ELSEVIER

European Journal of Operational Research 92 (1996) 1-27

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Invited Review

Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines

H.T. Papadopoulos [a,1], C. Heavey [b,*]

[a] Department of Mathematics, University of the Aegean, GR-832 00 - Karlovassi, Samos, Greece

[b] Department of Manufacturing and Operations Engineering, University of Limerick, Limerick, Ireland

Received September 1995

Abstract

Queueing network modeling of manufacturing systems has been addressed by a large number of researchers. The purpose of this paper is to provide a bibliography of material concerned with modeling of production and transfer lines using queueing networks. Both production and transfer lines have a product-flow layout and are used in mass manufacturing. We denote production lines as flow lines with asynchronous part transfer, while transfer lines have synchronous part transfer. As well as providing a bibliography of material, a contribution of this paper is also the systematic categorization of the queueing network models based on their assumptions. This, it is hoped, will be of use to researchers of queueing networks and also manufacturing system designers. A number of suggestions are also given for further research. The basic source for this work is the book by Papadopoulos, Heavey and Browne, with the addition of the newly published papers and books (from 1992 to early 1995).

Keywords: Queueing; Production line; Transfer line

1. Introduction

A wide range of modeling techniques are available to address manufacturing system design and operational problems. From the user's point of view it is important to differentiate between generative and evaluative models. Generative models provide the user with an 'optimal solution' that satisfies the user's objective function. Evaluative models, unlike generative models, do not provide the user with an 'optimal solution'. Instead, these models evaluate a given set of decisions by providing the user with performance measures. Although generative models have the advantage of providing the user with an 'optimal solution', they are, in general, restrictive in terms of their structural assumptions, which are usually only transparent to the model developer. Evaluative models do not guarantee the user an 'optimal solution', but they usually provide the user with valuable insights into the problems being addressed. Evaluative and generative models can be combined by closing the loop between them; that is, one can use feedback from an evaluative model to modify the decisions taken by the generative model (see Ref. [224]). This paper is concerned with queueing network models which are evaluative models.

There are available several schemes to classify discrete manufacturing systems, one classification scheme is volume of production. Using volume of production as a classification scheme, three main sys-

* Corresponding author. E-mail: cathal.heavey@ul.ie

[1] E-mail: hpap@aegean.gr

SSDI 0377-2217(95)00378-9

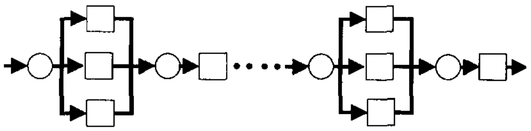

Fig. 1. Product-flow layout of mass manufacturing systems.

tem types for manufacturing can be identified: mass, batch, and job shop. The queueing network models discussed in this paper are used to analyse mass manufacturing systems. Mass manufacturing is characterised by very high production rates. Companies that use this type of manufacturing system, therefore, place great emphasis on efficiency and economy of scale in order to remain competitive. To this end, mass manufacturing systems have a product-flow layout, as is illustrated in Fig. 1.

Product-flow layout systems can be further classified according to the method of part transfer, and according to the number of part types produced by the system. There are three ways of transferring parts between work stations: synchronous transfer, asynchronous transfer and continuous transfer.

In systems with *synchronous part transfer*, parts are transported between work stations simultaneously. We refer to this type of system as a *transfer line*. Systems with *asynchronous part transfer*, are referred to as *production lines* or 'power-and-free systems'. In production lines each part moves independently of other parts. Hence, this type of system does not need to be fully balanced. Therefore, manually operated stations with cycle variations can be easily incorporated into them. As a result, they are commonly used for assembly operations. In systems with *continuous part transfer*, parts move continuously at constant speed. These systems place an upper limit on the time an operation can take.

Before reviewing queueing network models for the analysis of production and transfer lines, we list a number of books on the analysis of manufacturing systems in general which have recently been published. These are (ordered by the year of publication):

- Viswanadham and Narahari [233] offer a unique effort in presenting a unified and systematic treatment of various modeling methodologies and analysis techniques for performance evaluation of automated manufacturing systems. They begin with an overview of automated manufacturing systems, and then move on to provide a comprehensive discussion of three principal analytical modeling paradigms: Markov chains, queues and queueing networks and Petri nets. They also deal with the transient analysis of manufacturing systems performance and the important topic of performability of automated manufacturing systems (see Chapter 4 of their book, where many references relative to this issue are given).
- Buzacott and Shanthikumar [44], who provide a comprehensive treatment of stochastic models of manufacturing systems, and develop stochastic models that evaluate the performance, address issues in the design, control and operation of these systems, and provide an understanding of how different components of a manufacturing system can be coordinated. The authors present new treatments of such classical issues as workload allocation and new models of assembly systems with strict job sequence requirements. In addition, they have developed a new framework to describe the interaction between information and material flow in manufacturing.
- Askin and Standridge [15] provide an introduction to the analysis of manufacturing systems using analytical and experimental models. They bring together useful models and modeling approaches that address a wide variety of manufacturing system design and operation issues.
- Papadopoulos, Heavey and Browne [196] describe the modeling of manufacturing systems using queueing network models and two other closely related modeling techniques, simulation modeling and generative models for the buffer allocation problem. The main purpose of this work is to provide an overview of past research in this area.
- Gershwin [88] gives a fundamental description and analysis of some of the most important phenomena in material flow in manufacturing systems. He describes some potentially disruptive events that affect production plus the control actions that managers can take in anticipation of or in response to the disruptions.
- Yao [252] provides a collection of chapters which reflect recent developments of probabilistic models and methodologies, which have either been motivated by manufacturing systems research or have been demonstrated to have significant potential in such

research.

General introductory books in the area of queueing theory and stochastic processes are by Cinlar [50], Cohen [51], Gnedenko and Kovalenko [93], Gross and Harris [96], Kleinrock [132], whereas Takagi and Boguslavsky [225] provide a bibliography of books on queueing analysis and performance evaluation. Computational algorithmic approaches of computer and stochastic systems, in general, are treated by Neuts [186], Conway and Georganas [53] and Tijms [229,230].

The remainder of this paper is organised as follows: the next section (Section 2) reviews queueing network models of production lines, and Section 3 is concerned with queueing network models of transfer lines. At the beginning of each section the assumptions of the models are described. Then within each section, the different models are grouped according to their assumptions. The main source for this paper is the book by Papadopoulos et al. [196], with the addition of papers and books published in the last 2-3 years (from 1992 to early 1995). We apologize, in advance, if we have omitted any important papers in this area. This has not been done on purpose, but rather due to the difficulty in searching the wide range of sources where papers in this area are published.

2. Modeling production lines using queueing networks

Production lines are used to produce parts which have a high volume turnover, and they are characterised by a product-flow layout, a low product flexibility (the line is restricted to producing a small variety of part types) and *asynchronous* part transfer. Blocking and starving of parts are the main causes of inefficiency in production lines. These phenomena are mainly caused by variable processing times, and by disruptions to the line caused by the unreliability of stations. To increase the efficiency of the lines, buffers are placed between stations. The queueing network (QN) models mentioned in this section allow the system designer to evaluate system performance against part processing time variability, station unreliability and the size of buffers placed between stations. Other review papers in this area were given by Bitran and Dasu [24], Buzacott and Shanthikumar [43], Koenigsberg [136], Muth [181] and Perros [202], among others.

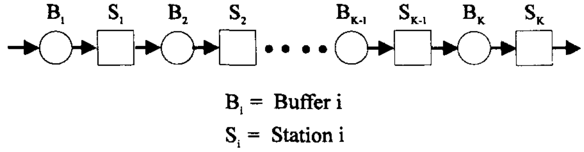

Fig. 2. The basic queueing network model of a production line.

2.1. *A classification of models*

Open queueing networks (OQN) are used to model production lines. The basic model studied is illustrated in Fig. 2. The production line consists of K stations arranged in series. Each station (S_i) has a buffer (B_i) preceding it. The buffer before the first station may be infinite or finite, all inter-station buffers are finite. Parts enter the system at station 1, pass through all stations in order where an operation is performed on the parts by the single machine in each station. The part leaves the Kth station in finished form. The common underlying assumptions of all the models are listed below:

- the line is operated at steady-state conditions;
- all random variables are independent;
- the transport times between stations are all zero;
- all failures are single-machine failures, and they are also operation-dependent (i.e. they can only fail while they are operating);
- no parts are scrapped;
- only a single part type is modeled;
- all parts queue according to queueing discipline first-in-first-out (FIFO);
- there are ample repair personnel.

The main assumptions of the models found in the literature, which distinguish one model from another, are given in Fig. 3. Manufacturing systems analysts, such as Wild [246], have identified two basic categories of manufacturing system types, namely continuous process manufacturing and discrete part manufacturing. Continuous manufacturing involves continuous production of a product, i.e. chemical plant, whereas discrete manufacturing involves the production of individual parts. Queueing network models of production lines reflect this, and so we have models for discrete part flow, and models for continuous part flow. The majority of the queueing network models of

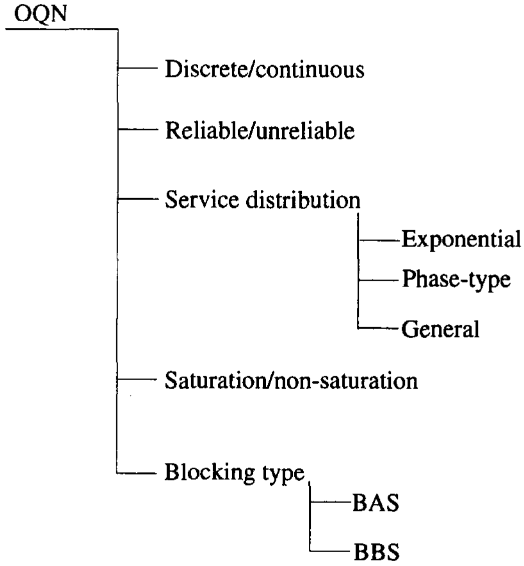

Fig. 3. Classification of OQN for modeling production lines.

production lines are for discrete parts flow. It should be noted that continuous models provide good approximations to some discrete part production lines.

Possibly, the primary assumption is the reliable/unreliable assumption. In reliable QN models of production lines the service stations are assumed to be perfectly reliable, that is, they operate without any breakdowns. These models are used to study the blocking and starving of parts caused by the variable processing and finite buffers. In unreliable models the service stations are allowed to fail, and the failure type is nearly always assumed to be operation dependent. When a service station fails it undergoes a repair. Failed service stations cannot process parts. The inclusion of unreliable service stations allows one to study the effect unreliable service stations have on the efficiency of the production line. Times to failure are assumed to be either exponential or geometric. Repair times are usually assumed to be exponentially distributed, however some models use a phase-type distribution.

As is to be expected, exponentially distributed services predominate. Recently though, a number of approximate models have been developed that allow phase-type distributions. The motivation for this is that phase-type distributions give rise to a Markovian state description while allowing general distributions to be modeled. Research work has also being carried out into asynchronous production lines with deterministic processing times. Results for these models are given in the next section, as these models are quite similar to transfer line models.

Two types of blocking exist in tandem QN models: blocking-before-service (BBS) and blocking-after-service (BAS). BAS is the prevalent blocking type in manufacturing, and so most QN models of production lines assume this type. For both production and transfer lines there is an equivalence between the two types of blocking. A line with buffer capacity vector N and BAS is equivalent to a similar line with blocking-before-service with position non-occupied (BBS-PNO) and capacity vector $N + 1$. These two lines have exactly the same behaviour, in particular, they have the same production rate [59]. This result was established by Onvural and Perros [188] in the case of exponentially distributed processing times. This equivalence was also shown by Dallery, Liu and Towsley [60] to hold for generally distributed service times. The blocking mechanisms BAS and BBS-PO are only equivalent for two-station lines. A two-station line with BAS and buffer capacity $N - 1$ is equivalent to a similar line, but with blocking-before-service with position occupied (BBS-PO) and buffer capacity N.

Another characteristic of QN models of production lines is whether the line is operated under saturation or not. In a line operated under saturation the first station is never starved, and the last station is never blocked, i.e. the line is operated at maximum capacity. When this is not the case it is usually assumed that a buffer exists in front of the first station to which parts arrive in a random pattern. A saturated line may be operated as a non-saturated one by assuming that parts enter the line after service at the first station. If not stated otherwise, models mentioned in this section operate under saturation.

The blocking caused by finite buffers greatly increases the complexity of the analysis of a QN model. Consequently, exact results for QN models of production lines exist for only a small limited number of cases, and are from a practical point of view of little use for manufacturing system design purposes. Even so, exact result models are important. Research into exact result techniques may 'ultimately' lead to efficient methods for obtaining exact results. However of more immediate importance, exact result models provide useful qualitative insight into the behaviour of these systems, they also provide results for comparison purposes against approximate results. Also, some of these models form the basis of approximate algorithms.

Evaluative models deal with the prediction of various performance measures such as the throughput, the mean queue lengths and the mean sojourn time, among others. It would be very useful if an analytical expression for the mean throughput of a production line would be available. Unfortunately, this is almost impossible, due to the mathematical complexities involved. However, several authors have dealt with this issue of trying to derive such a formula. A few of them have succeeded in obtaining an analytical formula for some cases, under certain assumptions (Hunt [113], Muth [180]). Some others provide predictive formulae for this measure, utilizing both analytical methods and curve fitting with experimentation (Baker, Powell, and Pyke [19], Blumenfeld [28] who extended the work of Muth and Alkaff [182] and Muth [181], Hira and Pandey [109], Knott [135], Martin [163] and Papadopoulos [193]).

The next subsection describes exact result models. Because of the difficulty of obtaining exact results, research has been focused on developing approximate models. As the underlying assumptions of the models described here will rarely closely fit real systems, approximate models which give reasonably accurate results are of use for engineering design purposes. The approximate models are given in three subsections. The first subsection describes those models that assume reliable machines with exponentially distributed service times. The second subsection describes reliable lines with generally distributed service times. Unreliable lines are given in the third subsection. The proceeding subsection describes continuous models, and the following subsection gives results on the reversibility of production lines.

2.2. *Exact result models*

Exact analytical results are difficult to obtain, and are only available for short production lines. These are concerned mainly with two queues in tandem with a finite buffer between them and are analysed using Markov processes, and in general via the use of transform methods (generating functions and Laplace transforms). Into this category fall the papers Neuts [184], and [185], Konheim and Reiser [137], Langaris and Conolly [145], Latouché and Neuts [147], Langaris [143], [144] and Langaris and Conolly [146].

Three methods for obtaining exact results are of significance in this subsection. Firstly, models exist which provide closed form solutions. Results of this type are available for only very small systems, i.e. systems with two to three stations. However, these results are important as they form a building block for algorithms used to obtain approximate results. The second method is the state model method. The state model method is the classical analysis of the network of queues, where all the feasible states of the Markov chain describing the model are identified, and after the steady-state probabilities are solved for, the various performance measures are calculated. Being a numerically based technique, the size of system that can be solved is dependent on the efficiency of the numerical and storage procedures used, and on the computational power of the computer. The size of the system that can be solved is severely restricted by the dimensionality of the problem. The third method is a method developed by Muth [180]. This method considers the sequence of holding times for successive work pieces at each station. It is very efficient from a computational point of view when compared to the state model method.

Closed-form models

A number of closed-form results have been developed for two-station production lines. The model by Buzacott [40] is representative of a number of these models, which use generating functions to obtain closed-form solutions. Buzacott studied the effect of station breakdowns and random processing times on the capacity of flow lines with in-process storage. The assumptions of the model are that processing and repair times are exponentially distributed, and the time between successive breakdowns of a station, measured as the number of completed operations, is geometrically distributed. For the convenience, the two stations are assumed identical.

Gershwin and Berman [89] modeled a system consisting of two unreliable machines separated by a finite buffer. The repair and work piece completion models of Gershwin and Berman are similar to those of Buzacott [40]. However, in Buzacott's failure model, the probability of a failure before a completion is independent of the time spent on the part, while in Gershwin and Berman's model, the longer an operation takes, the more likely it is that a failure occurs before the work is complete. The latter model better represents failures in mechanisms that are vulnerable during an entire operation, such as cutting tools or drivers. Buzacott's model is appropriate where the predominant cause of failure is the transfer mechanism, clamping, or some other action that takes place exactly once during an operation. Altiok and Stidham [9] remarked that the blocking type in Gershwin and Berman [89] is BBS-PO. This type of blocking does not seem to represent the physical reality of blocking in production lines. The blocking type BBS-PO is more appropriate for the modeling of communication systems [142], rather than production lines. However, as stated in the previous section this blocking type can be transformed (for a two-station line only) into BAS by simple redefinition.

Berman [23] extended the applicability of this model by replacing the exponential probability distributions of processing times with Erlang distributions. He presented an efficient method, which is independent of the buffer size, to solve analytically the steady-state probabilities of the system. Berman also studied the behaviour of some system performance measures such as the efficiency of the two machines, and the production rate of the system.

The state model method

The idea behind the state model is very simple and straightforward. There are, in general, three steps involved. First, all the feasible states of the Markov chain describing the model are identified. In the second step, the transition matrix is generated from analysing the states of the model. Once the transition matrix is obtained, the stationary equations together with the boundary conditions can be used to solve for the stationary distribution. As the number of states is very large for any realistically sized model, the state model method involves developing computer programs to automate steps 1 and 2 and the application of a solution procedure.

Two models that use the state model method are those by Hillier and Boling [106] and Altiok [6] and both use fundamentally the same method for generating the transition matrix. Hillier and Boling's [106] model assumes reliable stations, BAS blocking and the service times are exponential or Erlang distributed. Altiok [6] modeled a K-station production line with finite interstage buffers and unreliable machines. The type of blocking is BAS blocking. Processing times at station i are independent, and of phase type with S_i number of stages. The time until breakdown in a working station, i, is exponentially distributed and repair times in station i are independent and of phase type with R_i number of stages. Whenever a station is repaired, the service restarts at the phase where it has been interrupted.

The procedure developed by Heavey, Papadopoulos and Browne [104] differs considerably from that of the basic state model method. In summary, Heavey et al. examine more carefully the structure of the transition matrix from which a recursive algorithm for generating the transition matrix is developed. In Ref. [104] unreliable stations are modeled. The model described, is an extension of a previous model by Papadopoulos, Heavey and O'Kelly [197,198] to production lines with unreliable machines. The processing times at each station i is Erlang type P_i distributed with P_i, the number of phases, allowed to vary for each station. Buffers of non-identical capacities are allowed between successive stations. In general, a station may be reliable, or unreliable. Time to failure is exponentially distributed and repair times are Erlang type R_i distributed with R_i allowed to vary at each station. The performance measure of interest is the maximum throughput of the system.

The holding time model (HTM) method

The HTM method was first introduced by Muth [176], and was refined in Refs. [178,180]. Muth's method considers the sequence of holding times for successive work pieces at each station. This method is very efficient from the computational point of view, since the number of equations to be solved is greatly reduced in comparison to the state model method. The application of Muth's holding time model method yields not only the throughput, but also the cumulative distribution functions (CDFs) of the holding, starving and blocking periods. These are obtained from the recursive relationships that hold among these random variables.

The convergence of the solution procedure was found to be very fast. Details on the derivation of the various equations of the solution procedure for exponential service times are given in Ref. [4]. The balanced line case has only been treated, due to the computational complexities that occur in the derivation of the various mathematical expressions when the line is unbalanced. Comparing Muth's holding time model with the traditional Markovian state model, it is seen that in the latter method and for $K = 6$, $B = 0$ the number of states is 144, and this grows to 6765 for $K = 10$. Therefore, using the state model method with $K = 6$ and $B = 0$ a system of 144 simultaneous equations are required to be solved to obtain the steady-state probabilities. In Alkaff and Muth's method there are $\frac{1}{2}(K-2)(K-1)$ simultaneous non-linear equations to be solved to calculate the throughput. A FORTRAN program that implements this method is also available from the authors. Muth's method was also extended by Alkaff and Muth [3], to unreliable production lines. They also suggest two approximation methods in case the exact method becomes too complex, namely, (i) the single breakdown approximation, and (ii) the two-moment approximation.

Muth and Alkaff [182] applied the holding time model to the special case of a three-station line with a Laplace transformable distribution of the middle station and phase-type distribution of service times of the two outside stations.

The generality of the solution presented by Muth and Alkaff is based on the generality of the class of service time distributions that the method admits. The phase-type distribution comprises the general Erlang distribution, the hyperexponential distribution, and arbitrary mixtures of general Erlang distributions. On the other hand, the Laplace transformable distribution constitutes a rich class which, in addition to distributions of special phase type, includes deterministic (fixed) service times, uniformly distributed service times and gamma distributed service times.

The results are unifying because they comprise as special cases a number of published results [106,169,180,210]. All these cases are discussed in detail in Alkaff's thesis [2].

Other works utilizing the holding time method are Refs. [192-195] where some extensions of Alkaff and Muth's work have been made.

Other references for exact result models are works by Dallery [56] who worked on the modeling of failure and repair times, in stochastic models of manufacturing systems, using generalised exponential distributions, and Papadopoulos and O'Kelly [200] who attempted to derive the marginal probabilities and exact analytical expressions for the mean queue lengths of production lines with no intermediate buffers.

Special attention should be given to the work of Meerkov and his team, who worked on the process of continuous improvement of production systems and provided fundamental properties and guidelines (see Refs. [121,154]; in the latter paper the authors deal with homogeneous, asymptotically reliable transfer lines as defined here).

2.3. *Exponential service times, reliable machines – approximate models*

Since exact results for open queueing networks with blocking are very difficult to obtain, this has resulted in researchers developing methods that give reasonably good approximate results. From a manufacturing system designer's perspective these models are just as useful as the exact models, since the underlying assumptions of the queueing networks are hardly ever valid for real systems, i.e. processing times will not be exponentially distributed. There are many different methods for obtaining approximate results, however, most models are based on the *decomposition* method (see Refs. [21,211]).

In the decomposition method the common idea is to decompose the analysis of the original model into the analysis of a set of smaller subsystems which are easier to deal with [59]. Each decomposition method involves three steps: (i) characterizing the subsystems, (ii) deriving a set of equations that determine the unknown parameters of each subsystem, and (iii) developing an algorithm to solve these equations. In general, the set of equations can be expressed in the form of a fixed-point equation:

$$\boldsymbol{x} = f(\boldsymbol{x}),$$

which can be solved by deriving an iterative procedure of the type:

$$\boldsymbol{x}^{(k)} = f\left(\boldsymbol{x}^{(k-1)}\right),$$

where $\boldsymbol{x}^{(k)}$ is the estimate of $\boldsymbol{x}$, the vector of unknown parameters, at the kth step of the iteration procedure.

Almost all of the decomposition methods in the literature decompose a K-machine production line into a set of $K-1$ subsystems, each subsystem being associated with a buffer of the original line. Decomposition methods are approximations because (a) the subsystems are always simpler than the whole line, and thus cannot exhibit the same behaviour, and (b) some of the equations used to determine the parameters may be approximate, even within their assumptions. In decomposition methods, there is a trade-off between complexity and accuracy. A more complex characterization of subsystems will generally lead to a better approximation of the behaviour of the original line and, as a result, to more accurate results. However, obtaining the exact solution of subsystems will also be more complex and since each subsystem must usually be solved several times, the overall computational complexity will be greater. The basic principles of decomposition methods were derived by Hillier and Boling [106] in the context of flow lines with reliable machines. An excellent discussion of the decomposition method, as well as an extensive review of the literature covering the various algorithms that make use of this method, is given in Dallery and Gershwin [59].

Four representative decomposition algorithms are those by Hillier and Boling [106], Altiok [5], Perros and Altiok [204], and Brandwajn and Jow [33,32]). The algorithm by Brandwajn and Jow is a two-node decomposition algorithm because the queueing network is broken up into two-node subsystems, and it requires much more cpu time, but it has better accuracy. Recently, Yannopoulos and Alfa [250] presented a three-node decomposition algorithm for the analysis of queues in series with exponential servers, Poisson arrivals and blocking. The remaining algorithms are single-node decomposition algorithms. All these algorithms except the second one, are contained in Perros [203]. The main difference between the models is the type of subsystem the original line is decomposed into. The single-station queueing model (SSQM) used to analyse the subsystems is different in each of the algorithms. Hillier and Boling use a $M/M/1/N+1$ SSQM, Altiok uses a $M/C_2/1/N+1$ SSQM and Perros and Altiok use a $M/PH_{K-i+1}/1/N+1$ SSQM, where $i = 1, \ldots, K$ and N is the capacity of buffer i (including the part in service) of the production line. Notice that the system capacity of the isolated SSQM is $N+1$. The additional space is fictitious. It is used to accommodate a part in station $i-1$ blocked by station i. This then allows one to approximate the blocking probabilities of the stations by the probability that $N+1$ parts reside in the SSQM. The above transformation is used if the blocking type is BAS, while in the case of BBS blocking, N is used in the isolated SSQM. Another common assumption of these models is that arrivals to the isolated SSQM are assumed to be Poisson.

Other approximate models are by Takahashi, Miyahara and Hasegawa [226], Boxma and Konheim [30], Mitra and Mitrani [171-173], Bocharov and Rokhas [29], Bitran and Tirupati [25], Cheng [48], De Kok [63], De Koster [64], Feinberg and Chiu [74], Harrison and Nguyen [101], Labetoulle and Pujolle [141], Lee and Pollock [151,152], Pollock, Birge, and Alden [206], Pourbabai [207-209], and Reiser and Kobayashi [212].

2.4. *General service times, reliable machines – approximate models*

Only approximate approaches are appropriate for efficient analysis of open finite queueing networks with general processing times. There are two main approaches used to approximately model production lines with general service times and reliable machines. The first is the phase-type approach given by Altiok and Stidham and the second one is the generalised expansion method as given by Kerbache and MacGregor Smith. The 'flow-equivalent' approach of Yao and Buzacott also falls into this category but it is used to model flexible manufacturing systems.

The phase-type approach

Altiok and Stidham [10] (contained also in [6], plus see [7]) studied the approximation of general distributions, with known squared coefficient of variation, by phase-type distributions (particularly the two-stage Coxian distribution) using the first three moments. For any general distribution with the first three moments m_1, m_2, m_3 known, and with squared coefficient of variation c greater than 1, Altiok proposed the following test:

$$\frac{m_3}{m_1^3} > \frac{3}{2}(c+1)^2. \qquad (1)$$

- If condition (1) is satisfied, one can use the approximate two-stage, phase-type distribution having the same first three moments.
- If condition (1) is not satisfied, Altiok proposed choosing the closest acceptable third moment to the original third moment. This value can be found by adding a very small ϵ to the value of $3(c+1)^2/2$ corresponding to the given c. Hence, m_3 can be found using the following expression:

$$m_3 = \frac{3m_2^2}{2m_1 + \epsilon m_1^3}. \tag{2}$$

For the case where the squared coefficient of variation is less than 1, the two-moment approximation of Sauer and Chandy [215] and Marie [161] may be used. In this approximation the generalised Erlang distribution is used. The number of stages k should be such that

$$\frac{1}{k} \leqslant c \leqslant \frac{1}{k-1}. \tag{3}$$

Given that k is found from condition (3), the other parameters μ and b can be obtained from the following formulae:

$$b = \frac{2kc + k - 2 - \sqrt{k^2 + 4 - 4kc}}{2(c+1)(k-1)}, \tag{4}$$

$$\mu = \frac{b + k(1-b)}{m_1}. \tag{5}$$

Once the general service distribution has been approximated by a phase-type distribution, an appropriate model is required to analysing the system. A number of approximate models have been developed. Altiok [7] proposed an approximate algorithm for analysing tandem configurations with phase-type services. The algorithm decomposes the line into individual nodes, and is similar to the one by Perros and Altiok [204].

Two other models by Jun and Perros [125] and Gun and Makowski [99] are also applicable to analysing phase-type open queueing networks. Jun and Perros [125] analyse an open tandem configuration with BAS blocking, where each server has a C_2 service (two-stage Coxian) distribution. All nodes, except the first one, are assumed to be finite. The first node may be finite or infinite. The arrival process at the first node is Poisson. The algorithm is based on the algorithm given by Perros and Altiok [204]. That is, the system is decomposed into individual nodes with revised arrival and service processes, and augmented capacity. Each node is then analysed in isolation. In particular, the capacity of each decomposed node is augmented by one, the service process is revised to include the additional blocking delays, and the arrival process is described by a C_2 distribution (rather than a Poisson distribution). The algorithm yields the steady-state queue-length distribution of each node. The Gun and Makowski [99] algorithm is a two-node decomposition algorithm, similar to the Jun and Perros algorithm.

The generalised expansion method

The generalised expansion method is an extension of the expansion method for OQN with blocking, developed by Kerbache and MacGregor Smith [130], to non-exponential networks by the same authors [129]. This method, as well as being applicable to open tandem queueing networks, can handle split and merge configurations. The method consists of three steps: (i) network reconfiguration; (ii) parameter estimation; and (iii) feedback elimination.

Other references for production lines with general service times and reliable machines are Hildebrand [105] who considered a tandem configuration with saturated first queue, general service times and BAS mechanism. He showed that the system's throughput is the reciprocal of the effective service time at the first node. He was also able to obtain an expression for it in some cases. Knott [135] obtained an approximation for the throughput of a tandem configuration with saturated first queue, under a variety of different service time distributions. Kelly [126] considered a tandem configuration in which a unit receives the same service at each node. The service time of a unit is drawn from a general distribution. The first queue is always saturated, BAS or BBS may be assumed, and all nodes have the same finite capacity. An asymptotic expression for the throughput of the system was obtained as the number of nodes increases. Also, Kelly [127] investigated the effect of re-ordering the sequence of units on the system's throughput for a similar tandem configuration with no intermediate buffers, and Kuehn [139] analysed general queueing networks by decomposition.

2.5. Unreliable production lines – approximate models

In unreliable production lines, stations are prone to failure. Both decomposition and aggregation methods are used to approximately model production lines with unreliable stations. The decomposition method for these production lines is similar in principle to that for lines with perfect reliable stations. However, when unreliable stations are modeled the emphasis is on modeling the starvation and blocking caused by failures of stations rather than variations in processing times, as is the case for reliable lines.

The basic idea of aggregation [59] is to first replace a two-station one-buffer sub-line by a single equivalent station. Then this equivalent station is combined with a buffer and station of the original line to form a new two-station one-buffer sub-line, which is then aggregated into a single equivalent station. This process is repeated until the last or first station is reached, depending on the direction the aggregation is performed. Thus, an aggregation method to analyse a line with K stations consists of $K-1$ single aggregation steps. The aggregation method is generally not as accurate as the decomposition method. This is so because in each aggregation step, the downstream station (assuming the aggregation is performed in the order $1 \rightarrow K$) in each two-station one-buffer sub-line is represented by a station from the original line, thus ignoring the blocking of this station. In a sense the aggregation method can be looked on as a simplified decomposition method [59], i.e. the downstream station in the two-station one-buffer sub-lines of the decomposition method is represented by an original station in the aggregation method. Examples of the aggregation method used in the analysis of continuous production lines and transfer lines, are given in the next subsection and in the next section, respectively. For a discussion of the decomposition and aggregation method in the analysis of production and transfer lines the reader is directed to Dallery and Gershwin [59] and to Gelenbe and Pujolle [81], Chapter 5, for a more general treatment of these methods.

Two approximate models for the analysis of unreliable production lines are given by Choong and Gershwin [49] and Gershwin [85]. The first model [49] is based on the decomposition method, and is an extension to production lines of an earlier developed decomposition method [84] for the modeling of transfer lines (this method is mentioned in the next section). In the second model [85], a means for representing an unreliable line with stations having different (mean) service rates, with an unreliable line with equal (mean) service rates is used. This transformation allows the analyst to obtain approximate results for unreliable production lines by solving an approximately equivalent unreliable transfer line using Gershwin's decomposition algorithm [84].

Gershwin [86] also presented an efficient decomposition algorithm for the unreliable tandem queueing systems with finite buffers. Hong, Glassey, and Seong [110] developed an efficient analytical method, based on the decomposition technique proposed by Gershwin [84], for the analysis of n-machine production lines with unreliable machines and random processing times. Finally, Jafari and Shanthikumar [123] offered exact and approximate solutions to two-stage unreliable transfer lines with general up-time and down-time distributions, whereas Lipset, Sengupta, and Van Til [156] considered the effect of *buffer unreliability* upon the steady-state performance of serial transfer lines.

2.6. Continuous part flow models

In these models, the material is treated as a continuous fluid. These models are used to study the effects of failures, repairs, and finite buffers on system performance. Whilst continuous models can appropriately model manufacturing systems with continuous part flow, they also can be used to approximate discrete systems. They are mainly of use as approximations of unreliable asynchronous models with deterministic service times. The various models are divided into exact and approximate models.

Exact results

Again the exact methods can handle only short production lines (with two or three stations). Yeralan, Franck and Quasem [254] presented a general model of a continuous materials flow production line with two unreliable work stations, and an intermediate buffer storage with finite capacity. No restrictive assumptions were made about the distributions of the station breakdown and repair times when stations are blocked or starved. The production rate and the expected level of the buffer were given in closed form. Other models include the work of Finch (unpublished paper), Vladzievskii [234], and Sevast'yanov [219], all of which are described in Refs. [136,41]. Wijngaard [245] found the production rate of a production line with continuous materials flow, and derived an analytic formulation and a closed form solution. Fox and Zerbe [77] developed a production line model based on rather restrictive assumptions that buffers are empty while all stations are operating and that only one station can be repaired at any given instant. Murphy [175] developed a recursive equation to estimate the expected improvement in the production rate due to added buffers. Gershwin and Schick [90] computed performance measures, including production rate, the average level of the buffer and the probabilities of starving and blocking.

Richart [213] in her Ph.D. thesis extended these results to the two-station production line with an intermediate buffer in which the breakdown and repair times are n-stage and m-stage special Erlang distributed, respectively. The service rates of the two stations may be different. Closed form solutions for the production rate and for the expected buffer level were obtained. The model was validated against a simulation model.

Approximate results

De Koster [65] considered open networks consisting of combinations of split and merge configurations with finite intermediate buffers. The layout of the networks are arbitrary, but with the exception that no loops are allowed. A single product is manufactured and each station stores its output in at most one buffer and receives its input from at most one buffer.

De Koster used an approximation method to evaluate the throughput of the system. This algorithm consists of two steps: (a) the approximation of two multi-server stations in series with intermediate buffer by an *aggregated* single-server station and (b) the approximation of a two-station line by an *aggregated* station.

In Ref. [66] De Koster also provided an approximation method for the analysis of production lines with a continuous product flow. This algorithm is applied to lines in which the stations are unreliable with exponentially distributed times to failure and repair times. All intermediate buffers have finite capacity. The algorithm consists of repeated decomposition and aggregation steps, in which two-station lines are approximated by a single station.

Another approximation method for unreliable lines with continuous parts flow was given by Glassey and Hong [92], which appears to be better for state-dependent failures, while De Koster's method seems to give more accurate (for the throughput) results for time-dependent failures. A more recent paper is that by Yeralan and Tan [255], who in their paper provide a building block for the decomposition of continuous flow production systems.

Analysis of continuous flow lines using aggregation was proposed by Ancelin and Semery [14] and Terracol and David [228] in the case of operation-dependent failures. The up-times and down-times are exponentially distributed. The stations may have different processing rates. Thus, the stations of the original line as well as the equivalent stations have three parameters.

2.7. Reversibility of production lines

The production line has been the object of a good deal of study in the past. Much of the past work has involved the development of analytical solutions or empirical formulae for the production rate. Another aspect of production lines that has received attention is the optimization of the production rate. An important question in this context is how the production rate is affected by a rearrangement of the order of the stations. A special case of rearrangement is that of *line reversal*, that is, when each item passes through the work stations in the order $K, K-1, \ldots, 2, 1$ (the item enters the line at station K and departs from it at station 1).

Dattatreya [62] defined C-reversible and D-reversible tandem blocking queueing systems, as follows: A tandem queueing system is said to be C-*reversible* if the original system has the same capacity (production rate) as its reversed system. Both systems are empty at time $t = 0$. A blocking system is said to be D-*reversible* if the distributions of times of the nth departure epochs from both systems are identical for every n. C-reversibility was first shown analytically by Makino [160] for some simple systems.

It has been conjectured by Fujii, Tanioka, and Narutaki [80], Hillier and So [107,108], and Knott [134] that the production rate does not change under line reversal. Such a conjecture is strongly motivated by a few special cases for which it is known to be true, and by simulation results for more general cases. The conjecture was proved by Yamazaki and Sakasegawa [249] who also proved the stronger property, "D-reversibility". Separately, Dattatreya [62] and Muth [179] gave a similar proof of the reversibility property. The connection with the machine scheduling problem was discussed in Muth [177]. Theoretical extensions and applications were given by Yamazaki and Kawashima [247]. Subsequently this subject was discussed by Dattatreya [62].

Muth [179] considered a line consisting of K dissimilar work stations labeled $1, 2, \ldots, K$, and arranged in series in that order. The first station is assumed to be always busy, while the last station can never be blocked. Each raw production item enters the line at station 1, passes through all stations in order, and leaves station K in a finished form. The service time of any item at station j is a non-negative random variable denoted S_j. The random variables $S_1, \ldots, S_K$ are statistically independent, their distributions are arbitrary and not identical in general. The service time of item i at station j is S_{ij}, and the sequence S_{ij}, $i = 1, 2, \ldots$, is identically and independently distributed. It is assumed that a station cannot break down, and that each station can service only one item at a time. However, the case of station breakdown, under certain assumptions, can be reduced to the case of reliable lines, whereby periods of station breakdown become a part of the service time. Every station is at any time in one of three possible states: busy, blocked or idle. Muth did not consider buffer spaces between stations, since a buffer which can hold m items is equivalent to m stations in series, where each station has zero service time. Muth used the holding time model, to show that: (i) for any deterministic sequence of service times the production time of n items is invariant under line and time reversal, and (ii) he extended this result to the stochastic case, where the expected value of production time is invariant under line reversal alone.

The problem of whether or not reversibility can be extended to blocking systems with multi-server stations of non-deterministic service times has been considered by Yamazaki, Kawashima and Sakasegawa [248]. They were able to show that two-station blocking systems with multi-server stations of non-deterministic service times are C-reversible but are not D-reversible. The same authors also proved that C-reversibility cannot be extended to three or more station blocking systems with multi-server stations of non-deterministic service times.

Melamed [167] offered an alternative viewpoint based on a concept of duality. This led to a more intuitive methodology, and provided additional insight into the original and reversed systems. The same author [166] related the reversibility of certain discrete state Markovian queueing networks, the class of quasi-reversible networks, to the reversibility of the underlying switching process. Quasi-reversible networks are characterised by a product form equilibrium state distribution. When the state can be represented by 'customer' totals at each node, the reversibility of the state process is equivalent to the reversibility of the switching process. More complicated quasi-reversible networks require additional conditions to ensure the reversibility of the network state process. Walrand and Varaiya [238] dealt with the interconnections of Markov chains and quasi-reversible queueing networks. They were able to show that if Markov chains are coupled in a certain way, then to the resulting chain can be associated a queueing network which is itself quasi-reversible and the stationary distribution of the chain takes the product form. The product form for mixed networks is derived from the result for open networks. Walrand [236] analysed open networks of quasi-reversible nodes with a single class of 'customers' and in equilibrium. He showed, under a stability condition, that a flow on a link of such a network is Poisson, if and only if, the link is not part of a loop. This loop criterion was shown to apply to the usual quasi-reversible networks with bounded service rates.

Other references on the reversibility of tandem queues and manufacturing networks are, among others, papers by Chao and Pinedo [46], Corten and De Koster [54], and Van Dijk and Van Der Wal [232] who provided simple bounds and monotonicity results for finite multi-server exponential tandem queues. Finally, Ammar and Gershwin [12] considered a K-station transfer line with $K-1$ buffers with capacities N_i, $i = 1, 2, \ldots, K-1$. Using duality and equivalence ideas they were able to prove the known result that if the transfer line is reversed, its production rate remains unchanged and that the mean in-process inventory, $\bar{n}_i$, of the original line FTL (forward transfer line), is given by

$$\bar{n}_i^{\mathrm{FTL}} = N_i - \bar{n}_i^{\mathrm{RTL}}, \quad \forall i \tag{6}$$

where RTL denotes the reversed transfer line.

3. Modeling transfer lines using queueing networks

Most of this section deals with the modeling of transfer lines using queueing network models. Transfer lines are similar to production lines except that part transfer is *synchronous*. Therefore, the processing times at all the stations (stages) are equal and deterministic. The blocking and starving in these systems is caused by the failure of stations. Some models and results are also given for lines with unequal deterministic service times, even though in these models part transfer is asynchronous. Some results are also included for assembly/disassembly networks.

3.1. A classification of models and section overview

Transfer lines operate in a similar manner to production lines, with the only exception that part transfer is synchronous. The underlying assumptions for queueing network models of production lines, also hold for models of transfer lines. These assumptions are listed in Subsection 2.1. Assumptions that distinguish the different models of transfer lines, are given in Fig. 4. Two types of failures are considered in the models: time-dependent and operation-dependent failures. Recall that operation-dependent failures cannot occur when the station is forced down, while time-dependent failures can. Operation-dependent failures are suited to modeling failures of equipment that are dependent on the operation of the equipment, i.e. tool wear. Time-dependent failures are suited to modeling failures, such as computer network failure, which do not depend on equipment being operational.

In the next subsection, results are given for transfer lines with no intermediate buffers between stations, for both time-dependent and operation-dependent failures. For these cases the analysis is not difficult. In Subsections 3.3 and 3.4, models which allow buffers between stations, and which have operation-dependent and time-dependent failures, respectively, are mentioned. In Subsection 3.5, models and results are given for lines with unequal deterministic service times, and in the last subsection, assembly/disassembly network models are briefly summarised.

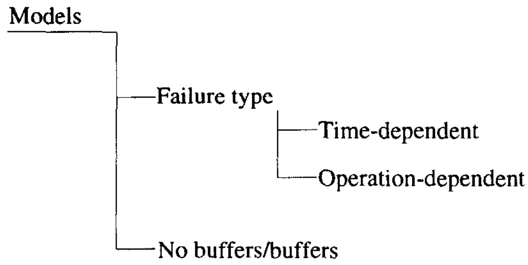

Fig. 4. Classification of models for modeling transfer lines.

3.2. Transfer line models with no buffers

In the case where transfer lines with no buffers between any two successive stations are considered, no assumptions about the form of the distributions of the 'up times' and 'down times' is required. These distributions must only have finite means and variances.

Time-dependent failures

Let U_i be the mean 'up time' of station i, which equals the difference between the clock time at which a failure of the station occurs and the clock time at which repair of the previous failure of that station ended, and let D_i be the mean 'down time' of station i.

The efficiency, E^{T}, of a transfer line with K stations and time-dependent failures, is given by

$$E^{\mathrm{T}} = \prod_{i=1}^{K} e_i^{\mathrm{T}}, \tag{7}$$

where e_i^{T} is the efficiency of a single-station, i, line, (with time-dependent failures), given by

$$e_i^{\mathrm{T}} = \frac{U_i}{U_i + D_i} = \frac{1}{1 + x_i^{\mathrm{T}}},$$

where

$$x_i^{\mathrm{T}} = \frac{D_i}{U_i}.$$

Formula (7) resembles the formula of reliability theory, giving the availability of K independent units in series.

Operation-dependent failures

Let T_i be the mean 'up-time' of station i, measured in completed cycles, and D_i be the mean 'down-time' of station i, in minutes, and finally let $1/c$ be the cycle time in minutes, when the system is running.

The efficiency, E^{O}, of a transfer line with K stations and operation-dependent failures, is given by

$$E^{\mathrm{O}} = \frac{1}{1 + \sum_{i=1}^{K} e_i^{\mathrm{O}}}, \tag{8}$$

where e_i^{O} is the efficiency of a single-station, i, line, (with operation-dependent failures), given by

$$e_i^{\mathrm{O}} = \frac{T_i}{T_i + cD_i} = \frac{1}{1 + x_i^{\mathrm{O}}},$$

and where

$$x_i^{\mathrm{O}} = \frac{cD_i}{T_i}.$$

Buzacott [37] gave a heuristic proof of formula (8), while a formal proof may be found in Barlow and Proschan [20]. Considering the two formulas (7) and (8), one may easily conclude that the efficiency of long transfer lines, with no intermediate buffers, is higher with operation-dependent failures than with time-dependent failures.

3.3. Transfer line models with buffers – operation-dependent failures

The models briefly summarised in this subsection all have the assumption of operation-dependent failures, where the down-times and up-times are geometrically distributed. Also, each model assumes that the first station is never starved and the last station is never blocked. Three models are mentioned, from those falling into this category.

Exact results models

The first model by Buzacott [38,35], (see also Ref. [36]) provides equations for the efficiency (availability) of a two-station transfer line under different assumptions. In this model, a cycle is defined to begin with a transition in the buffer level, and end with a transition in the station operating conditions.

Buzacott and Hanifin [41] showed how to derive the efficiency of a transfer line E' where operation-dependent total line failures are included, from E, the efficiency of the line without total line failure. One has

$$E' = \frac{1}{1 + G + \eta} \tag{9}$$

where G is given by

$$G = \frac{1}{E} - 1,$$

and η is defined as

$$\eta = \frac{\text{mean time of a total line failure in cycles}}{\text{mean 'up time' between total line failures in cycles}}.$$

Another exact results model is by Gershwin and Schick [91]. This model provides results for two- and three-station transfer lines. In this model, a cycle is defined to begin with a transition in the station operating conditions, and end with a transition in the buffer level, i.e. the opposite from Buzacott's model. Another difference between these two models is that Buzacott's model ignores events of small probability, such as the simultaneous failure and repair of two stations, whereas Gershwin and Schick's model takes them into account.

The system is modeled via the use of a Markov chain. The Markov chain can be solved using a state model method, as for production lines. This process requires the solution of a system of M linear equations in M unknowns, where M is the number of states of the system. However, M is large for any reasonably sized system. Gershwin and Schick observed that the Markov chain has a structure that can be exploited to reduce the number of equations that need to be solved.

An approximate decomposition model

Gershwin [84] uses the decomposition method to obtain approximate results for longer lines. An improved solution procedure to Gershwin's model, has been proposed by Dallery, David and Xie [57]. Other approximate models of long transfer lines were given by Dallery, David, and Xie [58], Shick and Gershwin [217] and Zimmern [257].

Dallery and Gershwin [59] provide an excellent review of models and analytical results both for production lines and transfer lines, whereas Gershwin in his book [88] treated, in detail, both the transfer line systems (in Chapter 3, the two-stage lines, and in Chapter 4, the decomposition of long transfer lines), and the assembly/disassembly systems (in Chapter 5).

3.4. Transfer line models with buffers – time-dependent failures

In this subsection, two models with time-dependent failures are included as summarised by Buzacott and Hanifin [41]. These are the Finch model (see Ref. [136]), and Sheskin's [222,223] model. These models have, in addition to the common assumptions, the following assumptions:

(i) Down times and up times are geometrically distributed.

(ii) For any stage i, the failure probability per cycle, p_i, and the repair probability per cycle, q_i, are such that

$$p_i + q_i = 1.$$

This assumption is very unlikely to be valid.

(iii) Transfer is carried out in *two steps*: On the *first step*, a job is transferred from each (operating) stage to the following buffer provided this is not full, and on the *second step*, a job is transferred to each operating stage from the preceding buffer, provided this is not empty.

(iv) All failures and repairs are synchronised to occur immediately after completion of the first step of transfer and prior to the beginning of the second step of transfer.

All the results are for two-station transfer lines. Other common assumptions of these models are, once a failure occurs repair starts immediately (i.e. a sufficient number of repair personnel are available) and more than one event can occur in a cycle.

3.5. Unequal deterministic service time models

The research work described here assumes unequal deterministic processing times and single machine random failures. Part transfer is asynchronous and therefore, strictly speaking, these models do not fall into the category of transfer lines, as we have defined. However, as in transfer lines service times are deterministic. A few fundamental results will be mentioned as given by Friedman [79], Avi-Itzhak and Yadin [17], Avi-Itzhak [16] and Muth [176].

Friedman [79] dealt with the following two models. The first model consists of n stations in series, $A_1, \ldots, A_n$, where station $A_i = (S_i; m_i)$ consists of m_i parallel servers, each having the same constant service time S_i, $i = 1, 2, \ldots, n$. Assume a sequence of production items (customers) $C_0, C_1, \ldots$ indexed in order of arrival, such that C_k arrives at the first station A_1 at time $t_{0,k}$ and departs from each station A_i at time $t_{i,k}$. Model 1 is described by

$$t_{0,k} \to (S_1; m_1) \to t_{1,k} \to \cdots \to (S_n; m_n) \to t_{n,k}\,. \tag{10}$$

The second model is similar to the first, except that one of the stations, say $A_\ell = (S_{\ell,k})$ is a single server ($m_\ell = 1$) having variable service times, $S_{\ell,k}$, such that $S_{\ell,k} \geqslant S_i/m_i$, $\forall k$ and $\forall i = 1, 2, \ldots, n$, $i \neq \ell$.

Friedman also assumed fixed but arbitrary arrival time sequences and in model 2, fixed service time sequences, and he derived the following results (proofs are omitted):

Result 1. Given any input sequence $\{t_{0,k}\}$ to any model 1 or model 2 system (10), then the output sequence $\{t_{n,k}\}$ is independent of the order of the stations. The proof for the two-station case of model 1 was given by Passman [79], while Friedman proved the two-station case for model 2 and he extended the proof for the n-station case for both models 1 and 2.

A similar result was obtained by Avi-Itzhak [16] and Muth [176].

Result 2. Station $A_1 = (S_1; m_1)$ *dominates* station $A_2 = (S_2; m_2)$, denoted by $A_1 \supseteq A_2$, iff (if and only if) $S_2 \leq \lfloor m_2/m_1 \rfloor S_1$, where $\lfloor m_2/m_1 \rfloor$ stands for the greatest integer number not exceeding m_2/m_1. For model 2 $A_1 = (S_{1,k})$, and $m_1 = 1$.

Result 3. Dominance is *transitive*, i.e. if $A_1 \supseteq A_2$ and $A_2 \supseteq A_3$ then $A_1 \supseteq A_3$.

Result 4. Dominance is *persistent*, i.e. if $A_i \supseteq A_j$ are non-adjacent stations in a system

$$A_1 \to A_2 \to \cdots A_i \to \cdots A_{j-1} \to A_j \to \cdots A_n$$

then no waiting ever occurs at A_j, regardless of the input sequence to the system.

These results are easily extended to permit arbitrary interarrival distributions, and arbitrary but dominant, processing time distributions.

Friedman stated that any model 1 or model 2 system

$$t_{0,k} \to A_1 \to A_2 \to \cdots \to A_n \to t_{n,k} \tag{11}$$

reduces to the single-station form

$$t_{0,k} \to A_j \to [S] \to t_{n,k}, \tag{12}$$

where S is the sum of the processing times of all dominated stations, each taken once. Any model 2 system can always be reduced to (12) by placing the variable service station first. Also, any model 1 system with all $m_i = 1$ can be reduced to (12) by placing the station with the longest service time first.

Friedman proposed the following reduction procedure based on Results 1-4:

- Order the stations by increasing multiplicity m_i, and in case of equal multiplicities, by decreasing processing time.
- Within each equal multiplicity sequence, the first station dominates all other stations. Move the dominated stations to the rear of the system.
- Apply Result 2 (the dominance test) to the remaining stations and move all dominated stations to the rear.
- Replace the dominated stations by $[S]$, where S is the sum of their processing times, each station taken once. The result is called the *reduced form*

$$t_{0,k} \to A_i \to \cdots \to A_p \to [S] \to t_{n,k} \tag{13}$$

of the original system (11).

This reduction procedure has some important applications. For example, certain steady-state time- and customer-dependent problems for the system and for its stations, involving waiting time, queue length, and busy period, can be reduced for any interarrival distribution to corresponding problems for a system of fewer stations.

Avi-Itzhak and Yadin [17] considered a production system consisting of a sequence of two service stations with infinite queue allowable before the first station, and no queue allowable between the stations. They were able to obtain the moment generating functions of the steady-state queueing times, as well as the generating functions of the steady-state numbers of items in the various parts of the system under assumptions of Poisson process of arrivals and arbitrarily distributed processing times at both stations. They investigated the cases of deterministic and exponential processing times, and they extended their results to include the system with a sequence of two stations with a finite intermediate queue allowable between them, infinite queue allowed before the first station, Poisson process of arrivals and deterministic processing times at both stations.

Avi-Itzhak [16] studied a production system with an ordered sequence of K stations under the following assumptions: (i) parts arrive according to an arbitrary process of arrivals, (ii) they are processed on a basis of FIFO, (iii) every one of them is processed through all the stations according to the order of the stations, (iv) processing times at all stations are deterministic given by $S_i \geqslant 0$, $i = 1, 2, \ldots, K$ respectively, and (v) an unlimited queue is allowable in front of the first station, and queues of arbitrary sizes are allowed before other stations, that is, if by N_j is denoted the buffer size of the buffer before the j^{th} station, it holds: $N_1 = \infty$ and $0 \leq N_j < \infty$ for $j = 2, 3, \ldots, K-1$. Each station in the sequence consists of one server. For this system, Avi-Itzhak proved Result 1 of Friedman and the following theorem:

Theorem 1. *Let $S_N = \max\{S_1, \ldots, S_K\}$, $(1 \leq N \leq K)$. Then for any specified process of arrivals the time spent in the system by the nth arriving job equals $\sum_{i=1}^{K} S_i$ plus the time that the same job would have been waiting in the queue of a single server system, $(K = 1)$, with deterministic processing time equalling S_N, assuming the same arrival process for this system.*

Avi-Itzhak proved that Result 1 of Friedman holds for any queue disciplines by which a station is never empty while customers are present in the queue before it (excluding preemptive disciplines). He also was able to show that Result 1 of Friedman and Theorem 1 are true for the case where each station in the sequence consists of r machines in parallel and processing times are equal at all machines belonging to one station.

Finally, an interesting conclusion is that the assumption of exponential processing times might lead to extremely large waiting times when dealing with a sequence of stations. In the case of single server stations with deterministic processing times and Poisson arrivals, the expected steady-state waiting time (not including service) is given as

$$\frac{\rho_N^2}{2\lambda(1-\rho_N)},$$

and is finite even in the case where $K \to \infty$, where $\rho_i = \lambda S_i$, for any $i = 1, 2, \ldots, K$. When assuming exponential processing times at all stations and unlimited queues, the steady-state waiting time is given as

$$\frac{\sum_{i=1}^{K} \rho_i^2}{\lambda(1-\rho_i)}.$$

This expression tends to ∞ when $K \to \infty$.

Newell [187] analysed a similar configuration, under the assumption that the first node is saturated. He obtained an explicit expression for the cumulative number of units to depart from server i, $i = 1, 2, \ldots, M$, by time t, assuming initially an empty system. Also, Altiok and Kao [8] obtained bounds for the throughput of a tandem configuration which is similar to the one by Avi-Itzhak discussed above, but assuming Poisson arrivals and a finite capacity at the first node.

Muth [176] also gave results for production lines with deterministic service times. Consider again a K-station production line with the same assumptions as above. It is assumed that a station cannot break down and that each station can service only one item at a time. A station can be either busy, blocked or idle, and the system is operated under saturation. There may be buffers provided between work stations to diminish the occurrence of blocking. The capacity of the buffer following station i is denoted by b_i, $i = 1, 2, \ldots, K-1$. Each buffer location can be regarded as a special type of work station, with zero service time. Let B denote the total number of buffer spaces in the system, i.e.

$$B = \sum_{i=1}^{K-1} b_i.$$

Muth proved the following results, as well as Result 1 of Friedman:

Result 1. Let work station m be the station with the greatest service time, i.e.

$$S_m = \max\{S_1, S_2, \ldots, S_K\}.$$

If this holds for several values of the index m, let m represent the index of smallest value, so that

$$S_i < S_m \quad \text{for } i = 1, 2, \ldots, m-1.$$

By partitioning the production line into two sub-lines, the first consisting of stations 1 through m, and the second consisting of the remaining stations, it can be shown that the *production time*, P, which is the total time a part spends in the production line, is given by

$$P = mS_m + \sum_{i=m+1}^{K} S_i, \tag{14}$$

and it does depend on the order of the work stations. The minimum production time is achieved by placing the station with the greatest service time first. The order of all other stations is immaterial. The production time corresponding to the optimal arrangement is the sum of the service times of all stations.

Result 2. The production rate of a production line with deterministic (fixed) service times is not changed by the addition of buffers.

3.6. Modeling assembly/disassembly networks

This subsection is concerned with *assembly* systems which can be viewed as extensions of the flow line structure. An assembly system is a manufacturing system in which some machines perform assembly operations. There are two kinds of assembly systems (Liu [157]): (i) those that add components to a work piece (such as printed circuit or surface mount assembly), and (ii) those that assemble different entities (work pieces) that have themselves already been processed within the manufacturing system. With respect to modeling and analysis, the first case is not different from a flow line. The second can form a network and is what we deal with here.

Assembly/Disassembly (A/D) network models

An A/D system consists of a set of A/D stations interconnected by a set of buffers such that each buffer has exactly one upstream server and one downstream station. The terms fork and join are sometimes used to denote disassembly and assembly, respectively. An A/D station has a set of input buffers and a set of output buffers. The number of input and output buffers can be different. An A/D station pulls one entity from each of its upstream buffers and delivers one entity to

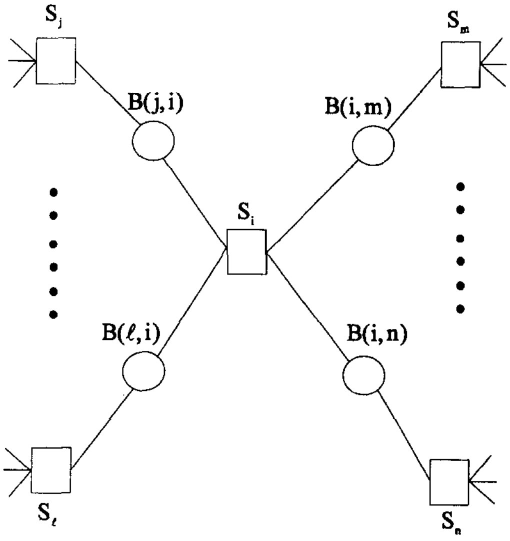

Fig. 5. An assembly/disassembly network

each of its downstream buffers. An example of an A/D network appears in Fig. 5. Each buffer is connected to exactly one upstream station and one downstream station. Each station is connected to at least one buffer.

As well as defining the blocking type, the mechanism for loading/unloading parts to and from the stations needs to be specified. Dallery, Liu, and Towsley [60] defined two different types of loading and unloading mechanisms, which they refer to as *independent* and *simultaneous* loading and unloading. Independent loading (or unloading) means that the different subparts can be loaded on (or unloaded from) the station independently of one another, while simultaneous loading (or unloading) means that all subparts must be loaded on (or unloaded from) the station at the same time.

Many of the results derived for production lines can be extended to the case of A/D models. The sample path behaviour of A/D networks can be described by means of evolution equations (see Ref. [60]). As a result, properties like conservation of flow [11,60], monotonicity (see Ref. [1] for assembly systems, Ref. [221] for closed loop systems, and Ref. [60] for general A/D networks), and reversibility [60,157] can be established using these evolution equations. Ammar and Gershwin [13] provided equivalence relations in queueing models of A/D queueing networks with blocking.

Methods for analysing A/D networks

Gershwin [82,83,87] developed a decomposition method for A/D networks, which is based on his earlier flow line decomposition method [84]. With the help of Otero [189] he developed a new and faster algorithm for the analysis of A/D networks. All approximate methods for A/D systems pertain to unreliable machines. These methods are extensions of the decomposition and aggregation methods used to analyse production and transfer lines. All decomposition methods again decompose the original system into a set of two-machine, one-buffer system.

Di Mascolo, David, Dallery [68] studied continuous material flow A/D networks. Liu [157] used sample path methods to justify decomposition. Hopp and Simon [111] showed that the throughput rate of a three-machine, two-buffer reliable assembly system with exponential processing times, is less than those of the M/M/1 queues (with finite buffers) obtained if one of the machines is replaced by an infinitely fast machine. Other references in the area of modeling assembly systems are, among others, works by Baker, Powell and Pyke [18], Blumenfeld [27], Bulgak and Sanders [34], Daganzo and Blumenfeld [55], Figour, Jollès, and Mascle [75], Harrison [100], Hopp and Simon [112], and Johri [124].

4. Areas for further research

It is amazing to see how many papers have been published in the area of queueing network modeling of production and transfer lines. However, there still remains a range of issues which need to be addressed in this area. Research areas that merit further work fall into two categories: application of the models to manufacturing systems design and extensions to the models.

Application

• Very few papers exist (the authors are only aware of the paper by Buzacott and Hanifin [41]) on the application of QN models of production and transfer lines. The main conclusion of Buzacott and Hanifin [41] was that:

"The available analytic models do not yet enable precise prediction of the effect of the inventory bank because their assumptions on the distributions are too elementary for the real world situation."

Since this paper was published a number of extensions to QN models of production and transfer lines have been developed. Good approximate models for reliable, unreliable and general service times have been developed. Papers evaluating these new models within an industrial environment would be of great benefit.

• "To integrate statistical process control (SPC) into teaching and application of queueing for manufacturing' (T. Chang's presentation at the ORSA/TIMS Joint National Meeting of Detroit, 1994). We agree with Prof. Chang's recommendation, as in queueing modeling: (a) arrival times and service times are often derived from historical data; (b) observed variation in data may not be all inherent system noise (common cause), and (c) some variation in data may be due to the occurrences of abnormal events (special causes). Therefore, an operation should be running in *statistical control* before it can be properly modelled. This has not been a practice in OR modeling. Data reflects both common and special cause variations. Data which has a mix of these two variations is likely to be fitted by a skewed probability distribution or a uniform distribution. An inflated variance estimate from such data may lead to over-design and/or under-control, which are both costly.

Without on-line control and analysis, an OR-designed system could not be expected to provide satisfactory results. In queueing modeling, the OR analysts should use only the data of the operation/process in statistical control for system design. Equally important, the performance of a designed system should be evaluated by the data of the operation in statistical control.

Extensions to models

• To try and derive a general analytical formula for the mean throughput of a general long production line with K stations and unequal buffers between any two successive stations of the line. We know that it is almost impossible to obtain such a general exact formula, but, for some special cases, with certain simple assumptions concerning the distributions of the processing times, the repair times and the times to failure of the individual machines of the line, we strongly believe that this is feasible. Otherwise, very accurate approximate expressions could be given, utilising both analytical methods and numerical curve fitting or even simulation (see the works of Muth, Blumenfeld, Papadopoulos, Martin and Liu and Lin [159].

• To use appropriate distributions to fit the processing times, the repair times and the times to failure of the real manufacturing systems, instead of the exponential distribution, or any other distribution with high coefficient of variation, such as Coxian or hyperexponential, etc., which are mathematically convenient but unrealistic in most cases. It is well known that the service time distributions have an effect on system performance.

• Other extensions include scrapping of parts [220], parallel machines [114,70,76,170], limited repair personnel [69,71], non-zero transports times [52,157], and multi-part types. These references were obtained from Dallery and Gershwin [59].

• All the models described in this paper deal with steady-state mean outputs. Dallery and Gershwin [59] also make the point that the *variance* of output during a time period is also important. Variability is an inherent characteristic of these systems. Therefore, it is important to be able to quantify this characteristic. Dallery and Gershwin [59] give references for three papers which deal with this problem, Refs. [168,150] and an unpublished paper by Ou and Gershwin [190]. These papers only treat two-machine lines, and they obtain results that are difficult to use and understand intuitively.

• The extension of the results for production lines to assembly/disassembly systems is also an important area of research.

Closing, we would like to express our sincere thanks to the various authors of the papers and books, included in this review paper, who have sent us the relevant material and made their comments and recommendations, and we apologize, for having omitted several other good and instructive papers. This has not been done on purpose, but rather due to the infeasibility in getting the information.

References

[1] Adan, I., and van der Wal, J. (1989), "Monotonicity of the throughput in single server production and assembly networks with respect to the buffer sizes", in: H. Perros T.

Altiok (eds)., *Queueing Networks with Blocking*, North-Holland, Amsterdam, 345-356.

[2] Alkaff, A. (1985), "The throughput rate of multistation production lines", PhD thesis, Dept. of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

[3] Alkaff, A., and Muth, E.J. (1986), "Analysis of multistation production lines with station breakdown", Technical Report, Dept. of Ind. Engineering and Operations Research, Syracuse University, Syracuse, NY 13210.

[4] Alkaff, A., and Muth, E.J. (1987), "The throughput rate of multistation production lines with stochastic servers", *Probability in the Engineering and Informational Sciences* 1, 309-326.

[5] Altiok, T. (1982), "Approximate analysis of exponential tandem queues with blocking", *European Journal of Operational Research* 11, 390-398.

[6] Altiok, T. (1985), "Production lines with phase-type operations and repair times and finite buffers", *International Journal of Production Research* 23(3), 489-498.

[7] Altiok, T. (1989), "Approximate analysis of queues in series with phase-type service and blocking", *Operations Research* 37/4, 601-610.

[8] Altiok, T., and Kao, Z. (1987), "Bounds for throughput in production/inventory systems in series with deterministic processing times", Technical Report, Industrial Engineering Rept., Rutgers University.

[9] Altiok, T., and Stidham, S. (1982), "A note on transfer lines with unreliable machines, random processing times, finite buffers", *IIE Transactions* 14, 125-127.

[10] Altiok, T., and Stidham, S. (1983), "The allocation of interstage buffer capacities in production lines", *IIE Transactions* 15, 292-299.

[11] Ammar, M.H. (1980), "Modelling and analysis of unreliable manufacturing assembly networks with finite storages", Technical Report LIDS-TH-1004, MIT Laboratory for Information and Decision Systems.

[12] Ammar, M.H., and Gershwin, S.B. (1980), "Equivalence relations in queueing models of manufacturing networks", in: *Proceedings of the 19th IEEE Conference on Decision and Control.*

[13] Ammar, M.H., and Gershwin, S.B. (1989), "Equivalence relations in queueing models of fork/join queueing networks with blocking", *Performance Evaluation* 10, 233-245.

[14] Ancelin, B., and Semery, S. (1987), "Calcul de la productivité d'une ligne intégrée de fabrication: calif, un logiciel industriel basé sur une nouvelle heuristique", *APII* 21, 209-238.

[15] Askin, R.G., and Standridge, C.R. (1993), *Modeling and Analysis of Manufacturing Systems*, Wiley, New York.

[16] Avi-Itzhak, B. (1965), "A sequence of service stations with arbitrary input and regular service times", *Management Science* 11, 565-571.

[17] Avi-Itzhak, B., and Yadin, M. (1965), "A sequence of two servers with no intermediate queue", *Management Science* 11, 553-564.

[18] Baker, K.R., Powell, S.G., and Pyke, D.F. (1989), "Buffered and unbuffered assembly systems with variable processing times", Technical Report 246, Amos Tuck School of Business Administration, Dartmouth College, October.

[19] Baker, K.R., Powell, S.G., and Pyke, D.F. (1994), "A predictive model for the throughput of unbalanced, unbuffered three-station serial lines", *IIE Transactions* 26/4, 62-71.

[20] Barlow, R.E., and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York.

[21] Bell, P.C. (1982), "The use of decomposition techniques for the analysis of open restricted queueing networks", *Operations Research Letters* 1/6, 230-235.

[22] Berkley, B.J. (1991), "Tandem queues and kanban-controlled lines", *International Journal of Production Research* 29/10, 2057-2081.

[23] Berman, O. (1982), "Efficiency and production rate of a transfer line with two machines and a finite storage buffer", *European Journal of Operational Research* 9, 295-308.

[24] Bitran, G.R., and Dasu, S. (1992), "A review of open queueing network models of manufacturing systems", *Queueing Systems* 12, 95-134.

[25] Bitran, G.R., and Tirupati, D. (1988), "Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference", *Management Science* 34/1, 75-100.

[26] Bitran, G.R., and Tirupati, D. (1989), "Tradeoff curves, targeting and balancing in manufacturing queueing networks", *Operations Research* 37/4, 547-564.

[27] Blumenfeld, D.E. (1989), "Performance comparison of assembly systems with fixed and flexible cycle times", *The International Journal of Flexible Manufacturing Systems* 1, 207-222.

[28] 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* 28/6, 1163-1182.

[29] Bocharov, P.P., and Rokhas, G.P. (1980), "On an exponential queueing system in series with blocking", *Problems of Control and Information Theory* 9/6.

[30] Boxma, O.J., and Konheim, A.G. (1981), "Approximate analysis of exponential queueing systems", *Acta Informatica* 15, 19-66.

[31] Brandon, J., and Yechiali, U. (1991), "A tandem Jackson network with feedback to the first node", *Queueing Systems* 9, 337-352.

[32] Brandwajn, A., and Jow, Y.L.L. (1985), "Tandem exponential queues with finite buffers", in: *Proc. Int. Seminar on Computer Networking and Performance Evaluation*, September.

[33] Brandwajn, A., and Jow, Y.L. (1988), "An approximation method for tandem queues with blocking", *Operations Research* 36/1 (January-February), 73-83.

[34] Bulgak, A.A., and Sanders, J.L., "Approximate analytical performance models for automatic assembly systems with

statistical process control and automated inspection", *Journal of Manufacturing Systems* 10/2, 121-133.

[35] Buzacott, J.A. (1967), Markov chain analysis of automatic transfer lines with buffer " stock and reliability of systems with in service repair", PhD thesis, Dept. of Engineering Production, University of Birmingham.

[36] Buzacott, J.A. (1967), "Automatic transfer lines with buffer stocks", *International Journal of Production Research* 5/3, 183-200.

[37] Buzacott, J.A. (1968), "Prediction of the efficiency of transfer lines without internal storage", *International Journal of Production Research* 6/3.

[38] Buzacott, J.A. (1969), "Methods of reliability analysis of production systems subject to breakdowns", in: *Operations Research and Reliability, Proc. of a NATO Conf.*, 211-232.

[39] Buzacott, J.A. (1971), "The role of inventory banks in flow-line production systems", *International Journal of Production Research* 9, 425-435.

[40] Buzacott, J.A. (1972), "The effect of station breakdowns and random processing times on the capacity of flow lines with in-process storage", *AIIE Transactions* 4/4 (December), 308-312.

[41] Buzacott, J.A., and Hanifin, L.E. (1978), "Models of automatic transfer lines with inventory banks a review and comparison", *AIIE Transactions* 10/2 (June), 197-207.

[42] Buzacott, J.A., and Kostelski, D. (1987), "Matrix-geometric and recursive algorithm solution of a two-stage unreliable flow line", *IIE Transactions* 19/4, 429-438.

[43] Buzacott, J.A., and Shanthikumar, J.G. (1992), "Design of manufacturing systems using queueing models", *Queueing Systems* 12, 135-213.

[44] Buzacott, J.A., and Shanthikumar, J.G. (1993), *Stochastic Models of Manufacturing Systems*, Prentice Hall, Englewood Cliffs, NJ.

[45] Caseau, P., and Pujolle, G. (1979), "Throughput capacity of a sequence of queues with blocking due to finite waiting room", *IEEE Transactions on Software Engineering*, 5/6, 631-642.

[46] Chao, X., and Pinedo, M. (1992), "On reversibility of tandem queues with blocking", *Naval Research Logistics* 39, 957-974.

[47] Cheng, D.W. (1992), "Second order properties in a tandem queue with general blocking", *Operations Research Letters* 12, 139-144.

[48] Cheng, D.W. (1994), "On the design of a tandem queue with blocking: Modeling, analysis, and gradient estimation", *Naval Research Logistics* 41, 759-770.

[49] 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* 19/2, 150-159.

[50] Cinlar, E. (1975), *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, NJ.

[51] Cohen, J.W. (1969), *The Single Server Queue*, North-Holland, Amsterdam.

[52] Commault, C., and Semery, A. (1990), "Taking into account delays in buffers for analytical performance evaluation of transfer lines", *IIE Transactions* 22/2, 133-142.

[53] Conway, A.E., and Georganas, N.D. (1988), *Queueing Networks - Exact Computational Algorithms: A Unified Theory Based on Decomposition and Aggregation*, MIT Press, Cambridge, MA.

[54] Corten, M., and De Koster, R. (1990), "On the reversibility of manufacturing networks", *Journal of Mathematical Analysis and Applications* 145, 309-327.

[55] Daganzo, C.F., and Blumenfeld, D.E. (1994), "Assembly system design principles and tradeoffs", *International Journal of Production Research* 32/3, 669-681.

[56] Dallery, Y. (1994), "On modeling failure and repair times in stochastic models of manufacturing systems using generalized exponential distributions", *Queueing Systems* 15, 199-209.

[57] Dallery, Y., David, R., and Xie, X. (1987), "An efficient algorithm for analysis of transfer lines with unreliable machines and finite buffers", Technical Report LAG No. 87-17, Laboratoire d'Automatique de Grenoble (UA 228 CNRS), ENSIEG-INPG), Saint Martin d'Heres, France, February.

[58] 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* 34/9, 943-953.

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

[60] Dallery, Y., Liu, Z., and Towsley, D. (1994), "Equivalence, reversibility, symmetry and concavity properties in fork/join queueing networks with blocking", *Journal of the ACM* 41/5, 903-942.

[61] Das, T.K., and Wortman, M.A. (1993), "Reliability of multistage production systems controlled by automated visual inspection", *IEEE Transactions on Reliability* 42/3, 362-368.

[62] Dattatreya, E.S. (1978), "Tandem queueing systems with blocking", PhD thesis, IEOR Dept., University of California, Berkeley.

[63] De Kok, A.G. (1990), "Computationally efficient approximations for balanced flowlines with finite intermediate buffers", *International Journal of Production Research* 28/2, 401-419.

[64] De Koster, M.B.M. (1987), "Estimation of line efficiency by aggregation", *International Journal of Production Research* 25/4, 615-626.

[65] De Koster, M.B.M. (1988), "Approximate analysis of production systems", *European Journal of Operational Research* 37, 214-226.

[66] De Koster, M.B.M. (1988), "An improved algorithm to approximate the behaviour of flow lines", *International Journal of Production Research* 26/4, 691-700.

[67] De Souza E Silva, E., and Lavenberg, S.S. (1989), "Calculating joint queue-length distributions in product-form queueing networks", *Journal of the Association for Computing Machinery* 36/1, 194-207.

[68] Di Mascolo, M., David, R., and Dallery, Y. (1991), "Modeling and analysis of assembly systems with unreliable machines and finite buffers", *IIE Transactions* 23/4, 315-330.

[69] Dudick, A. (1979), "Fixed-cycle production systems with in-line inventory and limited repair capability", PhD thesis, Columbia University.

[70] Elsayed, E.A., and Hwang (1984), "Analysis of manufacturing systems with buffer storage and redundant machines", Technical Report, IE Working Paper 84-101, Department of Industrial Engineering, Rutgers University.

[71] Elsayed, E.A., and Turley, R.E. (1980), "Reliability analysis of production systems with buffer storage", *International Journal of Production Research* 18/5.

[72] Evans, R.V. (1967), "Geometric distribution in some two dimensional queueing systems", *Operations Research* 15, 830-846.

[73] Faddev, D.K., and Faddeeva, V.N. (1963), *Computational Methods of Linear Algebra*, W.H. Freeman and Company, San Francisco.

[74] Feinberg, B.N., and Chiu, S.S. (1987), "A method to calculate steady-state distributions of large Markov chains by aggregating states",*Operations Research* 35/2, 282-290.

[75] Figour, J., Jollès, B., and Mascle, C. (1991), "Free-transfer lines with unreliable assembly stations and small buffers", *International Journal of Production Research* 29/5, 979-996.

[76] Forestier, J.P. (1980), "Modélisation stochastique et comportement asymptotique d'un systeme automatisé de production", *RAIRO Automatique* 14/2, 127-144.

[77] Fox, R.J., and Zerbe, D.R. (1974), "Some practical system availability calculations", *AIIE Transactions* 6/3, 228.

[78] Fredericks, A.A. (1986), "Performance analysis modeling for manufacturing systems", *AT&T Technical Journal* 65/4, 25-34.

[79] Friedman, H.D. (1965), "Reduction methods for tandem queuing systems", *Operations Research* 13, 121-131.

[80] Fujii, S., Tanioka, T., and Narutaki, R. (1974), "On the balance equations for tandem queues" (in Japanese), Reported at Japan Operations Research Meeting, Spring.

[81] Gelenbe, E., and Pujolle, G. (1987), *Introduction to Queueing Networks*, Wiley, New York.

[82] Gershwin, S.B. (1986), "Modeling and analysis of assembly/disassembly networks", *Presented at the IEEE Conference on Systems, Man and Cybernetics*, Atlanta, GA, October.

[83] Gershwin, S.B. (1986), "Representation and analysis of transfer lines", Technical Report LIDS-P-1579, Massachusetts Institute of Technology, July.

[84] Gershwin, S.B. (1987), "An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking", *Operations Research* 35/2 (March-April), 291-305.

[85] Gershwin, S.B. (1987), "Representation and analysis of transfer lines with machines that have different processing rates", *Annals of Operations Research* 9, 511-530.

[86] Gershwin, S.B. (1989), "An efficient decomposition algorithm for the unreliable tandem queueing systems with finite buffers", in: *Queueing Networks with Blocking*, Raleigh, NC, May 20-21.

[87] Gershwin, S.B. (1991), "Assembly/disassembly systems: An efficient decomposition algorithm for tree-structured networks", *IIE Transactions* 23/4, 302-314.

[88] Gershwin, S.B. (1994), *Manufacturing Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ.

[89] Gershwin, S.B., and Berman, O. (1981), "Analysis of transfer lines consisting of two unreliable machines with random processing times and finite storage buffers", *AIIE Transactions* 13/1, 2-11.

[90] Gershwin, S.B., and Schick, I.C. (1980), "Continuous model of an unreliable two-stage material flow system with finite interstage buffer", Technical Report LIDS-R-1039, Massachusetts Institute of Technology, Cambridge, MA.

[91] Gershwin, S.B., and Schick, I.C. (1983), "Modeling and analysis of three-stage transfer lines with unreliable machines and finite buffers", *Operations Research* 31/2, 354-380.

[92] Glassey, C.R., and Hong, Y. (1986), "The analysis of behaviour of an unreliable N-stage automatic transfer line with $(N-1)$ inter-stage buffer storages", Technical Report OPC-86-10, Dept. of IE and OR, University of California, Berkeley.

[93] Gnedenko, B.V., and Kovalenko, I.N. (1968), *Queueing Theory*, Weiner Bindery Ltd.

[94] Gopalan, M.N., and Kannan, S. (1994), "Expected duration analysis of a two-stage transfer-line production system subject to inspection and rework", *J. Opl Res. Soc.* 45/7, 797-805.

[95] Gopalan, M.N., and Anantharaman, N. (1991), "On the utilization analysis of a two-stage transfer line production system subject to inter-stage inspection and initial buffer", *Microelectron. Reliabilty* 31/1, 59-64.

[96] Gross, D., and Harris, C.M. (1985), *Fundamentals of Queueing Theory*, Wiley, New York, 2nd edn.

[97] Gross, D., and Miller, D.R. (1984), "The randomization technique as a modeling tool and solution procedure for transient Markov processes", *Operations Research* 32/2, 343-361.

[98] Gross, D., Gu, B., and Soland, R.M. (1993), "Iterative solution methods for obtaining the steady-state probability distributions of Markovian multi-echelon repairable item inventory systems", *Computers & Operations Research* 20/8, 817-828.

[99] Gun, L., and Makowski, A.M. (1989), "An approximation method for general tandem queueing systems subject to blocking", in: H. Perros and T. Altiok (eds.), *Queueing Networks with Blocking*, North-Holland, Amsterdam, 147-174.

[100] Harrison, J.M. (1973), "Assembly-like queues", *Journal of Applied Probability* 10, 354-367.

[101] Harrison, J.M., and Nguyen, V. (1990), "The QNET method for two-moment analysis of open queueing networks", *Queueing Systems* 6, 1-32.

[102] Haydon, B.J. (1972), "System states for a series of finite queues", *Operations Research* 20, 1137-1141.

[103] Heavey, C. (1990), "The throughput rate of multistation unreliable production lines", Technical Report, Department of Industrial Engineering, University College Galway, Ireland.

[104] Heavey, C., Papadopoulos, H.T., and Browne, J. (1993), "The throughput rate of multistation unreliable production lines", *European Journal of Operational Research* 68, 69–89.

[105] Hildebrand, D.K. (1968), "On the capacity of tandem server, finite queue, service systems", *Operations Research* 16, 72–82.

[106] Hillier, F.S., and Boling, R.W. (1967), "Finite queues in series with exponential or Erlang service times – A numerical approach", *Operations Research* 4, 286–303.

[107] Hillier, F.S., and So, K.C. (1989), "The assignment of extra servers to stations in tandem queueing systems with small or no buffers", *Performance Evaluation* 10, 219–231.

[108] Hillier, F.S., and So, K.C. (1991), "The effect of machine breakdowns and interstage storage on the performance of production line systems", *International Journal of Production Research* 29/10, 2043–2055.

[109] Hira, D.S., and Pandey, P.C. (1987), "Efficiency of manual flow line systems – Predictive equations", *International Journal of Production Research* 25/4, 603–614.

[110] Hong, Y., Glassey, C.R., and Seong, D. (1992), "The analysis of a production line with unreliable machines and random processing times", *IIE Transactions* 24/1, 77–83.

[111] Hopp, W.J., and Simon, J.T. (1989), "Bounds and heuristics for assembly-like queues", *Queueing Systems* 4, 137–156.

[112] Hopp, W.J., and Simon, J.T. (1993), "Estimating throughput in an unbalanced assembly-like flow system", *International Journal of Production Research* 31/4, 851–868.

[113] Hunt, G.C. (1956), "Sequential arrays of waiting lines", *Operations Research* 4, 674–683.

[114] Ignall, E., and Silver, A. (1977), "The output of a two-stage system with unreliable machines and limited storage", *AIIE Transactions* 9(2), 183–188.

[115] Iyama, T., and Odawara, T. (1990), "A study of an automatic transfer line system with a sub-line", *IIE Transactions* 22(3), 204–214.

[116] Jackman, J., and Johnson, E. (1993), "The role of queueing network models in performance evaluation of manufacturing systems", *Journal of the Operational Research Society* 44/8, 797–807.

[117] Jackson, J.R. (1957), "Networks of waiting lines", *Operations Research* 5/4.

[118] Jackson, J.R. (1963), "Jobshop-like queueing systems", *Management Science* (October) 131–142.

[119] Jackson, R.R.P. (1954), "Queueing systems with phase-type service", *Operational Research Quarterly* 5, 109–120.

[120] Jackson, R.R.P. (1956), "Random queueing processes with phase-type service", *J.R.Stat. Soc. series B* 18/1, 129–132.

[121] Jacobs, D., and Meerkov, S.M. (1993), "On the process of continuous improvement in production systems: Fundamentals, properties and guidelines", Report No. CGR-93-10, Systems Science and Electrical Engineering Division, Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI.

[122] Jain, S., and Smith, J.M. (1994), "Open finite queueing networks with $M/M/C/K$ parallel servers", *Computers & Operations Research* 21/3, 297–317.

[123] Jafari, M.A., and Shanthikumar, J.G. (1987), "Exact and approximate solutions to two-stage transfer lines with general uptime and downtime distributions", *IIE Transactions* 19/4, 412–420.

[124] Johri, P.K., "Engineering a circuit board assembly line for a desired capacity and flowtime", *Journal of Manufacturing Systems* 10/6, 492–500.

[125] Jun, K.P., and Perros, H.G. (1990), "An approximate analysis of open tandem queueing networks with blocking and general service times", *European Journal of Operational Research* 46, 123–135.

[126] Kelly, F.P. (1982), "The throughput of a series of buffers", *Advances in Applied Probability* 14, 633–653.

[127] Kelly, F.P. (1984), "Blocking, reordering, and the throughput of a series of servers", *Stochastic Processes and Their Applications* 17, 327–336.

[128] Kerbache, L. (1984), "Analysis of open finite queueing networks", PhD thesis, Dept. of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA.

[129] Kerbache, L. and MacGregor Smith, J. (1987), "The generalized expansion method for open finite queueing networks", *European Journal of Operational Research* 32, 448–461.

[130] Kerbache, L., and MacGregor Smith, J. (1988), "Asymptotic behavior of the expansion method for open finite queueing networks", *Computers & Operations Research* 15/2, 157–169.

[131] Kiessler, P.C., and Disney, R. (1988), "Further remarks on queueing network theory", *European Journal of Operational Research* 36, 285–296.

[132] Kleinrock, L. (1975), *Queueing Systems, Volume I: Theory*, Wiley, New York.

[133] Knesll, C. (1993), "On the sojourn time distribution in a finite capacity processor shared queue", *Journal of the Association for Computing Machinery* 40/5, 1238–1301.

[134] Knott, A.D. (1969), "Aggregation laws for a series of finite queues", *41st Congress ANZAAS*.

[135] Knott, A.D. (1970), "The inefficiency of a series of work stations – A simple formula", *International Journal of Production Research* 8/2, 109–119.

[136] Koenigsberg, E. (1959), "Production lines and internal storage – A review", *Management Science* 5, 410–433.

[137] Konheim, A.G., and Reiser, M. (1976), "A queueing model with finite waiting room and blocking", *Journal of the ACM* 23, 328–341.

[138] Kouikoglou, V.S. (1991), "An exact discrete-event model and control policies for production lines with buffers", *IEEE Transactions on Automatic Control* 36/5, 515–527.

[139] Kuehn, P.J. (1979), "Approximate analysis of general queueing networks by decomposition", *IEEE Transactions on Communications* 27/1, 113–126.

[140] Kumar, S., and Kumar, P.R. (1994), "Performance bounds for queueing networks and scheduling policies", *IEEE Transactions on Automatic Control* 39/8, 1600-1611.

[141] Labetoulle, J., and Pujolle, G. (1980), "Isolation method in a network of queues", *IEEE Transactions on Software Engineering* 6, 373-381.

[142] Lam, S., and Reiser, M. (1977), "Congestion control of store-and-forward networks by input buffer limits", in: *Proceedings of the National Telecommunications Conference*, IEEE, New York.

[143] Langaris, C. (1986), "The waiting-time process of a queueing system with gamma-type input and blocking", *Journal of Applied Probability* 23, 166-174.

[144] Langaris, C. (1989), "On a two-stage birth and death queueing process", *Operations Research* 37/3, 488-497.

[145] Langaris, C., and Conolly, B. (1984), "On the waiting time of a two-stage queueing system with blocking", *Journal of Applied Probability* 21, 628-638.

[146] Langaris, C., and Conolly, B. (1985), "Three stage tandem queue with blocking", *European Journal of Operational Research* 19, 222-232.

[147] Latouché, G., and Neuts, M.F. (1980), "Efficient algorithmic solutions to exponential tandem queues with blocking", *SIAM Journal on Algebraic and Discrete Methods* 1/1, 93-106.

[148] Lau, H.-S., and Martin, G.E. (1986), "A decision support system for the design of unpaced production lines", *International Journal of Production Research* 24/3, 599-610.

[149] Lau, H.-S., and Martin, G.E. (1987), "The effects of skewness and kurtosis of processing times in unpaced lines", *International Journal of Production Research* 25/10, 1483-1492.

[150] Lavenberg, S.S. (1975), "The steady-state queueing time distribution for the M/G/1 finite capacity queue", *Management Science* 21/5, 501-506.

[151] Lee, H.-S., and Pollock, S.M. (1989), "Approximate analysis for the merge configuration of an open queueing network with blocking", *IIE Transactions* 21/2, 122-129.

[152] Lee, H.-S., and Pollock, S.M. (1990), "Approximation analysis of open acyclic exponential queueing networks with blocking", *Operations Research* 38/6, 1123-1134.

[153] Lee, Y.-J., and Zipkin, P. (1992), "Tandem queues with planned inventories", *Operations Research* 40/5, 936-947.

[154] Lim, J., Meerkov, S.M., and Top, F. (1990), "Homogeneous, asymptotically reliable serial production lines: Theory and a case study", *IEEE Transactions on Automatic Control* 35/5, 524-534.

[155] Lin, L., and Cochran, J.K. (1990), "Metamodels of production line transient behaviour for sudden machine breakdowns", *International Journal of Production Research* 28/10, 1791-1806.

[156] Lipset, R., Sengupta, S., and Van Til, R.P. (1994), "The effect of buffer unreliability upon the steady-state performance of serial transfer lines", Paper presented at the *ORSA/TIMS Joint National Meeting*, Detroit, MI, October 23-26.

[157] Liu, X.G. (1990), "Toward Modeling assembly systems: Application of queueing networks with blocking", PhD thesis, Dept. of Management Science, University of Waterloo, Waterloo.

[158] Liu, X.G., and Buzacott, J.A. (1993), "A decomposition-related throughput property of tandem queueing networks with blocking", *Queueing Systems* 13, 361-383.

[159] Liu, C.M., and Lin, C.L. (1994), "Performance evaluation of unbalanced serial production lines", *International Journal of Production Research* 32/2, 2897-2914.

[160] Makino, T. (1964), "On the mean passage time concerning some queueing problems of the tandem type", *Journal of the Operations Research Society of Japan* 7, 17-47.

[161] Marie, R. (1980), "Calculating equilibrium probabilities for $\lambda(n)/C_k/1/N$ queues", in: *Proceedings of Performance 80, Int. Symp. on Computer Performance Modelling*, May.

[162] Marshall, K.T. (1968), "Some inequalities in queueing", *Operations Research* 16, 651-665.

[163] Martin, G.E. (1993), "Predictive formulae for unpaced line efficiency", *International Journal of Production Research* 31/8, 1981-1990.

[164] Martin, G.E., and Lau, H.-S. (1990), "Dynamics of the output interval distribution in unpaced lines", *Computers ind. Engng* 18/3, 391-399.

[165] Meester, L.E., and Shanthikumar, J.G. (1990), "Concavity of the throughput of tandem queueing systems with finite buffer storage space", *Advances in Applied Probability* 22, 764-767.

[166] Melamed, B. (1982), "On the reversibility of queueing networks", *Stochastic Processes and their Applications* 13, 227-234.

[167] Melamed, B. (1986), "A note on the reversibility and duality of some tandem blocking queueing systems", *Management Science* 32/12, 1648-1650.

[168] Miltenburg, G.J. (1987), "Variance of the number of units produced on a transfer line with buffer inventories during a period of length t", *Naval Research Logistics* 34, 811-822.

[169] Mishra, A., Acharya, D., Rao, N.P., and Sastry, G.P. (1985), "Composite stage effects in unbalancing of series production systems", *International Journal of Production Research* 23/1, 1-20.

[170] Mitra, D. (1988), "Stochastic theory of a fluid model of multiple failure-susceptible producers and customers coupled by a buffer", *Advances in Applied Probability*, September.

[171] Mitra, D., and Mitrani, I. (1988), "Analysis of a novel discipline for cell coordination in production lines I", Technical Report, AT&T Bell.

[172] Mitra, D., and Mitrani, I. (1990), "Analysis of a Kanban discipline for cell coordination in production lines", *Management Science* 36/12, 1548-1566.

[173] Mitra, D., and Mitrani, I. (1991), "Analysis of a Kanban discipline for cell coordination in production lines, II: Stochastic demands", *Operations Research* 39/5, 807-823.

[174] Mitra, D., and Tsoucas, P. (1988), "Relaxations for the numerical solutions of some stochastic problems", *Stochastic Models* 4, 387-419.

[175] Murphy, R.A. (1975), "The effect of surge on system availability", *AIIE Transactions* 7/4, 439–443.
[176] Muth, E.J. (1973), "The production rate of a series of work stations with variable service times", *International Journal of Production Research* 11/2, 159–169.
[177] Muth, E.J. (1975), "Some results concerning production lines and the machine scheduling problem", *Bulletin of the Operations Research Society of America* 23 (Supplement 2).
[178] Muth, E.J. (1977), "Numerical methods applicable to a production line with stochastic servers", *TIMS Studies in the Management Sciences* 7, 143–159.
[179] Muth, E.J. (1979), "The reversibility property of production lines", *Management Science* 25, 152–158.
[180] Muth, E.J. (1984), "Stochastic processes and their network representations associated with a production line queueing model", *European Journal of Operational Research* 15, 63–83.
[181] Muth, E.J. (1987), "An update on analytical models of serial transfer lines", Research Report 87-15, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.
[182] Muth, E.J., and Alkaff, A. (1987), "The throughput of three-station production lines: A unifying solution", *International Journal of Production Research* 25, 1405–1413.
[183] Narahari, Y., and Viswanadham, N. (1994), "Transient analysis of manufacturing systems performance", *IEEE Transactions on Robotics and Automation* 10/2, 230–244.
[184] Neuts, M.F. (1968), "Two queues in series with a finite intermediate waiting room", *Journal of Applied Probability* 5, 123–142.
[185] Neuts, M.F. (1970), "Two servers in series, studied in terms of a Markov renewal branching process", *Advances in Applied Probability* 2, 110–149.
[186] Neuts, M.F. (1981), *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore, MD.
[187] Newell, G.F. (1979), "Approximate behaviour of tandem queues", in: *Lecture Notes in Economics and Mathematical Systems* 171, Springer-Verlag, Berlin.
[188] Onvural, R.O., and Perros, H.G. (1986), "On equivalencies of blocking mechanisms in queueing networks with blocking", *Operations Research Letters* 5/6, 293–297.
[189] Otero, E. (1987), "A new algorithm for the analysis of tree-structured assembly/disassembly networks", Unpublished Private Communication, July.
[190] Ou, J., and Gershwin, S.B. (1989), "The variance of the lead time distribution of a two-machine transfer line with a finite buffer", Unpublished paper.
[191] Papadopoulos, H.T. (1988), "Mathematical modelling of reliable production lines", PhD thesis, Department of Industrial Engineering, University College Galway, Ireland.
[192] Papadopoulos, H.T. (1995), The throughput rate of "multistation reliable production lines with no intermediate buffers", *Operations Research* (July/August) in press.
[193] Papadopoulos, H.T. (1996), "An analytic formula for the mean throughput of K-station production lines with no intermediate buffers", *European Journal of Operational Research* 91, 481–494.
[194] Papadopoulos, H.T. (1994), "An algorithm for calculating the mean sojourn time of K-station production lines with no intermediate buffers", submitted for publication.
[195] Papadopoulos, H.T. (1995), "Mean throughput and mean sojourn time of production lines with Coxian service times", submitted for publication.
[196] Papadopoulcs, H.T., Heavey, C., and Browne, J. (1993), *Queueing Theory in Manufacturing Systems Analysis and Design*, Chapman and Hall, London.
[197] Papadopoulos, H.T., Heavey, C., and O'Kelly, M.E.J. (1989), "The throughput rate of multistation reliable production lines with inter station buffers: (I) Exponential case", *Computers in Industry* 13, 229–244.
[198] Papadopoulos, H.T., Heavey, C., and O'Kelly, M.E.J. (1990), "The throughput rate of multistation reliable production lines with inter station buffers: (II) Erlang case", *Computers in Industry* 13, 317–335.
[199] Papadopoulos, H.T., and O'Kelly, M.E.J. (1988), "The throughput rate of multistation series reliable production lines: An analytical approach", Technical Report 88-1, Dept. Ind. Eng., University College, Galway, Ireland.
[200] Papadopoulos, H.T. and O'Kelly, M.E.J. (1993), "Exact analysis of multistation series production lines with no intermediate buffers", *European Journal of Operational Research* 65, 118–137.
[201] Park, T., and Steudel, H.J. (1991), "A model for determining job throughput times for manufacturing flow line workcells with finite buffers", *International Journal of Production Research* 29/10, 2025–2041.
[202] Perros, H.G. (1984), "Queueing networks with blocking: A bibliography", *Performance Evaluation Review, ACM SIGMETRICS* 12, 8–12.
[203] Perros, H.G. (1992), "Approximate algorithms for queueing networks with blocking", Unpublished paper.
[204] Perros, H.G., and Altiok, T. (1986), "Approximate analysis of open networks of queues with blocking: Tandem configurations", *IEEE Transactions on Software Engineering* 12/3, 450–461.
[205] Perros, H.G., and Altiok, T. (1989), (Editors), *Queueing Networks with Blocking*, North Holland, Amsterdam.
[206] Pollock, S.M., Birge, J.R., and Alden, J.M. (1985), "Approximation analysis for open tandem queues with blocking: Exponential and general distributions", Technical Report IOE Rept. 85-30, Univ. of Michigan.
[207] Pourbabai, B. (1988), "Tandem behavior of an $M/M/1/N \to G/M/1$ queueing system", *Computers and Mathematics with Applications* 16/3, 215–220.
[208] Pourbabai, B. (1989), "Approximating the tandem behaviour of a $G/M/1$ queueing system", *International Journal of Systems Science* 20/6, 1035–1051.
[209] Pourbabai, B. (1990), "Tandem behavior of a finite capacity $G/M/1$ queueing system: An algorithm", *European Journal of Operational Research* 46, 380–387.
[210] Rao, N.P. (1976), "A generalization of the bowl phenomenon in series production systems", *International Journal of Production Research* 14, 437–443.

[211] Reiman, M.I. (1990), "Asymptotically exact decomposition approximations for open queueing networks", *Operations Research Letters* 9, 363-370.

[212] Reiser, M., and Kobayashi, H. (1973), "On the accuracy of the diffusion approximation for queueing systems", IBM Research Report RC4546, September.

[213] Richart, S.E. (1985), "Analysis of sequential manufacturing systems", PhD thesis, University of Missouri, Columbia.

[214] Saboo, S., Wang, L., and Wilhelm, W.E. (1989), "Recursion models for describing and managing the transient flow of materials in generalized flowlines", *Management Science* 35/6, 722-742.

[215] Sauer, C.H., and Chandy, K.M. (1975), "Approximate analysis of central server models", *IBM Journal of Research and Development* 19, 301-313.

[216] Savsar, M., and Cogun, C. (1994), "Analysis and modelling of a production line in a corrugated box factory", *International Journal of Production Research* 32/7, 1571-1589.

[217] Schick, I.C., and Gershwin, S.B. (1979), "Modeling and analysis of unreliable transfer lines with finite interstage buffers", In *Complex Materials Handling and Assembly Systems, Vol. VI*, Report ESL-FR-834-6, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.

[218] Seelen, L.P. (1986), "An algorithm for *Ph/Ph/c* queues", *European Journal of Operational Research* 23, 118-127.

[219] Sevast'yanov, B.A. (1962), "Influence of stage bin capacity on the average standstill time of a production line", *Theory of Probability and its Applications* 7/4, 429-438.

[220] Shanthikumar, J.G., and Tien, C.C. (1983), "An algorithmic solution to two-stage transfer lines with possible scrapping of units", *Management Science* 29/9, 1069-1086.

[221] Shanthikumar, J.G., and Yao, D.D. (1989), "Monotonicity and concavity properties in cyclic queueing networks with finite buffer sizes", in: H. Perros and T. Altiok (eds.), *Queueing Networks with Blocking*, North-Holland, Amsterdam, 325-345.

[222] Sheskin, T.J. (1976), "Allocation of interstage storage along an automatic production line", *AIIE Transactions* 8, 146-152.

[223] Sheskin, T.J. (1977), "Letter to the editor", *AIIE Transactions* 9/1, 120.

[224] Suri, R. (1985), "An overview of evaluative models for flexible manufacturing systems", *Annals of Operations Research* 3, 13-21.

[225] Takagi, H., and Boguslavsky, L.B. (1991), "A supplementary bibliography of books on queueing analysis and performance evaluation", *Queueing Systems* 8, 313-322.

[226] Takahashi, Y., Miyahara, H., and Hasegawa, T. (1980), "An approximation method for open restricted queueing networks", *Operations Research* 28/3, 594-602.

[227] Tayur, S.R. (1993), "Structural properties and a heuristic for Kanban-controlled serial lines", *Management Science* 39/11, 1347-1368.

[228] Terracol, C., and David, R. (1987), "Performance d'une ligne composée de machines et de stocks intermédiaires", *APII* 21, 239-262.

[229] Tijms, H.C. (1986), *Stochastic Modelling and Analysis: A Computational Approach*, Wiley, New York.

[230] Tijms, H.C. (1994), *Stochastic Models: An Algorithmic Approach*, Wiley, New York.

[231] Tsiotras, G.D. (1991), "Exact analysis of the joint production quantity-quality of a two-station stable tandem manufacturing system with restricted capacity", *International Journal of Computers and Mathematics with Applications*, 21/4, 23-31.

[232] Van Dijk, N.M., and Van der Wal, J. (1989), "Simple bounds and monotonicity results for finite multi-server exponential tandem queues", *Queueing Systems* 4, 1-16.

[233] Viswanadham, N., and Narahari, Y. (1992), *Performance Modeling of Automated Manufacturing Systems*, Prentice Hall, Englewood Cliffs, NJ.

[234] Vladzievskii, A.P. (1953), "Losses of working time and the division of automatic lines into sections", *Stanki i Instrument* 24/10, 9-15.

[235] Wallace, V. (1973), "Algebraic techniques for numerical solution of queueing networks", *Proc. of a Conference at Western Michigan University*, May 10-12.

[236] Walrand, J. (1982), "Poisson flows in single class open networks of quasireversible queues", *Stochastic Processes and their Applications* 13, 293-303.

[237] Walrand, J. (1988), *An Introduction to Queueing Networks*, Prentice Hall, Englewood Cliffs, NJ.

[238] Walrand, J., and Varaiya, P. (1980), "Interconnections of Markov chains and quasi-reversible queueing networks", *Stochastic Processes and their Applications* 10, 209-219.

[239] Ward, J.E. (1980), "TI-59 calculator programs for three two-machine one-buffer transfer line models", Technical Report LIDS-R-1009, Massachusetts Institute of Technology, Cambridge, MA.

[240] Wein, L.M. (1989), "Capacity allocation in generalized Jackson networks", *Operations Research Letters* 8, 143-146.

[241] Whitt, W. (1982), "Approximating a point process by a renewal process: The view through a queue, an indirect approach", *Management Science* 27/6, 619-636.

[242] Whitt, W. (1983), "The queueing network analyzer", *The Bell System Technical Journal* 62/9, 2779-2815.

[243] Whitt, W. (1984), "Open and closed models for network of queues", *AT&T Technical Journal* 63/9, 1911-1979.

[244] Whitt, W. (1993), "Large fluctuations in a deterministic multiclass network of queues", *Management Science* 39/8, 1020-1028.

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

[246] Wild, R. (1971), *The Techniques of Production Management*, Holt, Rinehart and Winson.

[247] Yamazaki, G., and Kawashima, T. (1978), "Production rate estimated with flow shop reversibility" (in Japanese), *Transactions of the Japan Society of Mechanical Engineers* (1976); translated into English, *Bulletin of the Japan Society of Mechanical Engineers* 21, 167.

[248] Yamazaki, G., Kawashima, T., and Sakasegawa, H. (1985), "Reversibility of tandem blocking queueing systems", *Management Science* 31/1, 78-83.

[249] Yamazaki, G., and Sakasegawa, H. (1975), "Properties of duality in tandem queueing systems", *Annals of the Institue of Statistics and Mathematics* 27, 201-212.

[250] Yannopoulos, E., and Alfa, A.S. (1994), "An approximation method for queues in series with blocking", *Performance Evaluation* 20, 373-390.

[251] Yao, D.D. (1987), "Majorization and arrangement orderings in open queueing networks", *Annals of Operations Research* 9, 531-543.

[252] Yao, D.D. (1994), (Editor) *Stochastic Modeling and Analysis of Manufacturing Systems*, Springer-Verlag, Berlin.

[253] Yeralan, S., Franck, W.E., and Quasem, M.A. (1984), "Production lines with continuous materials flow and station breakdown", Technical Report 8410711, Dept. of Ind. Engineering, University of Missouri, Columbia.

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

[255] Yeralan, S., and Tan, B. (1994), "A building block for the decomposition of continuous flow production systems", Paper presented at the *ORSA/TIMS Joint National Meeting*, Detroit, MI, October 23-26.

[256] Zangwill, W.I., and Garcia, C.B. (1981), *Pathways to Solutions, Fixed Points, and Equilibria*, Prentice Hall, Englewood Cliffs, NJ.

[257] Zimmern, B. (1956), "Ètudes de la propagation des arrêts aléatoires dans les chaines de production", *Revue de Statistique Appliquée* 4, 85-104.