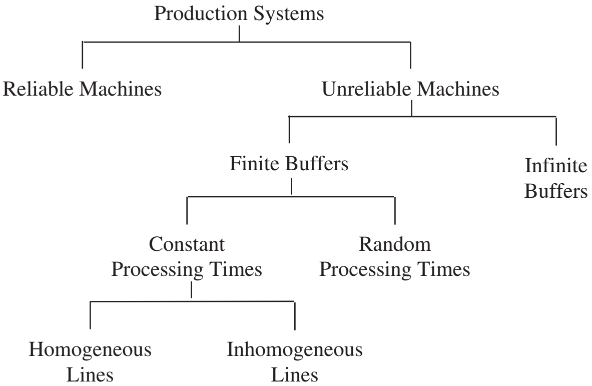

Figure 1. Categorisation of production systems in throughput analysis.

The topics covered in this paper (illustrated in Figure 3) include: serial lines (or flow lines, transfer lines, etc.), assembly/disassembly systems, parallel systems, split, merge, scrapping, and closed-loop systems (rework loops, loops with a constant number of carriers), etc. Such systems are widely encountered in the automotive and other manufacturing industries. Table 1 illustrates some typical line structures in automotive manufacturing plants.

Due to space limitation, only the main idea and a list of references are presented in this paper. Also note that this paper focuses on large-volume manufacturing systems, thus job shops are not in the scope of the paper. The intention of the paper is not to compare the superiority of the approximation methods or to justify the conclusions of the papers referred to, but to provide a platform or forum to list the available results and discuss future opportunities from the automotive industry perspective. Therefore, algorithm preferences or a comparison of the accuracies are not pursued in this paper. For a specific

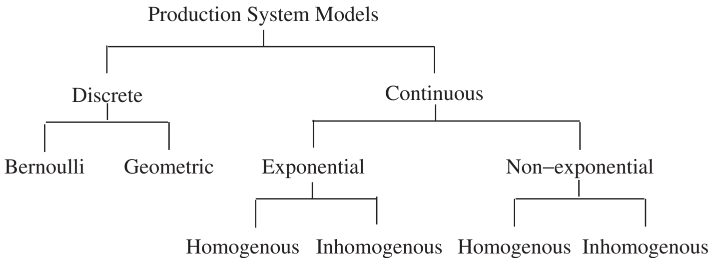

Figure 2. Machine reliability models in throughput analysis.

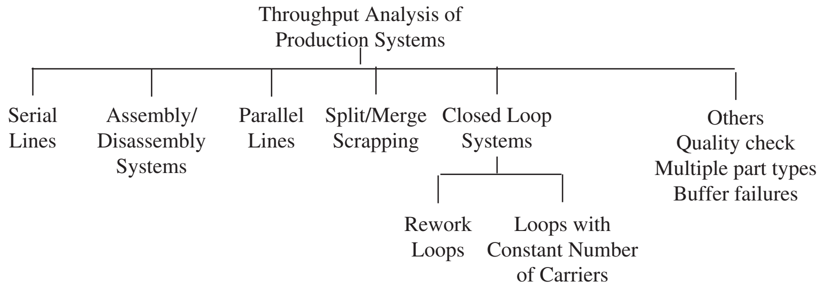

Figure 3. Topics covered.

Table 1. Typical line structures in automotive manufacturing plants.

	Serial lines	Assembly systems	Parallel lanes	Finite buffers	Closed loops with carriers	Repair loops	Split/merge
Stamping	✓		✓	✓			✓
Machining	✓	✓	✓	✓	✓	✓	✓
Engine/transimission	✓	✓	✓	✓	✓	✓	✓
Body shop	✓	✓	✓	✓	✓		✓
Paint shop	✓		✓	✓	✓	✓	✓
General assembly	✓	✓		✓	✓		✓

algorithm, the reader is encouraged to check the details in the corresponding references. Li *et al.* (2006) summarise eight two-machine line models and the assumptions, formulae, and numerical results are compared. It is shown that the differences among the models are small. Clearly, a comparison for more complex systems would be of interest and could form part of future work.

The remainder of the paper is structured as follows. Section 2 discusses serial production lines. An assembly system is introduced in Section 3. Section 4 considers closed-loop production systems. Parallel lines, split and merge are summarised in Section 5. Finally, future topics are introduced in Section 7 and conclusions are formulated in Section 8.

2. Serial lines

Serial production lines are the most practical production systems in many manufacturing organisations. An illustration of an M machine, $M-1$ buffer serial line is shown in Figure 4, where the circles represent the machines and the rectangles are the buffers, denoted as $m_i, i=1,\dots,M$, and $b_i, i=1,\dots,M-1$, respectively.

Exact solutions for system throughput can only be derived in two-machine line models (Dallery and Gershwin 1992, Li *et al.* 2006). For systems with more machines, approximate solutions are investigated. Generally speaking, most of the approximation methods are either based on *aggregation* or on *decomposition* approaches (Dallery and Gershwin 1992).

The general idea of aggregation is to replace a two-machine line by a single equivalent machine. The equivalent machine has the same throughput in isolation as that of the two-machine line. The aggregation is performed through recursive procedures along the serial line in forward and backward orders. When the procedure converges, the system production rate and other performance measures are obtained.

The basic principles of decomposition methods were first introduced in the 1960s (Sevast'yanov 1962), and efficient algorithms were developed in the 1980s. The idea of decomposition is to decompose the original model into the analysis of a set of smaller subsystems where analysis is available. A decomposition method usually involves three steps: (1) decomposing the longer line into a set of two-machine lines; (2) deriving a set of equations with unknown parameters for each two-machine line; and (3) developing an algorithm to solve the unknowns in these equations.

In this paper we discuss both aggregation and decomposition methods. More specifically, two methods, the aggregation approach proposed by Meerkov (denoted as Meerkov aggregation in the subsequent text) and the decomposition approach originated by Gershwin (denoted as Gershwin decomposition) are summarised. In addition, other decomposition and aggregation methods are briefly addressed.

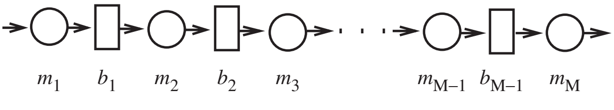

Figure 4. Serial production lines.

2.1 *Meerkov aggregation method*

Meerkov and his group studied aggregation methods for Bernoulli, geometric and exponential machine reliability models. The main advantage of the Meerkov aggregation method is that all the aggregation procedures have been analytically proved to be convergent. In addition, system-theoretic properties (e.g., improvability, reversibility, monotonicity, inverted bowl, bottleneck identification, etc.) have been proved and analysed. The accuracy of the aggregation procedures has been determined either analytically or numerically and it has been shown that, in most cases, these procedures result in good accuracy. Below, we summarise the key concepts of Meerkov aggregation methods. Detailed descriptions can be found in Lim *et al.* (1990), Jacobs (1993), Jacobs and Meerkov (1995a,b), Jacobs *et al.* (1996), Chiang (1999), Chiang *et al.* (2000a,b,c, 2001), Li (2000, 2004a), and Li and Meerkov (2003, 2007).

The Meerkov aggregation method consists of two principal components, backward and forward aggregation (denoted by superscripts 'b' and 'f', respectively (see Figure 5)). In backward aggregation, the last two machines, m_M and m_{M-1}, and the buffer b_{M-1} are aggregated into a single machine $m^{\rm b}_{M-1}$. Then, machine $m^{\rm b}_{M-1}$ is aggregated with machine m_{M-2} and buffer b_{M-2} to result in $m^{\rm b}_{M-2}$, and so on until all machines and buffers are aggregated into $m^{\rm b}_1$. In forward aggregation, the first machine m_1 is aggregated with machine $m^{\rm b}_2$ and buffer b_1 to produce $m^{\rm f}_2$. Then, $m^{\rm f}_2$ is aggregated with $m^{\rm b}_3$ and b_2 to result in $m^{\rm f}_3$, and so on until all machines and the intervening buffers are aggregated into $m^{\rm f}_M$. The process is then repeated. It has been proved that the recursive procedure is convergent and the throughput of $m^{\rm b}_1$ or $m^{\rm f}_M$ represents an estimate of the system throughput.

2.2 *Gershwin decomposition method*

In this subsection we summarise the key ideas of the Gershwin decomposition method. Detailed descriptions can be found in the monograph of Gershwin (1994) and the papers of Gershwin (1987a,b, 1989), Choong and Gershwin (1987), Dallery *et al.* (1988, 1989), Hong *et al.* (1992), Glassey and Hong (1993), Alvariz-Vargas *el al.* (1994), Burman (1995), Dallery and Le Bihan (1999) and Le Bihan and Dallery (2000).

The decomposition method is based on the representation of an $M-1$-buffer system by $M-1$ single-buffer systems, i.e. $M-1$ two-machine, one-buffer systems (see Figure 6 for an illustration of a four-machine, three-buffer line). Specifically, a set of two-machine lines is considered. Pseudo-machine $M_{\rm U}(i)$ models the line upstream of b_i, and $M_{\rm D}(i)$ models the line downstream from b_i. The parameters of the pseudo-machines are chosen such that the conservation of flow, flow rate–idle time, and resumption of flow relationships are satisfied. These relationships and the boundary conditions provide a total of $4(M-1)$ equations with $4(M-1)$ unknowns. Computation algorithms, referred to as Dallery–David–Xie (DDX) algorithms (Dallery *et al.* 1988, 1989) and the accelerated DDX (ADDX, Burman 1995) algorithm, have been introduced to solve these equations. Although not proved analytically, it is claimed that, in most cases, the ADDX algorithm converges, and it can provide a faster speed and more accurate estimates (Burman 1995). Other extensions and modifications of the DDX algorithm can be found in Hong *et al.* (1992), Glassey and Hong (1993), Dallery and Le Bihan (1999), Le Bihan and Dallery (2000), etc.

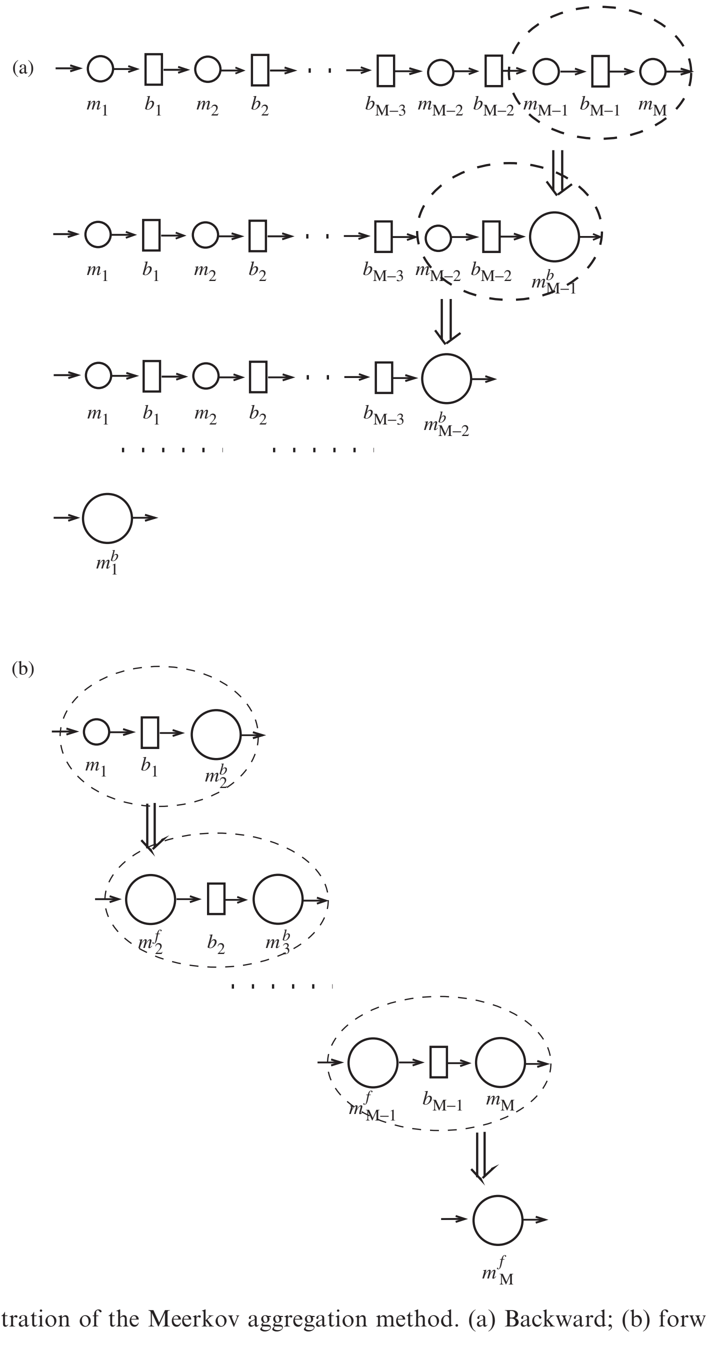

Figure 5. Illustration of the Meerkov aggregation method. (a) Backward; (b) forward.

2.3 *Serial lines with non-exponential machines*

In addition to the exponential assumption of machine up- and downtimes, the phase-type distribution of the operation and repair times (or completion times) has been studied by Buzacott and Shanthikumar (1993) and Altiok (1997). Decomposition methods based on

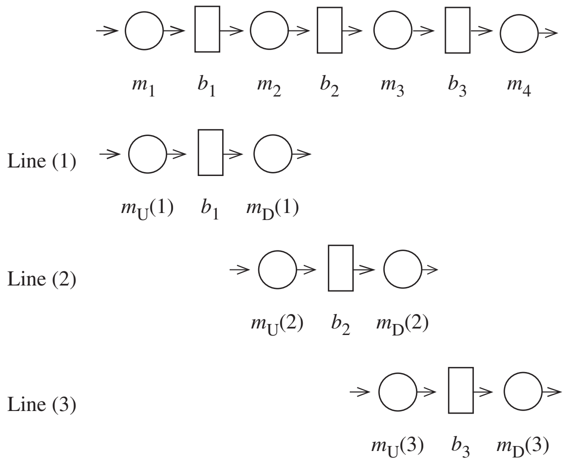

Figure 6. Illustration of the Gershwin decomposition method.

the Markov chain approach have been used to approximate system performance. However, such approaches are typically computationally intensive and are difficult to extend to larger systems. Detailed descriptions of the decomposition method and phase-type distribution can be found in Altiok (1997).

A simple formula for the estimation of the throughput of a serial line of identical machines and buffers with random processing times is presented by Blumenfeld (1990), where the variability of the machine processing times is addressed by the coefficient of variation. Tempelmeier and Burger (2001) report an unbalanced flow line with generally distributed processing times, unreliable machines with scrapping and finite buffers. An approximation procedure based on decomposition is presented, where a reliable two-machine line is used as a building block and machine failures are considered in terms of the completion time. Özdoğru and Altiok (2003) analyse a two-valve fluid flow system with a phase-type distribution of repair times. A continuous-time Markov chain approach is developed to study the steady-state behaviour of the valves and the material in storage.

Li and Meerkov (2005a) present empirical laws to approximate the performance of serial lines where the machines have different types of up- and down-time distributions. First, the line performance under the assumption of exponential up- and down-times is analysed. Then the coefficients of variation (CV) of the machines' up- and down-times are used to adjust the approximation. It shows that when $CV < 1$, which is in the majority of cases (Inman 1999, Li and Meerkov 2005b), the system throughput is practically independent of the type of machine up- and down-time distribution, and mostly depends on the CV.

More work on lines with non-exponential machines can be found in Buzacott and Kostelski (1987), Buzacott *et al.* (1993), Papadopoulos (1998), Vidalis and Papadopoulos (1999), Tan and Karabati (2001), Helber (2005), etc.

2.4 *Other approximation methods*

In addition to the above aggregation and decomposition methods, many other methods have been proposed to approximate the system throughput.

Jafari and Shanthikumar (1987, 1989) consider synchronous flow lines with the possible scrapping of parts and apply the results to determine the optimal allocations of buffer capacity. Liu and Buzacott (1989) consider a general Buzacott-type model (see Buzacott and Shanthikumar (1993)) of synchronous transfer lines. The general principle is again to decompose the original line into a set of two-machine lines. De Koster (1987) proposes an aggregation approach for a time-dependent model. However, the correlations among the buffers are not taken into account, i.e. aggregation is only proceeded forward. An improved, but more complex, algorithm that modifies the buffer capacity for certain situations by taking into account the average buffer content is also presented by De Koster (1988). For small-size discrete asynchronous serial lines (up to four machines), Yang *et al.* (2000) obtain an exact numerical solution using mixed vector–scalar operations. However, this method becomes intractable for longer lines.

Additional decomposition and aggregation methods for the analysis of serial production lines can be found in papers by Terracol and David (1987), De Kok (1990), Chen and Thinphangnga (1996), Tan and Yeralan (1997), Yeralan and Tan (1997a), Huang and Blumenfeld (1998), Kim and Jung (2000), Kuhn (2003), Sadr and Mahame (2004a,b) and, Blumenfeld and Li (2005), and in reviews by Dallery and Gershwin (1992) and Papadopoulos and Heavey (1996).

3. Assembly systems

By generalising the results of the throughput analysis of serial lines, approximate solutions can be obtained for assembly systems. In assembly systems (or feeder lines), the assembly machine will process a part only when all its upstream buffers are non-empty. Many of the results of the aggregation and decomposition methods presented in Section 2 for serial lines can be extended to the analysis of assembly systems.

3.1 *Meerkov aggregation method*

Similar to the serial line case, an aggregation method was developed by Meerkov *et al.* to study assembly systems (Kuo 1996, Kuo *et al.* 1997, Chiang *et al.* 2000a,b, Li 2000, 2005). The idea behind the approximation is as follows. Consider an 'upper' line consisting of a one-component line and the assembly line (Figure 7(a)). Assume that machine m_{01} is modified so as to account for the existence of the other component line. Then, using a serial line evaluation method, Prob{buffer b_{1M_1} is not empty} can be calculated. Now consider a line composed of the other component line and the assembly line, where m_{01} is again modified by considering the probability that buffer b_{1M_1} is empty. Then we calculate Prob{buffer b_{2M_2} is not empty}. Now use this probability for another iteration and continue the process, alternating between the two lines. The iterations are convergent and result in estimates of the production rate.

Note that, in this procedure, all machines and buffers in the assembly line overlap. Li (2005) extends this idea by introducing three, rather than two, serial lines with only one overlapping machine, m_{01} (see Figure 7(b)). It can be shown that the two methods result in exactly the same result. In both cases, convergence of the iteration procedure has been proved analytically. Clearly, such aggregation methods can readily be extended to an assembly system with multiple component lines and to multiple assembly systems

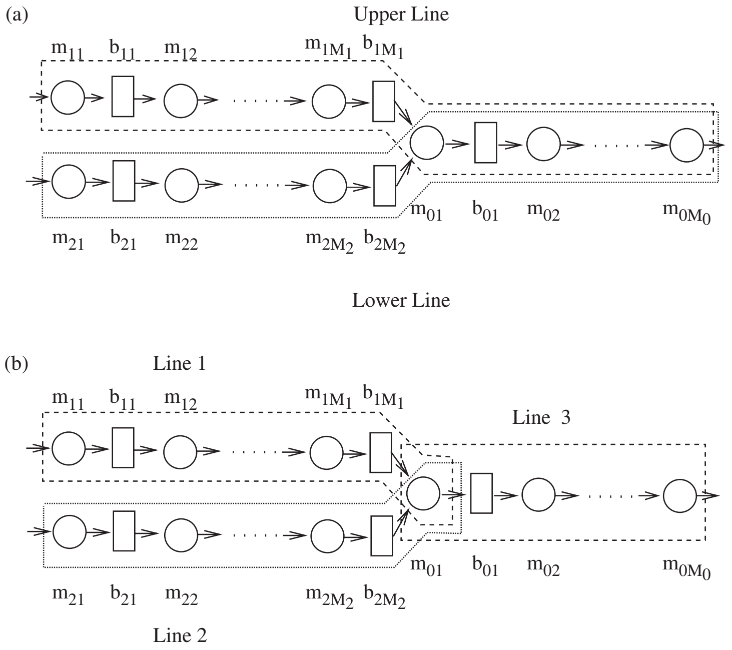

Figure 7. Illustration of assembly system aggregation.

(Kuo 1996, Kuo *et al.* 1997, Li 2000), and also to disassembly systems (where the disassembly machine will operate only when all the downstream buffers are available) (Li 2005).

3.2 *Gershwin decomposition method*

Efficient decomposition methods for calculating the throughput of tree-structured assembly/disassembly systems (Figure 8(a)) have been presented by Gershwin (1991, 1994), Di Mascolo *et al.* (1991), Gershwin and Burman (2000), etc. The models are extensions of the transfer line models described in Section 2.2.

The method decomposes the assembly/disassembly system into two-machine, one-buffer lines (Figure 8(b)). Pseudo-machine $M_U(j, i)$ models the part of the line upstream of buffer $b_{j,i}$ and $M_D(j, i)$ models the part of the line downstream from $b_{j,i}$. Again, based on the equations of the conservation of flow, the flow rate–idle time relationship, and the resumption of flow, one can obtain a total of $4(k-1)$ equations with $4(k-1)$ unknowns. DDX algorithms are typically used to solve the decomposition equations. The advantage of this decomposition method is that it can deal with complex tree-structured systems with many nodes (i.e. multiple assemblies/disassemblies) with less difficulty.

3.3 *Other assembly system models*

Liu and Buzacott (1990) adapt the results from transfer lines and apply them to assembly systems. They transformed an assembly system into an equivalent system with all the machines arranged in series. Jeong and Kim (1998) consider a tree-structured

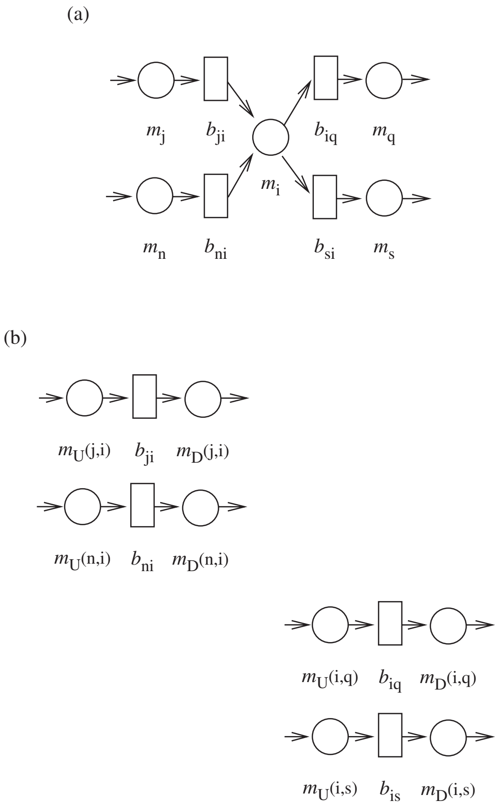

Figure 8. Illustration of Gershwin assembly system decomposition. (a) Assembly/disassembly system; (b) A/D decomposition.

assembly/disassembly system with finite buffer capacity using decomposition. They also extend the method to the case where each station consists of multiple identical machines (Jeong and Kim 1999).

Further papers addressing assembly/disassembly systems are those of Simon and Hopp (1991), Duenyas and Hopp (1992), Daganzo and Blumenfeld (1994), Duenyas (1994), Dallery (1999), Kouikolou (2000), etc.

4. Closed-loop production systems

A closed-loop system is a serial line with one or more loops attached to it. Not much work has been done with respect to the analysis of this kind of system. In this paper, two problems encountered with closed-loop system analysis are introduced: closed lines with a constant number of carriers and production systems with rework loops.

4.1 *Closed-loop system with a constant number of carriers*

A closed-loop serial production line with a constant number of carriers has been studied by Lim and Meerkov (1993), Frein *et al.* (1996), Gershwin *et al.* (2001), Werner (2001),

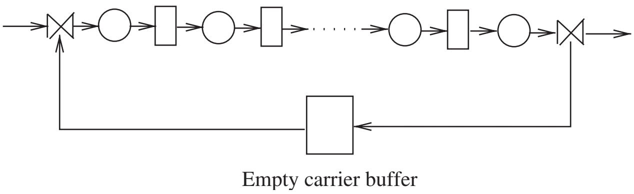

Figure 9. Closed serial production line.

Kim *et al.* (2002), Biller *et al.* (2007) and Gershwin and Werner (2007), where parts are loaded on and attached to the pallets at the first machine to undergo all the operations. Upon completion of the operations, the finished parts are unloaded and the pallets are released and sent back to the first machine through an empty carrier buffer (Figure 9).

Lim and Meerkov (1993) analyse an asymptotic reliable two-machine, two-buffer closed serial line. The closed-loop system is reduced to an open serial production line where the *effective* buffer capacity depends on the relationship between the actual buffer capacity and the number of pallets. A case study at an automotive paint shop is described. Frien *et al.* (1996) present a decomposition approach to approximate the system production rate for homogeneous production lines. It is assumed that the sum of the average amount of material in each buffer is equal to the total quantity of material available. However, this method does not account for the correlation between the actual number of pallets in the buffers. Gershwin *et al.* (2001) extend the decomposition approach to study closed-loop systems with geometric machines. The *virtual* failure mode is introduced to account for the blocking or starving effect due to downstream and upstream machines, respectively. A constant number of carriers is modeled by defining the range of starvation and range of blocking. However, due to computational efficiency, such a method is only applicable to small systems. Werner (2001) and Gershwin and Werner (2007) extend this idea and develop an approximation method to evaluate the performance of large closed-loop flow systems. Kim *et al.* (2002) develop an empirical formula for the optimal number of carriers, which is equal to the total number of machines plus half of the total buffer capacity. The formula is proved for the three-machine case and it is hypothesized that this formula is also applicable to the general case. However, it is difficult to extend the method to longer lines. In addition, Biller *et al.* (2007) introduce improvability measures in closed-loop systems with Bernoulli models. They provide a method to determine if the empty carrier buffer and the number of carriers impede line performance and how to make improvements if such impeding occurs. Moreover, bottlenecks in closed lines are discussed and an approach to select the smallest, i.e. lean, number of carriers and the capacity of the empty carrier buffer, which results in no impediment, is described.

4.2 *Production systems with rework loops*

Rework loops are widely used in many manufacturing systems for quality improvement. A production system with a rework loop is illustrated in the upper part of Figure 10. At machine m_k, parts with good quality will be sent to buffer b_k for continuing operations, and parts with defects will be sent to buffer b_{M_1} in the rework loop for repair, and then rejoin the main line at machine m_j for reprocessing.

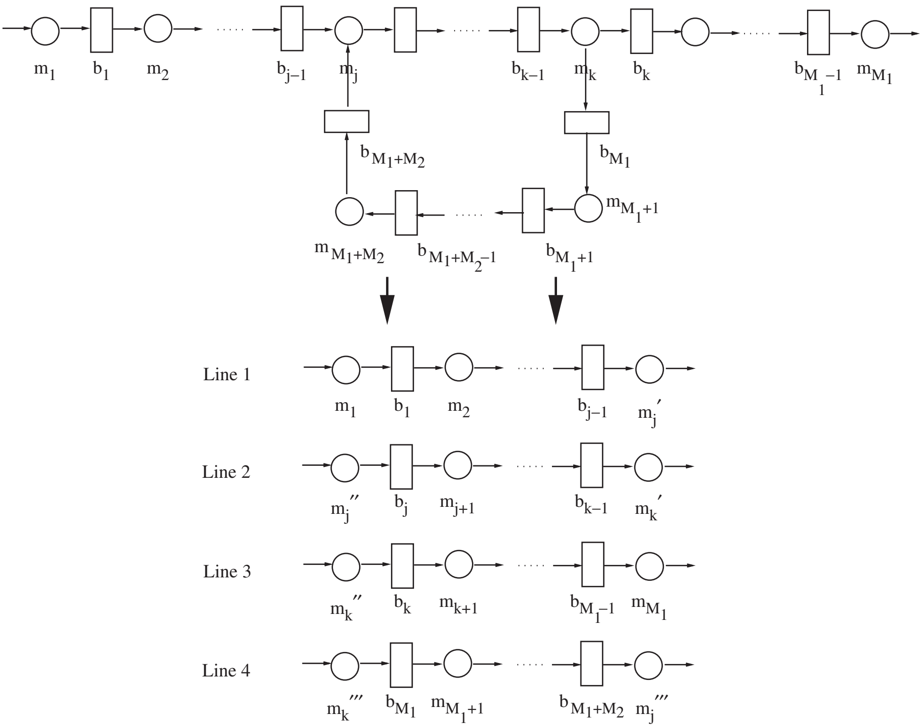

Figure 10. Production system with rework loop.

For such systems, using overlapping decomposition, Li (2004a) presents a recursive procedure to estimate the throughput, where the original system is decomposed into four overlapping serial lines (see Figure 10). The probabilities of starvation and blockage of the first and last machines of a serial line (m_j and m_k) are used to modify the machines to take into account the effects from other lines (where the primes indicate modifications). A recursive procedure is introduced to iterate among the four lines. It is shown that the iterations are convergent. This method has also been extended by Li (2004c) to study multiple rework loop systems in automotive paint shops.

In addition, Yerelan and Tan (1997a) have developed a flexible decomposition framework and extend it to analyse a work cell consisting of one workstation and one rework station and finite buffers fed by a Poisson stream of discrete parts. Helber (1999) discusses a production system with a rework loop using the decomposition approach where the geometric reliability of machines is assumed.

5. Parallel systems, split, merge and scrapping

5.1 *Parallel systems*

Parallel systems are often used to increase the production capacity. Most analyses study parallel lines by equivalence, i.e. aggregating parallel machines into an equivalent single machine.

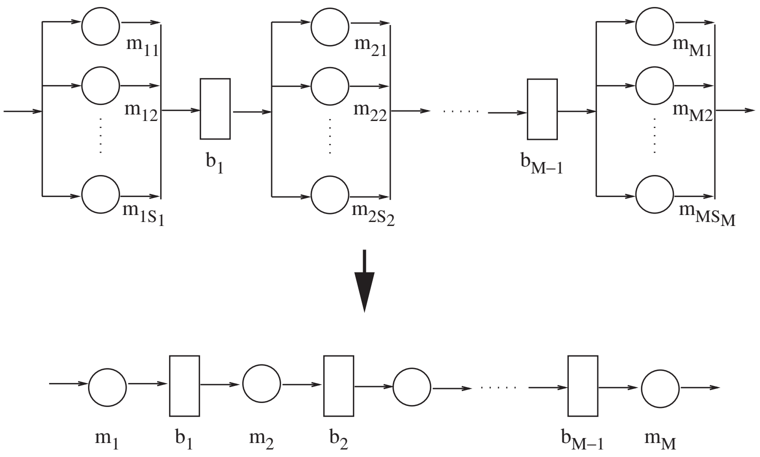

Figure 11. The parallel production systems discussed by Burman (1995), Arzi and Kalia (1996) and Patchong and Willaeys (2001).

Early work on parallel machines was by Ignall and Silver (1977), Iyama and Ito (1987), and Mitra (1988). Burman (1995) considers series–parallel flow lines where each stage consists of multiple parallel machines and finite buffers (Figure 11). The multiple machines are approximated by an equivalent single machine. The system performance is analysed using the equivalent machine and the ADDX algorithm is used to estimate the throughput. Similar ideas are used by Arzi and Kalia (1996) and Patchong and Willaeys (2001), where Arzi and Kalia (1996) consider inhomogeneous lines and Patchong and Willaeys (2001) extend the study to the non-identical parallel machine case and present another way of approximating parallel machines by a single machine.

Jeong and Kim (1998) consider an assembly line where each station consists of multiple identical machines using the similar equivalent method. Yang *et al.* (1999) present a two-stage module by decomposition and aggregation. However, the equivalence of the system production rate between the multiple machines and the aggregated machine only holds when the machines are identical.

Li (2004b) considers a parallel system that has multiple parallel lines, each with multiple machines and buffers (see upper part of Figure 12). An overlapping decomposition method is used to estimate the throughput of such a system. Specifically, the parallel system is decomposed into $k+2$ serial lines (see Figure 12). Again, the first or last machine in each line is modified to include the effects from other lines. A recursive procedure is used to estimate the production rate of the system. Again, it is shown that the iterations are convergent.

The idea of overlapping decomposition used in this subsection is similar to that in other subsections (e.g., assembly, rework loop, etc.). It has been shown that this is a general methodology for the analysis of complex production systems (Li 2005). In addition, this method is independent of the serial line analysis method. Any serial line evaluation method (either aggregation or decomposition, or others) can be used. However, the convergence of the whole procedure is dependent on the convergence of serial line methods.

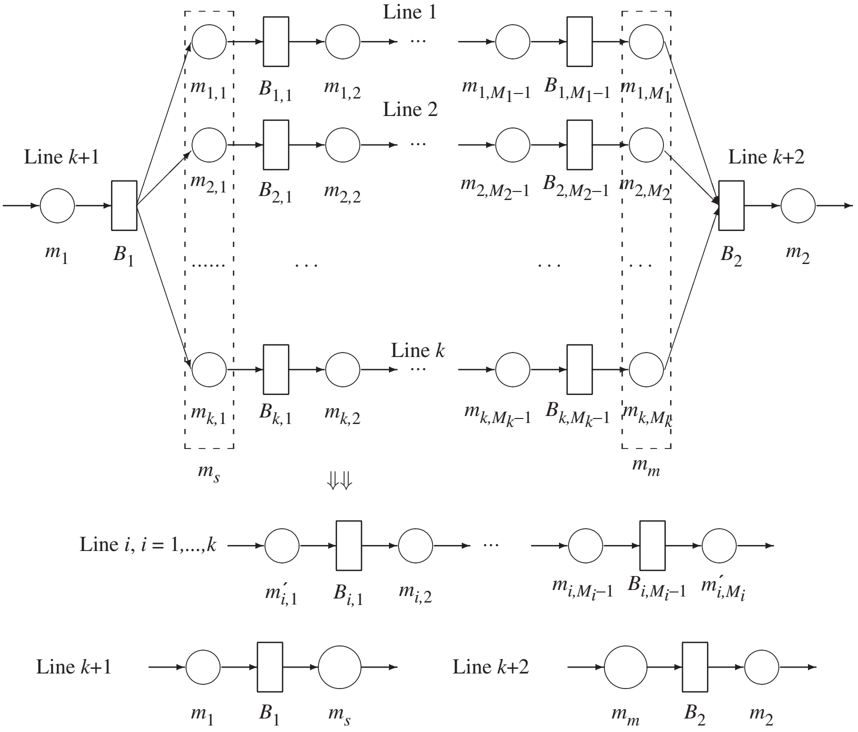

Figure 12. Overlapping decomposition of parallel systems.

5.2 *Split, merge and scrapping*

In this subsection we address split and merge modeling in production systems. Note that the split and merge discussed here are typically referred to as *production split and merge*, where parts are released to, or loaded from, different downstream, or upstream, buffers one by one, respectively. The assembly/disassembly systems discussed in Section 3 are often referred to as *assembly merge* or *disassembly split*, respectively, where parts are loaded at the same time to be composed into a new single part when both upstream buffers have parts available, or released at the same time to split into multiple parts when both downstream buffers have available space, respectively.

Yang *et al.* (1999) present a uniform framework for the approximate analysis of multiple-processing-route production lines, including scrapping (split), serial/parallel hybrid blocks. Dallery (1999) extends the continuous-flow model to incorporate random yield, split and merge mechanisms. Helber (2000) presents a Markov process model and an approximate decomposition technique for a discrete transfer line with finite buffers and a split of materials at some stations in the line. Li and Huang (2005) discuss split and merge of a feed-forward line producing two types of products. The overlapping decomposition method is used to decompose the system into several overlapping serial lines and a convergent iterative procedure is introduced to estimate the throughput. In addition, a three-machine system with two parallel machines merging in a shared buffer and another machine is addressed by Tan (2001), Helber and Mehrtens (2003), Helber and Jusic (2004), Diamantidis *et al.* (2004), and Diamantidis and Papadopoulos (2006).

The possible scrapping of workpieces can be viewed as a special case of split. Jafari and Shanthikumar (1987) assumed that parts are scrapped only due to machine breakdowns and that they will be detected and removed from the machine as soon as they are scrapped. Li (2005) uses the overlapping decomposition approach to study scrapping where two overlapping serial lines are considered through an iterative procedure.

6. Other complex systems

In addition to the above complex systems, other issues with a complicated nature have also been investigated. Here we briefly introduce three of them: production systems with quality check devices, multiple part types, and buffer transit time and failures.

6.1 *Production systems with quality check devices*

In order to improve product quality, most production lines implement quality control techniques with inspection stations allocated throughout the lines. The topic of inspection allocation has received much attention. Most of the studies address reliable serial production lines. Limited attention is paid to the investigation of unreliable lines with quality inspection devices. Jacobs and Meerkov (1991) study an asymptotically reliable two-machine line with quality inspection. A similar analysis is extended to longer lines by Han *et al.* (1998) and Chiang (2006). Reentrant production lines with inspection stations and probabilistic routing are addressed by Narahari and Khan (1996). Deliman and Feldman (1996) deal with serial lines with imperfect inspection and rework of defective items. Stochastic search techniques (generic algorithms and simulated annealing) are used by Viswanadham *et al.* (1996) to investigate the impact of inspection allocation from the cost perspective. A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing is presented by Mandroli *et al.* (2006).

The above studies do not consider the correlation between subsequent part quality. However, in many production systems, quality defects may exhibit Markovian behaviour. Kim (2004) and Kim and Gershwin (2005) introduce a quality and quantity model of a serial line with inspection and information feedback, based on the assumption that a machine will continue to produce defective parts until its operation is corrected (i.e. Markovian-type quality defects). In addition, Colledani and Tolio (2005a,b) develop a model to include a control chart in statistical process control in throughput analysis.

6.2 *Multiple part types*

Many manufacturing systems are flexible so that different products can be produced on the same line. Nemec (1999) presents a decomposition method for the analysis of a synchronous transfer line which processes multiple products through dedicated buffers. A similar problem is addressed by Syrowicz (1999) and Colledani *et al.* (2005). Li and Huang (2005) study a split and merge system, where two different products, *A* and *C*, are produced. Such a system consists of dedicated production lines *A* and *C* designed for specific processing of the respective products, and common processes before the split and after the merge are shared by both products.

6.3 *Buffer transit time and failures*

Buffers have been studied intensively in the production system literature, with most studies addressing the issues of optimal allocation of buffers or lean buffering to achieve the desired throughput. However, some other questions, such as the buffer transfer time and failures, are still unclear and deserve more investigation. Commault and Semery (1990) show that the delays in buffers influence the performance of serial lines of unreliable machines and this influence can be represented approximately by an equivalent line with the same machine characteristics but reduced buffer capacity. Buffer or transfer device failures were studied by Burman (1995), Lipset *et al.* (1999) and Suliman (1999). In addition, the shared buffer case was studied by Matta *et al.* (2005), where parts can be sent to a commonly shared buffer to avoid blockage. When machines and buffers are not limited to serial lines, Owen and Huang (2007) show that, in some cases, adding buffers may reduce the system throughput.

7. Future research topics

Although throughput analysis has achieved significant advances in recent years, there are still many problems deserving more in-depth investigation. Our interests are focused on analytical methods to estimate the mean throughput for production systems with complex features. We present below some future research topics of importance to the automotive industry based on our experience, and we believe such topics would also be of interest to other large-volume manufacturing industries.

Clearly, there are many topics that are important. Due to space limitations, we only present some of them. These topics are arranged in alphabetic order within three categories: floor-level operation and management; plant-level design and management; and enterprise-level planning. (Such a categorisation may not be unique and some topics may fall into more than one category.)

7.1 *Floor-level operation and management*

7.1.1 *Bottleneck identification and mitigation*

Identifying which machine is really the bottleneck still appears to be an important problem, in spite of the extensive work already performed in this area (see, for instance, Lawrence and Buss (1995), Kuo (1996), Kuo *et al.* (1996), Chiang *et al.* (1998, 2001), Chiang (1999), Li (2000), Li and Meerkov (2000), Roser *et al.* (2003), and Biller *et al.* (2007)). To establish how to identify and remove bottlenecks in both the long term (e.g., a practical bottleneck identification method in complex systems, such as parallel lanes, rework loops, etc.) and the short term is still an important research problem. In particular, a clear definition of a short-term bottleneck is needed. A short-term bottleneck machine should be the machine that impedes system performance in the short term, and its removal should help, rather than impede, the work of eliminating long-term bottlenecks. Research to integrate this issue with real-time maintenance scheduling, production control, etc., is desirable. Moreover, when there are several bottlenecks, prioritising the importance of the different bottlenecks is also an area requiring more research. For systems with complex layouts (e.g., repair loops, reentrant lines, etc.), these issues become more challenging (see Biller *et al.* (2007) for an example in a closed-loop system).

7.1.2 *Buffer behaviour*

Most throughput analysis studies do not consider buffer or conveyor failures and the buffer transit time. In practice, such failures do occur, and sometimes even affect system performance significantly. Typically, such issues are addressed by modeling buffers as machines (see Subsection 6.3). However, when buffer capacities are relatively large, and the buffer transit time and buffer failure are related to the buffer content, then the problem becomes tricky. Analogously, when the conveyor is working properly, all stations hooked on the conveyor work independently. When the conveyor fails, all stations stop working. Typical modeling approaches to include conveyor failure either increase each machine's average downtime, or model the conveyor as a machine feeding all the stations hooked on the conveyor. The first does not characterise the dependence nature, while the latter becomes messy when the system has multiple conveyors. In addition, analytical tools are needed to investigate the possible non-monotonic properties of throughput with respect to buffer capacity in a non-serial line environment. This is an area where few results are found, and yet would provide practical value.

7.1.3 *Closed-loop system with a constant number of carriers*

Although some studies have been carried out to analyse closed-loop systems with a constant number of carriers (see Subsection 4.1), almost all of them address closed serial lines only. How a system behaves as a function of the number of carriers in a more complex environment is still unclear. In addition, it has been shown in several studies that a plateau typically appears around the optimal number of carriers (to reach maximal throughput). In other words, there exists a range of the number of carriers that will deliver optimal or close-to-optimal performance. Then, what is the smallest, i.e. lean, number of carriers that will lead to similar performance? Biller *et al.* (2007) provide a preliminary study in this direction on improvability and leanness with respect to the number of carriers in the system. In addition, given that some production systems use a large loop and some use several connected small loops, which structure, under what conditions, and what kinds of control policy can provide better performance? All these questions deserve an in-depth investigation.

7.1.4 *Incomplete and inaccurate information*

Although modern technology provides more information for production management, in many cases managers and engineers may still be puzzled by incomplete and inaccurate information. For example, sensors may not be installed at every station or PLC logic may not be implemented to collect all the necessary information, due to cost, installation/operation difficulties, etc. In addition, sensors may not be reliable, and the data collected on the factory floor is often not accurate (5–10% error is typical). These issues will certainly have an impact on estimating production performance. How we can compensate for information incompleteness or inaccuracy is an important issue. Bonney *et al.* (1994), Denardo and Tang (1997), Feit and

Wu (2000), Li *et al*. (2005), and Kang and Gershwin (2005) illustrate some ideas of such work, and more detailed study is necessary.

7.1.5 *Multiple failure mode*

Machines may break down for various reasons, such as the failure of various components. Therefore, they may exhibit multiple failure characteristics. Tolio *et al*. (2002) consider a two-machine line where each machine has multiple failure modes. It is shown that such a system has lower throughput than the single failure mode case. Systems with multiple failure modes, random processing times and finite buffers are discussed by Levantesi *et al*. (2003a,b). The decomposition method was introduced to analyse lines with more than two machines. However, these methods are only applicable to small systems due to the computational intensity. The problem of multiple failure modes is becoming more critical where maintenance scheduling is integrated into throughput analysis, since different failure codes may introduce different machine behaviour and repair/maintenance times, and have different impacts on system performance. Clearly, more work is needed to model multiple failure modes of machines and systems.

7.1.6 *Multiple part types*

The available literature on production systems with multiple part types typically assumes dedicated buffers for each product (see Subsection 6.2). However, this is only a special case of multiple part type production systems. The more general cases, where different products occupy the same buffer spaces, have not yet been studied. In particular, when machines have different operating and reliability characteristics for different products, and changeovers may be needed during a product switch, these problems become more complicated.

Moreover, in many production systems, incoming parts are often arranged in sequence or grouped into batches for different products. For example, in automotive paint shops, vehicles with the same colour are often batched into groups to reduce the changeover time and cost, and to improve quality. In body shops and general assembly lines, vehicles are typically sequenced to meet different customer requirements. Often, long time average probabilities are used in throughput analysis without taking this into account explicitly. However, since machine reliability and product quality could be correlated to such sequencing, how to model the productivity and quality performance for specific batching and sequencing policies is important. An analytical model that could explicitly describe such operations would be beneficial, not only for performance evaluation, but also for analysis and decision-making of scheduling policies.

Such studies, combined with other studies described in this section, would be of particular interest to large-volume manufacturing industries.

7.1.7 *Production control*

Production control has been studied extensively in the manufacturing literature (see, for example, Kimemia and Gershwin (1983), Wein (1988), Gershwin (1994, 2000), Liberopoulos and Caramanis (1994), Perkins and Kumar (1995), Bonvik *et al*. (2000), and Tan (2002)). In modern manufacturing systems, more and more real-time information is available through wireless sensors, RFIDs, etc., which enables us to implement more advanced control logic and operations. Such control actions would certainly affect system throughput. How to integrate such different production control logics into throughput analysis is necessary and important. This topic is also coupled with other issues, such as transient analysis, material handling, bottleneck mitigation, part sequencing, maintenance scheduling, etc.

7.1.8 *Transient analysis*

Most of the throughput analysis literature addresses the steady-state behaviour of production systems. In contrast, the transient behaviour of systems is practically unexplored. However, transient analysis is also important, since the transient behaviour is directly related to production losses during transients, production control of work-in-process, maintenance scheduling, etc. For example, the following (and other) questions are of interest for production management. How long does it take for the system to reach the steady state? Which parameter controls the characteristics of the transient behaviour? How should the initial buffer occupancy be selected so that the production losses during the transient period are minimised?. Transient behaviour in a production system environment has been studied by Mitra (1988), Narahari and Viswanadham (1994) and Mocanu (2005). Meerkov and Zhang (2008) attempt to answer the above questions by considering a Bernoulli serial line. It was found that the second largest eigenvalue of the transition matrix determines the characteristics of the transients, and that work-in-process needs a longer settling time than the production rate to reach steady state. However, this is still an open area with many potential research topics, and more studies are needed to address complex systems and different machine reliability models.

7.2 ***Plant-level design and management***

7.2.1 *Hybrid models*

There are many production systems that have both manual and automatic operations in the automotive and other manufacturing industries. These two kinds of operations may have different characteristics with respect to productivity, quality, flexibility and almost every aspect of interest. Manual operations can typically be modeled using random processing times without downtime, whereas automatic machines have a constant speed subject to breakdowns. When we deal with hybrid systems with both manual and automatic operations, we tend to model manual operations using ‘equivalent’ automatic machines. But we know that such equivalency is an approximation and will cause inaccuracy in the results. More importantly, it does not provide us with any insight in terms of the trade-off between these two operations. For example, when we design a new production line, the question typically raised is: what is the percentage of automation? Or, in other words, how many and which operations should be automated to achieve maximised throughput, best quality and minimised cost? Future research that could provide an analytical tool for such hybrid systems would enable us to investigate the impact of manual/automatic operations on system performance measures of interest.

7.2.2 *Integrated productivity and quality models*

Both production system design and quality management are important parts of manufacturing systems. Extensive research and practice have been devoted to both, but independently. Little attention has been paid to the relationship between production system design and product quality. It has been shown that production system design also has a significant impact on product quality (Inman *et al.* 2003, Li *et al.* 2007, 2008). Therefore, an integrated model addressing both quality and productivity is required. In addition to the inspection allocation problems addressed in Subsection 6.1, Khouja *et al.* (1995) delineate the trade-off between throughput and quality for a robot whose repeatability deteriorates with speed. The competing effects of large or small batch sizes are discussed by Urban (1998) and a model is developed for the interaction between batch size and quality. Moreover, Li and Blumenfeld (2006) study a production system with Andon. Clearly, as reported by Inman *et al.* (2003), more research should be carried out to fully understand the coupling between production system design and product quality, which is a largely unexplored area with many promising research opportunities.

7.2.3 *Material handling*

The design of production systems needs to take into account the flow of parts delivered to workstations. Machines may be starved of parts due to late delivery of materials. A simple way to include this effect is to introduce additional starvation time as downtime in the corresponding machines. However, part deliveries are typically correlated (e.g., one delivery worker is usually responsible for several stations), and are coupled with line status. In addition, the production sequence (flow of jobs) interferes with the delivery schedule (flow of components or parts). Such interactions should be investigated in an integrated manner.

7.2.4 *Preventive maintenance*

Maintenance is important for production operation and management. In particular, preventive maintenance can increase machine reliability, reduce downtime, and improve quality. In the past, preventive maintenance has typically been carried out during weekends, off-shift hours, etc. Intuitively we may expect that performing more preventive maintenance during operational periods could improve productivity and quality. For example, due to a downstream machine breakdown, the upstream machines may be idle for a certain period of time. Could preventive maintenance be scheduled for some upstream machines if the workforce is available? If the answer is positive, then how to determine such opportunistic timing and schedule the activities within operations, what will be the impact on system throughput and quality, how to prioritise preventive and reactive maintenance jobs, etc., are important questions to be answered. The interference of the repair crew (for example, see Kuhn (2003), who takes into account the dependency between the production and repair system) could make the problem more complicated. Clearly, this topic is also related to bottleneck identification, multiple failure modes, etc., mentioned above.

This problem becomes more difficult when scheduled downtime (e.g., maintenance time) is considered, where the line is stopped and maintenance is carried out so that the reliability behaviour of the machine being maintained is changed. The fact that scheduled maintenance stops production but also reduces random failures (which, in turn, potentially increases production) makes the issue more challenging.

7.2.5 *Reentrant lines*

Reentrant lines are widely encountered in semiconductor manufacturing (Kumar 1993, Kumar and Kumar 2001, Hopp *et al.* 2002). A similar situation also occurs in the automotive industry. For example, in powertrain manufacturing plants, ignition components may be processed and washed multiple times. However, analysis of reentrant lines is almost non-existent. Although queueing networks have been used in some analyses, an accurate performance analysis is difficult to carry out and application is therefore limited. Therefore, developing an analytical tool similar to that for serial lines and other systems would be desirable and useful for many large-volume manufacturing industries.

7.3 *Enterprise-level planning and management*

7.3.1 *Flexibility*

Manufacturing systems are becoming more and more flexible. During the last 20 years, flexibility has been studied extensively. Most of the studies focus on flexibility measurement, investment cost, product scheduling, etc. From the point of view of throughput analysis, however, interest is focused on the analysis of flexible systems making multiple products with or without setup times. The issue becomes complex if product quality is also taken into consideration. Clearly, this issue overlaps with other topics discussed in this paper, such as multiple part types, integrated quality and quantity models, production control, etc. The similarities and differences, however, generate interest as well as challenges to researchers. Further development in this direction will certainly benefit both the automotive and other manufacturing industries.

7.3.2 *Integrating throughput analysis with supply chain management*

Both throughput and supply chain management have been studied extensively. However, only limited studies have addressed the two problems in an integrated manner. Production system research often ignores the shipping requirement from customers and raw material deliveries from suppliers, whereas studies of supply chain management typically assume countable production. In addition, throughput and supply chain analyses address the production–inventory–customer systems differently with respect to part flow (individual part vs. batches) and time scale (cycles vs. shipping periods). Such differences make an integrated study more difficult. Jacobs and Meerkov (1995c), Tan (1998), Li (2000), Li and Meerkov (2000a,b, 2001, 2003), and Li *et al.* (2004, 2006) have studied this issue from the point of view of due-time performance. More research on how to design an integrated production–inventory–customer system would be valuable.

8. Conclusions

Throughput analysis is important for the design, operation and management of production systems. This paper reviews and summarises the most important methods and publications dealing with the throughput analysis of production systems with unreliable machines and finite buffer capacities. In addition, future research topics are presented from the automotive industry perspective. We believe that a new paradigm in manufacturing systems is emerging (primarily due to new technology and real-time data availability) that provides both opportunities and challenges. The integration of productivity with quality and the flexibility to achieve the desired demand satisfaction with minimum cost, is the goal of any analytical study. Another equally important direction is to design a robust product system whose performance is resilient in response to variations in the manufacturing environment. Real-time data availability can provide a foundation to achieve a more sophisticated analysis for the control and scheduling of complex systems, from floor level to enterprise level. More in-depth and accurate analyses to understand the nature of production systems and provide production engineers and managers with quantitative tools for design and operation management is a key goal.

Although we present these topics from the perspective of automotive manufacturing systems, they are not only applicable to the automotive industry, but also to other manufacturing industries, such as semiconductors, appliances, electronics, packaging, etc.

Acknowledgement

The authors thank the anonymous reviewers for their constructive comments and suggestions.

References

Altiok, T., 1997. *Performance analysis of manufacturing systems*. Belin: Springer.

Alvarez-Vargas, R., Dallery, Y. and David, R., 1994. A study of the continuous flow model of production lines with unreliable machines and finite buffers. *Journal of Manufacturing Systems*, 13, 221–234.

Arzi, Y. and Kalir, A., 1996. Approximating throughput of non-homogeneous multiple-parallel machines' production lines. Tel Aviv, Israel: Department of Industrial Engineerin, Tel Aviv University, Technical Report.

Biller, S., Marin, S.P., Meerkov, S.M. and Zhang, L., 2008. Closed Bernoulli production lines: analysis, continuous improvement, and leaness. *IEEE Transactions on Automation Science and Engineering*, to appear.

Blumenfeld, D.E., 1990. A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity. *International Journal of Production Research*, 28, 1163–1182.

Blumenfeld, D.E. and Li, J., 2005. An analytical formula for throughput of a production line with identical stations and random failures. *Mathematical Problems in Engineering*, 2005, 293–308.

Bonney, M., Poplewell, K. and Matoug, M., 1994. Effects of errors and delays in inventory reporting on production system performance. *International Journal of Production Economics*, 35, 93–105.

Bonvik, A.M., Dallery, Y. and Gershwin, S.B., 2000. Approximate analysis of production systems operated by a CONWIP/finite buffer hybrid control policy. *International Journal of Production Research*, 38, 2845–2869.

Burman, M.H., 1995. *New results in flow line analysis*. Thesis (PhD). OR Center, MIT.

Buxey, G.M., Slack, N.D. and Wild, R., 1973. Production flow line systems design—a review. *AIIE Transactions*, 5, 37–48.

Buzacott, J.A. and Hanifin, L.E., 1978. Models of automatic transfer lines with inventory banks: a review and comarison. *AIIE Transactions*, 10, 197–207.

Buzacott, J.A. and Kostelski, D., 1987. Matrix–geometric and recursive algorithm solution of a two-stage unreliable flow line. *IIE Transactions*, 19, 429–438.

Buzacott, J.A. and Shantikumar, J.G., 1993. *Stochastic models of manufacturing systems*. Engelwood Cliffs, NJ: Prentice Hall.

Chen, F.F. and Thinphangnga, S., 1996. Analytical modeling and analysis of flexible manufacturing systems considering system component failure/repair rates. *Jounral of Manufacturing Systems*, 15, 143–154.

Chiang, S.-Y., 1999. *Bottlenecks in production systems with Markovian machines: theory and applications*. Thesis (PhD). University of Michigan.

Chiang, S.-Y., 2006. Bernoulli serial production lines with quality control devices: theory and application. *Mathematical Problems in Engineering*, 2006, Article No. 81950.

Chiang, S.-Y., Kuo, C.-T. and Meerkov, S.M., 1998. Bottlenecks in Markovian production lines: a systems approach. *IEEE Transactions on Robotics and Automation*, 14, 352–359.

Chiang, S.-Y., Kuo, C.-T. and S.M. Meerkov, S.M., 2000a. DT-bottleneck in serial production lines: theory and application. *IEEE Transactions on Robotics and Automation*, 16, 567–580.

Chiang, S.-Y., Kuo, C.-T., Lim, J.-T. and Meerkov, S.M., 2000b. Improvability theory of assembly systems. I: Problem formulation and performance evaluation. *Mathematical Problems in Engineering*, 6, 321–357.

Chiang, S.-Y., Kuo, C.-T., Lim, J.-T. and Meerkov, S.M., 2000c. Improvability theory of assembly systems. II: Improvability indicators and case study. *Mathematical Problems in Engineering*, 6, 359–393.

Chiang, S.-Y., Kuo, C.-T. and Meerkov, S.M., 2001. c-Bottleneck in serial production lines. *Mathematical Problems in Engineering*, 7, 543–578.

Choong, Y.F. and Gershwin, S.B., 1987. A decomposition method for the approximate evaluation of capacitated transfer lines with unreliable machines and random processing times. *IIE Transactions*, 19, 150–159.

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.

Colledani, M. and Tolio, T., 2005a. Impact of statistical process control (SPC) on the performance of production systems—Part 1 (small systems). *In: 5th international conference on analysis of manufacturing systems—production management*, Zakynthos Island, Greece.

Colledani, M. and Tolio, T., 2005a. Impact of statistical process control (SPC) on the performance of production systems—Part 2 (large systems). *In: 5th international conference on analysis of manufacturing systems—production management*, Zakynthos Island, Greece.

Commault, C. and Semery, A., 1990. Taking into account delays in buffers for analytical performance evaluation of transfer lines. *IIE Transactions*, 22, 133–142.

Daganzo, C.F. and Blumenfeld, E.E., 1994. Assembly system design principles and tradeoffs. *International Journal of Production Research*, 32, 669–681.

Dallery, Y., 1999. Extending the scope of analytical methods for performance evaluation of manufacturing flow systems. *In: 2nd internationl conference on analysis and modeling of manufacturing systems*.

Dallery, Y., David, R. and Xie, X.-L., 1988. An efficient algorithm for analysis of transfer lines with unreliable machines and finite buffers. *IIE Transactions*, 20, 280–283.

Dallery, Y., David, R. and Xie, X.-L., 1989. Approximate analysis of transfer lines with unreliable machines and finite buffers. *IEEE Transactions on Automatic Control*, 34, 943–953.

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

Dallery, Y. and Le Bihan, H., 1999. An improved decomposition method for the analysis of production lines with unreliable machines and finite buffers. *International Journal of Production Research*, 37, 1093–1117.

De Kok, A.G., 1990. Computationally efficient approximations for balanced flowlines with finite intermediate buffers. *International Journal of Production Research*, 28, 401–419.

De Koster, M.B.M., 1987. Estimation of the line efficiency by aggregation. *International Journal of Production Research*, 25, 615–626.

De Koster, M.B.M., 1988. An improved algorithm to approximate the behavior of flow lines. *International Journal of Production Research*, 26, 691–700.

Deliman, N.C. and Feldman, R.M., 1996. Optimization of process improvement and inspection location for serial manufacturing. *International Journal of Production Research*, 34, 395–405.

Denardo, E.V. and Tang, C.S., 1997. Control of a stochastic production system with estimated parameters. *Management Science*, 43, 1296–1307.

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

Di Mascolo, M., David, R. and Dallery, Y., 1991. Modeling and analysis of assembly systems with unreliable machines and finite buffers. *IIE Transactions*, 23, 315–330.

Duenyas, I., 1994. Estimating the throughput of a cyclic assembly system. *International Journal of Production Research*, 32, 1403–1419.

Duenyas, I. and Hopp, W.J., 1992. CONWIP assembly with deterministic processing and random outages. *IIE Transactions*, 24, 97–109.

Feit, E.M. and Wu, S.D., 2000. Transfer line design with uncertain machine performance information. *IEEE Transactions on Robotics and Automation*, 16, 581–587.

Frein, Y., Commault, C. and Dallery, Y., 1996. Modeling and analysis of closed-loop production lines with unreliable machines and finite buffers. *IIE Transactions*, 28, 545–554.

Gershwin, S.B., 1987a. An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Operations Research*, 35, 291–305.

Gershwin, S.B., 1987b. Representation and analysis of transfer lines with machines that have different processing rates. *Annals of Operations Research*, 9, 511–530.

Gershwin, S.B., 1989. An efficient decomposition algorithm for unreliable tandem queuing system with finite buffers, *In*: H.G. Perros and T. Altiok, eds. *Queueing networks with blocking*. Amsterdam: Elsevier, 127–147.

Gershwin, S.B., 1991. Assembly/disassembly systems: an efficient decomposition algorithm for tree-structured networks. *IIE Transactions*, 23, 315–331.

Gershwin, S.B., 1994. *Manufacturing systems engineering*. Engelwood Cliffs, NJ: PTR Prentice Hall.

Gershwin, S.B., 2000. Design and operation of manufacturing systems—the control-point policy. *IIE Transactions*, 32, 891–906.

Gershwin, S.B. and Burman, M.H., 2000. A decomposition method for analyzing inhomogeneous assembly/disassembly systems. *Annals of Operations Research*, 93, 91–115.

Gershwin, S.B., Dallery, Y., Papadopoulos, C.T. and Smith, J.M., 2003. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic.

Gershwin, S.B., Maggio, N., Matta, A., Tolio, T. and Werner, L.M., 2001. Analysis of loop networks by decomposition. *In: 3rd Aegean international conference on design and analysis of manufacturing systems*, Tinos Island, Greece.

Gershwin, S.B. and Werner, L.M., 2007. An approximate analytical method for evaluating the performance of closed-loop flow systems with unreliable machines and finite buffers. *International Journal of Production Research*, 45, 3085–3111.

Glassey, C.R. and Hong, Y., 1993. Analysis of behaviour of an unreliable n-stage transfer line with $(n-1)$ inter-storage buffers. *International Journal of Production Research*, 31, 519–530.

Govil, M.C. and Fu, M.C., 1996. Queueing theory in manufacturing: a survey. *Journal of Manufacturing Systems*, 18, 214–240.

Han, M.-S., Lim, J.-T. and Park, D.-J., 1998. Performance analysis of serial production lines with quality inspection machines. *International Journal of System Science*, 29, 939–951.

Helber, S., 1999. Performance analysis of flow lines with non-linear flow of material, *In: Lecture notes in economics and mathematical systems*. Berlin: Springer.

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., 2005. Analysis of flow lines with Cox-2-distributed processing times and limited buffer capacity. *OR Spectrum*, 27, 221–242.

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 a continuous material merge system with limited buffer capacity and three stations, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and modleing of manufacturing systems*. Dordrecht: Kluwer Academic, 85–121.

Hong, Y., Glassey, C.R. and Seong, D., 1992. The analysis of a production line with unreliable machines and random processing times. *IIE Transactions*, 24, 77–83.

Hopp, W.J., Spearman, M.L., Chayet, S., Donohue, K.L. and Gel, E.S., 2002. Using an optimized queueing network model to support wafer fab design. *IIE Transactions*, 34, 119–130.

Huang, N. and Blumenfeld, D.E., 1998. *Estimating throughput of serial production lines with identical stations*. Warren, MI: General Motors Research & Development Center, Report OR-37.

Ignall, E. and Silver, A., 1977. The output of a two-stage system with unreliable machines and limited storage. *AIIE Transactions*, 9, 183–188.

Inman, R.R., 1999. Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. *Production and Operation Management*, 4, 409–432.

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

Iyama, T. and Ito, S., 1987. The maximum production rate for an unbalanced multi-server flow line system with finite buffer storage. *International Journal of Production Research*, 25, 1157–1179.

Jacobs, D.A., 1993. *Improvability in production systems: theory and case studies*. Thesis (PhD). University of Michigan.

Jacobs, D.A. and Meerkov, S.M., 1991. Asymptotically reliable serial production lines with a quality control system. *Computers Mathematics and Applications*, 21, 85–90.

Jacobs, D.A. and Meerkov, S.M., 1995a. A system-theoretic property of serial production lines: improvability. *International Journal of System Science*, 26, 95–137.

Jacobs, D.A. and Meerkov, S.M., 1995b. Mathematical theory of improvability for production systems. *Mathematical Problems in Engineering*, 1, 95–137.

Jacobs, D.A. and Meerkov, S.M., 1995c. System-theoretic analysis of due-time performance in production systems. *Mathematical Problems in Engineering*, 1, 225–243.

Jacobs, D.A., Kuo, C.-T., Lim, J.-T. and Meerkov, S.M., 1996. Improvability of serial production lines: theory and application, *In*: G. Yin and Q. Zhang, eds. *Recent advances in control and optimization of manufacturing systems*. Berlin: Springer, 121–143.

Jafari, M.A. and Shanthikumar, J.G., 1987. An approximate model of multistage automatic transfer lines with possible scrapping of workpieces. *IIE Transactions*, 19, 1157–1170.

Jafari, M.A. and Shanthikumar, J.G., 1989. Determination of optimal buffer storage capacities and optimal allocation in multistage automatic transfer lines. *IIE Transactions*, 21, 130–135.

Jeong, K.-C. and Kim, Y.-D., 1998. Performance analysis of assembly/disassembly systems with unreliable machines and random processing times. *IIE Transactions*, 30, 41–53.

Jeong, K.-C. and Kim, Y.-D., 1999. An approximation method for performance analysis of assembly/disassembly systems with parallel-machine stations. *IIE Transactions*, 31, 391–394.

Kang, Y. and Gershwin, S.B., 2005. Information inaccuracy in inventory systems. *IIE Transactions*, 37, 843–859.

Khouja, M., Rabinowitz, G. and Mehrez, A., 1995. Optimal robot operation and selection using quality and output trade-off. *International Journal of Advanced Manufacturing Technology*, 10, 342–355.

Kim, D.S. and Jung, B.L., 2000. The equivalence of duplicate automated serial workstations and two-workstation tandem systems and its use in serial production line analysis. *International Journal of Production Research*, 38, 1525–1538.

Kim, D.S., Kulkarny, D.M. and Lin, F., 2002. An upper bound for carriers in a three-workstation closed serial production system operating under production blocking. *IEEE Transactions on Automatic Control*, 47, 1134–1138.

Kim, J., 2004. *Integrated quality and quantity modelling of a production line*. Thesis (PhD). MIT.

Kim, J. and Gershwin, S.B., 2005. Integrated quality and quantity model of a production line. *OR Spectrum*, 27, 287–314.

Kimemia, J.G. and Gershwin, S.B., 1983. An algorithm for the computer control of production in a flexible manufacturing systems. *IIE Transactions*, 15, 353–362.

Koenigsberg, E., 1959. Production lines and internal storage—a review. *Management Sciences*, 5, 410–433.

Kouikoglou, V.S., 2000. Optimal rate allocation in unreliable, assembly disassembly production networks with blocking. *IEEE Transactions on Robotics and Automation*, 16, 429–434.

Kuhn, H., 2003. Analysis of automated flow line systems with repair crew interference, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M Smith, eds. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic, 155–179.

Kumar, P.R., 1993. Re-entrant lines. *Queueing systems*, 13, 87–110.

Kumar, S. and Kumar, P.R., 2001. Queueing network models in the design and analysis of semiconductor wafer fabs. *IEEE Transactions on Robotics and Automation*, 17, 548–561.

Kuo, C.-T., 1996. *Bottlenecks in production systems: a systems approach*. Thesis (PhD). Univesity of Michigan.

Kuo, C.T., Lim, J.-T. and Meerkov, S.M., 1996. Bottlenecks in serial production lines: a system-theoretic approach. *Mathematical Problems in Engineering*, 2, 233–276.

Kuo, C.-T., Lim, J.-T., Meerkov, S.M. and Park, E., 1997. Improvability theory for assembly systems: two component–one assembly machine case. *Mathematical Problems in Engineering*, 3, 95–171.

Lawrence, S.R. and Buss, A.H., 1995. Economic analysis of production bottlenecks. *Mathematical Problems in Engineering*, 1, 341–363.

Le Bihan, H. and Dallery, Y., 2000. A robust decomposition method for the analysis of production lines with unreliable machines and finite buffers. *Annals of Operations Research*, 93, 265–297.

Levantesi, R., Matta, A. and Tolio, T., 2003a. Performance evaluation of production lines with random processing times, multiple failure modes and finite buffer capacity—Part I: the building block, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic, 181–200.

Levantesi, R., Matta, A. and Tolio, T., 2003b. Performance evaluation of production lines with random processing times, multiple failure modes and finite buffer capacity—Part II: decomposition, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic, 201–219.

Li, J., 2000. *Production variability in manufacturing systems: a systems approach*. Thesis (PhD). University of Michigan.

Li, J., 2004a. Performance analysis of production systems with rework loops. *IIE Transactions*, 36, 755–765.

Li, J., 2004b. Modeling and analysis of manufacturing systems with parallel lines. *IEEE Transactions on Automatic Control*, 49, 1824–1829.

Li, J., 2004c. Throughput analysis in automotive paint shops: a case study. *IEEE Transactions on Automation Sciences and Engineering*, 1, 90–98.

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

Li, J., Alden, J.M. and Rabaey, J.A., 2005. Approximating feeder line reliability statistics in assembly systems with partial data collection. *Computers & Industrial Engineering*, 48, 181–203.

Li, J. and Blumenfeld, D.E., 2006. Quantitative analysis of transfer production line with Andon. *IIE Transactions*, 38, 837–846.

Li, J., Blumenfeld, D.E. and Alden, J.M., 2006. Comparisons of two-machine line models in throughput analysis. *International Journal of Production Research*, 44, 1375–1398.

Li, J., Blumenfeld, D.E. and Marin, S.P., 2007. Manufacturing system design to improve quality buy rate: an automotive paint shop application study. *IEEE Transactions on Automation Science and Engineering*, 4, 75–79.

Li, J., Blumenfeld, D.E. and Marin, S.P., 2008. Production system design for quality robustness: theory and application in automotive paint shops. *IIE Transactions*, 40, 162–176.

Li, J., Enginarlar, E. and Meerkov, S.M., 2004. Random demand satisfaction by unreliable production–inventory–customer systems. *Annals of Operations Research*, 126, 159–175.

Li, J., Enginarlar, E. and Meerkov, S.M., 2006. Conservation of filtering in manufacturing systems with unreliable machines and finished goods buffers. *Mathematical Problems in Engineering*, 2006, 1–12.

Li, J. and Huang, N., 2005. Modeling and analysis of a multiple product manufacturing system with split and merge. *International Journal of Production Research*, 43, 4049–4066.

Li, J. and Meerkov, S.M., 2000a. Production variability in manufacturing systems: Bernoulli reliability case. *Annals of Operations Research*, 93, 299–324.

Li, J. and Meerkov, S.M., 2000b. Bottlenecks with respect to due-time performance in pull serial production lines. *Mathematical Problems in Engineering*, 5, 479–498.

Li, J. and Meerkov, S.M., 2001. Customer demand satisfaction in production systems: a due-time performance approach. *IEEE Transactions on Robotics and Automation*, 17, 472–482.

Li, J. and Meerkov, S.M., 2003. Due-time performance in production systems with Markovian machines, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic, 221–253.

Li, J. and Meerkov, S.M., 2005a. Evaluation of throughput in serial production lines with non-exponential machines, *In*: E.K. Boukas and R. Malhame, eds. *Analysis, control and optimization of complex dynamic systems*. Berlin: Springer, 55–82.

Li, J. and Meerkov, S.M., 2005b. On the coefficients of variation of up- and downtime of manufacturing equipment. *Mathematical Problems in Engineering*, 2005, 1–6.

Li, J. and Meerkov, S.M., 2007. *Production systems engineering*. Livermore, CA: Wingspan Press.

Liberopoulos, G. and Caramanis, M., 1994. Production control of manufacturing systems with production rate-dependent failure rates. *IEEE Transactions on Automation Science and Engineering*, 39, 889–895.

Lim, J.-T., Meerkov, S.M. and Top, F., 1990. Homogeneous, asymptotically reliable serial production lines: theory and a case study. *IEEE Transactions on Automatic Control*, 35, 524–534.

Lim, J.-T. and Meerkov, S.M., 1993. On asymptotic reliable closed serial production lines. *Control Engineering Practices*, 1, 147–152.

Lipset, R., Van Til, R.P. and Sengupta, S., 1999. Steady-state performance analysis of serial transfer lines subject to machine and buffer failure. *IEEE Transactions on Automatic Control*, 44, 319–325.

Liu, X.-G. and Buzacott, J.A., 1989. A zero-buffer equivalence technique for decomposing queueing networks with blocking, *In*: H.G. Perros and T. Altiok, eds. *Queueing networks with blocking*. Amsterdam: North-Holland, 87–104.

Liu, X.-G. and Buzacott, J.A., 1990. Approximate models of assembly systems with finite inventory banks. *European Journal of Operational Research*, 45, 143–154.

Mandroli, S.S., Shrivastava, A.K. and Ding, Y., 2006. A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing processes. *IIE Transactions*, 38, 309–328.

Matta, A., Runchina, M. and Tolio, T., 2005. Automated flow lines with shared buffer. *OR Spectrum*, 27, 265–286.

Meerkov, S.M. and Zhang, L., 2008. Transient behavior of serial production lines with Bernoulli machines. *IIE Transactions*, 40, 297–312.

Mitra, D., 1988. Stochastic theory of a fluid model of producers and consumers coupled by a buffer. *Advanced Applied Probabilities*, 20, 646–676.

Mocanu, S., 2005. Numerical algorithms for transient analysis of fluid queues. *In: 5th international conference on analysis of manufacturing systems—production management*, Zakynthos Island, Greece.

Narahari, Y. and Khan, L.M., 1996. Modeling reentrant manufacturing system with inspection stations. *Journal of Manufacturing Systems*, 15, 367–378.

Narahari, Y. and Viswanadham, N., 1994. Transient analysis of manufacturing systems performance. *IEEE Transactions on Robotics and Automation*, 10, 330–345.

Nemec, J.E., 1999. *Diffusion and decomposition approximations of stochastic models of multiclass processing networks*. Thesis (PhD). Sloan School of Management.

Owen, J.H. and Huang, N., 2007. Local improvements that degrade system performance: case studies and discussion for throughput analysis. *International Journal of Production Research*, 45, 2351–2364.

Özdoğru, U. and Altiok, T., 2003. Analysis of two-valve fluid-flow systems with general repair times, *In*: S.B. Gershwin, Y. Dallery, C.T. Papadopoulos and J.M. Smith, eds. *Analysis and modeling of manufacturing systems*. Dordrecht: Kluwer Academic, 255–288.

Papadopoulos, H.T., Heavey, C. and Browne, J., 1993. *Queueing theory in manufacturing systems analysis and design*. London: Chapman & Hall.

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.

Papadopoulos, H.T., 1998. Analysis of production lines with Coxian service times and no intermediate buffers. *Naval Research Logistics*, 45, 669–685.

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

Perkins, J.R. and Kumar, P.R., 1995. Optimal control of pull manufacturing systems. *IEEE Transactions on Automatic Control*, 40, 2040–2051.

Proceedings of the international workshop on performance evaluation and optimization of production lines. 1997. Samos Island, Greece.

Proceedings of the 2nd Aegean international conference on analysis and modeling of manufacturing systems. 1999. Tinos Island, Greece.

Proceedings of the 3rd Aegean international conference on design and analysis of manufacturing systems. 2001. Samos Island, Greece.

Proceedings of the 4th Aegean international conference on analysis of manufacturing systems. 2003. Samos Island, Greece.

Proceedings of the 5th international conference on analysis of manufacturing systems—production management. 2005. Zakynthos Island, Greece.

Proceedings of the 5th international conference on analysis of manufacturing systems. 2007. Lunteren, the Netherlands.

Roser, C., Nakano, M. and Tanaka, M., 2003. Constraint management in manufacturing systems. *JSME International Journal Series C—Mechanical Systems Machine Elements and Manufacturing*, 46, 73–80.

Sadr, J. and Malhame, R.P., 2004a. Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines. *IEEE Transactions on Automatic Control*, 49, 68–81.

Sadr, J. and Malhame, R.P., 2004b. Unreliable transfer lines: decomposition/aggregation and optimization. *Annals of Operations Research*, 125, 167–190.

Sevastyanov, B.A., 1962. Influence of storage bin capacity on the average standstill time of a production line. *Theory of Probability Applications*, 6, 429–438.

Simon, J.T. and Hopp, W.J., 1991. Availability and average inventory of balanced assembly-like flow systems. *IIE Transactions*, 23, 161–168.

Suliman, S.M.A., 1999. A mathematical model for a buffered two-stage manufacturing cell with an unreliable transfer device. *International Journal of Production Economics*, 63, 69–81.

Syrowicz, D., 1999. *Decomposition analysis of a deterministic, multiple-part-type, multiple-failure-mode production lines*. Thesis (SM). Department of EECS: MIT.

Tan, B., 1998. Effects of the variability on the due-time performance of production lines. *International Journal of Production Economics*, 54, 87–100.

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

Tan, B., 2002. Production control of a pull system with production and demand uncertainty. *IEEE Transactions on Automatic Control*, 47, 779–783.

Tan, B. and Karabati, S., 2001. Modelling and scheduling of an asynchronous cyclic production line with multiple parts. *Journal of the Operational Research Society*, 51, 1296–1308.

Tan, B. and Yeralan, S., 1997. A decomposition model for continuous materials flow production. *International Journal of Production Research*, 35, 2759–2772.

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

Terracol, C. and David, R., 1987. An aggregation technique for performance evaluation of transfer lines with unreliable machine and finite buffers. *In: Proceedings of the IEEE international conference on robotics and automation*, 1333–1338.

Tolio, T., Matta, A. and Gershwin, S.B., 2002. Analysis of two-machine lines with multiple failure modes. *IIE Transactions*, 34, 51–62.

Urban, T.L., 1998. Analysis of production systems when run length influences product quality. *International Journal of Production Research*, 36, 3085–3094.

Vidalis, M.I. and Papadopoulos, H.T., 1999. Markovian analysis of production lines with Coxian-2 service times. *International Transactions in Operational Research*, 6, 495–524.

Viswanadham, N. and Narahari, Y., 1992. *Performance modeling of automated manufacturing system*. Englewood Cliffs, NJ: Prentice Hall.

Viswanadham, N., Sharma, S.M. and Taneja, M., 1996. Inspection allocation in manufacturing systems using stochastic techniques. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 36, 222–230.

Wein, L., 1988. Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1, 115–130.

Werner, L., 2001. *Analysis and design of closed loop manufacturing systems*. Thesis (Masters). Or Center, MIT.

Yang, S., Hu, S.J. and Koren, Y., 1999. Approximate analysis of production lines with loop, scrapping and serial/parallel blocks. *In: 2nd international conference on analysis and modeling of manufacturing systems*, Tinos Island, Greece.

Yang, S., Wu, C. and Hu, S.J., 2000. Modeling and analysis of multi-stage transfer lines with unreliable machines and finite buffers. *Annals of Operations Research*, 93, 405–421.

Yeralan, S. and Tan, B., 1997a. Analysis of Multistation production systems with limited buffer capacity. Part 1: The subsystem model. *Mathematical Computer Modelling*, 25, 109–122.

Yerelan, S. and Tan, B., 1997b. Analysis of multistation production systems with limited buffer capacity. Part 2: The decomposition method. *Mathematics and Computer Modeling*, 25, 109–123.