Journal of the Operations Research
Society of Japan
Vol. 39, No. 1, March 1996

CAPACITY DESIGN IN A FLEXIBLE MANUFACTURING SYSTEM WITH LIMITED WAREHOUSE CAPACITY AND THE LIMITED NUMBER OF PALLETS

Hiroyuki Nagasawa
Osaka Prefecture University

(Received February 21, 1994; Revised January 27, 1995)

Abstract Capacity design problems in a flexible manufacturing system (FMS) with a Poisson arrival process and with limited system-capacity are formulated to determine the number of boxes to store arrival jobs in a supplementary warehouse, called the "warehouse capacity," and the number of pallets available in the FMS. The throughput of the FMS is performed through an approximate closed queueing network model. The limits, and the first- and the second-order properties of the throughput function with respect to the warehouse capacity and the number of pallets available in the FMS are derived and exploited to develop efficient solution methods for making the near-optimal capacity design.

1. Introduction

A typical flexible manufacturing system (FMS) consists of machining stations, a loading/unloading station, a material handling system and a supplementary automatic warehouse as shown in Fig. 1. The warehouse which is usually positioned adjacent to the loading/unloading station is composed of a stacker crane and many small boxes (spaces) stacked like shelves. In this paper, we focus on the capacity design of the FMS for determining the number of boxes in the warehouse and the number of pallets available in the system.

The number of jobs simultaneously circulating in the FMS, which is usually called "work-in-process" inventory, is limited to the sum of the number of boxes in the warehouse and the number of pallets in the "shop" (from this point on, we refer to the FMS excluding the warehouse as the "shop"). The local buffer capacity at each station in the shop is so large that no blocking occurs at any station in the shop.

Consider a Poisson arrival process. Arriving jobs are permitted to enter the FMS whenever the supplementary warehouse is not fully occupied. These jobs, temporarily stored at the warehouse, are immediately supplied to the shop (first, to the loading/unloading station) whenever a pallet becomes available (empty). In other words, these jobs are blocked to enter the shop and wait at the warehouse as long as the shop is full. On the other hand, if the FMS is full, all arriving jobs are blocked and lost (or transferred to the other manufacturing systems if possible). Increasing the number of available pallets and boxes will result in a decrease in the "blocked-and-lost" probability, and hence an increase in the throughput of the FMS.

Increasing the number of available pallets reduces the probability that each machining station becomes idle due to shop congestion, and then results in increasing the throughput of the FMS. However, it requires more horizontal spaces at each machining station and more fixtures to palletize jobs. On the other hand, increasing the warehouse capacity by adding another layer to the warehouse usually requires neither much more horizontal space nor much more cost compared with increasing the number of pallets.

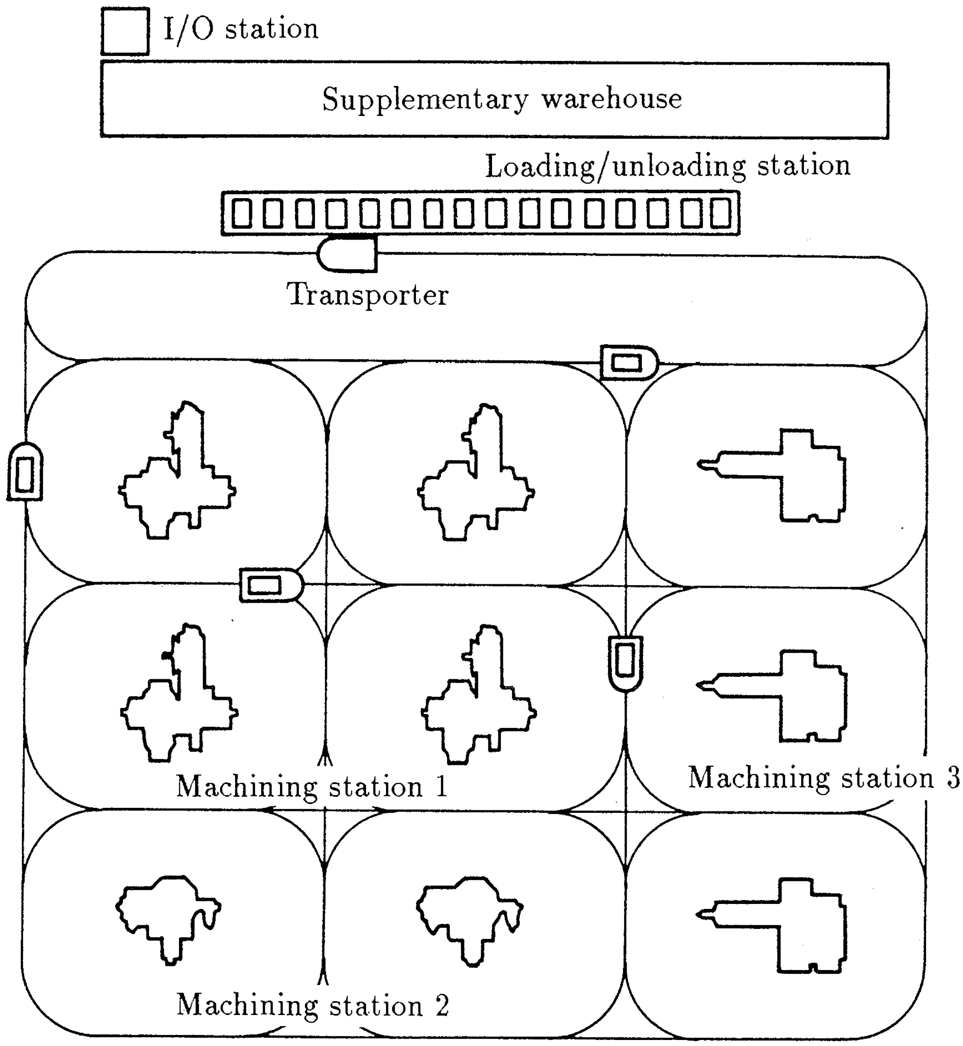

Figure 1. A typical FMS with a supplementary warehouse

The warehouse capacity and the number of pallets, therefore, should be carefully chosen to maximize the throughput of the FMS. The optimal design of the FMS with respect to these capacities will be addressed in this paper.

Up to now, there are numerous papers analyzing blocking phenomena occurring at each individual station with finite local buffer [6] [7], but there is little literature dealing with the blocking in this type of FMS with a supplementary warehouse. Shanthikumar and Stecke [9] deal with the case in which the warehouse capacity is infinite. Yao [15] and Yao and Shanthikumar [17] formulate some storage models for the FMSs as lot sizing models in which a batch of jobs periodically arrive at the FMS. Since in our FMS model as well as in the previous models the equilibrium probabilities of queue length at each station and the throughput of the FMS can not be exactly calculated, we use an approximation similar to the one given in these references [9] [15] [17].

In section 2, we formulate a mathematical model of the FMS using a closed queueing network model and give a heuristic derivation of the equilibrium probabilities and the throughput of the FMS. Section 3 gives the limits, and the first- and the second-order properties of the throughput function with respect to both the warehouse capacity (the number of boxes) and the number of pallets. In section 4, we propose solution methods to several capacity design problems in the FMS model exploiting these properties.

2. Model formulation

Consider an FMS consisting of M machining stations, a material handling system (MHS), a loading/unloading station and a supplementary automatic warehouse. A typical example of this FMS is illustrated in Fig. 1. The conditions of this system are as follows:

(1) Each station has c_i servers ($i = 0, 1, \cdots, M+1$) and jobs are served in a first-come-first-served (FCFS) order. Service times at each station are exponentially distributed with queue-length dependent service rate, $\mu_i(n_i), i = 0, 1, \cdots, M+1$, where n_i denotes queue length at station $i; i = 0$ denotes the loading station (for palletizing/ refixturing operations), $i = M+1$, the MHS, and $i = 1, 2, \cdots, M$, the machining stations.

(2) Jobs follow a Markov routing with $r(i,j)$, $i(\text{or } j) = -1, 0, 1, \cdots, M$, denoting the probability that a job is routed from station i to station j. The notations $r(i,0)$ and $r(i,-1)$ denote the routing probability from station i to the loading station for palletizing/refixturing operations and unloading station for depalletizing operations, respectively. We model the loading/unloading station as the two stations: station 0 modeling the loading operation and station-1 modeling the unloading operation. We assume that the operational time at the unloading station is small enough to be negligible.

(3) There are N_S pallets simultaneously available in the shop and the number of local buffer spaces at each station is unlimited (or there are N_S spaces). The warehouse has N_H boxes (spaces) to store arriving jobs until a pallet becomes available. Define $N = N_S + N_H$, called the "maximum population in the FMS." An accepted job will be temporarily stored in the warehouse and supplied in the FCFS order to the shop whenever a pallet becomes empty.

(4) Consider a Poisson arrival process with rate λ. If an arriving job finds the number of jobs in the FMS being just N, the job is refused entry and lost.

Although the FMS can be represented by a restricted open queueing network as shown in Jackson [4], we construct an equivalent closed queueing network (CQN) as shown in Fig. 2, so that many results for CQNs can be directly applied. In Fig. 2, the notation "I/O" represents the input/output process and is normally called the "input/output" station with rate λ. This I/O station in the CQN represents the "blocked-and-lost" mechanism because the "service" (corresponding to the "arrival" process) at this station is not implemented until any "customer" (corresponding to any "empty position" in the FMS) arrives at this station. The notation "H" and the arc"$\rightarrow\!\!\!\!+\!\!\rightarrow$" denote the automatic warehouse and the "blocked-and-hold 0" mechanism, respectively – the warehouse is considered as a station with infinite service rate and we assume that the blocking occurs at service initiation (this means "blocked-and- hold 0"). The arrow with "$\bullet$" denotes the transportation through the MHS (station $M+1$).

The flow equations corresponding to Fig. 2 are given as follows:

(1a) $$e_i = \sum_{j=0}^{M} e_j \gamma(j,i),\ i = 1, 2, \cdots, M,$$

(1b) $$e_0 = e_{-1} + \sum_{j=1}^{M} e_j \gamma(j,0),$$

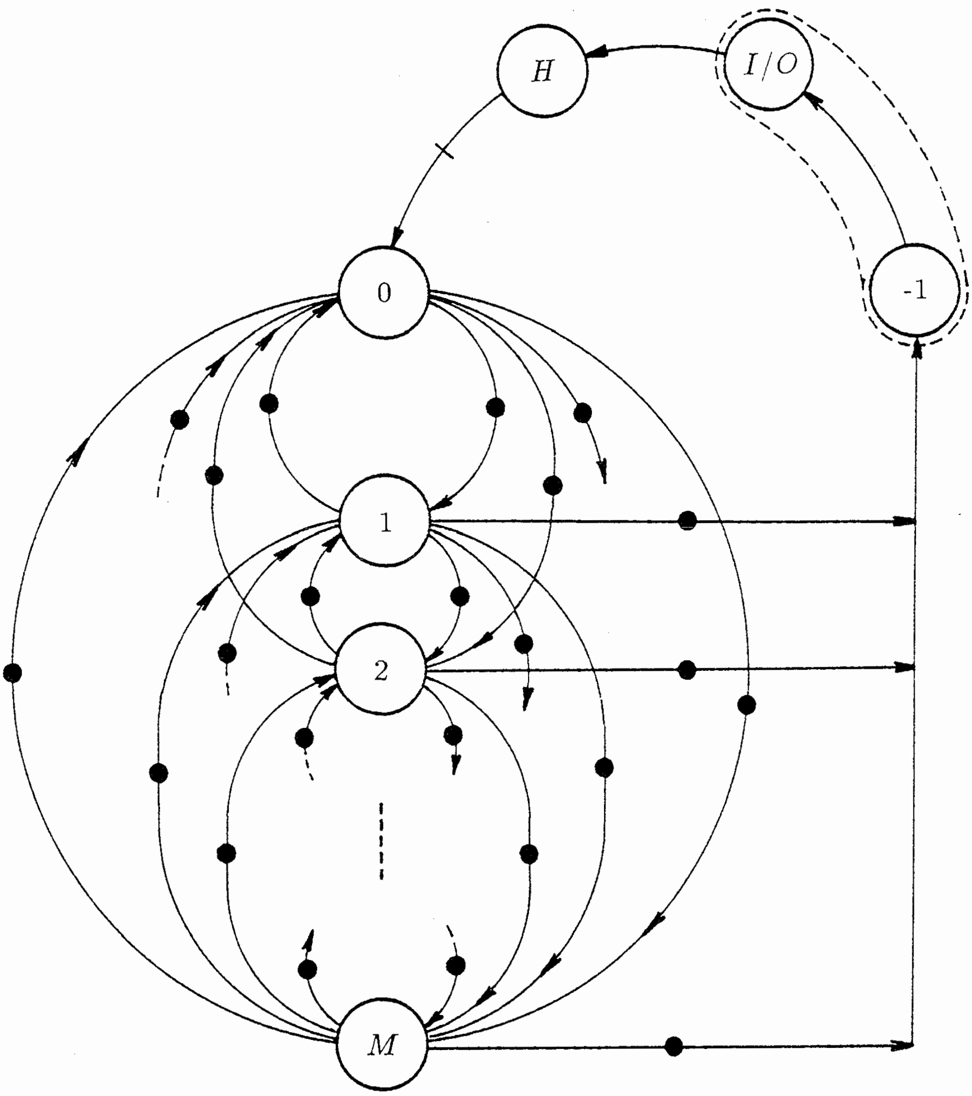

Figure 2. A queueing network model of the FMS with a supplementary warehouse

(1c) $$e_{-1} = \sum_{j=1}^{M} e_j \gamma(j, -1),$$

(1d) $$e_{M+1} = \sum_{j=0}^{M} e_j,$$

(1e) $$e_{-1} = 1,$$

where e_i denotes the visit ratio of an arriving job to station $i, i = -1, 0, \cdots, M+1$.

Eqns (1a) through (1e) have the unique solution, denoted by $e_i^*, i = -1, 0, \cdots, M+1$, which represents the expected visit times of an arriving job to each station.

If eqn. (1e) is removed, the remaining equations provide a set of solutions, denoted by $e_i, i = -1, 0, \cdots, M+1$, called the "relative visit ratio" of an arriving job. In this case, multiplication by an arbitrary constant does not affect the solution. For instance, the

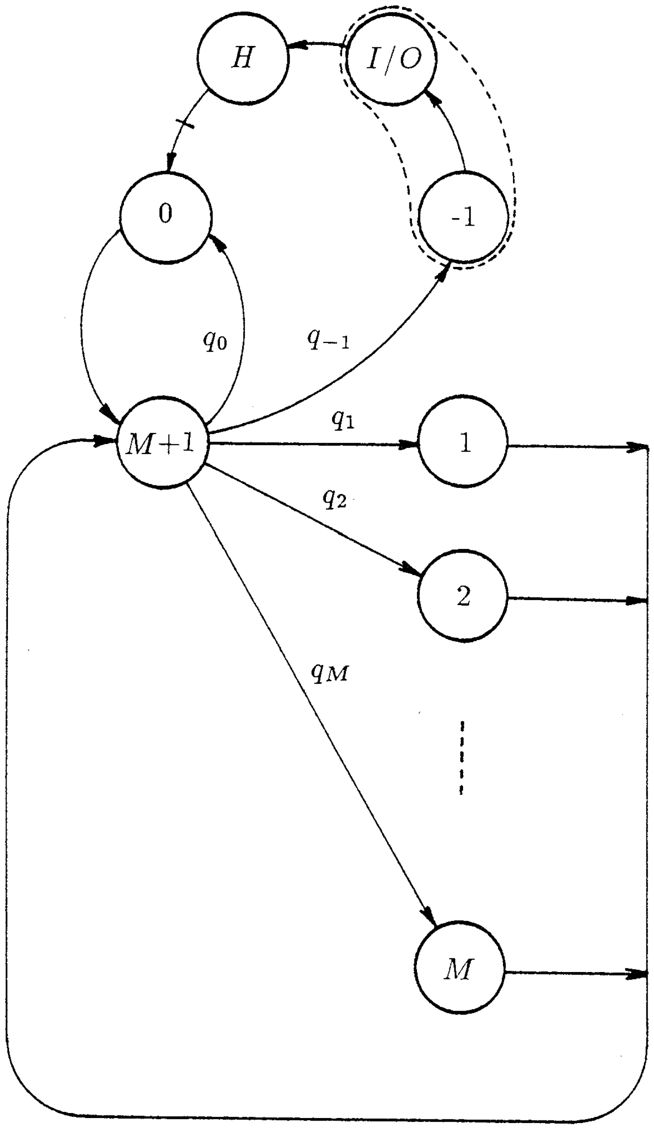

Figure 3. A simplified queueing network model of the FMS

following relative visit ratios, denoted by $q_i, i = -1, 0, \cdots, M+1$, also satisfy eqns (1a) through (1d) (e_0 corresponds to $q_0 + q_{-1}$ in this case):

(2a) $$q_i = \sum_{j=0}^{M} e_j \gamma(j,i)/\{\sum_{j=0}^{M} e_j\} = e_i/e_{M+1}, i = 1, 2, \cdots, M,$$

(2b) $$q_0 = \sum_{j=1}^{M} e_j \gamma(j,0)/\{\sum_{j=0}^{M} e_j\} = \{e_0 - e_{-1}\}/e_{M+1},$$

(2c) $$q_{-1} = \sum_{j=1}^{M} e_j \gamma(j,-1)/\{\sum_{j=0}^{M} e_j\} = e_{-1}/e_{M+1},$$

(2d) $$q_{M+1} = 1.$$

Using these ratios which satisfy $\sum_{i=-1}^{M} q_i = 1$, we can transform the original FMS model shown in Fig. 2 to the equivalent one shown in Fig. 3. This simplified CQN is helpful to intuitively understand the behavior of the FMS because the MHS (station $M+1$) is explicitly represented and there is a one-to-one relationship between the stations in the model and the physical system. In Fig. 2, there is no loop at each station, that is, the destination of each transportation through the MHS cannot be the just departing station. However, the simplified CQN shown in Fig. 3 permits a loop through the MHS at station $i, i = 0 \sim M$, that is, any job can return to the just departing station through the MHS, denoted by $M+1$ in Fig. 3. This is no problem for performing the equilibrium probability of any queue lengths and then the throughput of the FMS, because eqns (2a) through (2d) provide the relative visit ratios satisfying eqns (1a) through (1d); these equations hold for $e_0 = (q_0 + q_{-1})e_{M+1}$ and $e_i = q_i e_{M+1}, i \neq 0$. Similar CQN models for an FMS with no warehouse are presented by Solberg [14] and Nagasawa, Jeong and Nishiyama [5].

Since the I/O station has the same visit ratio as station -1 has and since station -1 has an infinite service rate, we combine these two stations to make a single station with service rate λ and visit ratio e_{-1}. The station is enclosed by the dashed line in Figs 2 and 3. The "blocked-and-lost" mechanism in the arrival process is completely represented by this I/O station.

The difficulty in obtaining the equilibrium probability of the queue lengths at each station lies in incorporating the "blocked-and-hold 0" mechanism at the warehouse. Unfortunately, we can not give an exact formulation of the equilibrium probability in this case but derive an approximation.

First, we consider the case $N_H = 0$, that is, there is no warehouse and no "blocked-and-hold 0" mechanism. The equilibrium probability of the queue lengths denoted by $P(\mathbf{n})$, where $\mathbf{n} \equiv (n_{-1}, n_0, \cdots, n_{M+1})$, is exactly derived as follows:

(3a)
$$P(\mathbf{n}) = \frac{1}{G(M+3, N, \mathrm{e}, \lambda)} \prod_{i=-1}^{M+1} h_i(n_i),$$

(3b)
$$G(M+3, N, \mathrm{e}, \lambda) \equiv \sum_{\sum_{i=-1}^{M+1} n_i = N} \sum_{i=-1}^{M+1} h_i(n_i),$$

(3c)
$$h_i(n_i) \equiv \frac{e_i}{\mu_i(n_i)} h_i(n_i - 1), h_i(0) \equiv 1, i = -1, 0, \cdots, M+1,$$

where $\mathbf{e} \equiv (e_{-1}, e_0, \cdots, e_{M+1})$ and $\mu_{-1}(k) \equiv \lambda \min(k, 1)$ for $k \geq 0$.

It should be noted that the e_i' s in these equations are relative visit ratios so that the efficient computational algorithms provided by Buzen [2] can be exploited for calculating the value of $G(M+3, N, \mathrm{e}, \lambda)$; Dubois [3], and Buzacott and Yao [1] use the unique value e_i^* (instead of e_i), which makes it hard to tune the value of e_i's in order to avoid overflow and computational error.

The throughput of station $i, TH_i(N, \lambda)$, the throughput of the FMS, $TH_{FMS}(N, \lambda)$, and the throughput of the CQN, $TH_{CQN}(N, \lambda)$ are derived as follows:

(4a)
$$TH_i(N, \lambda) \equiv \sum_{k=1}^{N} \mu_i(k) \mathrm{P}\{n_i = k\}$$

(4b)
$$= e_i \frac{G(M+3, N-1, \mathrm{e}, \lambda)}{G(M+3, N, \mathrm{e}, \lambda)}, i = -1, 0, \cdots, M+1,$$

$$TH_{FMS}(N,\lambda) \equiv \sum_{i=1}^{M} TH_i(N,\lambda)\gamma(i,-1) \tag{5a}$$

$$= TH_{-1}(N,\lambda) \tag{5b}$$

$$= e_{-1}\frac{G(M+3,N-1,\mathrm{e},\lambda)}{G(M+3,N,\mathrm{e},\lambda)}, \tag{5c}$$

$$TH_{CQN}(N,\lambda) \equiv \sum_{i=-1}^{M+1} TH_i(N,\lambda) \tag{6a}$$

$$= \{\sum_{i=-1}^{M+1} e_i\}\frac{G(M+3,N-1,\mathrm{e},\lambda)}{G(M+3,N,\mathrm{e},\lambda)} \tag{6b}$$

$$= \frac{G(M+3,N-1,\tilde{\mathrm{e}},\lambda)}{G(M+3,N,\tilde{\mathrm{e}},\lambda)}, \tag{6c}$$

where the relative visit ratios, denoted by e, are scaled as $\tilde{\mathrm{e}} \equiv (\tilde{e}_{-1}, \tilde{e}_0, \cdots, \tilde{e}_{M+1})$ and $\tilde{e}_i \equiv e_i/\{\sum_{j=-1}^{M+1} e_j\}$ so that $\sum_{i=-1}^{M+1} \tilde{e}_i = 1$ holds.

Therefore, we get·

$$TH_{FMS}(N,\lambda) = \tilde{e}_{-1} TH_{CQN}(N,\lambda), \tag{7}$$

that is, the throughput of the FMS is $\tilde{e}_{-1}$ times as large as the throughput of the CQN.

We shall provide another representation of the FMS model using the following function:

$$TH(n) \equiv \lim_{\lambda\to\infty} TH_{FMS}(n,\lambda) \tag{8a}$$

$$= e_{-1}\frac{G(M+2,n-1,\mathrm{e})}{G(M+2,n,\mathrm{e})}, \tag{8b}$$

where

$$G(M+2,n,\mathrm{e}) \equiv \sum_{\sum_{i=0}^{M+1} n_i = n} \prod_{i=0}^{M+1} h_i(n_i). \tag{8c}$$

The function $TH(n)$ denotes the throughput of the FMS provided that the population in the FMS is always n(in this case, $N_H = 0$ and therefore the population in the shop is always n). Using this function, we can simplify the FMS model as shown in Fig. 4, where stations in the FMS are aggregated into a single station with service rate governed by the expression of $\min\{TH(n),\ TH(N_S)\}$.

From the product form solution given by eqns (3a) through (3c), the equilibrium probability that the number of jobs in the FMS is exactly n, denoted by $P(n)$, is given by

$$P(n) = \begin{cases} \frac{\lambda}{TH(n)} P(n-1), & \text{if } n \leq N; \\ 0, & \text{if } N > n, \end{cases} \tag{9a}$$

$$P(0) = \left\{1 + \sum_{n=1}^{N} \prod_{k=1}^{n} \frac{\lambda}{TH(k)}\right\}^{-1}. \tag{9b}$$

Similar results have been derived by Buzacott and Yao [1] and Shanthikumar and Stecke [9]; the only difference is that they defined the function $TH(n)$ as the throughput of the CQN instead of the throughput of the FMS.

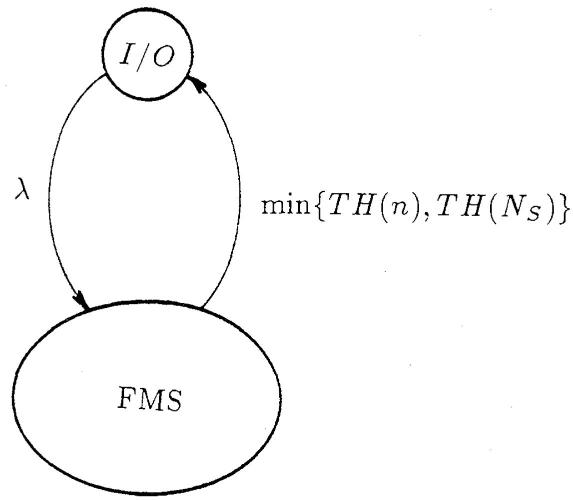

Figure 4. A simplified expression of the FMS with $N_H = 0$

Since $P(N)$ denotes the probability that the I/O station is idle, the throughput of the FMS is also given by

$$TH_{FMS}(N, \lambda) \equiv \lambda\{1 - P(N)\}, \tag{10}$$

which is equivalent to the expressions in eqns (5a) through (5c).

Let us consider the case $N_H > 0$, next. It is intractable to formulate the "blocked-and-hold 0" mechanism exactly. There are numerous approximation methods dealing with various blocking phenomena between each connected pair of stations with limited local buffer capacities [6] [7]. In our case, blocking occurs only at the warehouse when the population in the shop reaches the limit N_S and no blocking occurs at any station in the shop because of the unlimited local buffer capacities. This type of blocking has not been dealt with by any researcher.

We present the following approximation, denoted by $\tilde{P}(n)$, to get the equilibrium probability $P(n)$ for the case of $N_H > 0$:

(11a)
$$\tilde{P}(n) = \begin{cases} \frac{\lambda}{TH(n)}\tilde{P}(n-1), & \text{if } 1 < n \le N_S; \\ \frac{\lambda}{TH(N_S)}\tilde{P}(n-1), & \text{if } N_S < n \le N_S + N_H; \\ 0, & \text{if } N_S + N_H < n, \end{cases}$$

(11b)
$$\tilde{P}(0) = \left\{1 + \sum_{n=1}^{N_S-1}\prod_{k=1}^{n}\frac{\lambda}{TH(k)} + \left[\prod_{k=1}^{N_S}\frac{\lambda}{TH(k)}\right]\sum_{j=0}^{N_H}\left[\frac{\lambda}{TH(N_S)}\right]^j\right\}^{-1}.$$

Using $\tilde{P}(n)$, we define the approximate throughput of the FMS as

(12)
$$\widetilde{TH}_{FMS}(N_S, N_H, \lambda) \equiv \lambda\{1 - \tilde{P}(N_S + N_H)\}.$$

While these equations give the approximate values for the equilibrium probabilities and the throughput of the FMS, the exact values are also obtained through these equations in

Table 1. Accuracy of the approximation compared with the SIMAN simulation run ($\lambda/\{\sum_{i=1}^{M} c_i/\sum_{j=1}^{M}(e_j^*/\mu_j)\} = 0.7$)

$(M, \sum_{i=1}^{M} c_i)$	$(c_1, c_2, \ldots, c_M)$	(N_S, N_H)	SIMAN	CQN	Error
(1, 5)	(5)	(5, 0)	0.293	0.296	0.9
		(10,20)	0.347	0.350	0.9
		(15,30)	0.347	0.350	0.9
(1,10)	(10)	(10, 0)	0.641	0.645	0.7
		(15,20)	0.697	0.700	0.5
		(20,30)	0.697	0.700	0.5
(1,20)	(20)	(20, 0)	1.367	1.358	−0.6
		(30,20)	1.385	1.400	1.1
		(40,30)	1.385	1.400	1.1
(3, 5)	(1,2,2)	(5, 0)	0.260	0.259	−0.1
		(10,20)	0.348	0.349	0.3
		(15,30)	0.350	0.350	−0.1
(3,10)	(1,1,8)	(10, 0)	0.589	0.589	0.0
		(15,20)	0.701	0.699	−0.2
		(20,30)	0.699	0.700	0.1
(3,20)	(3,7,10)	(20, 0)	1.312	1.318	0.4
		(30,20)	1.390	1.400	0.7
		(40,30)	1.392	1.400	0.6
(5,10)	(1,1,1,1,6)	(10, 0)	0.548	0.549	0.2
		(15,20)	0.697	0.692	−0.7
		(20,30)	0.693	0.700	0.9
(5,20)	(1,2,3,4,10)	(20, 0)	1.261	1.263	0.1
		(30,20)	1.384	1.399	1.1
		(40,30)	1.407	1.400	−0.5
(5,20)	(4,4,4,4,4)	(20, 0)	1.271	1.276	0.4
		(30,20)	1.408	1.400	−0.6
		(40,30)	1.390	1.400	0.7
(10,10)	(1,1,1,1,1,1,1, 1,1,1)	(10, 0)	0.473	0.475	0.5
		(15,15)	0.620	0.618	−0.2
		(20,30)	0.677	0.677	0.0
(10,20)	(1,1,1,1,1,1,1, 1,1,11)	(20, 0)	1.119	1.124	0.4
		(30,20)	1.377	1.376	−0.1
		(40,20)	1.388	1.398	0.7
(10,20)	(1,1,1,2,2,2,2, 3,3,3)	(20,0)	1.156	1.155	−0.1
		(30,20)	1.386	1.383	−0.2
		(40,20)	1.398	1.398	0.0

Error≡100 × (CQN−SIMAN)/SIMAN [%]

SIMAN : the throughput of the FMS obtained by the simulation run

CQN : the approximate throughput of the FMS calculated using eqn.(12)

Table 2. Accuracy of the approximation compared with the SIMAN simulation run ($\lambda/\{\sum_{i=1}^{M} c_i / \sum_{j=1}^{M}(e_j^*/\mu_j)\} = 0.9$)

$(M, \sum_{i=1}^{M} c_i)$	$(c_1, c_2, \ldots, c_M)$	(N_S, N_H)	SIMAN	CQN	Error
		(5, 0)	0.340	0.341	0.3
(1, 5)	(5)	(10,20)	0.445	0.447	0.6
		(15,30)	0.445	0.449	0.9
		(10, 0)	0.743	0.749	0.8
(1,10)	(10)	(15,20)	0.887	0.895	1.0
		(20,30)	0.900	0.899	−0.1
		(20, 0)	1.589	1.603	0.9
(1,20)	(20)	(30,20)	1.777	1.796	1.1
		(40,30)	1.775	1.799	1.4
		(5, 0)	0.283	0.283	0.0
(3, 5)	(1,2,2)	(10,20)	0.406	0.406	0.0
		(15,30)	0.435	0.432	−0.7
		(10, 0)	0.648	0.646	−0.3
(3,10)	(1,1,8)	(15,20)	0.832	0.829	−0.4
		(20,30)	0.877	0.870	−0.8
		(20, 0)	1.481	1.483	0.1
(3,20)	(3,7,10)	(30,20)	1.761	1.750	−0.7
		(40,30)	1.793	1.788	−0.3
		(10, 0)	0.586	0.587	0.2
(5,10)	(1,1,1,1,6)	(15,20)	0.749	0.751	0.2
		(20,30)	0.808	0.808	0.0
		(20, 0)	1.369	1.374	0.4
(5,20)	(1,2,3,4,10)	(30,20)	1.672	1.663	−0.5
		(40,30)	1.751	1.744	−0.4
		(20, 0)	1.392	1.395	0.3
(5,20)	(4,4,4,4,4)	(30,20)	1.674	1.671	−0.2
		(40,30)	1.756	1.748	−0.5
		(10, 0)	0.494	0.495	0.2
(10,10)	(1, 1, 1, 1, 1, 1, 1,	(15,15)	0.623	0.625	0.4
	1, 1, 1)	(20,30)	0.688	0.690	0.2
		(20, 0)	1.174	1.176	0.2
(10,20)	(1, 1, 1, 1, 1, 1, 1,	(30,20)	1.447	1.452	0.4
	1, 1, 11)	(40,20)	1.569	1.574	0.3
		(20, 0)	1.212	1.213	0.0
(10,20)	(1, 1, 1, 2, 2, 2, 2,	(30,20)	1.465	1.476	0.7
	3, 3, 3)	(40,20)	1.587	1.589	0.1

Error≡100 × (CQN−SIMAN)/SIMAN [%]

SIMAN : the throughput of the FMS obtained by the simulation run

CQN : the approximate throughput of the FMS calculated using eqn.(12)

Table 3. Accuracy of the approximation compared with the SIMAN simulation run ($\lambda/\{\sum_{i=1}^{M} c_i/\sum_{j=1}^{M}(e_j^*/\mu_j)\} = 1.0$)

$(M, \sum_{i=1}^{M} c_i)$	$(c_1, c_2, \ldots, c_M)$	(N_S, N_H)	SIMAN	CQN	Error
(1, 5)	(5)	(5, 0)	0.353	0.358	1.2
		(10,20)	0.478	0.482	1.0
		(15,30)	0.485	0.489	0.6
(1,10)	(10)	(10, 0)	0.778	0.785	1.0
		(15,20)	0.966	0.966	0.0
		(20,30)	0.966	0.978	1.2
(1,20)	(20)	(20, 0)	1.666	1.682	1.0
		(30,20)	1.931	1.945	0.7
		(40,30)	1.946	1.964	0.9
(3, 5)	(1,2,2)	(5, 0)	0.289	0.291	0.7
		(10,20)	0.409	0.409	−0.1
		(15,30)	0.436	0.437	0.3
(3,10)	(1,1,8)	(10, 0)	0.665	0.663	−0.3
		(15,20)	0.840	0.838	−0.2
		(20,30)	0.875	0.884	1.0
(3,20)	(3,7,10)	(20, 0)	1.532	1.524	−0.5
		(30,20)	1.805	1.800	−0.3
		(40,30)	1.851	1.864	0.7
(5,10)	(1,1,1,1,6)	(10, 0)	0.599	0.598	−0.2
		(15,20)	0.751	0.752	0.1
		(20,30)	0.808	0.810	0.2
(5,20)	(1,2,3,4,10)	(20, 0)	1.397	1.400	0.3
		(30,20)	1.669	1.677	0.5
		(40,30)	1.769	1.768	−0.1
(5,20)	(4,4,4,4,4)	(20, 0)	1.418	1.422	0.3
		(30,20)	1.681	1.686	0.3
		(40,30)	1.766	1.773	0.4
(10,10)	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)	(10, 0)	0.499	0.500	0.2
		(15,15)	0.625	0.625	−0.1
		(20,30)	0.687	0.690	0.4
(10,20)	(1, 1, 1, 1, 1, 1, 1, 1, 1, 11)	(20, 0)	1.182	1.188	0.5
		(30,20)	1.443	1.452	0.7
		(40,20)	1.577	1.577	0.0
(10,20)	(1, 1, 1, 2, 2, 2, 2, 3, 3, 3)	(20,0)	1.221	1.226	0.3
		(30,20)	1.473	1.476	0.2
		(40,20)	1.590	1.592	0.1

Error≡100 × (CQN−SIMAN)/SIMAN [%]

SIMAN : the throughput of the FMS obtained by the simulation run

CQN : the approximate throughput of the FMS calculated using eqn.(12)

Figure 1. A typical FMS with a supplementary warehouse

the following cases: (1) the FMS model which consists of only one machining station with rate $\mu_i(n) = \min\{\mu_i(n), \mu_i(N_S)\}$, provided that a loading/unloading station and a MHS have infinite service rates; (2) the FMS model in which the number of available pallets N_S is infinite (or very large); (3) the FMS model with $N_H = 0$ (no supplementary warehouse).

It is hard to investigate the accuracy of this approximation for all cases of the FMS. Instead, we provide some comparison with the SIMAN simulation run, where the conditions are the same as given in this section except that the process route of arriving jobs are previously given randomly so that each job visits all stations only once (therefore, $e_i^* = 1, i = 1 \sim M$). We set $\mu_0(n) = \mu_{M+1}(n) = \infty, \mu_{-1}(n) = \lambda \min\{n, 1\}$ and $\mu_i(n) = \mu_i \min\{n, c_i\}, i = 1 \sim M$, where $e_i^*/(\mu_i c_i) = e_j^*/(\mu_i c_j), i \neq j, i, j = 1 \sim M$, for balancing workload. The arrival rate λ takes the following values: $\lambda / \{\sum_{i=1}^{M} c_i / \sum_{j=1}^{M} (e_j^*/\mu_j)\} = 0.7,\ 0.9$ and 1.0, where the term $\sum_{i=1}^{M} c_i / \sum_{j=1}^{M} (e_j^*/\mu_j)$ represents the limit value of the throughput of the FMS because the λ^* defined in Theorem 1 can not exceed the value of $\sum_{i=1}^{M} c_i / \sum_{j=1}^{M} (e_j^*/\mu_j)$ for any configuration of c_i and for any loading e_i^*/μ_i, given the total number of machines $\sum_{i=1}^{M} c_i$ and the total work load $\sum_{i=1}^{M} (e_i^*/\mu_i)$. The throughput of the FMS is calculated through the SIMAN simulation run by dividing 10,000 [jobs] by the time interval between the arrival of the 1,001st job and the departure of the 11,000th job. Tables 1 through 3 summarize the results, making it clear that the approximation gives the precise values (only 1.2% error at most) for the throughput of the FMS with balanced loading. We also investigated the accuracy for some cases of the FMS with unbalanced loading. The results, omitted here, are similar to the case of balanced loading (we cannot generalize the results because we can not perform all cases of the FMS).

Similar approximation methods have been presented by Shanthikumar and Stecke [9] for the case $N_H = \infty$ and suggested by Yao [15] and Yao and Shanthikumar [17] for the case $N_H > 0$.

3. Properties of the throughput

We derive properties of the throughput of the FMS given in section 2. These properties are exploited for making optimal design as will be shown later.

Theorem 1. *Consider the FMS model with the warehouse defined in section 2. Suppose service rate of each station satisfies $\mu_i(n) = \mu_i(c_i)$ for $n \geq c_i$ and $\mu_i(n) \leq \mu_i(c_i)$ for $n < c_i, i = 0, 1, \cdots M+1$. Then, the approximate throughput of the FMS defined in eqn. (12) has the following limits:*

$$\text{(i)} \lim_{\lambda \to \infty} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) = TH(NS);$$

$$\text{(ii)} \lim_{N_S \to \infty} \lim_{\lambda \to \infty} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) = \lambda^*;$$

$$\text{(iii)} \lim_{N_S \to \infty} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) = \min(\lambda, \lambda^*);$$

$$\text{(iv)} \lim_{N_H \to \infty} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) = \min\{\lambda, TH(N_S)\},$$

where $e_i^*, i = -1, 0, \cdots, M+1$, *is the unique solution to eqns (1a) through (1e) and*

$$\lambda^* \equiv \min_{0 \le i \le M+1} \{\mu_i(c_i)/e_i^*\}.$$

(Proof)

Property (i) holds because $\tilde{P}(N_S + N_H) \to 1$ and $\tilde{P}(n) \to 0 (n \ne N_S + N_H)$ as $\lambda \to \infty$, that is, there are always N_S jobs in the shop when λ is very large.

Substituting property (i) into property (ii) yields

$$\lim_{N_S \to \infty} TH(N_S) = \lambda^*,$$

or equivalently,

$$\lim_{N_S \to \infty} e_{-1}^* \frac{G(M+2, N_S - 1, \mathrm{e}^*)}{G(M+2, N_S, \mathrm{e}^*)} = \min_{0 \le i \le M+1} \{\frac{\mu_i(c_i)}{e_i^*}\}$$

where $e_{-1}^* = 1$ (Theorem A in appendices shows that this equation is valid).

From the definitions of $P(n)$ and $\tilde{P}(n)$, we have, for any λ,

$$\lim_{N_S \to \infty} \tilde{P}(n) = \lim_{N_S \to \infty} P(n),$$

and then

$$\lim_{N_S \to \infty} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) = \lim_{N_S \to \infty} TH_{FMS}(N_S, \lambda).$$

It is noticeable that the limit of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ with respect to N_S does not depend on N_H.

Therefore, to prove property (iii), it suffices to show

$$\lim_{N_S \to \infty} TH_{FMS}(N_S, \lambda) = \min(\lambda, \lambda^*).$$

Consider the equivalent CQN without warehouse where station -1 has service rate λ and the expected visit times e_{-1}^*. Applying Theorem A to the equivalent CQN yields

$$\begin{aligned} TH_{FMS}(N_S, \lambda) &= e_{-1}^* \frac{G(M+3, N_S - 1, \mathrm{e}^*, \lambda)}{G(M+3, N_S, \mathrm{e}^*, \lambda)} \\ &\to \min\left\{\frac{\lambda}{e_{-1}^*}, \min_{0 \le i \le M+1} \frac{\mu_i(c_i)}{e_i^*}\right\} = \min(\lambda, \lambda^*) \text{ as } N_S \to \infty. \end{aligned}$$

To prove (iv), from the definitions of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$, $\tilde{P}(n)$ and $TH(n)$, we get

$$\begin{aligned} \widetilde{TH}_{FMS}(N_S, N_H, \lambda) &= \lambda \sum_{n=0}^{N_S + N_H - 1} \tilde{P}(n) \\ &= \frac{\sum_{n=0}^{N_S - 1} \lambda^{n+1} G(M+2, n, \mathrm{e}^*) + \lambda^{N_S + 1} G(M+2, N_S, \mathrm{e}^*) \sum_{j=0}^{N_H - 1} \rho_S^j}{\sum_{n=0}^{N_S - 1} \lambda^n G(M+2, n, \mathrm{e}^*) + \lambda^{N_S} G(M+2, N_S, \mathrm{e}^*) \sum_{j=0}^{N_H} \rho_S^j} \end{aligned}$$

where $\rho_S = \lambda / TH(N_S)$.

If $\rho_S = 1$, then $\widetilde{TH}_{FMS}(N_S, N_H, \lambda) \to \lambda (= TH(N_S))$ as $N_H \to \infty$;

$$\text{if } \rho_S < 1, \text{ then } \widetilde{TH}_{FMS}(N_S, N_H, \lambda) \to \lambda (< TH(N_S)) \text{ as } N_H \to \infty;$$

$$\text{if } \rho_S > 1, \text{ then } \widetilde{TH}_{FMS}(N_S, N_H, \lambda) \to \lambda/\rho_S (= TH(N_S) < \lambda) \text{ as } N_H \to \infty.$$

Therefore, for any ρ_S, $\widetilde{TH}_{FMS}(N_S, N_H, \lambda) \to \min\{\lambda, TH(N_S)\}$ as $N_H \to \infty$. Q. E. D.

Although the exact throughput of the FMS is not analytically tractable, considering the practical and physical meaning of these limits, we can verify that the exact throughput of the FMS approaches these limits shown in Theorem 1. It is therefore important to note that both the approximate and exact throughputs of the FMS have the same limits as given in Theorem 1. This remark implies that the approximation provides a precise measure for the throughput of the FMS if at least one of the following conditions holds: (1) very large arrival rate; (2) a sufficiently large amount of pallets available in the shop; or (3) sufficiently large capacity of the warehouse.

The limiting behavior of the throughput with respect to N_S and N_H can be determined using the first- and second-order properties given as follows:

Theorem 2. *Consider the approximate FMS model defined in section 2. Suppose that service rate of each station, $\mu_i(n_i)$, is nondecreasing and concave in the queue length $n_i, i = 0, 1, \cdots, M+1$. Then,*
(i) $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ is nondecreasing concave in N_H;
(ii) $\widetilde{TH}_{FMS}(N_S, N - N_S, \lambda)$ is nondecreasing concave in N_S;
(iii) $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ is nondecreasing in N_S.

(Proof)

Consider a CQN which consists of two nodes. Node 1 is the approximate FMS with mean service rate $\gamma(n)$ such that $\gamma(n) = \min\{TH(n), TH(N_S)\}$ for any $n \geq 0$. Node 2 is the I/0 node with service rate λ. This two-node CQN represents the approximate CQN with $N_H > 0$, where the equilibrium probability is defined by eqns (11a) and (11b).

Consider, first, two networks ($p = 1, 2$) such that $N_S^1 = N_S^2$ and $N_H^1 < N_H^2 = N_H^1 + 1$, that is, $N^1 = N_S^1 + N_H^1 < N_S^2 + N_H^2 = N^2 = N^1 + 1$. The queue-length dependent service rates at node 1 in these networks are the same, i. e., $\gamma^1(n) = \gamma^2(n) = \gamma(n)$, because $N_S^1 = N_S^2$. The two networks differ only in job population.

Since all stations in the FMS have service rates, $\mu_i(n), i = 0, 1, \cdots, M+1$, which are nondecreasing concave in n, the throughput of the FMS with population n, denoted by $TH(n)$, is also nondecreasing concave in n as shown in Shanthikumar and Yao [12] [13]. That is, $\gamma(n)$ is nondecreasing concave in n. Hence, the throughput of the CQN under discussion is also nondecreasing concave in N.

The throughputs of nodes 1 and 2 in the CQN are the same and equal to a half of the throughput of the CQN – "throughput of a CQN" is usually defined as the sum of throughputs of all nodes included in the CQN like the definition given in eqn. (6a). Therefore, the throughput of node 1 in the CQN, corresponding to $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$, is nondecreasing concave in N_H.

To prove (ii), consider two networks ($p = 1, 2$) such that $N_S^1 < N_S^2 = N_S^1 + 1$ and $N_H^1 > N_H^2 = N_H^1 - 1$, that is, $N^1 = N_S^1 + N_H^1 = N_S^2 + N_H^2 = N^2$. These networks differ from each other only in the queue-length dependent service rates. If we set $TH(n) = n\mu$, that is, $\gamma(n) = \mu \min(n, N_S)$, then we can find that the throughput of the CQN, and hence $\widetilde{TH}_{FMS}(N_S, N - N_S, \lambda)$ is nondecreasing concave in N_S as shown in Shanthikumar and Yao [11]. Based on the proof of this special case, we can derive the same result for the general case $\gamma(n) = \min\{TH(n), TH(N_S)\}$. The proof is given in Theorem B in appendices.

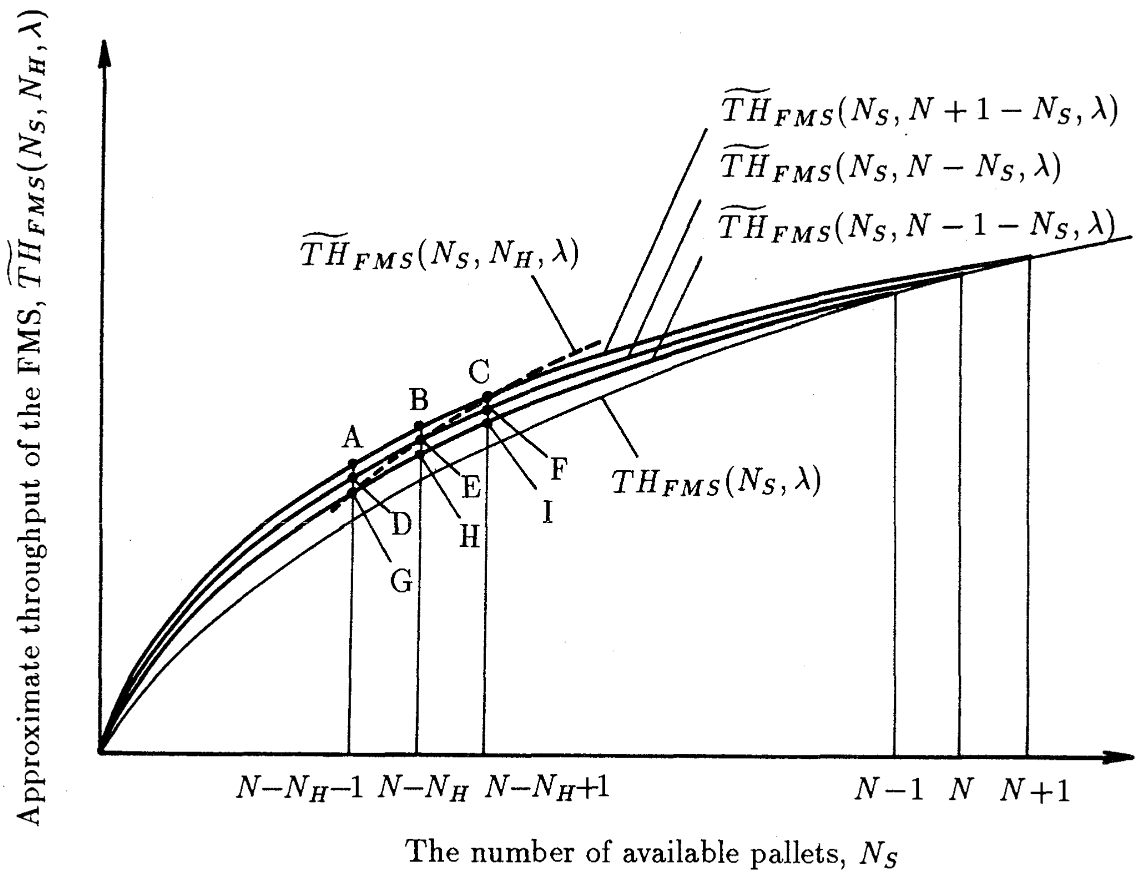

Figure 5. Illustration for explaining the second order property of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ with respect to N_S

Property (iii) is directly obtained from (i) and (ii). That is,

$$\widetilde{TH}_{FMS}(N_S, N_H, \lambda) \leq \widetilde{TH}_{FMS}(N_S+1, N_H-1, \lambda) \leq \widetilde{TH}_{FMS}(N_S+1, N_H, \lambda),$$

where the first and second inequalities are due to properties (ii) and (i), respectively.

Q. E. D.

We can show that the concavity of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ in N_S is not likely to hold. As illustrated in Fig. 5, the concavity is expressed by the concavity of the curve $G-E-C$. From Theorem 2 (i) and (ii), the curves $A-B-C$, $D-E-F$ and $G-H-I$ are all concave, and $\overline{AD} \leq \overline{DG}, \overline{BE} \leq \overline{EH}$ and $\overline{CF} \leq \overline{FI}$. Therefore, if $\overline{AD} \geq \overline{BE} \geq \overline{CF}$ holds, the curve $G-E-C$ becomes concave. If N_H is small enough and N_S is large enough, then the relation $\overline{AD} \geq \overline{BE} \geq \overline{CF}$, and therefore, the concavity of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ in N_S is likely to hold. However, if N_H is very large and N_S is very small, the relation does not hold any more.

Combining Theorems 1 and 2 yields the following corollary:

Corollary 1. *Consider the approximate FMS model with the warehouse defined in section 2. Suppose that service rate $\mu_i(n_i)$ is nondecreasing concave in n_i and $\mu_i(n_i) = \mu_i(c_i)$ for $n_i > c_i, i = 0, 1, \cdots, M+1$. Then,*

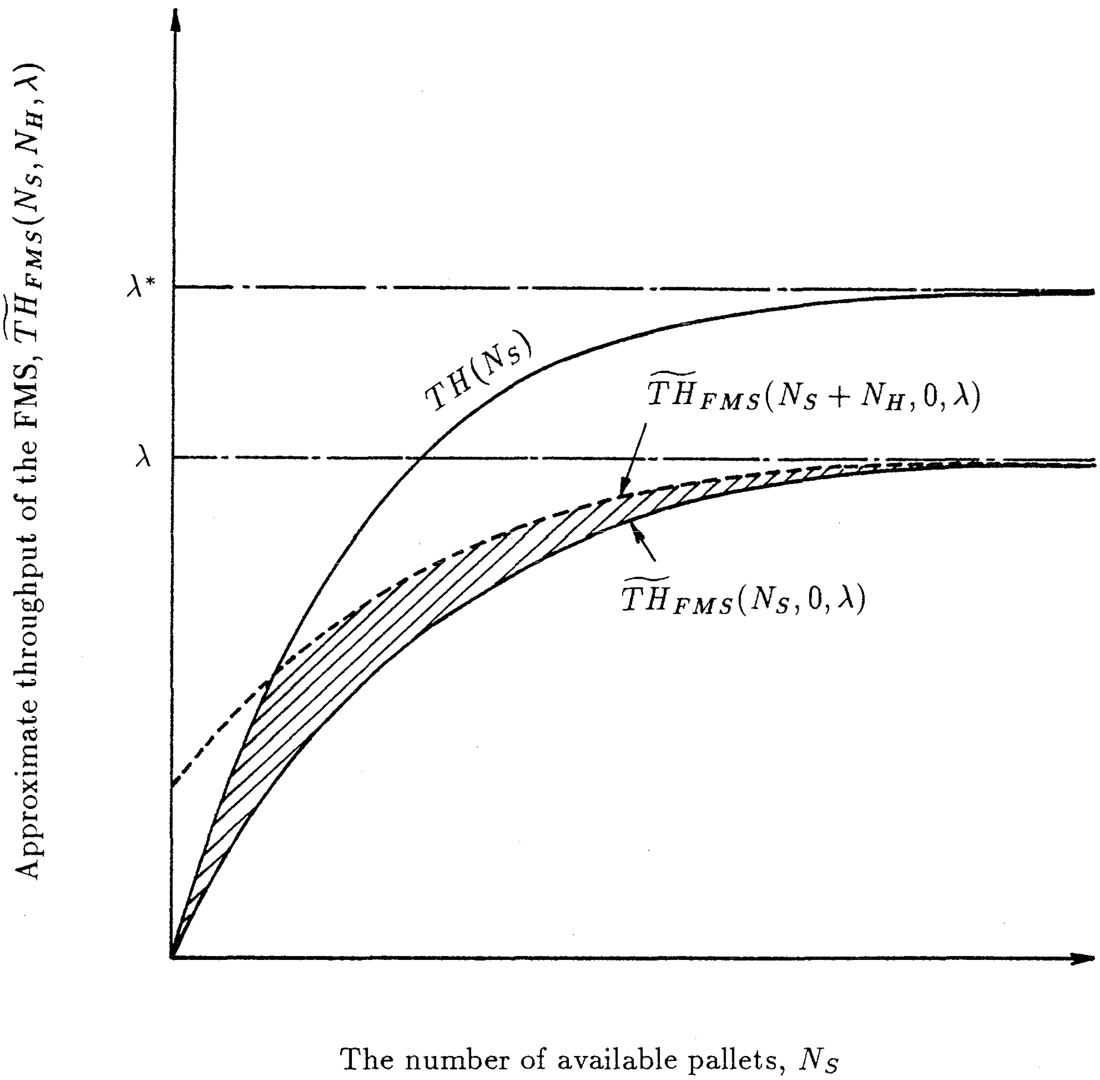

Figure 6. Relationships between the approximate throughput of the FMS and its bounds ($\lambda^* \geq \lambda$)

(i) $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ *is nondecreasing concave in* N_H *and converges to* $\min\{\lambda, TH(N_S)\}$ *as* $N_H \to \infty$*;*

(ii) $\widetilde{TH}_{FMS}(N_S, 0, \lambda) \leq \widetilde{TH}_{FMS}(N_S, N_H, \lambda) \leq \min\{\widetilde{TH}_{FMS}(N_S + N_H, 0, \lambda), TH(N_S)\}$*;*

(iii) $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ *is nondecreasing in* N_S *and converges to* $\min(\lambda, \lambda^*)$ *where* $\lambda^* \equiv \min_{0 \leq i \leq M+1}\{\mu_i(c_i)/e_i^*\}$.

The relationship between the throughput of the FMS and its bounds in illustrated in Figs 6 and 7. In these figures, the horizontal and vertical axes represent the number of pallets available in the shop and the throughput of the FMS, respectively. $\widetilde{TH}_{FMS}(N_S + N_H, 0, \lambda)$, expressed by a dotted curve, is obtained by translating the curve of $\widetilde{TH}_{FMS}(N_S, 0, \lambda)$ (or equivalently, $TH_{FMS}(N_S, \lambda)$) to the left by the quantity N_H. The hatched area represents the region where the value of $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ can be determined (as specified in Corollary 3 (ii)).

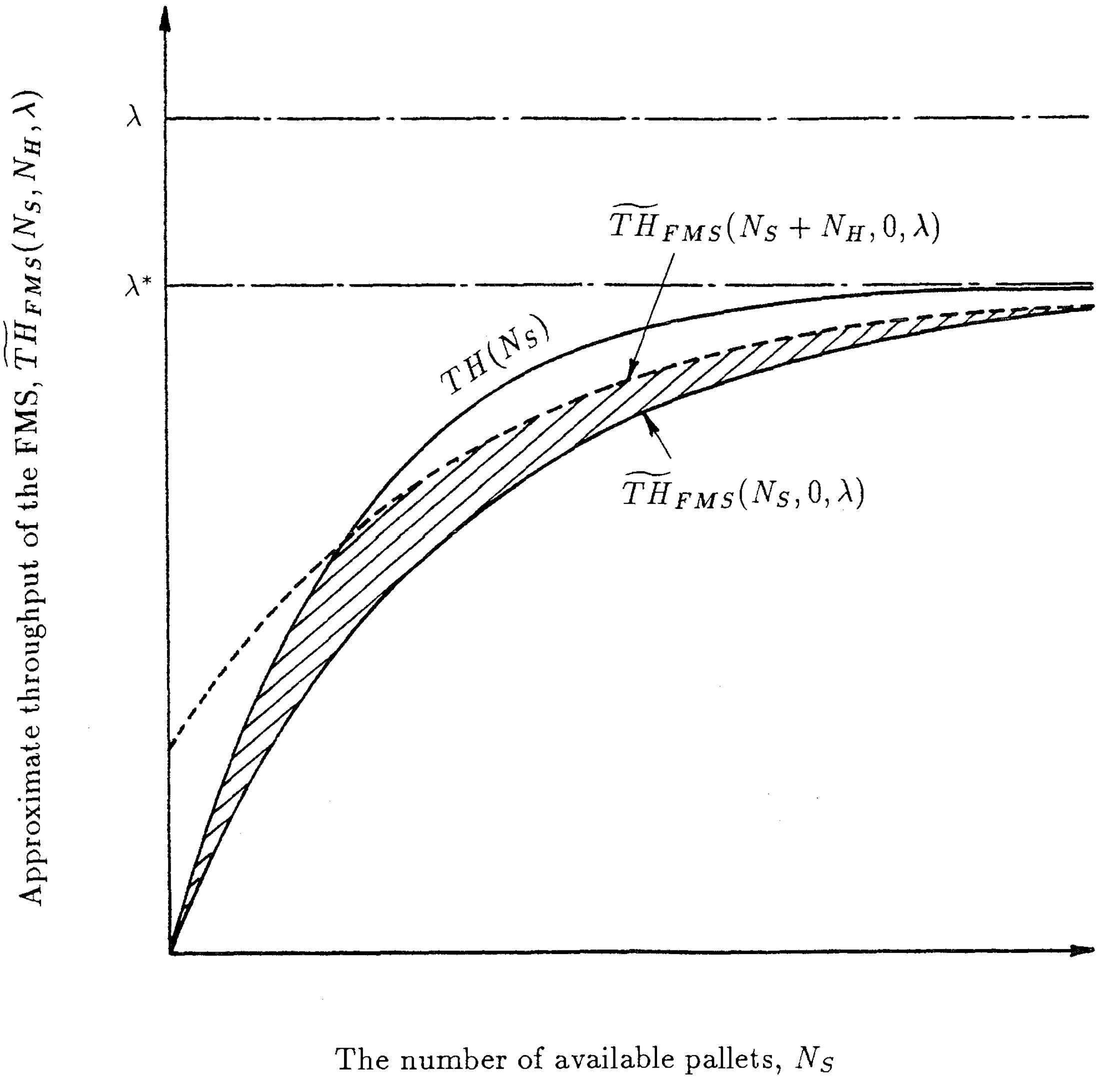

Figure 7. Relationships between the approximate throughput of the FMS and its bounds ($\lambda^* < \lambda$)

4. Capacity design problems

We provide solution methods to a few examples of capacity design problems for determining warehouse capacity, N_H, and the number of pallets available in the shop, N_S. Suppose that the service times of both the loading/unloading station and the MHS are small enough to have a negligible effect on the throughput of the FMS. Machining station i has c_i-parallel servers (machines) with queue-length dependent service rate $\mu_i(n) = \mu \min(n, c_i), i = 1 \sim M$.

We consider the two cases of the FMS: a single-station FMS with one machining station and a multiple-station FMS with several machining stations. The equilibrium probability given by eqns (11a) and (11b) is.approximation for the multiple-station FMS but exact for the single-station FMS. Therefore, the following differences between the single- and multiple-station FMSs are important in developing the solution methods: (1) The throughput of the FMS defined by $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ is exact for the single-station FMS but approximate for the multiple-station FMS; (2) The function $TH(n)$ defined by eqns (8a) and (8b) does not have any break point except at $n = N_S$ in the multiple-station FMS while in the single-station FMS the function $TH(n) = \mu\min(n,c)$ has the break point at $n = c$ and $TH(n) = TH(c)$ for any $n > c$ if $c < N_S$, where c denotes the number of parallel servers at the machining station in the single-station FMS.

In both cases, Theorems 1 and 2 and Corollary 1 are very useful for obtaining the optimal solution efficiently.

4.1 *A single-station FMS*

Consider first the single-station FMS. Note that $\tilde{P}(n)$ given by eqns (11a) and (11b) provide the exact value for $P(n)$ in this case. Substituting $TH(n) = \mu\min(n,c)$ to eqns (11a) and (11b), we get the exact expression for $P(n)$ as

$$\text{(13a)} \qquad P(n) = \begin{cases} \frac{\lambda}{\mu n} P(n-1), & \text{if } 0 \le n \le \min(N_S, c); \\ \frac{\lambda}{\mu \min(N_S,c)} P(n-1), & \text{if } \min(N_S,c) \le n \le N_S + N_H; \\ 0, & \text{if } N_S + N_H < n, \end{cases}$$

$$\text{(13b)} \qquad P(0) = \left\{ \sum_{n=0}^{N_c-1} \frac{\rho^n}{n!} + \frac{\rho^{N_c}}{N_c!} \sum_{k=0}^{N_S+N_H-N_c} \left(\frac{\rho}{N_c}\right)^k \right\}^{-1},$$

where $N_c = \min(N_S, c)$ and $\rho = \lambda/\mu$.

Then, the exact expression for the throughput of the FMS is also obtained as

$$\text{(14)} \qquad \begin{aligned} TH_{FMS}(N_S, N_H, \lambda) &\equiv \lambda\{1 - P(N_S + N_H)\} \\ &= \lambda \frac{\sum_{n=0}^{N_c-1} \rho^n/n! + (\rho^{N_c}/N_c!) \sum_{k=0}^{N_S+N_H-N_c-1} (\rho/N_c)^k}{\sum_{n=0}^{N_c-1} \rho^n/n! + (\rho^{N_c}/N_c!) \sum_{k=0}^{N_S+N_H-N_c} (\rho/N_c)^k}. \end{aligned}$$

Note that in this case we use $TH_{FMS}(N_S, N_H, \lambda)$ instead of using $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$ because the eqn. (12) gives the exact values for the throughput of the FMS. It is obvious that the throughput of the FMS expressed by this $TH_{FMS}(N_S, N_H, \lambda)$ satisfies Theorems 1 and 2 and Corollary 1. From Corollary 1 (ii), the lower and upper bounds of the throughput are obtained as follows:

$$\text{(15)} \qquad \begin{aligned} TH_{FMS}(N_S, 0, \lambda) &\le TH_{FMS}(N_S, N_H, \lambda) \\ &\le \min\{TH_{FMS}(N_S + N_H, 0, \lambda), \mu N_c\}. \end{aligned}$$

In this single-station FMS, we consider the following design problems:

P1 : $\max_{N_S, N_H} \{TH_{FMS}(N_S, N_H, \lambda) \mid N_S + N_H \le N \text{ and } N_S \ge 0, N_H \ge 0\}$,

P2 : $\max_{N_S, N_H} \{TH_{FMS}(N_S, N_H, \lambda) \mid aN_S + bN_H \le A \text{ and } N_S \ge 0, N_H \ge 0\}$.

Constraints in these two design problems represents the two kinds of tradeoff between the warehouse capacity and the number of available pallets. The constraint in the first problem means the space tradeoff under the fixed total spaces N, and the constraint in the second one, the cost tradeoff under the fixed total available investment A for installing the spaces.

On the basis of Theorems 1 and 2 and Corollary 1, we provide solution methods to these problems below.

Solution to **P1**

Since $TH_{FMS}(N_S, N_H, \lambda)$ is nondecreasing in both N_S and N_H from Theorems 2(i) and 2(iii), the optimal values for N_S and N_H, denoted by N_S^* and N_H^*, respectively, satisfy $N_S^* + N_H^* = N$. Hence, we can transform the original problem **P1** to

$$\mathbf{P1\prime} : \max_{N_S}\{TH_{FMS}(N_S, N - N_S, \lambda) \mid 0 \le N_S \le N\}.$$

From Theorem 2(ii), $N_S^* = N$ is clearly one of the optimal solutions to **P1′** and then **P1**. On the other hand, when $c \le N_S \le N$, substituting $N_c = c$ to eqn. (14), we get for any $N_S \in [c, N]$

$$TH_{FMS}(N_S, N_H, \lambda) = \lambda \frac{\sum_{n=0}^{c-1} \rho^n/n! + (\rho^c/c!) \sum_{k=0}^{N_S+N_H-c-1} (\rho/c)^k}{\sum_{n=0}^{c-1} \rho^n/n! + (\rho^c/c!) \sum_{k=0}^{N_S+N_H-c} (\rho/c)^k} \tag{16a}$$

$$= TH_{FMS}(N_S + N_H, 0, \lambda). \tag{16b}$$

This equation implies that any $N_S \in [c, N]$ is also optimal.

If $N < c$, then $N_S < c$. Substituting $N_c = N_S$ to eqn. (14), we get

$$TH_{FMS}(N_S, N_H, \lambda) = \lambda \frac{\sum_{n=0}^{N_S-1} \rho^n/n! + (\rho^{N_S}/N_S!) \sum_{k=0}^{N_H-1} (\rho/N_S)^k}{\sum_{n=0}^{N_S-1} \rho^n/n! + (\rho^{N_s}/N_s!) \sum_{k=0}^{N_H} (\rho/N_S)^k}. \tag{17}$$

This equation shows that the useful relation represented by eqn. (16b) does not hold in this case. We know, however, that $TH_{FMS}(N_S, N - N_S, \lambda)$ is nondecreasing in N_S. Therefore, we get $N_S^* = N$.

In conclusion, we get the optimal solution to **P1**, denoted by N_S^* and N_H^*, as any N_S and N_H satisfying

$$\min(c, N) \le N_S \le N \text{ and } N_H = N - N_S.$$

Solution to **P2**

Using an approach similar to that used for solving **P1**, we get $N_H^* = \lfloor (A - aN_S^*)/b \rfloor$ where $\lfloor x \rfloor$ denotes the maximal integer being less than or equal to x. Hence, the original problem can be transformed to

$$\mathbf{P2\prime} : \max_{N_S}\{TH_{FMS}(N_S, \lfloor (A - aN_S)/b \rfloor, \lambda) \mid 0 \le N_S \le \lfloor A/a \rfloor\}.$$

There are three possible cases:

(i) $a = b$

In this case, the constraint $aN_S + bN_H \le A$ is equivalent to $N_S + N_H \le \lfloor A/a \rfloor$. From the results obtained for **P1** we get N_S^* and N_H^* as any N_S satisfying

$$\min(c, N) \le N_S \le N \text{ and } N_H = N - N_S, \text{ where } N = \lfloor A/a \rfloor.$$

(ii) $a < b$

In this case, $N_S + N_H = N_S + \lfloor (A - aN_S)/b \rfloor \le N = \lfloor A/a \rfloor$ and $N_S = N$ (and $N_H = 0$) is a feasible solution. From the upper bound of the throughput given in eqn. (15) and the nondecreasing property of $TH_{FMS}(N_S, 0, \lambda)$ in N_S, we get $TH_{FMS}(N_S, \lfloor (A - aN_S)/b \rfloor, \lambda) \le TH_{FMS}(N_S + \lfloor (A - aN_S)/b \rfloor, 0, \lambda) \le TH_{FMS}(N, 0, \lambda)$, the bounds being satisfied with equality when $N_S = N$ (and $N_H = 0$). Therefore, we get $N_S^* = N$ and $N_H^* = 0$.

(iii) $a > b$

In this case, the total space $N_S + N_H$ is maximized when $N_H = \lfloor A/b \rfloor$ and $N_S = 0$. However, $N_S = 0$ implies that the throughput of the FMS is zero. On the other hand, while the maximal value of N_S is $\lfloor A/a \rfloor (\leq \lfloor A/b \rfloor)$ and then $N_H = 0$, the total space $N_S + N_H$ can be increased by sacrificing the space of pallets N_S. This fact implies that the throughput of the FMS may be increased in this manner.

Consider the case $c \leq \lfloor A/a \rfloor$ first.

For $c \leq N_S \leq \lfloor A/a \rfloor$, eqn. (16b) holds and $N_S + N_H \leq c + \lfloor (A - ac)/b \rfloor$ with equality when $N_S = c$ and $N_H = \lfloor (A - ac)/b \rfloor$. Hence, we get

$$\begin{aligned} TH_{FMS}(N_S, \lfloor (A - aN_S)/b \rfloor, \lambda) &= TH_{FMS}(N_S + \lfloor (A - aN_S)/b \rfloor, 0, \lambda) \\ &\leq TH_{FMS}(c + \lfloor (A - ac)/b \rfloor, 0, \lambda) \end{aligned}$$

with equality when $N_S = c$ and $N_H = \lfloor (A - ac)/b \rfloor$.

For $0 \leq N_S < c$, we do not know whether the throughput $TH_{FMS}(N_S, \lfloor (A - aN_S)/b \rfloor, \lambda)$ is less than $TH_{FMS}(c, \lfloor (A - ac)/b \rfloor, \lambda)$. However, from eqn. (15) we get for any $N_S \leq N_1 - 1$

$$\begin{aligned} TH_{FMS}(N_S \lfloor (A - aN_S)/b \rfloor, \lambda) &\leq \mu(N_1 - 1) \\ &\leq TH_{FMS}(c, \lfloor (A - ac)/b \rfloor, \lambda), \end{aligned}$$

where $N_1 = \lfloor TH_{FMS}(c + \lfloor (A - ac)/b \rfloor, 0, \lambda)/\mu \rfloor + 1$.

It is clear that N_S^* exists on $[N_1, c]$ and is obtained by solving

$$\mathbf{P2''} : \max_{N_S} \{\text{eqn.}(17) \mid N_1 \leq N_S \leq c \text{ and } N_H = \lfloor (A - aN_S)/b \}.$$

Consider the case $\lfloor A/a \rfloor < c$. Obviously, $N_S \leq \lfloor A/a \rfloor < c$ and from eqn. (15), for any $N_S \geq N_2 - 1$, we get

$$\begin{aligned} TH_{FMS}(N_S, \lfloor (A - aN_S)/b, \lambda) &\leq \mu(N_s - 1) \\ &\leq TH_{FMS}(\lfloor A/a \rfloor, 0, \lambda), \end{aligned}$$

where $N_2 = \lfloor TH_{FMS}(\lfloor A/a \rfloor, 0, \lambda)/\mu \rfloor + 1$.

It is also clear that N_S^* exists on $[N_2, \lfloor A/a \rfloor]$ and is obtained by solving

$$\mathbf{P2'''} : \max_{N_S} \{\text{eqn.}(17) \mid N_2 \leq N_S \leq \lfloor A/a \rfloor \text{ and } N_H = \lfloor (A - aN_S)/b \rfloor \}.$$

Combining **P2''** and **P2'''**, we get the optimal solution to problem **P2** by solving the following reduced problem:

$$\mathbf{RP2} : \max_{N_S} \{\text{eqn.}(17) \mid N_3 \leq N_S \leq N_4 \text{ and } N_H = \lfloor (A - aN_S)/b \rfloor \}$$

where $N_4 = \min(\lfloor A/a \rfloor, c)$ and $N_3 = \lfloor TH_{FMS}(N_4 + \lfloor (A - aN_4)/b \rfloor, 0, \lambda)/\mu \rfloor + 1$.

The value of $TH_{FMS}(N_S, N_H, \lambda)$ defined by eqn. (17) is easily calculated from the following recursive equations:

$$\begin{aligned} TH_{FMS}(N_S, N_H + 1, \lambda) =& \lambda \frac{\mu N_S}{\mu N_S + \lambda - TH_{FMS}(N_S, N_H, \lambda)} \\ & \text{for } N_H \geq 0, \text{ and } N_S \geq 1, \end{aligned} \tag{18}$$

and

$$\begin{aligned} TH_{FMS}(N_S + 1, 0, \lambda) =& \lambda \frac{\mu(N_S + 1)}{\mu(N_S + 1) + \lambda - TH_{FMS}(N_S, 0, \lambda)} \\ & \text{for } N_S \geq 0, \end{aligned} \tag{19}$$

where $TH_{FMS}(0,0,\lambda)=0$.

These recursive equations are stable in the sense that computational error is not increased by the successive iterations as shown in Theorem C (see appendices: $TH(n)$ in Theorem C should be replaced with $\mu\ n$ in this case). The optimal solution to problem **RP2** is obtained efficiently by calculating the values of $TH_{FMS}(N_S, \lfloor (A-aN_S)/b\rfloor, \lambda)$ using these recursive equations for each value of N_S on $[N_3, N_4]$.

4.2 *multiple-station FMS*

We consider, next, the following capacity design problems in the multiple- station FMS, denoted by **MP1** and **MP2** similar to **P1** and **P2**, respectively:

$$\mathbf{MP1}\ :\ \max_{N_S,N_H}\{\widetilde{TH}_{FMS}(N_S,N_H,\lambda)\mid N_S+N_H\le N \text{ and } N_S\ge 0, N_H\ge 0\},$$

$$\mathbf{MP2}\ :\ \max_{N_S,N_H}\{\widetilde{TH}_{FMS}(N_S,N_H,\lambda)\mid aN_S+bN_H\le A \text{ and } N_S\ge 0, N_H\ge 0\}.$$

In these case, it should be noted that eqns (11a), (11b) and (12) provide the approximation for the throughput of the FMS, and that $TH(n)$ given by eqns (8a) and (8b) does not have any break point except at $n=N_S$. Considering these facts, we provide solution methods to problems **MP1** and **MP2** below.

Solution to **MP1**

For problem **MP1**, the optimal solution is $N_S^*=N$ and $N_H^*=0$; in the multiple-station FMS, from the above fact (2), we can not get any relation similar to that shown in eqn. (16b) for $c\le N_S\le N$.

Solution to **MP2**

There are three cases to consider:

(i) $a=b$

From the argument similar to that in case (i) of **P2** and the result to **MP1**, we get $N_S^*=N$ and $N_H^*=0$ where $N=\lfloor A/a\rfloor$.

(ii) $a<b$

The derivation of the optimal solution to the case (ii) of **P2** is also applicable to this case as well. We get $N_S^*=N$ and $N_H^*=0$ where $N=\lfloor A/a\rfloor$.

(iii) $a>b$

Since there is no break point in the function $TH(n)$ for $n<N_S$, we can not specify the upper bound of N_S^* less than $\lfloor A/a\rfloor$ unlike case (iii) of **P2**. Using Corollary 3(ii) in determining the lower bound of N_S^* yields

$N_5=\min\{N_S\mid TH(N_S)\ge \widetilde{TH}_{FMS}(\lfloor A/a\rfloor,0,\lambda) \text{ and } 0\le N_S\le\lfloor A/a\rfloor\}$.

Then, the original problem is transformed into

$$\mathbf{MP2\prime}\ :\ \max_{N_S}\{\widetilde{TH}_{FMS}(N_S,\lfloor (A-aN_S)/b,\rfloor,\lambda)\mid N_5\le N_S\le\lfloor A/a\rfloor\}.$$

The value of $\widetilde{TH}_{FMS}(N_S,N_H,\lambda)$ is easily calculated by the following recursive equations:

$$\widetilde{TH}_{FMS}(N_S,N_H+1,\lambda)=\lambda\frac{TH(N_S)}{TH(N_S)+\lambda-\widetilde{TH}_{FMS}(N_S,N_H,\lambda)} \qquad (20)$$

for $N_H\ge 0$ and $N_S\ge 1$,

$$\widetilde{TH}_{FMS}(N_S+1,0,\lambda)=\lambda\frac{TH(N_S+1)}{TH(N_S+1)+\lambda-\widetilde{TH}_{FMS}(N_S,0,\lambda)} \quad \text{for } N_S\geq 0, \tag{21}$$

and $\widetilde{TH}_{FMS}(0,0,\lambda)=0$.

The function $TH(n)$ is defined in eqn. (8b) and the value of $G(M+2,n,\mathbf{e})$ in $TH(n)$ is efficiently obtained by Buzen's algorithm [2] using any scaling for $e_i, i=-1,0,\cdots,M+1$. The above recursive equations are stable as shown in Theorem C(see appendices). The optimal solution to problem **MP2′** is obtained by the efficient manner similar to that used for solving problem **RP2**.

5. Conclusions

We extended the existing FMS models through the addition of a supplementary automatic warehouse with finite capacity. Since the equilibrium probabilities of queue lengths at each station and the throughput of the FMS can not be exactly formulated in this case, we provided approximate expressions for the probabilities and the throughput, using a closed queueing network model. We showed the condition that the approximations yield the exact values. Comparing with the SIMAN simulation run, we also made clear that the accuracy of the approximation is very high especially in the FMS with balancing workload. We derived the limits, and the first- and second-order properties of the throughput with respect to both the warehouse capacity and the number of the pallets available in the shop. Exploiting these properties, we proposed efficient solution methods to several capacity design problems.

Capacity design problems presented here are essential ones and can be applied to various versions of these problems; for instance a design problem to maximize a profit function involving both the warehouse capacity and the number of available pallets. Furthermore, the following capacities should be modeled to be determined in design problems: (1) capacity at each station; the number of machines at each machining station, the number of carts in the MHS, and the number of machines (or labors) and fixtures at the loading/unloading station; (2) capacity of local buffer spaces at each station. Solution methods to design problems with these capacities will be topics of future research.

Appendices

Theorem A. *Consider a closed queueing network (CQN) with M stations at which service times are exponentially distributed with service rates, $\mu_i(n_i)$, $i=1,2,\cdots,M$, where n_i denotes the queue length at station i. Jobs follow a Markov routing with $\gamma(i,j)$, i(or j) $=0,1,\cdots,M$, where station 0 denotes a dummy station representing an input/ output station with an infinite service rate. The expected visit times of each job, $e_i^*, i=1,2,\cdots,M$, are obtained as a unique solution to the flow equations such that*

$$e_i=\gamma(0,i)+\sum_{j=1}^{M}e_j\gamma(j,i) \text{ for } i=1,2,\cdots,M.$$

Define the system throughput as

$$\widehat{TH}(N)\equiv\sum_{i=1}^{M}TH_i(N)\gamma(i,0)=\frac{G(M,N-1,\mathbf{e}^*)}{G(M,N,\mathbf{e}^*)}.$$

where N and $TH_i(N)$ are the population in the CQN and the throughput of station i, respectively, and

$$G(M,N,\mathbf{e}^*)\equiv\sum_{\sum_{i=1}^{M}n_i=N}\prod_{i=1}^{M}\left\{\prod_{k=1}^{n_i}\left(\frac{e_i^*}{\mu_i(k)}\right)\right\}.$$

If $\mu_i(n) = \mu_i(N_i)$ *for* $n \geq N_i$, *and* $\mu_i(n) \leq \mu_i(N_i)$ *for* $n < N_i, i = 1, 2, \cdots, M$, *then*

$$\lim_{N\to\infty} \widehat{TH}(N) = \min_{1\leq i\leq M} \left\{\frac{\mu_i(N_i)}{e_i^*}\right\}.$$

(Proof)

Schweitzer [8] has proved the above result for the special case of $\mu_i(n) = \mu_i$ for $i = 1, 2, \cdots, M$. We shall generalize his result below.

Consider two networks, denoted by superscript $p = 1, 2$, with service rates $\mu_i^p(n)$'s such that

$$\mu_i^1(n) = \mu_i(N_i)1\{n \geq N_i\} \text{ and } \mu_i^2(n) = \mu_i(N_i), i = 1, 2, \cdots, M,$$

where $1\{x\}$ is an indicator function; $1\{x\} = 1$ if x is true and 0 otherwise.

Since for any $n \leq m$, $\mu_i^1(n) \leq \mu_i(m)$ and $\mu_i(n) \leq \mu_i^2(m)$, $i = 1, 2, \cdots, M$, the corresponding system throughputs are related as follows according to the results given by Shanthikumar and Yao [10]:

$$\widehat{TH}^1(N) \leq \widehat{TH}(N) \leq \widehat{TH}^2(N).$$

We clearly have $\widehat{TH}^2(N) \rightarrow \min_{1\leq i\leq M}\{\mu_i(N_i)/e_i^*\}$ as $N \rightarrow \infty$ because $\mu_i(N_i)$'s are all constant.

When $N > \sum_{i=1}^{M} N_i - M$, the throughput of network 1 is equivalent to that of network 2 with the population $N - \sum_{i=1}^{M} N_i + M$, that is, $\widehat{TH}^1(N) = \widehat{TH}^2(N - \sum_{i=1}^{M} N_i + M)$ for $N > \sum_{i=1}^{M} N_i - M$. This relation implies that $\lim_{N\to\infty} \widehat{TH}^1(N) = \lim_{N\to\infty} \widehat{TH}^2(N)$.

Therefore, $\widehat{TH}(N) \rightarrow \min_{1\leq i\leq M}\{\mu_i(N_i)/e_i^*\}$ as $N \rightarrow \infty$. Q. E. D.

Theorem B. *Consider a CQN with M stations at which service times are exponentially distributed with queue-length dependent service rates* $\gamma_i(n_i), i = 1, 2, \cdots, M$. *Suppose the* $\gamma_i(n_i)$*'s are all nondecreading concave in* n_i *and* $\gamma_1(n_1) = \min\{\mu_1(n_1), \mu_1(c)\}$, *where* $\mu_1(n_1)$ *is also nondecreasing concave in* n_1. *Then, the throughput of the CQN,* $TH_{CQN}(c)$, *is nondecreasing concave in c.*

(Proof)

Shanthikumar and Yao [11] proved Theorem B for the case of $\gamma_1(n_1) = \mu \min(n_1, c)$, i. e., $\mu_1(n_1) = \mu n_1$. We shall prove Theorem B along a way similar to their proof below.

Without loss of generality, consider a two-node CQN with N jobs and visit rations $\nu_1 = \nu_2 = 1/2$. Since $\min\{\mu_1(n), \mu_1(c)\} \leq \min\{\mu_1(n), \mu_1(c+1)\}, TH_{CQN}(c) \leq TH_{CQN}(c+1)$ follows directly from the results derived by Shanthikumar and Yao [11]. We shall prove the concavity of $TH_{CQN}(c)$ i. e., $TH_{CQN}(c) + TH_{CQN}(c+2) \leq 2TH_{CQN}(c+1)$.

Construct four networks ($p = 1 \sim 4$) such that $r_1^p(n_1) = \min\{\mu_1(n_1), \mu_1(c)\}$ for $p = 1$; $\min\{\mu_1(n_1), \mu_1(c+1)\}$ for $p = 2, 3$; $\min\{\mu_1(n_1), \mu_1(c+2)\}$ for $p = 4$, and the other conditions in these networks are all the same. Let $Z_j^p(t)$ be a random variable which represents the queue length at node j of network p, and $D_j^p(t)$ be a random variable which denotes the number of departures from node j of network p in $(0, t)$. Since the throughput of node j is $TH_j(c) = \lim_{t\to\infty} E\{D_j^p(t)\}/t$ and the throughput of the CQN is $TH_{CQN}(c) = \sum_{j=1}^{2} TH_j(c)$, it suffices to prove $\mathbf{D}^1(t) + \mathbf{D}^4(t) \leq^{st} \mathbf{D}^2(t) + \mathbf{D}^3(t)$, where $\mathbf{D}^p(t) = (D_1^p(t), D_2^p(t))$ and $\mathbf{Z}^p(t) = (Z_1^p(t), Z_2^p(t))$ for $p = 1 \sim 4$.

We shall construct the process $\{\mathbf{Z}^p(t), \mathbf{D}^p(t)\}(p = 1 \sim 4)$ on a common probability space (Ω, F, P) such that

(i) $\{\hat{\mathbf{Z}}^p(t), \hat{\mathbf{D}}^p(t)\} =^{st} \{\mathbf{Z}^p(t), \mathbf{D}^p(t)\}, p = 1 \sim 4$;

(ii) $\hat{Z}_1^4(t) \leq \hat{Z}_1^2(t) \leq \hat{Z}_1^1(t)$ and $\hat{Z}_1^4(t) \leq \hat{Z}_1^3(t) \leq \hat{Z}_1^1(t)$ for $t \geq 0$; this implies $\hat{Z}_2^4(t) \geq \hat{Z}_2^2(t) \geq \hat{Z}_2^1(t)$ and $\hat{Z}_2^4(t) \geq \hat{Z}_2^3(t) \geq \hat{Z}_2^1(t)$ for $t \geq 0$;

(iii) $\hat{D}_j^1(t) + \hat{D}_j^4(t) \leq \hat{D}_j^2(t) + \hat{D}_j^3(t)$ for $t \geq 0$ and $j = 1, 2$;

(iv) $\hat{\mathbf{Z}}^p(0) = (N, 0)$ and $\hat{\mathbf{D}}^p(0) = (0, 0), p = 1 \sim 4$.

Let $0 = \tau_0 < \tau_1 < \cdots$ be the event epochs of a Poisson process with rate 2η, where $\eta = \max\{\gamma_1(c+2), \gamma_2(N)\}$, and $(u_k)_{k=0}^{\infty}$ be a sequence of i. i. d. uniform random variables with support on $(-\eta, \eta)$. These two processes $\{\tau_k\}$ and $\{u_k\}$ are independent of each other and are both defined on a common probability space (Ω, F, P). Consider a path $\omega \in \Omega$, and denote $\tau_k(\omega)$ and $u_k(\omega)$ simply as τ_k and u_k.

When $t = \tau_0 = 0, \hat{\mathbf{Z}}^p(0) = (N, 0)$ and $\hat{\mathbf{D}}^p(0) = (0, 0)$ satisfies (ii) $\sim$ (iv). Suppose $\hat{\mathbf{Z}}^p(\tau_{k-1}), \hat{\mathbf{D}}^p(\tau_{k-1}))$ has been specified for all p and (i) $\sim$ (iv) are valid at $t = \tau_{k-1}$. For $t \in (\tau_{k-1}, \tau_k)$, set $(\hat{\mathbf{Z}}^p(t), \hat{\mathbf{D}}^p(t)) = (\hat{\mathbf{Z}}^p(\tau_{k-1}), \hat{\mathbf{D}}^p(\tau_{k-1}))$ for $p = 1 \sim 4$.

To simplify notations for all p, let $\mathbf{z}^p = \hat{\mathbf{Z}}^p(\tau_{k-1})$, $\mathbf{d}^p = \hat{\mathbf{D}}^p(\tau_{k-1})$, $\gamma_j^p = \gamma_j^p(\hat{Z}_j^p(\tau_{k-1}))$, $\gamma^{12} = \min(\gamma_1^1, \gamma_1^2)$, $\gamma^{34} = \min(\gamma_1^3, \gamma_1^4)$, $R^{13} = \max(\gamma_1^1, \gamma_1^3)$ and $R^{24} = \max(\gamma_1^2, \gamma_1^4)$.

Set $\hat{\mathbf{Z}}^p(\tau_k) = \mathbf{z}^p + (\Delta_2^p - \Delta_1^p, \Delta_1^p - \Delta_2^p), \hat{\mathbf{D}}^p(\tau_k) = \mathbf{d}^p + (\Delta_1^p, \Delta_2^p)$, where

$$\Delta_1^1 = 1\{0 \leq u_k < \gamma^{12}\} + 1\{R^{13} - \gamma_1^1 + \gamma^{12} \leq u_k < R^{13}\},$$
$$\Delta_1^2 = 1\{0 \leq u_k < \gamma^{12}\} + 1\{R^{24} - \gamma_1^2 + \gamma^{12} \leq u_k < R^{24}\},$$
$$\Delta_1^3 = 1\{0 \leq u_k < \gamma^{24}\} + 1\{R^{13} - \gamma_1^3 + \gamma^{34} \leq u_k < R^{13}\},$$
$$\Delta_1^4 = 1\{0 \leq u_k < \gamma^{34}\} + 1\{R^{24} - \gamma_1^4 + \gamma^{34} \leq u_k < R^{24}\},$$

where $1\{x\}$ is the indicator function; $1\{x\} = 1$ if x is true and 0 otherwise.

Since $P\{\Delta_j^p = 1\} = \gamma_j^p/(2\eta)$ for all nodes j of network p, the transition rates of the constructed processes from a state $(\mathbf{z}^p, \mathbf{d}^p)$ to states $(\mathbf{z}^p + (-1, 1), \mathbf{d}^p + (1+0))$ and $(\mathbf{z}^p + (-1, 1), \mathbf{d}^p + (0, 1))$ are $2\eta\gamma_1^p/(2\eta) = \gamma_1^p$ and $2\eta\gamma_2^p/(2\eta) = \gamma_2^p$, respectively. These rates coincide with those of the original processes.

Hence, $\{\hat{\mathbf{Z}}^p(t), \hat{\mathbf{D}}^p(t)\} =^{st} \{\mathbf{Z}^p(t), \mathbf{D}^p(t)\}, p = 1 \sim 4$ at $t = \tau_k$, that is, (i) is satisfied. We can prove that (ii) is satisfied at $t = \tau_k$ using an approach similar to that in Shanthikumar and Yao [9].

We shall prove (iii) at $t = \tau_k$, that is, $d_j^1 + d_j^4 \leq d_j^2 + d_j^3 + \Delta_j^2 + \Delta_j^3 - \Delta_j^1 - \Delta_j^4, j = 1, 2$. Consider $j = 1$ first. There are the following possible cases:

(a) $\gamma_1^1 \leq \gamma_1^2$ and $\gamma_1^4 > \gamma_1^3$; then $\gamma^{12} = \gamma_1^1$, and $\gamma^{34} = \gamma_1^3$, and hence $\Delta_1^2 \geq \Delta_1^1, \Delta_1^3 \leq \Delta_1^4$;

(b) $\gamma_1^1 > \gamma_1^2$ and $\gamma_1^4 \leq \gamma_1^3$; then $\gamma^{12} = \gamma_1^2$ and $\gamma^{34} = \gamma_1^4$, and hence $\Delta_1^2 \leq \Delta_1^1, \Delta_1^3 \geq \Delta_1^4$;

(c) $\gamma_1^1 \leq \gamma_1^2$ and $\gamma_1^4 \leq \gamma_1^3$; then $\gamma^{12} = \gamma_1^1$ and $\gamma^{34} = \gamma_1^4$, and hence $\Delta_1^2 \geq \Delta_1^1, \Delta_1^3 \geq \Delta_1^4$;

(d) $\gamma_1^1 > \gamma_1^2$ and $\gamma_1^4 > \gamma_1^3$.

Case (d) is impossible because if $z_1^4 \geq c+2$ then $c+2 \leq z_1^1, z_1^2, z_1^3$, that is, $\gamma_1^1 = \mu_1(c) \leq \gamma_1^2 = \gamma_1^3 = \mu_1(c+1) \leq \gamma_1^4 = \mu_1(c+2)$, and if $z_1^4 \leq c+1$ then $\gamma_1^4 = \gamma_1^2(z_1^4) \leq \gamma_1^2 = \gamma_1^2(z_1^2)$ and $\gamma_1^3 = \gamma_1^2(z_1^3)$ (consider that $\gamma_1^2(z)$ is nondecreasing in z and that $z_1^4 \leq z_1^2, z_1^3$).

In case (c), $(\Delta_1^2 - \Delta_1^1) + (\Delta_1^3 - \Delta_1^4) \geq 0$, and hence $d_1^1 + d_1^4 \leq d_1^2 + d_1^3 + \Delta_1^2 + \Delta_1^3 - \Delta_1^1 - \Delta_1^4$, provided that $d_1^1 + d_1^4 \leq d_1^2 + d_1^3$.

In cases (a) and (b), $(\Delta_1^2 - \Delta_1^1) + (\Delta_1^3 - \Delta_1^4) \geq -1$. If $d_1^1 + d_1^4 < d_1^2 + d_1^3$, then $d_1^1 + d_1^4 \leq d_1^2 + d_1^3 - 1 \leq d_1^2 + d_1^3 + \Delta_1^2 + \Delta_1^3 - \Delta_1^1 - \Delta_1^4$. If $d_1^1 + d_1^4 = d_1^2 + d_1^3$(and $d_2^1 + d_2^4 \leq d_2^2 + d_2^3$), then $z_1^1 + z_1^4 = (N - d_1^1 + d_2^1) + (N - d_1^4 + d_2^4) \leq (N - d_1^2 + d_2^2) + (N - d_1^3 + d_2^3) = z_1^2 + z_1^3$. Provided that $z_1^4 \leq z_1^2 \leq z_1^1, z_1^4 \leq z_1^3 \leq z_1^1, z_1^1 + z_1^4 \leq z_1^2 + z_1^3$ and $\mu_1(n)$ is nondecreasing concave in n, we can show that $\gamma_1^1 + \gamma_1^4 \leq \gamma_1^2 + \gamma_1^3$; if $z_1^4 \geq c+2$ then $\gamma_1^1 + \gamma_1^4 = \mu_1(c) + \mu_1(c+2) \leq 2\mu_1(c+1) = \gamma_1^2 + \gamma_1^3$,

and if $z_1^4 \le c+1$ then $\gamma_1^1 - \gamma_1^2 = \gamma_1^1(z_1^1) - \gamma_1^2(z_1^2) \le \gamma_1^2(z_1^1) - \gamma_1^2(z_1^2) = \gamma_1^2(z_1^2 + (z_1^1 - z_1^2)) - \gamma_1^2(z_1^2) \le \gamma_1^2(z_1^4 + (z_1^1 - z_1^2)) - \gamma_1^2(z_1^4) \le \gamma_1^2(z_1^3) - \gamma_1^2(z_1^4) = \gamma_1^3(z_1^3) - \gamma_1^4(z_1^4) = \gamma_1^3 - \gamma_1^4$.
Therefore, we get
(a) $\Delta_1^2 + \Delta_1^3 - \Delta_1^1 - \Delta_1^4$
$= 1\{R^{24} - (\gamma_1^2 - \gamma_1^1) \le u_k < R^{24}\} - 1\{R^{24} - (\gamma_1^4 - \gamma_1^3) \le u_k < R^{24}\} \ge 0;$
(b) $\Delta_1^2 + \Delta_1^3 - \Delta_1^1 - \Delta_1^4$
$= 1\{R^{13} - (\gamma_1^3 - \gamma_1^4) \le u_k < R^{13}\} - 1\{R^{13} - (\gamma_1^1 - \gamma_1^2) \le u_k < R^{13}\} \ge 0;$

These results imply that $d_1^1 + d_1^4 \le d_1^2 + d_1^3 + \Delta_1^2 + \Delta_1^3 - \Delta_1^1 - \Delta_1^4$ for cases (a) and (b).

For $j = 2$, we can prove $d_2^1 + d_2^4 \le d_2^2 + d_2^3 + \Delta_2^2 + \Delta_2^3 - \Delta_2^1 + \Delta_2^4$ in the same manner as in Shanthikumar and Yao [11].

We have thus completed the induction. Q. E. D.

Theorem C. *Define* $\widetilde{TH}_{FMS}(N_S, N_H, \lambda) \equiv \lambda\{1 - B(N_S, N_H)\}$ *and*

$$B(N_S, N_H) = \frac{\left(\prod_{k=1}^{N_S} \frac{\lambda}{TH(k)}\right)\left(\frac{\lambda}{TH(N_S)}\right)^{N_H}}{1 + \sum_{n=1}^{N_S-1} \prod_{k=1}^{n} \frac{\lambda}{TH(k)} + \left(\prod_{k=1}^{N_S} \frac{\lambda}{TH(k)}\right) \sum_{j=0}^{N_H} \left(\frac{\lambda}{TH(N_S)}\right)^j}.$$

Then we get

$$(i)\ \widetilde{TH}_{FMS}(N_S, N_H + 1, \lambda) = \lambda \frac{TH(N_S)}{TH(N_S) + \lambda - \widetilde{TH}_{FMS}(N_S, N_H, \lambda)}$$

for $N_H \ge 0$ and $N_S \ge 1$;

$$(ii)\ \widetilde{TH}_{FMS}(N_S + 1, 0, \lambda) = \lambda \frac{TH(N_S + 1)}{TH(N_S + 1) + \lambda - \widetilde{TH}_{FMS}(N_S, 0, \lambda)} \text{ for } N_S \ge 0$$

and $\widetilde{TH}_{FMS}(0, 0, \lambda) = 0$;

$$(iii)\ |\Delta(N_S, N_H)| < |\Delta(N_S, N_H - 1)| \le |\Delta(N_S, 0)| < |\Delta(N_S - 1, 0)|$$

$$\text{and } \frac{|\Delta(N_S, N_H)|}{\widetilde{TH}^*_{FMS}(N_S, N_H, \lambda)} < \frac{|\Delta(N_S, N_H - 1)|}{\widetilde{TH}^*_{FMS}(N_S, N_H - 1, \lambda)} \le \frac{|\Delta(N_S, 0)|}{\widetilde{TH}^*_{FMS}(N_S, 0, \lambda)} < \frac{|\Delta(N_S - 1, 0)|}{\widetilde{TH}^*_{FMS}(N_S - 1, 0, \lambda)} \text{ for } N_S \ge 1 \text{ and } N_H \ge 1,$$

where "" and* $\Delta(N_S, N_H)$ *denote the exact value without computational error and the computational error from the calculation of* $\widetilde{TH}_{FMS}(N_S, N_H, \lambda)$*, respectively.*

(Proof)

We show the proof in an argument similar to that in Yao and Shanthikumar [16]. From the definition of $B(N_S, N_H)$, we get

$$\frac{1}{B(N_S, N_H + 1)} = \frac{1}{B(N_S, N_H)} \frac{TH(N_S)}{\lambda} + 1,$$

and

$$\frac{1}{B(N_S + 1, 0)} = \frac{1}{B(N_S, 0)} \frac{TH(N_S + 1)}{\lambda} + 1.$$

Substituting $\widetilde{TH}_{FMS}(N_S, N_H, \lambda) = \lambda\{1 - B(N_S, N_H)\}$ into these equations yields (i) and (ii).

Define $\Delta(N_S, N_H) \equiv \widetilde{TH}_{FMS}(N_S, N_H, \lambda) - \widetilde{TH}^*_{FMS}(N_S, N_H, \lambda) = \lambda\{B^*(N_S, N_H) - B(N_S, N_H)\}$. Since both $B(N_S, N_H)$ and $B^*(N_S, N_H)$ satisfy the above equations, we get

$$\frac{\Delta(N_S, N_H)}{B^*(N_S, N_H)} = \frac{\Delta(N_S, N_H - 1)}{B^*(N_S, N_H - 1)} \frac{1}{1 + \lambda B(N_S, N_H - 1)/TH(N_S)}$$

and

$$\frac{\Delta(N_S, 0)}{B^*(N_S, 0)} = \frac{\Delta(N_S - 1, 0)}{B^*(N_S - 1, 0)} \frac{1}{1 + \lambda B(N_S - 1, 0)/TH(N_S)}.$$

Since $\widetilde{TH}^*_{FMS}(N_S, N_H, \lambda)$ (and then $B^*(N_S, N_H)$) is nondecreasing (nonincreasing) in both N_S and N_H, we can show that (iii) holds. Q. E. D.

References

[1] Buzacott J. A. and Yao D. D.: On Queueing Network Models of Flexible Manufacturing Systems. *Queueing Systems*, Vol. 1 (1986), 5–27.

[2] Buzen J. P.: Computational Algorithms for Closed Queueing Networks with Exponential Servers. *Communications of the ACM*, Vol. 16 (1973), 527–531.

[3] Dubois D.: A Mathematical Model of a Flexible Manufacturing System with Limited In-process Inventory. *Eur. J. Oper. Res.*, Vol. 14 (1983), 66–78.

[4] Jackson J. R.: Jobshop Like Queueing Systems, *Mgmt. Sci.* Vol. 10 (1963), 131–142.

[5] Nagasawa H., Jeong S. C. and Nishiyama N.: A Generalized Representation of Flexible Manufacturing Systems Using a Closed Queueing Network Model. *Bulletin of the University of Osaka Prefecture (Series A)*, Vol. 38 (1989), 113–132.

[6] Onvural R. O. : Survey of Closed Queueing Networks with Blocking, *ACM Computing Surveys*, Vol. 22 (1990), 83–121.

[7] edited by Perros H. G. and Altiok T. : Queueing Networks with Blocking, North-Holland (1989).

[8] Schweitzer P. J. : Maximum Throughput in Finite-Capacity Open Queueing Networks with Product-Form Solutions. *Mgmt. Sci.* Vol. 24 (1977), 217–223.

[9] Shanthikumar J. G. and Stecke K. E. 1986. Reducing Work-in-process Inventory in Certain Class of Flexible Manufacturing Systems. *Eur. J. Oper. Res.* Vol. 26 (1986), 266–271.

[10] Shanthikumar J. G. and Yao, D. D.: The Preservation of Likelihood Ratio Ordering Under Convolution. *Stochastic Process and Their Applications*, Vol. 23 (1986), 259–267.

[11] Shanthikumar J. G. and Yao, D. D.: Optimal Server Allocation in a System of Multi-server Stations. *Mgmt. Sci.*, Vol. 33 (1987), 1173–1180.

[12] Shanthikumar J. G. and Yao D. D.: Second-Order Properties of the Throughput of a Closed Queueing Network. *Math. Oper. Res.*, Vol. 13 (1988), 524–534.

[13] Shanthikumar J. G. and Yao D. D.: Second Order Stochastic Properties in Queueing Systems. *Proceedings of the IEEE*, Vol. 77 (1989), 162–170.

[14] Solberg J. J.: A Mathematical Model of Computerized Manufacturing Systems. *4th International Conference on Production Research, Tokyo* (1977), 1265–1275.

[15] Yao D. D.: An Optimal Storage Model for a Flexible Manufacturing System. *Flexible Manufacturing Systems; Methods and Studies*, A. Kusiak (eds.). *North-Holland* (1986), 113–125.

[16] Yao D. D. and Shanthikumar J. G.: The Optimal Input Rates to a System of Manufacturing Cells. *INFOR*, Vol. 25 (1987), 57–65.
[17] Yao D. D. and Shanthikumar J. G.: Allocating a Joint Setup in a Multi-Cell System. *Annals of Operations Research*, Vol. 15 (1988), 155–167.

Hiroyuki Nagasawa
Department of Industrial Engineering
College of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka, 580, Japan
E-mail: ng@ie.osakafu-u.ac.jp