IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2022, accepted May 27, 2022, date of publication June 2, 2022, date of current version June 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3179711

A Queuing Network Model for Solving Facility Layout Problem in Multifloor Flow Shop

XIAOPENG LIU[1,2], HUIYU ZHANG[1], JUN LIN[2], XUANRUI CHEN[1], QINGXIN CHEN[1], AND NING MAO[1]

[1]Key Laboratory of Computer Integrated Manufacturing System, Guangdong University of Technology, Guangzhou 510006, China
[2]China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 510619, China

Corresponding author: Huiyu Zhang (hyzhang_henry@126.com)

This work was supported in part by the National Natural Science Foundation, China, under Grant 61973089; in part by the National Key Research and Development Program, China, under Grant 2018YFB1802403; and in part by the Guangdong Natural Science Foundation, China, under Grant 2022A1515011175.

ABSTRACT Based on comprehensive consideration of stochastic factors and the system performance, we presented an integrated queuing network model (IQNM) to effectively optimize the facility layout of the multi-floor flow shop (MFFS). In this paper, the arrival interval and service times of the workstations are described by the two-moment approximation method, and an iterative algorithm is proposed for estimating the system performance, including the mean throughput, work-in-process (WIP) and sojourn time. In the design process, an optimization model with the throughput and cycle time constraint is established, while the minimize transportation cost is taken as the objective function. After several iterations between system performance and optimization objective, the facility layout problem (FLP) of the MFFS is solved, which takes the maximize throughput and minimize cycle time as the constraint condition. Finally, the effectiveness of the proposed method is verified by a series of numerical experiments. The average error of system throughput is 2.11%. And the sojourn time errors are less than 5% when SCV = 1.5. Therefore, this research can provide relevant theoretical support for the facility layout of the MFFS.

INDEX TERMS Multi-floor flow shop, queuing networks, facility layout problem, system performance.

I. INTRODUCTION

Individualized production is the developing tendency of the manufacturing industry in the future [1]. The multi-floor flow shop (MFFS) with the intelligent material handling system (IMHS) and the intelligent manufacturing cells (IMCs) is shown in Fig 1, which is one of the typical workshop structures in individualized production. In this system, although space-developing workshop has the advantages of saving land resources, shortening transportation paths, and reducing energy for air-conditioning [2], it also exists some uncertainty factor that lead to the imbalance of workload distribution in space and time. Thus, it is important to design and plan a MFFS in such a way that can efficiently adapt to the dynamic demand environment, thereby reducing queuing phenomena and balancing cycle time. However, the success or failure of the design and operation of such systems is judged by the degree to which performance objectives are met. To validly optimal the facility layout of the MFFS, we present an queuing network modeling framework and solution methodology by considering its production characteristics.

This paper discusses the facility layout problem (FLP) of the MFFS, including rough machining layer, precision machining layer and assembly and debugging layer which is composed of multiple intelligent processing or assembly cells. The process of parts going through in two or more processing cells is completed by Automated Guided Vehicles (AGVs). Lifts provide vertical movement between floors, when a work-in-process (WIP) needs to be moved to another floor for processing. In order to produce an optimal layout, we divide the cells into a series of blocks, and determine the size, shape and position of the block in the workshop in turn. According to the above description, there are two difficult points in the facility layout of the MFFS:

(1) One is the optimal problem. The throughput and cycle time need to be taken as the constraint condition when the optimizing the workshop layout. However, in the MFFS of complex production environment, it is difficult to accurately calculate the system performance index.

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Zunino.

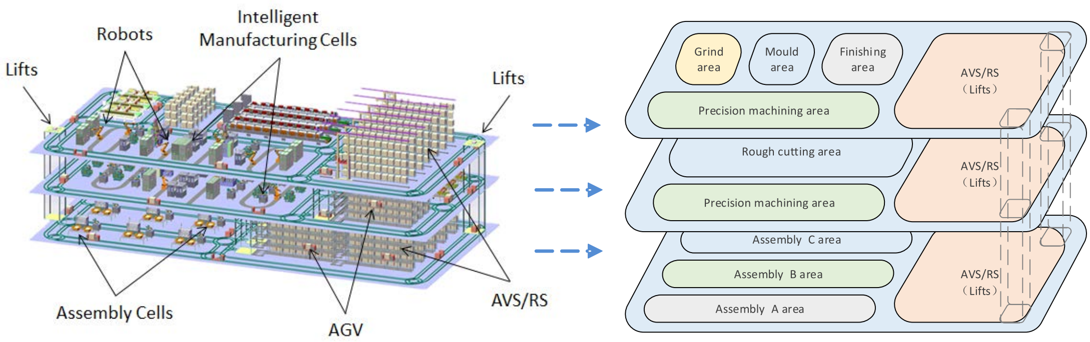


FIGURE 1. The typical structure of the MFFS.

(2) The other is the modeling problem. In course of the manufacture, the synchronization constraints between WIP and transportation resource must be fully considered, which is generally described by queuing network model. At present, there is no specific queuing network model has been found.

For traditional FLP, there are a lot of uncertain factors in the production process of the MFFS when the logistics intensity matrix is taken as the core. It is necessary to establish a comprehensive workshop layout optimization model which fully considers multiple inevitable waiting phenomena in the process of WIP transportation and various cells such as processing, assembly and storage. Chen [3] uses a M/G/1 queuing network model to describe the elevator handling system, calculates the elevator utilization and average waiting time of vertical logistics handling, takes the solution results as the material handling cost between departments on different floors, and solves the multi-floor facility layout problem (MFLP) with the minimum material handling cost as the optimization objective. Izadubua [4], [5] takes the first underground floor as a warehouse, and other units are placed in potential predetermined positions on other floors. Under the interval constraint of logistics intensity fluctuation between preset units, a mixed integer programming model and a set of robust layout optimization methods are proposed, and use an ant colony optimization algorithm to solve it. Although some achievements have been made in the research about MFLP [6]–[9], there are few papers involving uncertainty modeling, and no research reports considering waiting phenomena in the workshop production process, which account for a part of the cycle time and even exceed its processing and assembly time.

The techniques for evaluating and analyzing the performance of manufacturing system have received great concern since the early 1900s. Queuing network theory (QNT) is an effective mathematical modeling tool for the performance of stochastic systems, such as manufacturing and communication systems [10]. A queuing network can be described as customers arriving for service, waiting for service if it is not immediate, and if having waited for service, leaving the system after being served. By establishing relationships between the customers' arrival and departure in the network, it can characterize the inevitable queuing phenomenon in the processing process and estimate the network performance measures. As we know, the MFFS is a stochastic production system with a large numbers of uncertainty factors, such as the delivery time, the task time, storage and transportation time of parts. Therefore, it is essential to develop an effective queuing network model that could completely describe its production characteristics, and provide reasonably accurate estimates of the performance measures. We now review the existing queuing network studies about manufacturing systems from two directions.

General distribution models. Continuous-Time Markov Chain (CTMC) models are usually used for analyzing manufacturing systems in some research. However, a majority of these models make the restrictive assumption of exponentially distributed service times, or Poisson arrivals of demands. In reality, the assumption of general distribution is more suitable for practical production. There are more and more studies about stochastic systems with general service times because of the limitations of the exponential distribution. Whitt [11] uses two parameters to characterize the arrival processes and service times, one to describe the rate and the other to describe the variability. This approach has far been primarily quoted in analysis of open queuing networks (OQNs), which called parametric decomposition method or two-moment approximation method. Kerbache and Smith [12] proposed an effective approximation technique, the Generalized Expansion Method (GEM) for the analysis of general open finite queuing networks (OFQNs). This method is developed from the isolation method [13] and parametric decomposition method [14]. Afterward, the GEM was extended into analyses of the complex structure queuing networks such as the multi-objective routing within large scale facilities [15], [16], the flexible flow shop with match processing constrain [17] and the reentrant job-shops with finite buffers and feedback [18].

Synchronization constraint models. As far as we know, synchronization stations are commonly used to characterize the synchronization constraints in queuing network models.

Generally, there are three kinds of synchronization station models: 1) Assembly/Disassembly(A/D) network, 2) Kanban controlled production system, and 3) Semi-open queuing network (SOQNs). The A/D networks was first done by Harrison [19] who studied assembly-like queues with infinite buffers. Govil and Fu [20] reviewed the analysis methods of A/D networks with three different constraints: infinite buffers, transient performance and finite buffers. For transfer lines or flow lines with exponentially distribution, together with a Markov-chain-based analysis of the two-station subsystems [21], [22], the DDX algorithm [23] can be used as an appropriate decomposition approach, see Gershwin and Burman and Amir et al [24], [25]. In the case of general distribution, such an approach was applied to the assembly and A/D networks with finite buffers by Manitz [26], [27]. As for kanban controlled production system, Krishnamurthy and Suri [28] used parametric decomposition methods to analyze it with general distribution. Afterward, this method was extended into analyses of a multi-product kanban system [29] and even a multi-product kanban system with batch size constraints [30].

Semi-open queuing networks (SOQNs), which decouple the arriving customers from the network resources using a synchronization station, can potentially capture the customer waiting times more precisely and provide a rich network modelling construct. Roy et al [31] develop a SOQNs of the autonomous vehicle-based storage and retrieval system (AVS/RS) where the performance measures such as the expected number of transactions waiting in the external queue are of significance. Roy et al [32] expanded this model to also incorporate vehicle blocking delays by explicitly modelling the cross-aisle and the individual aisles as resources. Also see Ekren et al [33], [34] and Roy et al [35] for other variants of the SOQNs model for analyzing autonomous vehicle-based warehousing systems with multiple tiers. They also model the interactions between vehicles and lifts during vertical transport of unit load.

Overall, it is clear that the synchronization constraint problems have a certain research foundation in manufacturing and logistics enterprises, but there is no literatures analyze the synchronization constraint with the public transportation resource because of the analytical difficulties. The A/D networks aim at the matching problem in the production process of parts, which is different to the synchronization constraints between parts and public transportation resource in this paper. For the existing research about SOQNs, they only focus on the synchronization constraints during the storage/retrieval transactions of WIP, but ignore the impact of production fluctuations on transactions arrival time.

The main contribution of this paper is that we use an IQNM to estimate the system performance based on parametric decomposition methods. The IQNM can describe the synchronization constraint between WIP and transportation resource in MFFS. There is no relevant reports of such models have been found so far. Thus, we address the first question:

(1) How to estimate the performance index of the MFFS quickly and effectively?

Another contribution of this paper is the improvement of the MFFS layout scheme. According to the optimization model with workshop performance index constraints, our approach effectively solve the block layout problem of the MFFS. Hence, our second question is:

(2) Which optimization method can effectively improve the layout scheme of the MFFS?

At the end of this paper, we design a series of numerical experiments. The numerical studies shows the reasonably accurate performance evaluation of this analytical method which is helpful in system design, quick evaluation and management decision. In respect of this, our third question is addressed as follows:

(3) How to verify that the proposed analysis method has practical benefits?

The remainder of this paper is organized as follows. In Section II, we present a optimization model and a detailed description of the queuing network model we study. Then the proposed performance evaluation approach is presented in Section III. Section IV presents the results from various numerical experiments which indicate the accuracy of our proposed methods. Section V is devoted to a case study of an actual manufacturing industry and the paper is concluded in Section VI.

II. MODEL DESCRIPTION

A. QUEUING NETWORK MODEL OF THE MFFS

The queuing network model of the MFFS is shown in Fig 2. The production system is decomposed into several floors, each floor being comprised of a set of one or more work station. With each work station is associated a intermediate buffer. In course of manufacture, lifts provide vertical movement between floors. A part would wait for lifts to be free or available, when it needs to be moved to another floor for processing. Once transportation is complete, the lift would move to transport the next part or wait at the former place. AGVs provide horizontal movement between multiple work station. A part would wait for vehicles to be free or available, when it needs to be moved to the next work cell. Once transportation is complete, the vehicle would move to transport the next part or wait at the former place.

B. COMPARISON WITH EXISTING MODELS

The relevant researches on the synchronization constraint problem are shown in Table 1. [33], [34] studied the synchronization constraint problem in the warehouse system under the condition of considering multi-product. However, it ignores the product processing process and is a separate study of logistics problems. So the assumption of its exponential distribution is not suitable for the study of production system. [12] fully considered the queuing problem of WIP in the production system. They set the assumptions of finite

TABLE 1. The research on the synchronization constraint problem.

References	Year	Application scenario	Model[a]	Assumptions	Solution method
[12]	2006	Manufacturing system	OQN	General distribution Finite buffer Single-product	Generalized expansion
[18]	2017	Manufacturing system	OQN	General distribution Finite buffer Single-product	Generalized expansion
[28]	2006	Kanban controlled production system	SOQN	General distribution Finite buffer Single-product	Parametric decomposition
[30]	2013	Kanban controlled production system	SOQN	General distribution Finite buffer Multi-product	Parametric decomposition
[33]	2013	Warehousing system	SOQN	Exponential distribution Infinite buffer Multi-product	Aggregation
[34]	2014	Warehousing system	SOQN	Exponential distribution Infinite buffer Multi-product	Matrix geometric

[a] OQN: open queuing network, SOQN: semi-open queuing network.

buffer and general distribution to make the model close to the real situation, but took the transportation time as a fixed value and ignored the fluctuations in the transportation process. Base on [12], [18] solved the more complicated queuing problem which can be used as a reference for the model problem of the MFFS.

The SOQNs used in [28] describes the multi-stage Kanban production system, which fully considers the synchronization constraint of the transportation equipment and WIP. Although transportation resources is not shared, the modeling techniques and the solution methods can be used for the MFFS. Therefore, we make a profound study in this method, expand the scope of application and improve the accuracy of the solution results.

C. FACILITY LAYOUT OPTIMIZATION MODEL

In this paper, we establish a stochastic nonlinear integer programming model with minimizing the total transportation cost as the objective function. The following notations are employed in our model:

$Q(X)$: Total transportation cost.

$Q_r(X_r, \zeta)$: Transportation cost of floor r, $r = 1, 2, \cdots, s$.

$I(i)$: The possible locations of object i, $i = 1, 2, \cdots, m$.

$J_i(k)$: The set of locations occupied by object i, $\alpha_{ikt} \begin{cases} 1, \text{loaction } t \in J_i(k) \\ 0, \text{otherwise.} \end{cases}$

X: The result of layout, $x_{ik} \in X, x_{ik} \begin{cases} 1, \text{object} i \in k \\ 0, \text{otherwise,} \end{cases}$ $k = 1, 2, \cdots, I(i)$.

$T(X, \zeta)$: The mean cycle time of the entire system.

$\theta(X, \zeta)$: The mean throughput rate of the entire system.

ζ: The stochastic factors.

The specific optimization model is as follows:

$$\text{Min}Q(X) = \sum_{r=1}^{s} Q_r(X_r, \zeta); \quad (1)$$

$$\forall i: \sum_{k=1}^{I(i)} x_{ik} = 1; \quad (2)$$

$$\forall t: \sum_{i=1}^{m} \sum_{k=1}^{I(i)} \alpha_{ikt} x_{ik} \leq 1; \quad (3)$$

$$T(X, \zeta) \leq T^{\max}; \quad (4)$$

$$\theta(X, \zeta) \geq \theta^{\min}. \quad (5)$$

In this optimization model, Equation (2) requires that each object is assigned to one location and Equation (3) insures that each location is occupied by at most one object. Equation (4) and Equation (5) provides additional lower bound $T^{\max}$ and upper bound $\theta^{\min}$ for the result of the layout, because the actual operation capacity of the system may cause a specific limitation. Note that above $T(X, \zeta)$

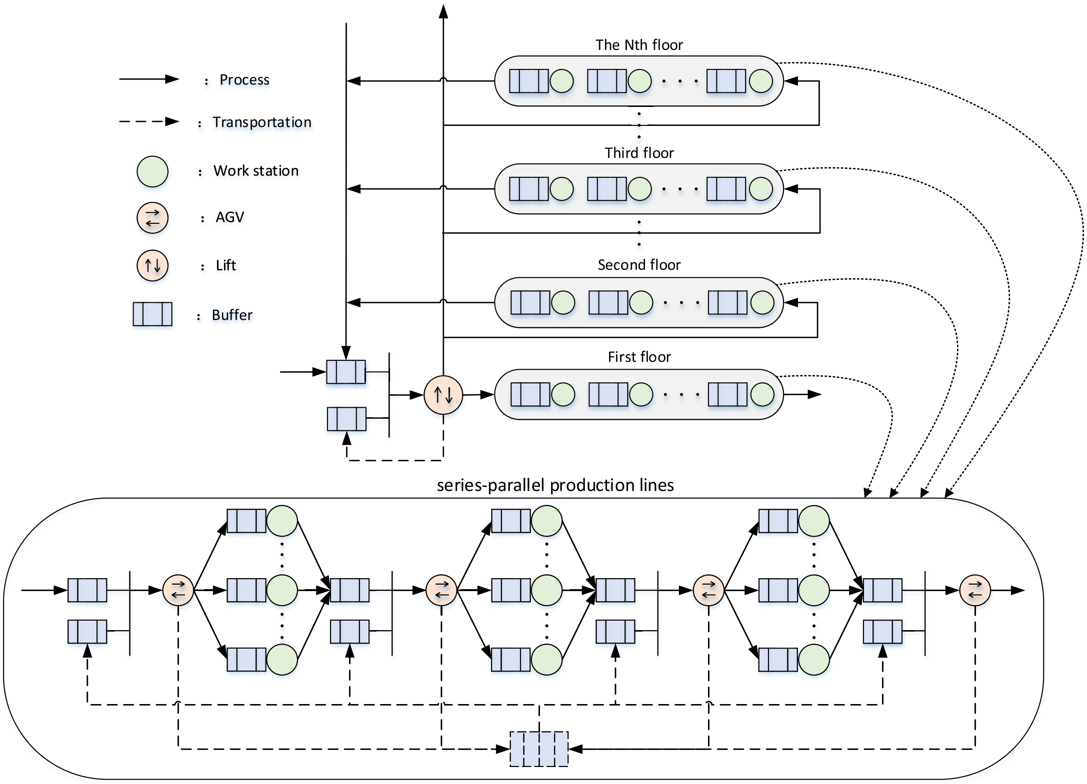


FIGURE 2. The queuing network model of the MFFS.

and $\theta(X, \zeta)$ are the complex functions without closed mathematical expression. Hence, we propose a new queuing model based on the parametric decomposition technique that enables the effective calculation of the system performance index.

III. SOLUTION METHOD

A. ASSUMPTIONS

This section first presents an IQNM to describe the operation of the production system and use the two-moment approximation method to describe the arrival interval and service times of the workstations, which denoted by mean and square coefficient of variation (SCV). Then, a new approach based on parametric decomposition to analyse the synchronization constraints between WIP and transportation resource is presented.

We first consider a queuing network model with three task nodes, including two transportation nodes, one processing node and two synchronization stations. Fig 3 shows the processes of a part from waiting for transportation to being transported and from waiting for processing to being processed. The synchronization station represents the queuing process of a part waiting for transportation, and the queuing time is determined by the idle interval of transportation equipment. In order to simplify the description, the following discussion will analyze the module based on this model. The specific assumptions are as follows:

(1) The arrival interval time type of parts enter the system are the general distribution. The arrival rate can be characterized by $\lambda_{d,i}$ and SCV, $c_{d,i}^2$.

(2) The arrival interval time type of parts enter the system are the general distribution. The arrival rate can be characterized by $\lambda_{d,i}$ and SCV, $c_{d,i}^2$.

(3) The service time type of each processing node are the general distribution. The service rate can be characterized by $\mu_{s,i}$ and SCV, $c_{s,i}^2$.

(4) The service time type for a transportation task are the general distribution. The service rate can be characterized by $\mu_{t,i}$ and SCV, $c_{t,i}^2$.

(5) As shown in Figure 4, the process of the transportation equipment returning to the dwell point is regarded as a virtual "collection queue". The return rate can be characterized by $\lambda_{a,i}$ and SCV, $c_{a,i}^2$.

(6) The "collection queue" represents the transportation equipment waiting for new tasks, and the probability of being assigned to a transportation task is α. According to the return rate, the idle interval time of transportation equipment can be characterized by $\lambda_{F,i}$ and SCV, $c_{F,i}^2$. The detailed analysis is clear in Section III.C.

(7) The synchronization station represents the synchronization constraint between the part and the transportation

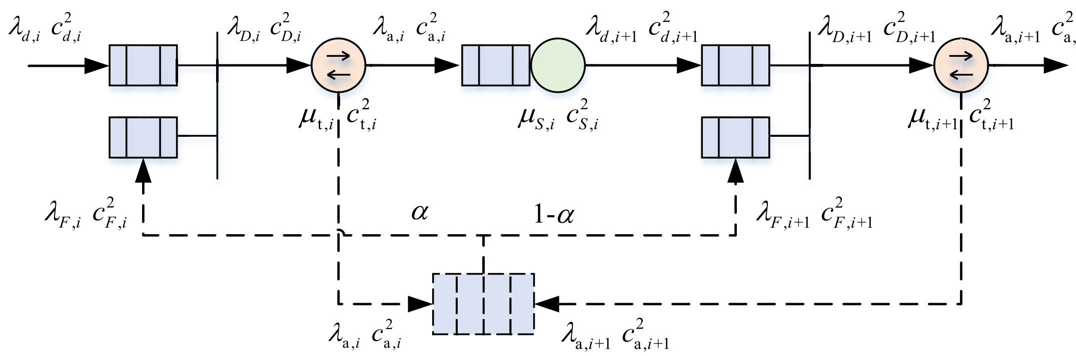


FIGURE 3. The basic analysis module of queuing network model.

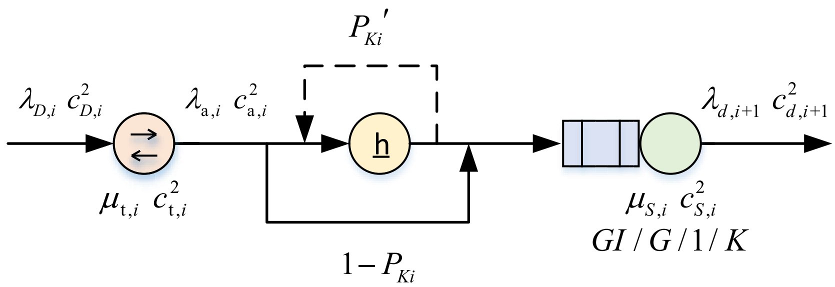


FIGURE 4. The analysis model of the task node.

equipment. The output rate of the synchronization station, $\lambda_{D,i}$ is always less than $\lambda_{d,i}$ and $\lambda_{F,i}$, because of the uneven arrival and then affect the node output. The detailed analysis is clear in Section III.C.

B. PARAMETER DECOMPOSITION METHOD

The parametric decomposition method has been applied extensively in the analysis of the queuing networks, which the nodes input and service process is the general time distribution. For different queuing network models, there are many specific solutions based on the idea of parameter decomposition technique, which are mainly different from the decomposition mode of subsystem and the iterative mode of solution algorithm. According to the production characteristics of the MFFS, our approach is based on the parameter decomposition method proposed in reference [28], and extended to combine with the GEM. The specific operation steps are as follows:

(1) Decomposition: The task nodes and synchronization stations are split, and the relationship between input and output is studied separately.

(2) Characterization: The input of the node and the service processes are approximate to the renewal processes, and the renewal interval distribution is characterized by the two-moment approximation method, which denoted by mean and SCV.

(3) Linkage: The parameter value relationship between the nodes is established, and a series of nonlinear equations are derived by using flow balance.

(4) Solution: An iterative algorithm is constructed to approximately solve the nonlinear equations, and the system performance indexes such as mean throughput and mean queue length of each node can be easily calculated.

C. DECOMPOSITION AND DESCRIPTION

Fig 4 shows the task node obtained from decomposition of the queuing network model, which composed of a GI/G/c transportation node and a GI/G/1/K processing node. Based on the principle of the flow conservation [10], the traffic equations are:

$$\lambda_{a,i} = \lambda_{D,i} \tag{6}$$

Let $\rho_t = \frac{\lambda_{D,i}}{c\mu_{t,i}}$ denote the utilization at node GI/G/c. The expression for the SCV of the output process distribution is determined based on the asymptotic method in Whitt (1982) [11] as follows:

$$c^2_{a,i} = 1 + \left(1-\rho_t^2\right)\left(c^2_{D,i}-1\right) + \frac{\rho_t^2}{\sqrt{c}}\left(\max\left\{c^2_{t,i}, 0.2\right\}-1\right) \tag{7}$$

The mean queue length of the transportation node is equal to its WIP, because of the buffer capacity is expressed by the synchronization station.

$$L_{t,i} = \rho_t \tag{8}$$

Note that the capacity of processing node is K_i. Since the node is affected due to the buffer blocking, an artificial holding node h is added to accommodate the blocked parts [18]. According to the GEM [12], the utilization at node GI/G/1/K is defined as $\rho_s = \frac{\lambda_{a,i}}{\mu_{s,i}}$. Based on the diffusion approximation methodology [13], the expression for blocking probability P_{Ki} are:

$$P_{Ki} = \frac{\rho_s (1-\rho_s)}{\hat{\rho}_s^{-(K_i-1)} - \rho_s^2} \tag{9}$$

where $\hat{\rho}_s = \exp\left\{-2\frac{1-\rho_s}{\rho_s \tilde{c}_{a,i}^2 + c_{d,i+1}^2}\right\}$. And, the return blocking probability of the artificial holding node h is given by:

$$P'_{Ki} = \left\{\frac{\mu_{s,i}+\mu_h}{\mu_h} - \frac{\lambda\left[\left(r_2^{K_i}-r_1^{K_i}\right)-\left(r_2^{K_i-1}-r_1^{K_i-1}\right)\right]}{\mu_h\left[\left(r_2^{K_i+1}-r_1^{K_i+1}\right)-\left(r_2^{K_i}-r_1^{K_i}\right)\right]}\right\}^{-1} \tag{10}$$

where r_1 and r_2 are roots to:

$$\mu_h x^2 - (\lambda + \mu_{s,i} + \mu_h)x + \lambda = 0 \tag{11}$$

while $\lambda = \tilde{\lambda}_{a,i} - \lambda_h (1 - P'_{Ki})$, $\tilde{\lambda}_{a,i}$ is the effective input rate to the processing node, which given by:

$$\tilde{\lambda}_{a,i} = \lambda_{a,i}(1-P_{Ki}) + \lambda_h (1-P'_{Ki})^{\rho_h} (1-P_{Ki})^{\rho_s} \tag{12}$$

Since the input of the processing node comes from one of the blocked traffic flow and the non blocked traffic flow, the expression for the SCV of the input process distribution is given by:

$$\tilde{c}_{a,i}^2 = c_{a,i}^2 (1-P_{Ki}) + P_{Ki} \tag{13}$$

When the system reaches the steady state, the flow conservation implies the following equation:

$$\lambda_{d,i+1} = \tilde{\lambda}_{a,i} \tag{14}$$

Finally, the SCV of output process distribution can be obtained by Marshall formula [36]:

$$c_{d,i+1}^2 = \tilde{c}_{s,i}^2 + 2\rho_s^2 c_{s,i}^2 - 2\rho_s(1-\rho_s)\mu_{s,i} W_{qi} \tag{15}$$

where W_{qi} is the mean waiting time in a GI/G/1/K queue and follows that, $W_{qi} = \frac{L_{s,i}}{\lambda_{a,i}} - \mu_{s,i}^{-1}$, while $L_{s,i}$ is the WIP at the processing node, and it can be obtained by calculating the distribution $q(x)$: $L_{s,i} = \int_1^{K_i-1} q(x)\,dx + K_i P_{Ki}$.

Fig 5 shows the synchronization station obtained from decomposition of the queuing network model. Firstly, a backward stream is added to the input stream of parts to form a closed queuing network. Then, a forward stream is added to the output stream of transportation equipment, and the synchronization station is transformed into a node composed of two closed queuing networks. Define $r = \frac{\lambda_{d,i}}{\lambda_{F,i}}$, from the derivation of literature [37], we have the following:

If $r = 1$,

$$\lambda_{D,i} = \lambda_{d,i}\left(\frac{c+n}{c+n+1}\right)\left(1 - \frac{0.5(c_i^2-1)}{2(c+n)+1}\right) \tag{16}$$

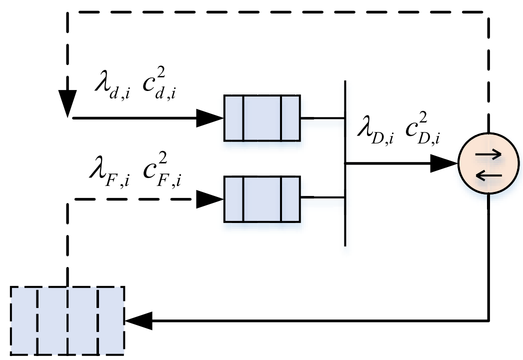


FIGURE 5. The analysis model of the synchronization station.

$$L_{d,i} = \left(\frac{n}{2}\right)\left(\frac{n+1}{c+n+1}\right) \tag{17}$$

$$L_{F,i} = \left(\frac{c}{2}\right)\left(\frac{c+1}{c+n+1}\right) \tag{18}$$

If $r \neq 1$,

$$\lambda_{D,i} = \lambda_{d,i}\left(\frac{1-r^{c+n}}{1-r^{c+n+1}}\right) \times \left[1 - 0.5\left(c_i^2-1\right)\left(\frac{(1-r)\,r^{c+n}}{1-r^{2(c+n)+1}}\right)\right] \tag{19}$$

$$L_{d,i} = \left[\left(\frac{r^{c+1}}{1-r}\right)\left(\frac{1-r^n}{1-r^{c+n+1}}\right) - \left(\frac{nr^{c+n+1}}{1-r^{c+n+1}}\right)\right] \times \left[1 + \left(\frac{1-r}{1+r}\right)\left(\frac{r^4}{1+r^8}\right)\left(c_i^2-1\right)\right] \tag{20}$$

$$L_{F,i} = \left[\left(\frac{c}{1-r^{c+n+1}}\right) - \left(\frac{r}{1-r}\right)\left(\frac{1-r^c}{1-r^{c+n+1}}\right)\right] \times \left[1 + \left(\frac{1-r}{1+r}\right)\left(\frac{r^4}{1+r^8}\right)\left(c_i^2-1\right)\right] \tag{21}$$

where $c_i^2 = 0.5\left(c_{d,i} + c_{F,i}\right)$, and c is the number of transportation equipment, n is the buffer capacity of the synchronization station.

Finally, the output process distribution of the synchronization station can be derived in Krishnamurthy et al [38] and we can see that the output rate of the synchronization station, $\lambda_{D,i}$ is always less than $min(\lambda_{d,i}, \lambda_{F,i})$ due to the influence of the synchronization constraint.

$$c_{D,i}^2 = \left[\left(\frac{\lambda_{d,i}^5}{\lambda_{d,i}^5 + \lambda_{F,i}^5}\right)c_{F,i}^2 + \left(\frac{\lambda_{F,i}^5}{\lambda_{d,i}^5 + \lambda_{F,i}^5}\right)c_{d,i}^2\right] \times \left[1 - \frac{1}{1+c+n} - \frac{1}{(1+c+n)^2}\right]\left[\frac{\sqrt{\lambda_{d,i}^4 + \lambda_{F,i}^4}}{\lambda_{d,i}^2 + \lambda_{F,i}^2}\right] \tag{22}$$

D. LINKAGE AND SOLUTION

In this section, the relationship between above parameters are linked to the traffic equations, which decomposed form the queuing network mode. As shown in Fig 3, the arrival interval time of parts enter the buffer, $\lambda_{d,i}$ is determined by the throughput of its upstream. In fact, the parts are affected by the synchronization constraint in this process due to the existence of the synchronization station, which resulted in the $\lambda_{d,i}$ not equal to the throughput of its upstream.Therefore, the idle interval time of the transportation equipment is determined by the its return rate $\lambda_{a,i}$, but not equal to $1/\lambda_{a,i}$ because of the effects of synchronization constraint. In Krishnamurthy and Suri [28], a proportion factor π is proposed to describe the impact of such problems and its derivation process is proved, which can be seen in chapter 4.3 of literature [28].

We introduce the improvement part of this method as below.

Define the proportion factor $\pi_{d,i}$, the arrival interval time of parts enter the buffer can be corrected as following:

$$\lambda'_{d,i} = \frac{\lambda_{d,i}}{1-\pi_{d,i}} \tag{23}$$

The SCV of the input process distribution is given by:

$$c_{d,i}^{2}{}' = \frac{c_{d,i}^2}{(1-\pi_{d,i})^2} - \frac{\pi_{d,i}}{(1-\pi_{d,i})^2}\left(\frac{\lambda_{d,i}}{\lambda_{F,i}}\right)\left(\frac{2c_{F,i}^2}{1+c_{F,i}^2}\right) \tag{24}$$

Define the proportion factor π_F, the return rate of the "collection queue", $\lambda_{a,i}$ need to be accumulated owing to the public transportation equipment. Then, the probability of being assigned to a transportation task, α combines the input rate of the synchronization station and is listed below:

$$\lambda'_{F,i} = \alpha\frac{\lambda_{a,i}+\lambda_{a,i+1}}{1-\pi_F} \tag{25}$$

$$\lambda'_{F,i+1} = (1-\alpha)\frac{\lambda_{a,i}+\lambda_{a,i+1}}{1-\pi_F} \tag{26}$$

The SCV of the input process distribution is given by:

$$c_{F,i}^{2}{}' = \frac{\alpha c_{a,i}^2}{(1-\pi_F)^2}\frac{\lambda_{a,i}}{\lambda_{a,i}+\lambda_{a,i+1}} - \frac{2\pi_F}{(1-\pi_F)^2} \times\left(\frac{\lambda_{a,i}+\lambda_{a,i+1}}{\lambda'_{d,i}+\lambda'_{d,i+1}}\right)\left(\frac{c_{d,i}^{2}{}'\lambda_{d,i}}{c_{d,i}^{2}{}'\lambda_{d,i}+\lambda_{d,i}+\lambda_{d,i+1}}\right) \tag{27}$$

$$c_{F,i+1}^{2}{}' = \frac{1-\alpha}{(1-\pi_F)^2}\frac{c_{a,i}^2\lambda_{a,i}}{\lambda_{a,i}+\lambda_{a,i+1}} - \frac{2\pi_F}{(1-\pi_F)^2} \times\left(\frac{\lambda_{a,i}+\lambda_{a,i+1}}{\lambda'_{d,i}+\lambda'_{d,i+1}}\right)\left(\frac{c_{d,i}^{2}{}'\lambda_{d,i}}{c_{d,i}^{2}{}'\lambda_{d,i}+\lambda_{d,i}+\lambda_{d,i+1}}\right) \tag{28}$$

In this IQNM, the relationship between above parameters are described by the nonlinear equations. Considering that the relevant parameter value of nodes in the processing direction are unknown, we propose a binary search iterative method to calculate the nonlinear equations. The initial values of the iterative process can be set by the flow relationship in queuing network theory. Therefore, according to the principle that the traffic density cannot be greater than 1 and the throughput must be less than the input rate, we set the reasonable initial value to reduce the search space, so as to speed up the whole iterative process. The detailed steps are as follows:

Step 1. ***Initialization***

1.1: Set boundary: $Low = \lambda_{d,i}$ and $High = 2\lambda_{d,i}$.

Step 2. ***Iteration***

2.1: Calculate initial values: $\lambda_{F,i}^{(j)} = (Low + High)/2$ and $\lambda_{F,i+1}^{(j)} = \lambda_{F,i}(1-\alpha)/\alpha$;

2.2: Calculate the synchronization station: $\left\{\lambda_{D,i}, c_{D,i}^2, L_{d,i}, L_{F,i}\right\}$ by Eqs. (16)—(24);

2.3: Calculate the transportation node: $\left\{\lambda_{a,i}, c_{a,i}^2, L_{t,i}\right\}$ by Eqs. (6)—(8);

2.4: Calculate the processing node: $\left\{\lambda_{d,i+1}, c_{d,i+1}^2, L_{s,i}\right\}$ by Eqs. (9)—(15);

2.5: Solve for the next node: $\{\lambda_{D,i+1}, c_{D,i+1}^2, L_{d,i+1}, L_{F,i+1}\}$ by Eqs. (16)—(24);

2.6: Calculate the transportation node: $\{\lambda_{a,i+1}, c_{a,i+1}^2, L_{t,i+1}\}$ by Eqs. (6)—(8);

2.7: Accumulate the return rate: $\{\lambda'_{F,i}, \lambda'_{F,i+1}, c_{F,i}^{2}{}', c_{F,i+1}^{2}{}'\}$ by Eqs. (25)—(28).

Step 3. ***Convergence***

3.1: Calculate $\delta = L_{F,i} + L_{F,i+1} + L_{t,i} + L_{t,i+1} - c$;

3.2: If $\delta < -\varepsilon$, set $Low = \lambda_{F,i}^{(j)}$ and $High = 2\lambda_{d,i}$, then repeat Step 2;

3.3: If $\delta > \varepsilon$, set $Low = \lambda_{d,i}$ and $High = \lambda_{F,i}^{(j)}$, then repeat Step 2;

Step 4. ***Performance measures***

4.1: Stop iteration and compute the mean throughput: $\theta = \lambda_{a,i+1}$, the mean work in process: $WIP = \sum_i (L_{d,i} + L_{t,i} + L_{s,i})$ the mean cycle time: $T = \sum_i \left(\frac{L_{d,i}}{\lambda_{d,i}} + \frac{1}{\mu_{t,i}} + \frac{L_{s,i}}{\lambda_{a,i}}\right)$.

IV. NUMERICAL RESULTS

To investigate the efficiency and quality of our proposed method, the results of the analytical approach and simulation are compared in a series of experiments with the various settings. All these experiments were implemented on a PC, Intel(R) CPU 2.390 GHz, 8.00 GB RAM, under the Windows 10 operating system. The analytical methods are programmed on MATLAB, and the simulation models are built on the Siemens Tecnomatix Plant Simulation (eM-Plant). For each experiment setting, the stable simulation results are gained from the average of 5 independent replications with 95% confidence intervals. Each run contained 1000-time units and 500-time units warm-up in order to avoid transient start-up effects.

A. DISCRETE EVENT SIMULATION

Discrete event simulation (DES) is a common computer simulation method based on Monte Carlo technology. Firstly, the system model is constructed through each module to simulate the reality, and the simulation logic, parameters and inputs are set. Then, the system is conceptualized by allowing entities to logically collect the required data through static processes. Finally, a statistical estimator is constructed to calculate the results of data samples and the average value

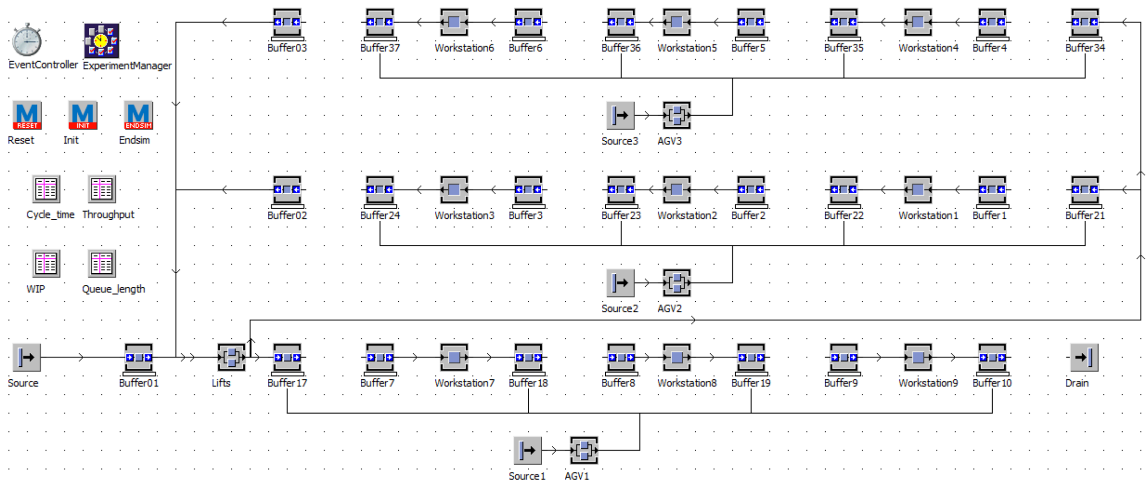


FIGURE 6. **The simulation model of the MFFS.**

of system performance measures are obtained. Fig 6 shows a MFFS simulation model which consists of three floors and nine processing procedures. In order to evaluate the accuracy of the proposed analytical approach, we use the percentage errors comparing the analytical approach(A) with the simulation (S). The percentage error for throughput is defined as:

$$\Delta 1 = (\theta^A - \theta^S)/\theta^A \times 100\%.$$

The same as the percentage error for cycle time and WIP, which is computed as:

$$\Delta 2 = (T^A - T^S)/T^A \times 100\%,$$
$$\Delta 3 = (WIP^A - WIP^S)/WIP^A \times 100\%.$$

B. GENERAL TIME DISTRIBUTION

In this section we analyze the performance of a MFFS queuing network, which consists of three floors and nine processing procedures, and study the sensitivity of system performance to variability in arrival rate of parts, transportation equipment configuration and buffer capacity. The specific queuing network model is based on the processing routing of parts, as shown in Fig 7. The parts first arrive at the half finishing area on second floor, the processing sequence is 1 to 3, and then is transported to the finishing area on third floor, the processing sequence is 4 to 6, and finally returns to the assembly area on first floor, the assembly sequence is 7 to 9.

In this experiment, we study the system performance by changing the arrival rate of parts on different distribution type. To focus on the effect of variability, we set the transportation rate, $\mu_{t,i} = 2$, the service rate of workstations 1 to 3, $\mu_{s,i} = 1.5$, the service rate of workstations 4 to 6, $\mu_{s,i} = 2$, the service rate of workstations 7 to 9, $\mu_{s,i} = 1.8$, the number of lifts are 2, the number of AGVs are 15, the buffer capacity $n_{d,i} = k_{s,i} = 3$ and the arrival rate of parts as shown in Table1. For the distribution type, we consider three distinct settings as follows: Erlang-2 (SCV = 0.5), exponential (SCV = 1) and two-stage hyper-exponential (SCV = 1.5).

From Table 2, we can find that, compared with the simulation results, the deviations of system mean throughput, system mean WIP and system mean cycle time are within an acceptable range, which calculated by the analytical approach.The calculation effect of the system mean throughput works well, and the errors are less than 3%, and the errors are less than 1% when SCV = 0.5. As shown in Fig 8, the system mean throughput, system mean WIP and system mean cycle time increase significantly with the increase of the arrival rate. However, the growth trend of system mean throughput is becoming smaller and smaller, which is caused by the insufficient production capacity.

For the production system with high arrival rate, in order to improve the system mean throughput and the cycle time, it is necessary to reasonably optimize the number of transportation equipment and buffer capacity of the system. Otherwise, the increase of the system WIP lead to the buffer blocking phenomenon. When the buffer capacity reaches the maximum, the processing of the previous procedure and the transportation process are affected, and finally the delivery time of the product is prolonged. The backlog of WIP leads to the increase of the mean cycle time. Although it can meet the needs of the production plan in theory, in fact, the cycle time of the product far exceeds the requirements of the planned lead time, so the production of the workshop cannot be carried out according to the plan.

C. EFFECT ON TRANSPORTATION EQUIOMENT CONFIGURATION

In this experiment, we study the system performance by changing the transportation equipment configuration on different distribution type. To focus on the effect of variability, we set the arrival rate of parts $\lambda_{d,i} = 1.5$, the transportation rate $\mu_{t,i} = 2$, the service rate of workstations 1 to 3, $\mu_{s,i} = 1.5$, the service rate of workstations 4 to 6, $\mu_{s,i} = 2$, the

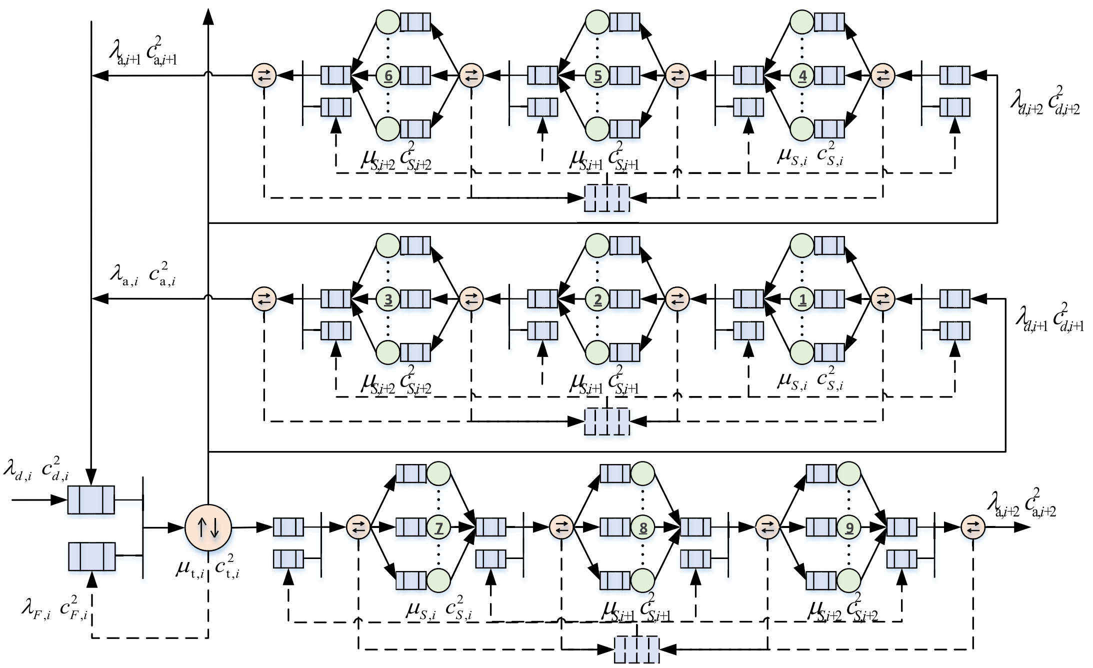


FIGURE 7. The queuing network model of the MFFS.

TABLE 2. System parameters and performance measures.

No.	SCV	$\lambda_{d,i}$	θ			T			WIP		
			A	S	Δ1/%	A	S	Δ2/%	A	S	Δ3/%
A1a	0.5	1.05	1.028	1.032	0.39	22.687	24.189	6.21	23.322	25.089	7.05
A1b		1.30	1.172	1.175	0.26	23.736	25.112	5.48	27.819	28.732	3.18
A1c		1.45	1.243	1.254	0.88	25.372	27.533	7.85	31.537	33.091	4.71
A2a	1	1.05	1.023	1.021	-0.21	24.408	25.512	4.33	24.969	26.048	4.14
A2b		1.30	1.095	1.101	0.54	26.368	25.464	-3.55	28.873	27.883	-3.81
A2c		1.45	1.191	1.209	1.49	27.520	29.907	7.98	32.776	35.157	6.78
A3a	1.5	1.05	0.995	1.023	2.74	27.032	27.144	0.41	26.899	27.768	3.13
A3b		1.30	1.087	1.102	1.36	29.536	28.536	-3.51	32.106	31.446	-2.09
A3c		1.45	1.101	1.119	1.61	30.888	31.317	1.37	34.007	35.043	2.96

service rate of workstations 7 to 9, $\mu_{s,i} = 1.8$, the buffer capacity $n_{d,i} = k_{s,i} = 3$ and the number of the transportation equipment as shown in Table 2. For the distribution type, we consider three distinct settings as follows: Erlang-2 (SCV = 0.5), exponential (SCV = 1) and two-stage hyper-exponential (SCV = 1.5).

From Table 3, we can find that, compared with the simulation results, the deviations of system mean throughput and system mean cycle time are within an acceptable range, which calculated by the analytical approach. The calculation effect of the system mean throughput works well, and the average errors are 3.11%, and the errors are less than 4% when SCV = 1.5. As shown in Fig 9, the system mean throughput increase significantly with the increase of the transportation equipment. However, the growth trend is becoming smaller and smaller, because the increase in the number of transportation equipment is easy to cause AGVs congestion in the limited track, and reducing the transportation efficiency. The mean cycle time increases sharply, because the increase in the number of transportation equipment accelerates the speed of parts transportation to the workstation, and increases the mean WIP and the mean queue length. Fig 9b illustrated that the processing efficiency is improved, but the length of the product is also prolonged.

According to the production characteristics of the MFFS, the transportation equipment must be configured to deal with the fluctuation of production system. Although increasing the number of transportation equipment can improve the mean

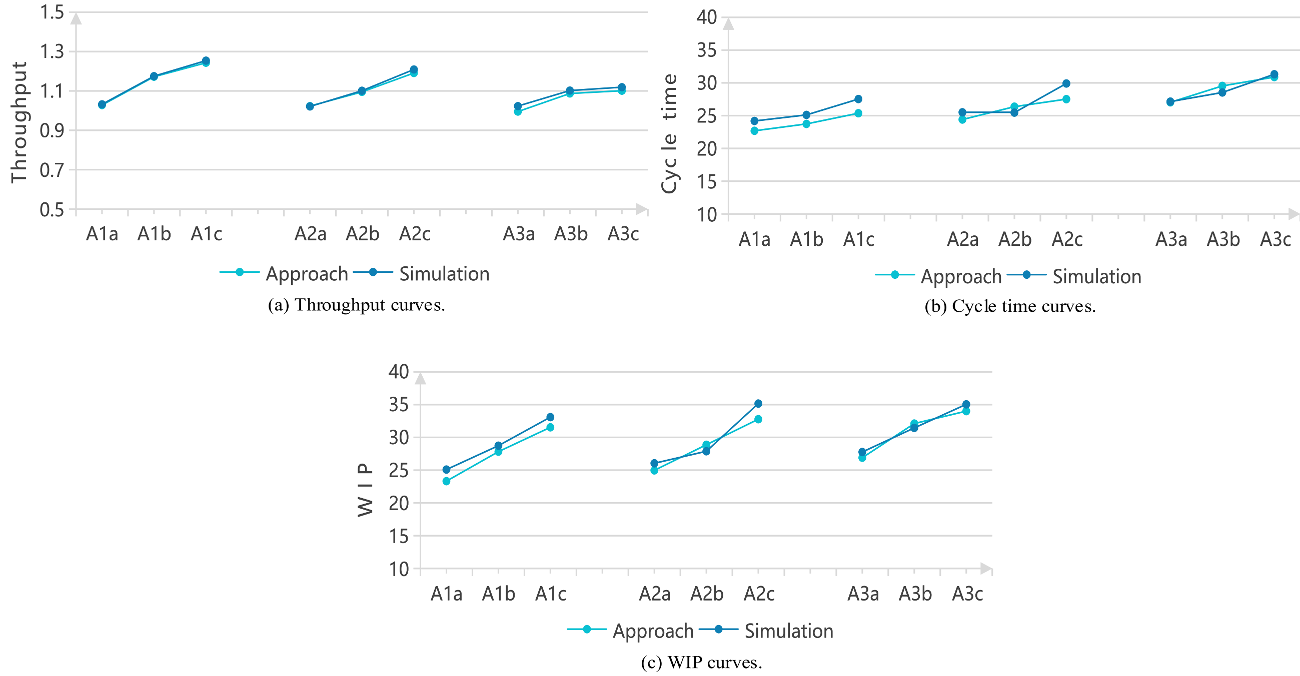


FIGURE 8. System performance measures curves for different arrival rate.

throughput of the system, it requires high cost and too many transportation equipment is easy to lead to congestion, which reduces the transportation efficiency. Thus, it is particularly important to reasonably set the buffer capacity. Next experiment analyses the impact of different buffer capacity on system performance.

D. EFFECT ON BUFFER CAPACITY

In this experiment, we study the system performance by changing the buffer capacity on different distribution type. To focus on the effect of variability, we set the arrival rate of parts $\lambda_{d,i} = 1.5$, the transportation rate $\mu_{t,i} = 2$, the service rate of workstations 1 to 3, $\mu_{s,i} = 1.5$, the service rate of workstations 4 to 6, $\mu_{s,i} = 2$, the service rate of workstations 7 to 9, $\mu_{s,i} = 1.8$, the number of lifts are 2, the number of AGVs are 15, and the buffer capacity as shown in Table1. For the distribution type, we consider three distinct settings as follows: Erlang-2 (SCV = 0.5), exponential (SCV = 1) and two-stage hyper-exponential (SCV = 1.5).

From Table 4, we can find that, compared with the simulation results, the deviations of system mean throughput and system mean cycle time are within an acceptable range, which calculated by the analytical approach. The calculation effect of the system mean throughput works well, except for the experimental group with buffer capacity of 1, the other errors are less than 5%.

The average errors of the mean cycle time are 2.05% when SCV = 1.5. The mean throughput is low because the buffer capacity is too small to cope with the production fluctuation of the MFFS, some prats are refused to enter the system due to blocking.

As shown in Fig 10, with the increase of buffer capacity, the mean throughput of the system has an obvious increasing trend, but the growth trend is getting smaller and smaller due to the processing and transportation resources are insufficient. Fig 10b illustrated that the mean cycle time of products increases sharply. As more and more parts are waiting for processing and transportation in the system, the average WIP and average queue length are gradually increasing.

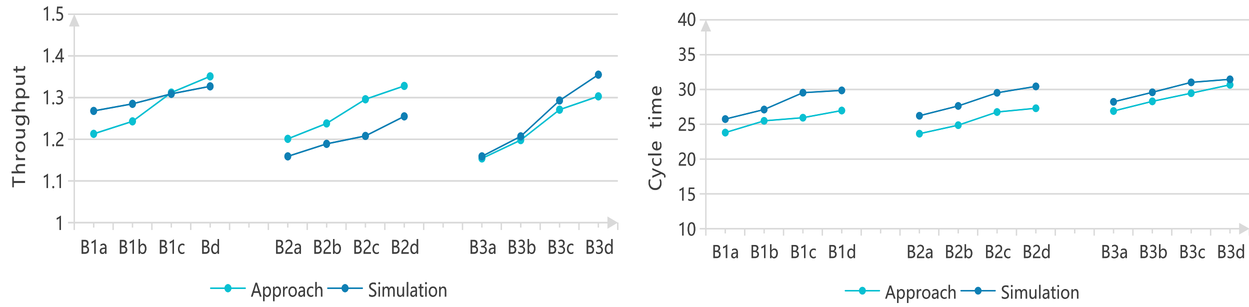
In summary, the system throughput are usually affected by three factors, including arrival rate of parts, transportation equipment configuration and the buffer capacity. For the production system with high arrival rate, it is particularly important to reasonably configure the number of transportation equipment and buffer capacity. In terms of cost, the increase of buffer capacity provides a cheaper method for the system. However, the increase of the system WIP when the buffer is too large, resulted in a longer cycle time, and affected the delivery time

V. CASE STUDY

In this section, we take a robot manufacturing enterprise as an actual case study and analyze the MFFS comprising of floor 3 and block 15. In view of the production problems existing in the workshop, such as unbalanced production capacity, low equipment utilization, congestion in material transportion process, we consider the impact of various performance indicators in the actual production of the MFFS. Therefore, we improve the original empirical layout and

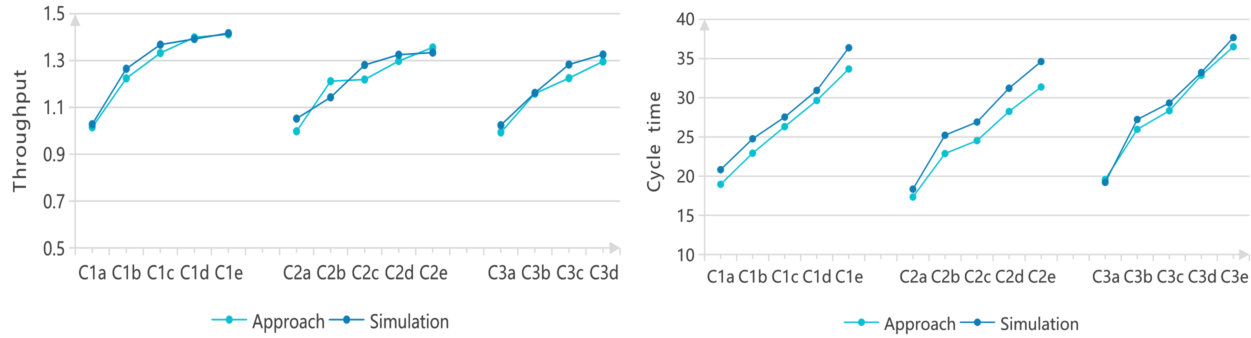
TABLE 3. System parameters and performance measures.

No.	*Lifts*	*AGVs*	θ			T		
			A	S	Δ1/%	A	S	Δ2/%
B1a	2	12	1.213	1.268	4.34	23.808	25.749	7.54
B1b	2	15	1.243	1.285	3.27	25.497	27.119	5.99
B1c	2	18	1.312	1.309	-0.23	25.946	29.542	11.12
B1d	2	20	1.351	1.327	-1.81	26.977	29.868	9.68
B2a	2	12	1.201	1.159	-3.62	23.648	26.225	9.82
B2b	2	15	1.238	1.189	-4.12	24.873	27.637	10.01
B2c	2	18	1.296	1.208	-7.28	26.762	29.525	9.36
B2d	2	21	1.328	1.255	-5.82	27.302	30.443	10.32
B3a	2	12	1.154	1.159	0.43	26.906	28.224	4.67
B3b	2	15	1.198	1.207	0.75	28.295	29.595	4.39
B3c	2	18	1.271	1.293	1.71	29.468	31.024	5.02
B3d	2	20	1.303	1.355	3.84	30.671	31.459	2.51



(a) Throughput curves. (b) Cycle time curves.

FIGURE 9. System performance measures curves for different transportation equipment configuration.



(a) Throughput curves. (b) Cycle time curves.

FIGURE 10. System performance measures curves for different buffer capacity.

propose a decomposition-coordination method [39], which embedding the queuing network model, to deal with the layout optimization problem of the MFFS.

The general flow direction of materials in the workshop is as follows: the raw materials first enter the rough processing area on the second floor, and then are transported to the finish processing area on the third floor by lifts when the processing is completed. After the finished products are processed, they return to the assembly area on the first floor for assembly and packaging. The specific floor division, block size, shape and block optional location are shown in Table 5.

TABLE 4. System parameters and performance measures.

No.	$n_{d,i}$	$k_{s,i}$	θ			T		
			A	S	Δ1/%	A	S	Δ2/%
C1a	1	1	1.015	1.028	1.26	18.953	20.825	8.98
C1b	2	2	1.224	1.265	3.24	22.934	24.768	7.40
C1c	3	3	1.332	1.368	2.63	26.321	27.539	4.42
C1d	4	4	1.399	1.392	-0.51	29.655	30.942	4.15
C1e	5	5	1.412	1.417	0.35	33.673	36.383	7.44
C2a	1	1	0.998	1.052	5.13	17.335	18.342	5.49
C2b	2	2	1.212	1.143	-6.03	22.881	25.221	9.27
C2c	3	3	1.219	1.281	4.83	24.529	26.903	8.82
C2d	4	4	1.298	1.325	2.03	28.254	31.218	9.49
C2e	5	5	1.356	1.334	-1.65	31.376	34.621	9.37
C3a	1	1	0.993	1.024	3.02	19.571	19.202	-1.9
C3b	2	2	1.159	1.162	0.25	25.965	27.228	4.63
C3c	3	3	1.225	1.283	4.52	28.352	29.329	3.33
C3d	4	4	1.296	1.326	2.26	32.872	33.227	1.06
C3e	5	5	1.325	1.351	1.92	36.512	37.682	3.10

A. DECOMPOSITION-COORDINATION METHOD

In order to optimal the facility layout of the MFFS with efficiency, we propose a new search algorithm based on the decomposition coordination mechanism. The optimization direction is changed from local optimization to global optimization by reasonably adjusting the constraints of the subsystem. The new original system layout is constantly formed, updated and verified, and finally the best layout scheme of multi-storey workshop is found. In the subsystem optimization problem, the object is multiple regular or irregular blocks in the floor. We use branch and bound algorithms to optimize the quadratic set covering problem (QSP), which can be seen in the literature for details. The detailed steps of the decomposition coordination method are as follow:

Step 1. *Decomposition*

1.1: $X^i \rightarrow \left(X_1^i, X_2^i, \cdots, X_s^i\right)$, $i = 0$, $n = 0$, $\beta = 0.01$;

1.2: Calculate initial values: $\theta_r^{\min} = \theta^{\min} + n\beta\,(\alpha_r)^{-2}$ and $T_r^{\max} = \frac{T^{\max}}{r} - n\beta\,(\alpha_r)^{-2}$.

Step 2. *Optimization*

2.1: Calculate arrival rate of subsystem: $\lambda \rightarrow \lambda_r$;

2.2: Optimize subsystem: $Localsoultion \rightarrow X_r^{i+1}$.

Step 3. *Performance measures*

3.1: Merge subsystem: $X_1^{i+1}, X_2^{i+1}, \cdots, X_s^{i+1} \rightarrow X^{i+1}$;

3.2: Calculate system performance: *Queuing network* $\rightarrow$ θ^{i+1} and T^{i+1}.

Step 4. *Coordination*

4.1: If $\theta^{i+1} < \theta^{\max}$ or $T^{i+1} > T^{\max}$, set $n = n + 1$, then update $\theta_r^{\min}$ and $T_r^{\max}$, go to step2;

4.2: If $\theta^{i+1} \geq \theta^{\max}$, $T^{i+1} \leq T^{\max}$ and $\beta \leq 0.01$, then go to step5; or set $\theta_r^{\min} = \theta_r^{\min} - \beta\,(\alpha_r)^{-2}$, $T_r^{\max} = T_r^{\max} + \beta\,(\alpha_r)^{-2}$ and $\beta = \frac{\beta}{10}$;

4.3: Set $\theta_r^{\min} = \theta_r^{\min} + \beta\,(\alpha_r)^{-2}$ and $T_r^{\max} = T_r^{\max} - \beta\,(\alpha_r)^{-2}$, then go to step2.

Step 5. *Convergence*

5.1: If $\max_{\forall n}\left|\frac{n\beta}{(\alpha_r)^2}\right| < \varepsilon$, stop the optimization; otherwise, go back to step2.

B. OPTIMIZATION RESULTS

To approach the actual production situation, we set the raw material arrival rate $\lambda = 2$, the service rate of each block $\mu_s = 2$, the transportation rate between blocks and lifts transportation rate $\mu_t = 2$, the buffer capacity $n_d = k_s = 3$, the throughput $\theta^{\min} = 1.2$ and the cycle time $T^{\max} = 30$. The case optimization results are shown in Table 6 and Fig 11.

From Table 6, we can draw a conclusion that our approach is effective in improving the layout of the MFFS, and the production cycle of products is significantly reduced. This is not only one of the important indicators to evaluate the performance of workshops, but also the guarantee of production plan and delivery time. As shown in Fig 11, this section takes a specific production workshop structure as the application scenario. For the QSP model [40], we divide the site area and make layout planning according to the actual size of the block. The total transportation cost is taken as the optimization objective including transportation time and waiting

TABLE 5. The block size, shape and detailed optional location.

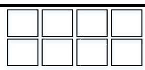
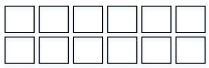
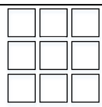
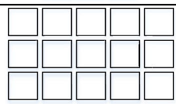
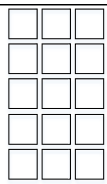
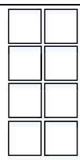
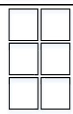
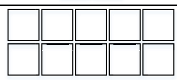
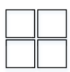
Block	*Shape*	*Floors*	*Location*	*Definition*
1，2，3		1	1—48	*Assembly*
4，5		3	1-6，9-14 17-22，25-30 33-38，41-46	*Precision machining*
6，7		2	1-32	*Rough machining*
8		2	9-13，17-21 25-29，33-37	*Rough machining*
9		1	1-48	*Storage area*
10		2	1-48	*Storage area*
11		3	7-8，15--16 23-24，31-32	*Storage area*
12		3	1-16，33-48	*Precision machining*
13		1-3	5-8，13-16 21-24，29-32 37-40，45-48	Lift
14，15		1-3	1，4，5，8 17，24，25 32，41，48	Lift

TABLE 6. The optimization results of the specific case.

Objects	*Results*
Transportation cost（hours)	17.92
Throughput（minutes^{-1})	1.256
Cycle time（minutes)	26.581
Optimization time（minutes)	3.81
Layout scheme	*Figure 11*

time, which considered the various queuing phenomena in the production process of the MFFS. The layout scheme is continuously optimized under the constraints of certain throughput and cycle time. Finally, the specific location of

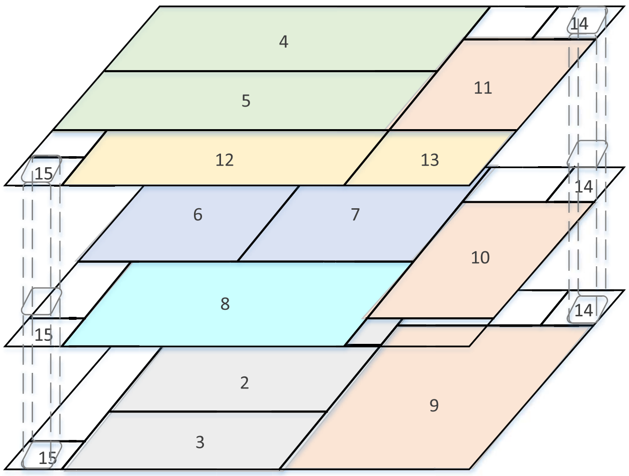


FIGURE 11. The block layout scheme of the MFFS.

each block is obtained and the performance index of the system is calculated. The results show that the application of decomposition coordination optimization algorithm in practice and the practical application value is highlighted.

VI. CONCLUSION

In this paper, we proposed an IQNM for estimating the performance measures of the MFFS, which considered its stochastic factors. In order to optimal the facility layout of the workshop, we take the total transportation cost as the objective function and establish a mathematical model with the system performance constraints. An iteration algorithm based on decomposition and coordination mechanism is used to optimize the overall layout through the repeated adjustments of subsystems. In addition, a series of experiments are designed to verify the effectiveness and accuracy of our approach and it is applied to the actual manufacturing industry for specific case analysis, which provides an important decision-making basis for the layout planning of this kind of production workshop.

The method of using queuing network model to quickly calculate performance indicators is an approximate solution method. It is prone to certain errors, resulting in inaccurate judgment of constraint conditions and slight deviation of layout scheme. However, from the optimization results, this deviation is within the acceptable error range. Therefore, our proposed method is a high potential for optimizing the complicated manufacturing systems, and it also needs to be further improved.

Our future research focus on the more advanced and complex multi-floor workshop, such as the tier-to-tier logistics transportation system and the tier-captive AGVs system, which limits the vehicles to transferring during the tiers. For the solution accuracy of analysis methods, the mathematical methods such as queuing network model can not reach the effect of simulation, but the production system needs a faster solution speed. Therefore, it is an opportunity and a challenge for mathematical analysis. If we can break through the shortcomings of existing methods in solution accuracy, the analysis method with high solution efficiency will play a decisive role in the resource allocation and layout planning of manufacturing industry.

REFERENCES

[1] A. Ahmadi, M. S. Pishvaee, and M. R. Akbari Jokar, "A survey on multi-floor facility layout problems," *Comput. Ind. Eng.*, vol. 107, pp. 158–170, May 2017.
[2] H. Hosseini-Nasab, S. Fereidouni, S. M. T. F. Ghomi, and M. B. Fakhrzad, "Classification of facility layout problems: A review study," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 1–4, pp. 957–977, 2018.
[3] X. Chen, "Study on multi-floor facility layout problem based on M/G/1 lift handling system," Ph.D. dissertation, College Mech. Eng., Zhejiang Univ. Technol., Hangzhou, China, 2009.
[4] N. Izadinia and K. Eshghi, "A robust mathematical model and ACO solution for multi-floor discrete layout problem with uncertain locations and demands," *Comput. Ind. Eng.*, vol. 96, pp. 237–248, Jun. 2016.
[5] N. Izadinia, K. Eshghi, and M. H. Salmani, "A robust model for multi-floor layout problem," *Comput. Ind. Eng.*, vol. 78, pp. 127–134, Dec. 2014.
[6] J. O. Ejeh, S. Liu, and L. G. Papageorgiou, "Optimal layout of multi-floor process plants using MILP," *Comput. Chem. Eng.*, vol. 131, Dec. 2019, Art. no. 106573.
[7] C. O. Díaz-Ovalle, V. Rico-Ramirez, and E. O. Castrejón-González, "An MINLP approach to the 3D process layout problem," *Chem. Eng. Res. Design*, vol. 165, pp. 137–149, Jan. 2021.
[8] R. Wang, H. Zhao, Y. Wu, Y. Wang, X. Feng, and M. Liu, "An industrial facility layout design method considering energy saving based on surplus rectangle fill algorithm," *Energy*, vol. 158, pp. 1038–1051, Sep. 2018.
[9] T. Lerher, M. Ficko, and I. Palčič, "Throughput performance analysis of automated vehicle storage and retrieval systems with multiple-tier shuttle vehicles," *Appl. Math. Model.*, vol. 91, pp. 1004–1022, Mar. 2021.
[10] D. Gross, J. F. Shorite, J. M. Thompsonm, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. New York, NY, USA: Wiley, 2008.
[11] W. Whitt, "Approximating a point process by a renewal process, I: Two basic methods," *Oper. Res.*, vol. 30, no. 1, pp. 125–147, 1982.
[12] L. Kerbachea and J. M. Smith, "The generalized expansion method for open finite queueing networks," *Eur. J. Oper. Res.*, vol. 32, no. 3, pp. 448–461, Dec. 1987.
[13] J. Labetoulle and G. Pujolle, "Isolation method in a network of queues," *IEEE Trans. Softw. Eng.*, vol. SE-6, no. 4, pp. 373–381, Jul. 1980.
[14] W. Whitt, "The queuing network analyzer," *Bell Syst. Tech. J.*, vol. 62, pp. 2779–2815, Nov. 1983.
[15] L. Kerbache and J. M. Smith, "Multi-objective routing within large scale facilities using open finite queueing networks," *Eur. J. Oper. Res.*, vol. 121, no. 1, pp. 105–123, 2000.
[16] L. Kerbache and J. M. Smith, "Queueing networks and the topological design of supply chain systems," *Int. J. Prod. Econ.*, vol. 91, no. 3, pp. 251–272, 2004.
[17] H.-Y. Zhang, Q.-X. Chen, and N. Mao, "System performance analysis of flexible flow shop with match processing constraint," *Int. J. Prod. Res.*, vol. 54, no. 20, pp. 6052–6070, Oct. 2016.
[18] H.-Y. Zhang, Q.-X. Chen, J. M. Smith, N. Mao, A.-L. Yu, and Z.-T. Li, "Performance analysis of open general queuing networks with blocking and feedback," *Int. J. Prod. Res.*, vol. 55, no. 19, pp. 5760–5781, Oct. 2017.
[19] J. M. Harrison, "Assembly-like queues," *J. Appl. Probab.*, vol. 10, no. 2, pp. 354–367, Jun. 1973.
[20] M. K. Govil and M. C. Fu, "Queueing theory in manufacturing: A survey," *J. Manuf. Syst.*, vol. 18, no. 3, pp. 214–240, 1999.
[21] M. van Vuuren, I. J. B. F. Adan, and S. A. E. Resing-Sassen, "Performance analysis of multi-server tandem queues with finite buffers and blocking," *OR Spectr.*, vol. 27, nos. 2–3, pp. 315–338, Jun. 2005.
[22] M. Colledani, F. Gandola, A. Matta, and T. Tolio, "Performance evaluation of linear and non-linear multi-product multi-stage lines with unreliable machines and finite homogeneous buffers," *IIE Trans.*, vol. 40, no. 6, pp. 612–626, Apr. 2008.
[23] Y. Dallery, R. David, and X.-L. Xie, "An efficient algorithm for analysis of transfer lines with unreliable machines and finite buffers," *IIE Trans.*, vol. 20, no. 3, pp. 280–283, Sep. 1988.

[24] S. B. Gershwin and M. H. Burman, "A decomposition method for analyzing inhomogeneous assembly/disassembly systems," *Ann. Oper. Res.*, vol. 93, pp. 91–115, Jan. 2000.

[25] A. Azaron, H. Katagiri, K. Kato, and M. Sakawa, "Modelling complex assemblies as a queueing network for lead time control," *Eur. J. Oper. Res.*, vol. 174, no. 1, pp. 150–168, Oct. 2006.

[26] M. Manitz, "Queueing-model based analysis of assembly lines with finite buffers and general service times," *Comput. Oper. Res.*, vol. 35, no. 8, pp. 2520–2536, Aug. 2008.

[27] M. Manitz, "Analysis of assembly/disassembly queueing networks with blocking after service and general service times," *Ann. Oper. Res.*, vol. 226, no. 1, pp. 417–441, Mar. 2015.

[28] A. Krishnamurthy and R. Suri, "Performance analysis of single stage Kanban controlled production systems using parametric decomposition," *Queueing Syst.*, vol. 54, no. 2, pp. 141–162, Oct. 2006.

[29] K. Satyam and A. Krishnamurthy, "Performance evaluation of a multi-product system under CONWIP control," *IIE Trans.*, vol. 40, no. 3, pp. 252–264, Jan. 2008.

[30] K. Satyam and A. Krishnamurthy, "Performance analysis of CONWIP systems with batch size constraints," *Ann. Oper. Res.*, vol. 209, no. 1, pp. 85–114, Oct. 2013.

[31] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, "Performance analysis and design trade-offs in warehouses with autonomous vehicle technology," *IIE Trans.*, vol. 44, no. 12, pp. 1045–1060, Dec. 2012.

[32] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, "Blocking effects in warehouse systems with autonomous vehicles," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 439–451, Apr. 2014.

[33] B. Y. Ekren, S. S. Heragu, A. Krishnamurthy, and C. J. Malmborg, "An approximate solution for semi-open queueing network model of an autonomous vehicle storage and retrieval system," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 1, pp. 205–215, Jan. 2013.

[34] B. Y. Ekren, S. S. Heragu, A. Krishnamurthy, and C. J. Malmborg, "Matrix-geometric solution for semi-open queuing network model of autonomous vehicle storage and retrieval system," *Comput. Ind. Eng.*, vol. 68, pp. 78–86, Feb. 2014.

[35] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, "Stochastic Models for unit-load operations in warehouse systems with autonomous vehicles," *Ann. Oper. Res.*, vol. 231, pp. 120–155, 2014a.

[36] K. T. Marshall, "Some inequalities in queuing," *Oper. Res.*, vol. 16, no. 3, pp. 651–668, 1968.

[37] A. Krishnamurthy, R. Suri, and M. Vernon, "Two-Moment approximations for throughput and mean queue length of a fork/join station with general input processes," in *Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains* (International Series in Operations Research and Management Science), vols. 87–126, J. G. Shanthikumar, D. D. Yao, and W. H. M. Zijm, Eds. Norwell, MA, USA: Kluwer, 2003.

[38] A. Krishnamurthy, R. Suri, and M. Vernon, "Two-Moment approximations for the variability parameter of the departure process from a fork/join station with general inputs," Dept. Ind. Eng., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2002.

[39] S. Xi, Q. Chen, J. MacGregor Smith, N. Mao, A. Yu, and H. Zhang, "A new method for solving buffer allocation problem in large unbalanced production lines," *Int. J. Prod. Res.*, vol. 58, no. 22, pp. 6846–6867, Nov. 2020.

[40] M. S. Bazaraa, "Computerized layout design: A branch and bound approach," *AIIE Trans.*, vol. 7, no. 4, pp. 432–437, 1975.

XIAOPENG LIU is currently pursuing the Ph.D. degree with the Department of Electromechanical Engineering, Guangdong University of Technology. His research interests include queueing network modelling and optimization of facility layout. He is responsible for writing articles, experimental analysis, and simulation verification.

HUIYU ZHANG is currently an Assistant Professor at the Department of Industrial Engineering, Guangdong University of Technology. His research interests include queueing network modelling and optimization of manufacturing, and service systems. He provides guidance on mathematical methods.

JUN LIN is currently a Senior Engineer and the Vice Director of the Software and System Research Unit, China Electronic Product Reliability and Environment Testing Research Institute. He is mainly engaged in computer software development and quality engineering technology research. He provides guidance on article writing.

XUANRUI CHEN is currently pursuing the Ph.D. degree with the Department of Electromechanical Engineering, Guangdong University of Technology. His research interests include design and operation optimization of manufacturing, and service systems. He participated in part of the experiment and gives relevant suggestion.

QINGXIN CHEN is currently a Full Professor with the Department of Electromechanical Engineering, Guangdong University of Technology. His research interests include production resource allocation, shop layout optimization, and production planning and control. He provides guidance on article writing.

NING MAO is currently a Full Professor at the Department of Electromechanical Engineering, Guangdong University of Technology. Her research interests include production resource allocation, production planning and control, and enterprise information management systems. She provides guidance on article writing.