

## OPTIMIZATION IN PRODUCTION SYSTEMS: A SURVEY ON QUEUING APPROACHES

Tadayoshi Shioyama  
*Kyoto Institute of Technology*

Hiroshi Kise  
*Kyoto Institute of Technology*

(Received October 30, 1987; Revised July 25, 1988)

**Abstract** This paper surveys recent works on queuing approaches to a number of optimization problems in production systems, especially modern jobbing production systems (or multi-item small lot production systems) represented by flexible manufacturing systems (FMS). This paper outlines optimization problems concerning with such aspects of the system design and control as workloading, job selection, production rate control, production/inventory control and system configuration.

### 1. Introduction

As a result of increased variety of demands in marketplace, the role of jobbing production (or multi-item small lot production) has become important. In fact, at least 75% of the metal products in industrialized countries have been manufactured in lot size of less than 50 [58]. This trend has induced requirement for the jobbing production system to be efficient. In response to the requirement, the concept of modern jobbing production system represented by flexible manufacturing systems (FMS) [7], [22], [38], [39] has been developed with the aid of computer and numerical control techniques. Implementing the concept has given rise to a number of problems relating to the system design, control and so on. These problems seem highly challenging to OR/MS researchers. The application of queuing theory is now widely recognized as one of the most useful approaches to these problems.

The purpose of this paper is to classify the recent queuing theoretic studies by type of optimization problem in the modern production system and to outline how the properties of optimal solutions are derived for each type of problem. The following types of optimization problems are discussed here:

(1) workloading, i.e., workload assignment among work stations;

- (2) job selection;
- (3) production rate control;
- (4) production/inventory control; and
- (5) system configuration, i.e., allocation of resources such as machines and storages, and location of a movable server (transporter).

If we restrict our attention to FMS's, there are some surveys [7], [25]. Buzacott and Yao [7] review works on the recent development of analytical queuing models of an FMS. Kalkunte, Sarin and Wilhelm [25] present a comprehensive review of modeling approaches related to the design and operation of an FMS.

## 2. Loading Problem

The loading problem is to optimally allocate the required workload among work stations. We consider the case in equilibrium [5], [50] and subsequently deal with the nonstationary case [11].

Using a closed queuing network model [14], [23], Stecké and Solberg [50] have solved the loading problem of maximizing the expected production rate of the system. In the closed queuing network model, a finished job is removed from the system and a new job is instantaneously put in the system. Therefore, the number of jobs in the system is fixed, corresponding to the fixed number of pallets on which jobs are mounted.

In their model, the production system consists of  $m$  machines partitioned into  $g$  work stations. For each work station  $i$ ,  $i=1, \dots, g$ , let  $s_i$  be the number of machines,  $t_i$  the average processing time, and  $q_i$  the arrival rates.  $q_i$  satisfies the following traffic equations

$$q_i = \sum_{j=1}^g p_{ji} q_j, \quad i=1, \dots, g,$$

where  $p_{ji}$  denotes the routing probability from station  $j$  to station  $i$ . The workload  $w_i$  of work station  $i$  is given by

$$w_i = q_i t_i, \quad i=1, \dots, g.$$

Define the workload measure  $X_i$  as

$$X_i \equiv q_i t_i / \left[ \sum_{j=1}^g q_j t_j / m \right].$$

Then,  $\sum_{i=1}^g X_i = m$  holds and  $X_i/s_i$  is the workload per machine at work station  $i$ .

By balancing the work load per machine, it holds that  $X_i/s_i = 1$ ,  $i=1, \dots, g$ .

Denote by  $n=(n_1, n_2, \dots, n_g)$  the state of the system where  $n_i$  represents the number of jobs waiting or in process at work station  $i$ . The steady state probability  $p(n)$  is assumed to be given by the product form solution [1],[14],[23]:

$$p(n) = \prod_{i=1}^g f_i(n_i) / G(g, n; S, X),$$

where  $S \equiv (s_1, s_2, \dots, s_g)$  and  $X \equiv (X_1, X_2, \dots, X_g)$ .  $G(g, n; S, X)$  is the normalizing constant:

$$G(g, n; S, X) = \sum_{\substack{n_1 \geq 0 \\ n_1 + \dots + n_g = n}} \dots \sum_{\substack{n_g \geq 0 \\ n_1 + \dots + n_g = n}} f_1(n_1) f_2(n_2) \dots f_g(n_g),$$

$$f_i(n_i) = X_i^{n_i} / n_i!, \quad n_i \leq s_i,$$

$$= X_i^{n_i} / (s_i! s_i^{n_i - s_i}), \quad n_i > s_i, \quad i=1, \dots, g,$$

and  $n$  is the total number of jobs present in the system. The production rate  $PR(g, n; S, X)$  is given [42] as

$$PR(g, n; S, X) = G(g, n-1; S, X) / G(g, n; S, X).$$

The problem is to maximize  $PR(g, n; S, X)$  over  $X$  with given  $g, m, S, n, q_j$  and

$$\sum_{j=1}^g q_j t_j.$$

The conclusions drawn from this study are summarized as follows:

- (1) for a system consisting of work stations with unequal number of machines, unbalanced workloads are better than balanced ones for purpose of increasing the expected production rate;
- (2) balancing the workload per machine is optimal only if all work stations consist of equal number of machines.

For a system consisting of work stations with unequal number of machines, it is conjectured that the production rate is maximized by a unique unbalanced allocation of the workload per machine. In particular, more (less) than the balanced amount of workload per machine is assigned to the work station with the larger (smaller) number of machines.

The similar results to the above are obtained by several previous studies on other types of stochastic production lines. They have pointed out that balancing the workload per machine is nonoptimal in serial systems of single-machine stations with finite storages [20],[29],[41]. This phenomenon is

related to the finite storage condition, whereas the result of Stecké and Solberg [50] is related to the multiple-server efficiency issue.

The case of parallel machines has been studied by Bell and Stidham [5]. Introducing waiting cost, they have dealt with optimization in the allocation of jobs to single machine work stations. Any station can process any of the jobs. In their model, it is assumed that the overall arrival rate  $\Lambda$  is fixed. Upon arrival, job must choose one of the  $M$  work stations. Jobs cannot observe the current storage at each work station. But they are aware of the service time distribution with mean  $1/\mu_i$ , waiting cost  $h_i$  per job at the  $i$ -th work station. Although service time distributions may differ from station to station and need not be exponential, they have the same coefficient of variation  $(b-1)^{1/2}$ . Bell and Stidham [5] have derived the following optimal arrival rate pattern  $(\lambda_1^*, \dots, \lambda_M^*)$  which minimizes the long-run average cost per unit time for the entire system:

$$\lambda_k^*(r) = \max[0, \mu_k - \mu_k \{b/(b+2r h_k^{-1} \mu_k - 2)\}^{1/2}], \quad k=1, \dots, M,$$

where  $r$  is uniquely determined with

$$\sum_{k=1}^M \lambda_k^*(r) = \Lambda.$$

Note that  $\lambda_k^*$  is decreasing in  $b$ , where  $b$  is the ratio of the second moment of the service time to the squared mean service time, and hence it is also decreasing in the coefficient of variation  $(b-1)^{1/2}$ .

The preceding works are related to the workload assignment in equilibrium, whereas Filipiak [11] has dealt with the nonstationary workload assignment to multiple work stations. In his model, it is assumed that the system consists of the  $M$  single machine work stations with different service rates subject to a load of jobs that varies from time to time. Any station is capable of processing any of the jobs. Let  $f(t)$  be total arrival rates into the system and  $f_m(t)$  the flow intensity to station  $m$  at time  $t$ . Define the flow pattern as  $\mathbf{f}(t) \equiv (f_1(t), \dots, f_M(t))$ . Denote by  $A_m(t)$  the cumulative number of arrivals until time  $t$  at station  $m$ , by  $D_m(t)$  the cumulative number of jobs that have left the server in time interval  $(0, t)$  and by  $Q_m(t)$  the number of jobs in the station at  $t$ . The following conservation principle holds

$$Q_m(t) = A_m(t) - D_m(t) + Q_m(0).$$

Assume that the time pattern of  $A_m(t)$ ,  $D_m(t)$  and  $Q_m(t)$  have been measured during successive days of normal use of the system, and then the averages  $\bar{A}_m(t)$ ,  $\bar{D}_m(t)$  and  $\bar{Q}_m(t)$  have been calculated. Using the averages  $\bar{A}_m(t)$  and

$\bar{Q}_m(t)$ ,  $f_m(t)$  and  $x_m(t)$  are represented by

$$f_m(t) \equiv d\bar{A}_m(t)/dt \text{ and } x_m(t) \equiv \bar{Q}_m(t).$$

By approximating the intensity of an outgoing flow by  $c_m \mu_m(x_m(t))$ , we have

$$(2.1) \quad dx_m(t)/dt = -c_m \mu_m(x_m(t)) + f_m(t), \quad 0 \leq \mu(x_m(t)) \leq 1, \quad m=1, \dots, M,$$

where  $c_m$  is the processor capacity. For a given  $f(t)$ , it holds that

$$(2.2) \quad \sum_{m=1}^M f_m(t) = f(t), \quad 0 \leq f_m(t) \leq f(t).$$

The total waiting time  $W$  of all jobs in the system during  $(0, T)$  is given by

$$(2.3) \quad W = \int_0^T \sum_{m=1}^M x_m(t) dt.$$

The optimization problem is to determine flow pattern  $f(t)$  to the system described by (2.1) so as to minimize the total waiting time  $W$  under constraint (2.2). From Pontryagin minimum principle, necessary conditions for optimality are derived. Using these necessary conditions and equations (2.1) and (2.2), an algorithm for obtaining  $f_m(t)$ ,  $m=1, \dots, M$  has been developed [11]. The performance obtained by the algorithm is compared with that of the policy of joining the shorter queue (JSQ).

It should be noted that JSQ is known to minimize each customer's individual expected delay and the long-run average delay per customer [59], [67] when the service time distribution has non-decreasing hazard rate [59] and identical customers arrive according to a stochastic process. However, Whitt [66] has shown that there are service time distributions for which it is not optimal for a customer to always join the shortest queue.

Shanthikumar [47], [48] has considered a loading problem on a production system modelled as M/G/1 queue. External arrival jobs are received in a storage in front of the system called dispatch area from which their releases to the system are controlled.

### 3. Job Selection

When the production system consists of a single work station with  $m$  machines, the system is represented by a queuing model A/B/m where A and B describe the interarrival time distribution and service time distribution, respectively. For simplicity, we use this notation A/B/m in the sequel of this paper.

When the capacity of the production system is limited, it may not be feasible to accept all jobs that demand service in the system. Then, it may be profitable to reject a job with small reward for the purpose of keeping the capacity available for a job with large reward. Thus, it is important to control the input to the system by accepting or rejecting external arrival jobs, which is called job selection or arrival control.

Selection of arrival jobs in an isolated service facility has first been studied by Naor [35] (see [24], [52] for survey on the single-facility problem). We consider the job selection problems in a single and two work stations, and show the monotonicity of the optimal control policies.

Matsui [30] has discussed a job selection problem in M/G/1 system in which both the marginal reward  $S$  and the processing time of each job are independent of the arrival process. A selection criterion  $C_u$  is a control variable depending on the number  $u$  of jobs in the system. An arriving job is accepted or rejected according as  $S \geq C_u$  or  $S < C_u$ . Let  $N$  be the storage capacity. Define a vector of selection criteria  $f$  as

$$f \equiv (C_0, C_1, \dots, C_{N-1}).$$

The problem is to decide the selection criteria  $f$  so as to maximize the expected reward rate. In periodic selection policy (PSP),  $C_u$  is decided at each service completion. In dynamic selection policy (DSP),  $C_u$  is decided at each arrival epoch. Numerical results [30] conjecture that

- (1) under the optimal policy,  $C_u$  monotonically increases with  $u$ ; and
- (2) the policy DSP is superior to the policy PSP.

Optimal job selection problems for GI/M/s queuing systems have been studied by Stidham [51]. Applying Schäl's results [44], Stidham [51] has shown that a control limit policy is optimal under convex nondecreasing cost function in the GI/M/s queuing system. Under the control limit policy, a job is admitted into the queue if and only if fewer than  $n$  jobs are present in the system where  $n$  is a given constant called a control limit.

Mendelson and Yechiali [31] have studied an alternative job selection for GI/M/1 queuing system. It is well known [51], [55], [70 - 72] that, when decision epochs are restricted to arrival instants, the reward-maximizing control policy is a control limit policy. By generalizing this simple control limit policy, Mendelson and Yechiali [31] have considered an  $(n, t)$ -policy such that the  $n$ -th job in the queue is accepted unless  $t$  units of time have already elapsed without any service completion where  $n$  is the control limit. They have developed conditions under which the  $(n, t)$ -policy is preferable to the simple control limit policy.

An optimal job selection problem allowing rejection of external arrivals in two work stations in series has been studied by Ghoneim and Stidham [13]. Their main results are as follows:

The optimal policies are monotonic, in the sense that the rejection region is an increasing set [52], that is, if the decision of rejection is optimal in a certain state of the system, then it is also optimal in a more congested state.

#### 4. Production Rate Control (Service Rate Control)

We consider the problem of controlling service (processing) rate or the number of active parallel machines in a production system (see [9]). Controls by adjusting the service rate for M/G/1 system have been studied by Mitchell [33] and Doshi [10]. Controls by turning servers on and off have been studied by Heyman [19], Bell [2] and Sobel [49] for a single server. For an M/M/s system, the control problems have been studied by Bell [3], Huang et al. [21], Bell [4] and Szarkowicz and Knowles [53].

Szarkowicz and Knowles [53] have shown a monotone form of optimal control policy for an M/M/s system. In their model, the system state  $(x_1, x_2)$  is defined by  $x_1$ , the number of jobs in the system, and  $x_2$ , the number of active machines. Let  $c_1(x_1)$  be the holding cost rate,  $c_2(x_2)$  the service operating cost rate, and  $u$  the control action, i.e., the number of active machines after the control. The start-up and shut-down cost  $C(x_2, u)$  is given by

$$\begin{aligned} C(x_2, u) &= C_1 \cdot (u - x_2), \quad u \geq x_2, \\ &= C_2 \cdot (u - x_2), \quad u \leq x_2, \end{aligned}$$

where  $C_1 \geq 0$  and  $C_2 \leq 0$ . The interarrival times are exponential random variables with parameter  $a$ . The service times are also exponentially distributed with parameter  $b$ . Define  $(x_1 \wedge x_2) \equiv \min(x_1, x_2)$ . Following the idea of Lippman's exponential clock [28], we define  $\ell(x, u) \equiv a + (x_1 \wedge u)b$  and  $L \equiv \max_{x, u} [\ell(x, u)] = a + Sb$  where  $S$  is the maximum number of operating machines.  $L$  is the exponential parameter that yields the minimum expected duration. The state transition probabilities  $p(x' | x, u)$  are

$$\begin{aligned} p(x' | x, u) &= [L - a - (x_1 \wedge u)b] / L && \text{if } x' = x, \\ &= a / L && \text{if } x' = (x_1 + 1, u), \\ &= (x_1 \wedge u)b / L && \text{if } x' = (x_1 - 1, u), \\ &= 0 && \text{otherwise.} \end{aligned}$$

Let  $V(n; x_1, x_2)$  be the minimum expected discounted cost for the last  $n$  transitions, given initial states  $x_1$  and  $x_2$ . An  $n$ -step functional equation of dynamic programming is given by

$$\begin{aligned} V(n; x_1, x_2) &= \min_{u \in U} [C(x_2, u) + W(n; x_1, u)], \\ W(n; x_1, u) &= \sum_{(x_1^-, u) \in X} P((x_1^-, u) | (x_1, x_2), u) \int_0^\infty \int_0^t \exp(-\alpha s) \\ &\quad \times (c_1(x_1) + c_2(u)) ds + \exp(-\alpha t) V(n-1; x_1^-, u) \} \\ &\quad \times \exp\{-\ell(x, u)t\} dt, \end{aligned}$$

where  $\alpha$  is a discount rate,  $X = (0, 1, 2, \dots) \times (0, 1, \dots, S)$  is the countable state space, and  $U = (0, 1, \dots, S)$  the finite control action space. Define  $v(n; x_1, x_2, u)$ ,  $v_1(n; x_1, x_2, u)$  and  $v_2(n; x_1, x_2, u)$  as

$$\begin{aligned} v(n; x_1, x_2, u) &\equiv v_1(n; x_1, x_2, u), \quad u \geq x_2, \\ &\equiv v_2(n; x_1, x_2, u), \quad u \leq x_2, \\ v_1(n; x_1, x_2, u) &\equiv C_1 \cdot (u - x_2) + W(n; x_1, u), \text{ and} \\ v_2(n; x_1, x_2, u) &\equiv C_2 \cdot (u - x_2) + W(n; x_1, u). \end{aligned}$$

Furthermore, define the following:

$$\begin{aligned} z(n; x_1) &\equiv \min_u [\arg \min [v_1(n; x, u)]], \text{ and} \\ Z(n; x_1) &\equiv \min_u [\arg \min [v_2(n; x, u)]], \end{aligned}$$

that is,  $z(n; x_1)$  is the smallest value of  $u$  that minimizes  $v_1(n; x, u)$ .

**Theorem 4.1.** [53] For each state  $x \in X$  and for each  $n=1, 2, \dots, N$ , the optimal action  $u^*$  is

$$\begin{aligned} u^* &= z(n; x_1), \quad x_2 \leq z(n; x_1), \\ &= u(n; x_1), \quad z(n; x_1) \leq x_2 \leq Z(n; x_1), \\ &= Z(n; x_1), \quad Z(n; x_1) \leq x_2, \end{aligned}$$

where  $u(n; x_1) = \arg \min_{u \in U} [v(n; x, u)]$ .

That is, the optimal policy is given by the control limit form. Note that the theorem shows the optimality of the control limit policy under less restrictive assumptions than those used by Huang et al. (e.g., convexity of the cost function) [21].



Graves [15] has considered the problem of controlling processing rate at work stations in an FMS. An FMS is a very flexible production facility that consists of a set of versatile work stations. It is capable of processing a wide variety of jobs. Because of the lack of a dominant work flow, production control is often very difficult in an FMS. For each work station we assign a planned lead time. While the actual time spent at the work station will deviate from the planned lead time, one of the objectives of the planned lead time control should be to minimize the variance of this deviation. Graves [15] has derived an expression for the variance.

Ohno and Ichiki [37] has considered an optimal control of service rates for a tandem queueing system that has a general cost structure. They have also proposed a procedure of modified policy iteration algorithm for finding the optimal control, and evaluated its efficiency in comparison with several variants of the algorithm.

## 5. Production/Inventory Control

In a Material Requirement Planning (MRP) System or a Dependent Demand System, it is important to control the production process depending on the inventory level of the products for which there are external demands. This kind of control is referred as production/inventory control.

Schmidt and Nahmias [45] have considered an MRP-type assembly system in which an end product is assembled from two components externally supplied. Demands for the end product are assumed to be random. Using the functional equation approach of dynamic programming, they have characterized the forms of both the optimal order policy for the components and the optimal assembly policy of the end product as outlined in the following.

Let  $\bar{x}$  be the inventory level of end product,  $x_{i0}$  be the sum of  $\bar{x}$  and the inventory level of component  $i$ ,  $i=1,2$ , and  $x_{ij}$  be the sum of  $x_{i0}$  and number of components scheduled to arrive in the next  $j$  periods. Define  $\underline{x}_i$  as

$$\underline{x}_i = (x_{i0}, x_{i1}, \dots, x_{i\ell_i-1}).$$

Let decision variables  $y$  be the sum of  $\bar{x}$  and the amount of new assembly, and  $y_i$  the sum of  $x_{i\ell_i-1}$  (this is the classical inventory level) and the amount of new order for component  $i$ . Denote by  $c$  the unit assembly cost of the end product, by  $h$  the unit holding cost of the end product, by  $p$  the unit penalty cost of the end product, by  $\ell$  the number of periods required to assemble the end product, by  $c_i$  the unit order cost of component  $i$ , by  $h_i$  the unit holding

cost of component  $i$ , by  $\ell_i$  the lead time of ordering component  $i$  and by  $\beta$  the one period discount factor. It is assumed that  $\ell_1 \leq \ell_2$ ,  $\beta^\ell (h_1 + h_2 + p) > c$  and  $(1 - \beta)c + \beta^\ell (h - h_1 - h_2) > 0$ . The problem is to decide  $y$  and  $y_i$ ,  $i=1,2$ , so as to minimize the expected cost.

Let  $C_n(\bar{x}, x_1, x_2)$  be the minimum expected cost for  $n$  periods when the current state of the system is  $(\bar{x}, x_1, x_2)$ . The functional equations defining an optimal policy are given by

$$\begin{aligned} C_n(\bar{x}, x_1, x_2) = \min \{ & c(y - \bar{x}) + L_n(y) + \sum_{i=1}^2 [c_i(y_i - x_{i\ell_i-1}) + h_i x_{i0}] \\ & + \beta \int_0^\infty C_{n-1}(y - \xi, x_{11} - \xi, \dots, y_1 - \xi, x_{21} - \xi, \dots, y_2 - \xi) d\phi(\xi) \mid \\ & \bar{x} \leq y \leq \min(x_{10}, x_{20}), y_i \geq x_{i\ell_i-1} \}, \end{aligned}$$

where

$$\begin{aligned} L_n(y) &= 0 \quad \text{for } n < \ell + 1, \\ L_n(y) &= \beta^\ell \int_0^\infty [h(y - \xi)^+ - (h_1 + h_2)(y - \xi)] d\phi^{(\ell)}(\xi) \\ &\quad + \beta^\ell \int_0^\infty p(\xi - y)^+ d\phi^{(\ell+1)}(\xi) \quad \text{for } n \geq \ell + 1, \\ c_0(\cdot) &\equiv 0, \\ x^+ &= \max(0, x), \end{aligned}$$

and  $\xi$  denotes demands.  $C_n$  can be shown to be separable in  $\bar{x}$  and  $(x_1, x_2)$ . That is,

$$C_n(\bar{x}, x_1, x_2) = D_n(\bar{x}) + g_n(x_1, x_2).$$

Let  $\hat{y}_n$  denote the minimum point of

$$cy + L_n(y) + \beta \int_0^\infty D_{n-1}(y - \xi) d\phi(\xi).$$

**Theorem 5.1.** [45] The optimal assembly policy has the following structure.

$$\begin{aligned} y_n^* &= \min(\hat{y}_n, x_{10}, x_{20}) \quad \text{if } \bar{x} < \hat{y}_n, \\ &= \bar{x} \quad \text{if } \bar{x} \geq \hat{y}_n. \end{aligned}$$

In order to derive the optimal order policy in period  $n$ , for  $\ell + \ell_1 \leq n \leq \ell + \ell_2$ , we define  $\beta_n(u)$  and  $\bar{h}_n(y)$  as

$$\begin{aligned}
\beta_n(u) &= (1-\beta)c_1 + \beta^{\ell_1} h_1 + \beta^{\ell_1} \int_0^\infty \tilde{\Lambda}'_{n-\ell_1}(u-\xi) d\phi^{(\ell_1)}(\xi) \\
&\quad + \beta \int_0^{(u-u_{1n-1})^+} \beta_{n-1}(u-\xi) d\phi(\xi) \quad \text{for } \bar{\ell} \equiv \ell + \ell_1 < n \leq \ell + \ell_2, \\
&= c_1 + (\beta^{\ell_1} + \dots + \beta^{\bar{\ell}+1}) h_1 + \beta^{\ell_1} \int_0^\infty \tilde{\Lambda}'_{\bar{\ell}}(u-\xi) d\phi^{(\ell_1)}(\xi) \quad \text{for } n = \bar{\ell},
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\Lambda}_n(y) &= c[\min(y, \hat{y}_n) - \hat{y}_n] + L_n[\min(y, \hat{y}_n)] - L_n(\hat{y}_n) \\
&\quad + \beta \int_0^\infty \{D_{n-1}[\min(y, \hat{y}_n) - \xi] - D_{n-1}(\hat{y}_n - \xi)\} d\phi(\xi),
\end{aligned}$$

where the symbol ' denotes the derivative, and  $u_{1n}$  is the unique zero of  $\beta_n(u)$ .

Theorem 5.2. [45] Assume that

$$c_1 + \beta^{\ell_1} (1-\beta)c + (\beta^{\ell_1} + \dots + \beta^{\ell+\ell_1}) h_1 < \beta^{\ell+\ell_1} (h_1 + h_2 + p).$$

The optimal order policy in period  $n$ , for  $\ell + \ell_1 \leq n \leq \ell + \ell_2$ , is

$$\begin{aligned}
y_1^* &= x_{1\ell_1-1} && \text{if } x_{1\ell_1-1} \geq x_{2\ell_1} \text{ or } u_{1n} \leq x_{1\ell_1-1} \leq x_{2\ell_1}, \\
&= u_{1n} && \text{if } x_{1\ell_1-1} \leq u_{1n} \leq x_{2\ell_1}, \\
&= x_{2\ell_1} && \text{if } x_{1\ell_1-1} \leq x_{2\ell_1} \leq u_{1n}, \\
y_2^* &= x_{2\ell_2-1}.
\end{aligned}$$

For the purpose of derivation of the optimal order policy in period  $n$ , for  $n \geq \ell + \ell_2 + 1 \equiv \hat{\ell}$ , we define

$$\begin{aligned}
\sigma_n(y_2) &= (1-\beta)c_2 + \beta^{\ell_2} h_2 + \beta^{\ell_2} \int_0^\infty \tilde{\Lambda}'_{n-\ell_2}(y_2-\xi) d\phi(\xi) \\
&\quad + \beta \int_0^{(y_2-v_{n-1})^+} \sigma_{n-1}(y_2-\xi) d\phi(\xi) \quad \text{for } n > \hat{\ell}, \\
\sigma_{\hat{\ell}}(u) &= c_2 + (\beta^{\ell_2} + \dots + \beta^{\hat{\ell}-1}) h_2 + \beta^{\ell_2} \int_0^\infty \tilde{\Lambda}'_{\hat{\ell}-\ell_2}(u-\xi) d\phi^{(\ell_2)}(\xi),
\end{aligned}$$

$$\begin{aligned}\gamma_n(y_2) &= (1-\beta)c_2 + \beta^{\ell_2}h_2 + \lambda_{n,\ell_2-\ell_1+1}(y_2) \\ &\quad + \beta \int_0^{(y_2-u_{2n-1})^+} \gamma_{n-1}(y_2-\xi) d\phi(\xi) \quad \text{for } n > \hat{\ell}, \text{ and} \\ \gamma_{\hat{\ell}}(u) &= c_2 + (\beta^{\ell_2} + \dots + \beta^{\hat{\ell}-1})h_2 + \lambda_{\hat{\ell},\ell_2-\ell_1+1}(u)\end{aligned}$$

where  $\lambda_{ij}(u) \equiv \beta^{j-1} \int_{(u-u_{1i+1-j})^+}^{\infty} \beta_{i+1-j}(u-\xi) d\phi^{(j-1)}(\xi)$ , and  $u_{2n}$  is the unique zero of  $\gamma_n(y_2)$ .

Theorem 5.3. [45] Assume

$$\begin{aligned}c_2 + \beta^{\ell_2}(1-\beta)c + (\beta^{\ell_2} + \dots + \beta^{\hat{\ell}-1})h_2 &< \beta^{\ell+\ell_2}(h_1+h_2+p); \text{ and} \\ c_2 + \beta^{\ell_2-\ell_1}c_1 + \beta^{\ell_2}(1-\beta)c + (\beta^{\ell_2} + \dots + \beta^{\hat{\ell}-1})(h_1+h_2) &\leq \beta^{\hat{\ell}-1}(h_1+h_2+p).\end{aligned}$$

The optimal policy for the components in period  $n \geq \hat{\ell}$  has the following form:

$$\begin{aligned}y_1^* &= x_{1\ell_1-1}, \quad (x_{1\ell_1-1} \geq x_{2\ell_1}) \text{ or } (u_{1n} \leq x_{1\ell_1-1} \leq x_{2\ell_1}), \\ &= u_{1n}, \quad x_{1\ell_1-1} \leq u_{1n} \leq x_{2\ell_1}, \\ &= x_{2\ell_1}, \quad x_{1\ell_1-1} \leq x_{2\ell_1} \leq u_{1n}, \\ y_2^* &= x_{2\ell_2-1}, \quad (x_{2\ell_2-1} \geq x_{1\ell_1-1} \text{ and } x_{2\ell_2-1} \geq u_{2n}) \\ &\quad \text{or } (x_{1\ell_1-1} \geq x_{2\ell_2-1} \geq v_n) \\ &= x_{1\ell_1-1}, \quad u_{2n} \leq x_{1\ell_1-1} \leq v_n \text{ and } x_{1\ell_1-1} \geq x_{2\ell_2-1}, \\ &= v_n, \quad x_{2\ell_2-1} \leq v_n \leq x_{1\ell_1-1}, \\ &= u_{2n}, \quad x_{1\ell_1-1} \leq u_{2n} \text{ and } x_{2\ell_2-1} \leq u_{2n},\end{aligned}$$

where  $v_n$  is the unique zero of  $\sigma_n(y_2)$ , and it can be shown that  $u_{2n} < v_n$ .

Note that under the optimal order policy,  $y_i^*$ ,  $i=1,2$ , is a nondecreasing function of the inventory level of the other component. The optimal assembly policy is a base stock policy (that is, assemble up to  $\hat{y}_n$ ) unless it exceeds the available quantity of components.

The management of a system, which combines standard inventory and queuing submodels, presents a rather complex trade-off: if there were unlimited production capacity, product inventories could be kept small by making products in small batches. However, as batch sizes are reduced, the workload in the system increases, causing greater congestion. Hence larger safety stocks are required to protect against production delays. Assume that batches are ordered according to a  $(q,r)$  policy, that is, when the inventory level reaches the reorder point  $r$ , a batch of size  $q$  is ordered. The optimality of the  $(q,r)$  policy in the basic inventory model has been proved by Veinott [56]. The inventory process can be represented by one of the continuous-time, single-location stochastic-leadtime models (e.g., see Hadley and Whitin [16], chapter 4). The production facility is represented as one or more servers in a queuing system in which a server serves one batch at a time. The sojourn times (total delay plus service times) in the queuing system in turn become the order lead times for the inventory submodel. Combining these submodels by treating the mean sojourn time in the queuing system as a linking variable, Zipkin [74] has derived a convex program.

Production/inventory problems, in which processing rate is controlled by choosing one of two processing rates, have been studied by Gavish [12] and Kok [26]. They have shown that the optimal policy is two-critical-number policy, that is, when the inventory level reaches one of the two critical numbers, the processing rate is changed.

Venkatesan [57] has combined the production/inventory control and the equipment replacement control. A single-product single-equipment production-inventory system with infinite storage and production capacity is reviewed periodically. He has provided sufficient conditions for the combined optimal policy to have a relatively simple monotonic structure. The production or replacement decisions are made on the basis of the inventory level in storage and the level of deterioration of the equipment. It is assumed that the deterioration process is a finite state Markov process with transition probabilities  $q_{ij}$  and states  $0, 1, \dots, F$ . The process is independent of the demand process. Replacement is instantaneous. The production cost is a function of the level of deterioration of the equipment.

The system state  $(i, x)$  is defined by  $i$ , the level of deterioration and  $x$ , the inventory level. An action  $(a, y)$  is defined by  $a=1(0)$  indicating replacement (no replacement), and  $y$ , the sum of  $x$  and the amount produced. Four types of costs are considered. Let  $R$  be a replacement cost for equipment and  $A$  a fixed penalty cost for replacement through failure. Define  $\delta(y-x)=1$  for  $y > x$ ,  $=0$  otherwise. Define the production cost for each  $i < F$  as

$$\delta(y-x)K_i + c_i(y-x),$$

where  $K_i$  denotes the start-up cost, and  $x$  and  $y$  are inventory levels before and after replenishment. Let  $L(y)$  be the one-period holding and shortage cost function which is assumed independent of  $i$ . Let  $f_0(j, x')$  be a cost for terminating the system in state  $(j, x')$ . One-period cost function for state  $(i, x)$  and action  $(a, y)$  for  $0 < i < F$  is

$$r((i, x), (1, y)) = R + \delta(y-x)K_0 + c_0(y-x) + L(y),$$

$$r((i, x), (0, y)) = \delta(y-x)K_i + c_i(y-x) + L(y).$$

The expected discounted cost function in state  $(i, x)$  by following policy  $\pi_n$  is

$$\begin{aligned} f_n(i, x | \pi_n) &= E_{\pi_n} \left\{ \sum_{m=0}^{n-1} \beta^{n-m-1} [r((J_m, X_m), (A_m, Y_m))] | X_n = x, J_n = i \right\} \\ &\quad + E_{\pi_n} \{ \beta^n f_0(j, x') \}, \end{aligned}$$

where  $(J_m, X_m)$  and  $(A_m, Y_m)$  are random vectors representing the state and the action at stage  $m$ . Let  $f_n(i, x)$  be the optimal cost function, that is

$$f_n(i, x) = f_n(i, x | \pi_n^*) = \inf_{\pi_n} f_n(i, x | \pi_n).$$

For notational convenience, we use the following notation:

$$w(i, y) \equiv c_i y + L(y), \text{ and}$$

$$g(i, y) \equiv (c_i - \beta c_{F-1})y + L(y) \text{ for } i < F.$$

Define  $\underline{s}_i$  as the infimum point  $y$  of  $w(i, y)$ ,  $Q_i$  as the infimum point  $y$  of  $g(i, y)$ ,  $\underline{s}_i (\leq \underline{s}_i)$  as the smallest number satisfying  $w(i, \underline{s}_i) = K_i + w(i, \underline{s}_i)$ , and  $\bar{s}_i$  and  $\bar{Q}_i$  as the upper bounds satisfying  $\underline{s}_i \leq \bar{s}_i \leq Q_i \leq \bar{Q}_i$ ,  $g(i, \bar{s}_i) = g(i, Q_i) + \beta K$  and  $g(i, \bar{s}_i) = g(i, Q_i) + (1 - \beta)K$ , respectively where  $K_i = K$  follows from condition 5 below.

Assume the following conditions:

Condition 1.  $r_q(i) \equiv \sum_{j=q}^F q_{ij}$  is nondecreasing in  $i < F$  for each  $q = 0, 1, \dots, F$ ,

Condition 2.  $0 \leq c_i < \infty$  and  $0 \leq K_i < \infty$  are nondecreasing in  $i < F$ ,

Condition 3. For all  $x$ ,  $f_0(i, x) = 0$ ,  $i < F$ ,  $= R + A$ ,  $i = F$ ,

Condition 4.  $L(y)$  is continuous and satisfies the following;

- (a)  $L(y)$ ,  $\{c_i y + L(y)\}$ , and  $\{(c_i - \beta c_{F-1})y + L(y)\}$  are quasiconvex [53] in  $y$  for all  $i < F$ ,
- (b)  $\{c_{F-1} y + L(y)\} \rightarrow \infty$  as  $y \rightarrow \infty$ ,
- (c)  $\{(c_0 - c_{F-1})y + L(y)\} \rightarrow \infty$  as  $y \rightarrow \infty$ ,

Condition 5.  $K_i = K \geq 0$  for all  $i$  and  $(1-\beta)K \geq g(i, S_i) - g(i, Q_i)$ , and

Condition 6. For all  $i < F$ ,  $R \geq \beta(q_{iF} - q_{0F})A / [1 - \beta(1 - q_{0F})]$ .

Let  $\delta_n^*(i, x)$  be the optimal decision rule at stage  $n$ .

**Theorem 5.4.** [57] Under conditions 1-6, for any period  $n \geq 1$ , if the equipment is observed to be in state  $i$ ,  $0 < i < F$ , and the on-hand inventory is  $x$ , then there exists an optimal policy of the form  $(\sigma_i(n), s_i(n), S_i(n), S_0(n))$  such that  $\sigma_i(n) \leq s_0(n)$ ,  $\underline{s}_i < s_i(n) \leq \bar{s}_i \leq S_i \leq S_i(n) \leq \bar{S}_i$  and either

(1)  $\sigma_i(n) < s_i(n)$  and

$$\begin{aligned}\delta_n^*(i, x) &= (1, S_0(n)), & x \in (-\infty, \sigma_i(n)], \\ &= (0, S_i(n)), & x \in (\sigma_i(n), s_i(n)], \\ &= (0, x), & x \in (s_i(n), \infty),\end{aligned}$$

or

(2)  $\sigma_i(n) \geq s_i(n)$  and

$$\begin{aligned}\delta_n^*(i, x) &= (1, S_0(n)), & x \in (-\infty, \sigma_i(n)], \\ &= (0, x), & x \in (\sigma_i(n), \infty).\end{aligned}$$

It is remarked that the dynamic programming recursion for finding  $s_i(n)$ ,  $S_i(n)$  and  $\sigma_i(n)$  can be made more efficient by using their boundaries  $\underline{s}_i$ ,  $\bar{s}_i$ ,  $\underline{S}_i$  and  $\bar{S}_i$  for reducing the search ranges.

## 6. System Configuration

An important layout problem for a production system is to determine the best location of two or more service stations in series with no precedence constraints. For given external arrival process and service time distributions, the objective is to determine the order of the stations that minimizes the expected equilibrium sojourn time per job. Weber's result [60] shows that the final departure process is independent of the order of the stations for arbitrary arrival process, given exponential or deterministic service time distributions at all stations with unlimited storages. Whitt [65] has dropped the exponential or deterministic assumption, and has derived approximation methods. In his methods, each station is treated as a GI/G/1 queuing system characterized by first two moments of arrival and departure processes.

The squared coefficient of variation (scv) of the renewal interval in the approximating renewal process is

$$C_d^2 = \rho^2 C_s^2 + (1-\rho^2) C_a^2, \quad \rho = \lambda\tau < 1,$$

where  $C_a^2$ ,  $C_s^2$  and  $C_d^2$  are scv's of arrival, service and departure processes, respectively,  $\lambda$  is the arrival rate, and  $\tau$  is the mean service time [62]-[64]. The approximate expected equilibrium queuing delay EW in a GI/G/1 queue is given [65] by

$$EW = \tau\rho(C_a^2 + C_s^2)/2(1-\rho).$$

Problem is to choose the smallest of the following

$$EW_1 + \dots + EW_n,$$

for  $n!$  permutations where  $n$  is the number of stations. He has obtained the following conjecture [65]:

If  $C_{s1}^2 \leq C_{s2}^2 \leq \dots \leq C_{sn}^2$ ,  $C_a^2 < 1$  and  $\tau_1 = \tau_2 = \dots = \tau_n$ , then,  $(1, 2, \dots, n)$  is the best order.

Consider the case of finite intermediate storage. In a saturated system with arbitrary service time distributions and finite intermediate storages in series, reversing of the order of stages does not influence the throughput. This, so-called reversibility property has been proved by Yamazaki and Sakasegawa [69], and independently by Muth [34] (see [68] for review of the reversibility). Thus, the production rate (throughput) of a saturated two-stage system can be affected by two factors: the proper workload division between stages and the capacity of intermediate storage. Rao [40] and Wolisz [68] have derived closed form expressions for maximal production rate of the saturated two-stage system for the case where the distribution of service time is exponential at one of the stages, but arbitrary at the other one.

Let  $E(b_i)$ ,  $i=1,2$ , be the mean of service time  $b_i$  with the service time distribution function  $B_i(t)$ . Let  $c_i$  be the coefficient of variation of  $b_i$ . It is assumed that  $E(b_1) + E(b_2) = \text{constant}$ ,  $E(b_1) = 1/\mu_1$ ,  $c_1 = 1$  and  $c_2 \geq 0$ . Rao [40] has shown that if  $0 \leq c_2 < 1$ , then the throughput for  $E(b_1) < E(b_2)$  is greater than that for  $E(b_1) = E(b_2)$ . He then pointed out that unbalancing of such a system in the direction of allotting a slightly higher load to the less variable stage increases system's throughput. Wolisz [68] has shown that if  $c_2 > 1$ , then the relation of Rao does not always hold. When both  $B_1(\cdot)$  and  $B_2(\cdot)$  are exponential, it is shown by Hillier and Boling [20] that the throughput is maximized at  $E(b_1) = E(b_2)$ . The delay time in the finite intermediate storages in series with general service times is discussed by Sakasegawa and Yamazaki [43].



Kubat and Sumita [27] have considered a tandem queuing system with  $N$  unreliable machines.  $M$  storages with unlimited capacity and  $K$  additional equipments to lengthen uptime are available. The problem is how to determine which machines should have additional equipments and where to locate storages, so as to maximize the throughput. They have developed a bivariate dynamic programming procedure.

Vinod and Solberg [58] have considered the problem of allocating machines to work stations in an FMS modelled as a closed network of queues so as to determine the system configuration that is optimal cost effective.

Berman, Larson and Chiu [6] have considered a problem of locating a facility for a single mobile server (transporter) on network so as to minimize the response time, i.e., the sum of mean queuing delay and mean travel time. Demands for service arise as a Poisson process only on the nodes of a network  $G$ . A single mobile server resides at a facility located on  $G$ . Using the local convexity of the expected response time, the algorithm for finding the optimal location has been developed in [6]. The key results are as follows [6]:

- (1) for very small or very large network-wide demand arrival rate  $\lambda$ , the optimal facility location is the Hakimi point that minimizes mean service time [17], [18];
- (2) for intermediate values of  $\lambda$ , the optimal location is typically not the Hakimi point, but is a point either on a link or a node "between" the Hakimi point and the point that minimizes the second moment of the service time.

## Conclusion

Various aspects of optimization problems in production systems have been surveyed. The combined problems such as MRP-type production/inventory problem, production/inventory/replacement problem, etc., are desired to be solved in practice. The anticipated computational complexity of exploding state spaces may prohibit the practical use of such models. Current research in approximation methods for solving large Markov chains such as decomposition method [8], and aggregation-disaggregation method [46], [54], and approximation methods for large scale Markov decision processes [32], [36], [61], [73] may lead to practical computational procedures for these more complex production systems.

## Acknowledgment

The authors are grateful to Professor H. Mine for his encouragement, to Professor K. Ohno for his kind advices, to Professor H. Sakasegawa for discussing some of his own work and to Professor T. Kushiya for general discussion. They also wish to thank Professor T. Ibaraki for his helpful remarks concerning the organization of the present paper.

## References

- [1] Baskett, F., Chandy, K. M., Muntz, R. R. and Palacios, F. G.: Open, Closed and Mixed Networks of Queues with Different Classes of Customers. *J. Assoc. Comput. Mach.* Vol.22 (1975), 248-260.
- [2] Bell, C. E.: Characterization and Computation of Optimal Policies for Operating and M/G/1 Queuing System with Removable Server. *Opns. Res.* Vol.19 (1971), 208-218.
- [3] Bell, C. E.: Turning Off a Server with Customers Present: Is This Any Way to Run an M/M/c Queue with Removable Servers. *Opns. Res.* Vol.23 (1975), 571-574.
- [4] Bell, C. E.: Optimal Operation of an M/M/2 Queue with Removable Servers. *Opns. Res.* Vol.28, No.5 (1980), 1189-1204.
- [5] Bell, C. E. and Stidham, S.: Individual versus Social Optimization in the Allocation of Customers to Alternative Servers. *Mgmt. Sci.* Vol.29, No.7 (1983), 831-839.
- [6] Berman, O., Larson, R. C. and Chiu, S. S.: Optimal Server Location on a Network Operating as an M/G/1 Queue. *Opns. Res.* Vol.33, No.4 (1985), 746-771.
- [7] Buzacott, J. A. and Yao, D. D.: Flexible Manufacturing Systems: A Review of Analytical Models. *Mgmt. Sci.* Vol.32, No.7 (1986), 890-905.
- [8] Courtois, P. J.: Decomposability, Queuing and Computer Systems Applications. Academic Press, New York 1977.
- [9] Crabill, T. B., Gross, D. and Magazine, M. J.: A Classified Bibliography of Research on Optimal Design and Control of Queues. *Opns. Res.* Vol.25 (1977), 219-232.
- [10] Doshi, B. T.: Optimal Control of the Service Rate in an M/G/1 Queuing System. *Adv. Appl. Prob.* Vol.10 (1978), 682-701.
- [11] Filipiak, J.: Dynamic Routing in a Queuing System with a Multiple Service Facility. *Opns. Res.* Vol.32, No.5 (1984), 1163-1180.

- [12] Gavish, B. and Graves, S. C.: A One-Product Production/Inventory Problem under Continuous Review Policy. *Opns. Res.* Vol.28, No.5 (1980), 1228-1236.
- [13] Ghoneim, H. A. and Stidham, S.: Control of Arrivals to Two Queues in Series. *Eur. J. Opns. Res.* Vol.21 (1985), 399-409.
- [14] Gordon, W. J. and Newell, G. F.: Closed Queuing System with Exponential Server. *Opns. Res.* Vol.15 (1967), 254-265.
- [15] Graves, S. C.: A Tactical Planning Model for a Job Shop. *Opns. Res.* Vol.34, No.4 (1986), 522-533.
- [16] Hadley, G. and Whitin, T.: Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs, N.J. 1963.
- [17] Hakimi, S. L.: Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Opns. Res.* Vol.12 (1964), 450-459.
- [18] Hakimi, S. L.: Optimum Distribution of Switching Centers on a Communications Network and Some Related Graph Theoretic Problems. *Opns. Res.* Vol.13 (1965), 462-475.
- [19] Heyman, D. P.: Optimal Operating Policies for M/G/1 Queuing Systems. *Opns. Res.* Vol.16 (1968), 362-382.
- [20] Hillier, F. S. and Boling, R. W.: The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Operation Time. *J. Industrial Engineering.* Vol.17 (1966), 651-658.
- [21] Huang, C. C., Brumelle, S. L. and Sawaki, K.: Optimal Control for Multi-Servers Queuing Systems under Periodic Reviews. *Naval Res. Logist. Quart.* Vol.24 (1977), 127-135.
- [22] Hutchinson, G. K. and Wynne, B. E.: A Flexible Manufacturing System. *Industrial Engineering.* (December 1973), 10-17.
- [23] Jackson, J. R.: Jobshop-like Queuing System. *Mgmt. Sci.* Vol.10 (1963), 131-142.
- [24] Johansen, S. G. and Stidham, S.: Control of Arrivals to a Stochastic Input-Output System. *Adv. Appl. Prob.* Vol.12 (1980), 972-999.
- [25] Kalkunte, M. V., Sarin, S. C. and Wilhelm, W. E.: Flexible Manufacturing Systems: a Review of Modeling Approaches for Design, Jastification and Operation, in Kusiak, A. (Ed.), *Flexible Manufacturing Systems: Methods and Studies* (North-Holland, Amsterdam, 1986), 3-25.
- [26] Kok, A. G.: Approximations for a Lost-Sales Production/Inventory Control Model with Service Level Constraints. *Mgmt. Sci.* Vol.31, No.6 (1985), 729-737.
- [27] Kubat, P. and Sumita, U.: Buffers and Backup Machines in Automatic

- Transfer Lines. *Int. J. Prod. Res.* Vol.23, No.6 (1985), 1259-1270.
- [28] Lippman, S.: Applying a New Device in the Optimization of Exponential Queuing Systems. *Opns. Res.* Vol.23 (1975), 687-710.
- [29] Makino, T.: On the Mean Passage Time Concerning Some Queuing Problems of the Tandem Type. *J. Opns. Res. Soc. Jpn.* Vol.7 (1964), 17-47.
- [30] Matsui, M.: Optimal Order-Selection Policies for a Job Shop Production System. *Int. J. Prod. Res.* Vol.23, No.1 (1985), 21-31.
- [31] Mendelson, H. and Yechiali, U.: Controlling the GI/M/1 Queue by Conditional Acceptance of Customers. *Eur. J. Opns. Res.* Vol.7 (1981), 77-85.
- [32] Mendelssohn, R.: An Iterative Aggregation Procedure for Markov Decision Processes. *Opns. Res.* Vol.30 (1982), 62-73.
- [33] Mitchell, B.: Optimal Service Rate Selection in an M/G/1 Queue. *SIAM J. Appl. Math.* Vol.24 (1973), 19-30.
- [34] Muth, E. J.: The Reversibility Property of Production Lines. *Mgmt. Sci.* Vol.25 (1979), 152-158.
- [35] Naor, P.: The Regulation of Queue Size by Levying Tolls. *Econometrica.* Vol.37 (1969), 15-24.
- [36] Ohno, K.: A Unified Approach to Algorithms with a Suboptimality Test in Discounted Semi-Markov Decision Processes. *J. Opns. Res. Jpn.* Vol.24 (1981), 296-323.
- [37] Ohno, K. and Ichiki, K.: Computing Optimal Policies for Controlled Tandem Queuing Systems. *Opns. Res.* Vol.35 (1987), 121-126.
- [38] Osborn, J.: Direct On-Line Computer Control of Machine Tools and Material Handling. in DeVries, M.A. (Ed.), *The Expanding World of NC*, Numerical Control Society, 1972, 260-268.
- [39] Perry, C. B.: Variable Mission Manufacturing Systems. in DeVries, M.A. (Ed.), *NC: 1971, The Opening Door to Productivity and Profits*, Numerical Control Society, 1971, 409-433.
- [40] Rao, N. P.: On the Mean Production Rate of a Two-Stage Production System of the Tandem Type. *Int. J. Prod. Res.* Vol.13 (1975), 207-217; Two-Stage Production System with Intermediate Storage. *AIIE Trans.* Vol.7 (1975), 414-421.
- [41] Rao, N.P.: A Generalization of the "Bowl Phenomenon" in Series Production System. *Int. J. Prod. Res.* Vol.14 (1976), 437-443.
- [42] Reiser, M. and Kobayashi, H.: Queuing Networks with Multiple Closed Chains: Theory and Computational Algorithms. *IBM J. Res. Develop.* Vol.19 (1975), 283-294.
- [43] Sakasegawa, H. and Yamazaki, G.: Inequalities and an Approximation

- Formula for the Mean delay Time in Tandem Queuing Systems. *Ann. Inst. Stat. Math.* Vol.29 (1977), 445-466.
- [44] Schäl, M.: Conditions for Optimality in Dynamic Programming and for the Limit of N-state Optimal Policies to be Optimal. *Zeit. Wahrsch. Verw. Gebiete.* Vol.32 (1975), 179-196.
- [45] Schmidt, C. P. and Nahmias, S.: Optimal Policy for a Two-Stage Assembly System under Random Demand. *Opns. Res.* Vol.33, No.5 (1985), 1130-1145.
- [46] Schweitzer, P. J.: Aggregation Methods for Large Markov Chains. in Iazeolla, G., Courtois, P. J. and Hordijk, A. (Ed.), *Mathematical Computer Performance and Reliability*, Elsevier Science Publishers B.V. (North-Holland) 1984.
- [47] Shanthikumar, J. G.: M/G/1 Queues with Scheduling within generations and Removable Server. *Opns. Res.* Vol.29, No.5 (1981), 1010-1018.
- [48] Shanthikumar, J. G.: Comparison of Dispatch Policies for a Single Server Queuing Model with Limited Operational Control. *Int. J. Prod. Res.* Vol.22, No.3 (1984), 389-403.
- [49] Sobel, M. J.: Optimal Average Cost Policy for a Queue with Start-up and Shut-down Costs. *Opns. Res.* Vol.17 (1969), 145-162.
- [50] Steckel, K. E. and Solberg, J. J.: The Optimality of Unbalancing Both Workloads and Machine Group Size in Closed Queuing Networks of Multi-server Queues. *Opns. Res.* Vol.33, No.4 (1985), 882-910.
- [51] Stidham, S.: Social and Individual Optimal Control of Arrivals to a GI/M/1 Queue. *Mgmt. Sci.* Vol.24 (1978), 1598-1610.
- [52] Stidham, S.: Optimal Control of Admission to a Queuing System. *IEEE Trans. on Automatic Control*, Vol.AC-30 (1985), 705-713.
- [53] Szarkowicz, D. S. and Knowles, T. W.: Optimal Control of an M/M/s Queuing System. *Opns. Res.* Vol.33, No.3 (1985), 644-660.
- [54] Takahashi, Y.: A Lumping Method for Numerical Calculations of Stationary Distributions of Markov Chains. Res. Report B-18, Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan 1975.
- [55] Teghem, J.: Properties of (0,k) policy in a M/G/1 Queue and Optimal Joining Rules in an M/M/1 Queue with Removable Server, in: Haley, K.B. (Ed.), *Operational Research '75* (North-Holland, Amsterdam, 1976), 229-259.
- [56] Veinott, A. F.: On the Optimality of (s,S) Inventory Policies: New Conditions and a New Proof. *SIAM J.* Vol.14 (1966), 1067-1083.
- [57] Venkatesan, M.: Production-Inventory with Equipment Replacement -PIER. *Opns. Res.* Vol.32, No.6 (1984), 1286-1295.

- [58] Vinod, B. and Solberg, J. J.: The Optimal Design of Flexible Manufacturing Systems. *Int. J. Prod. Res.* Vol.23, No.6 (1985), 1141-1151.
- [59] Weber, R. W.: On the Optimal Assignment of Customers to Parallel Servers *J. Appl. Prob.* Vol.15 (1978), 406-413.
- [60] Weber, R. R.: The Interchangeability of Tandem  $M/M/1$  Queues in Series. *J. Appl. Prob.* Vol.16 (1979), 690-695.
- [61] Whitt, W.: Approximations of Dynamic Programs I. *Mathematics of Opns. Res.* Vol.3 (1978), 231-243.
- [62] Whitt, W.: Approximating a Point Process by a Renewal Process, I: Two Basic Methods. *Opns. Res.* Vol.30 (1982), 125-147.
- [63] Whitt, W.: The Queuing Network Analyzer. *Bell Systems Technical Journal* Vol.62 (1983), 2779-2815.
- [64] Whitt, W.: Approximations for Departure Processes and Queues in Series. *Naval Res. Logist. Quart.* Vol.31 (1984), 499-521.
- [65] Whitt, W.: The Best Order for Queues in Series. *Mgmt. Sci.* Vol.31, No.4 (1985), 475-487.
- [66] Whitt, W.: Deciding Which Queue to Join: Some Counter Examples. *Opns. Res.* Vol.34, No.1 (1986), 55-62.
- [67] Winston, W.: Optimality of the Shortest Line Discipline. *J. Appl. Prob.* Vol.14 (1977), 181-889.
- [68] Wolisz, A.: Production Rate Optimization in a Two-Stage System with Finite Intermediate Storage. *Eur. J. Opns. Res.* Vol.18 (1984), 369-376.
- [69] Yamazaki, G. and Sakasegawa, H.: Properties of duality in Tandem Queuing Systems. *Annal. Inst. Stat. Math.* Vol.27 (1975), 201-212.
- [70] Yechiali, U.: On Optimal Balking Rules and Toll Charges in a  $GI/M/1$  Queuing Process. *Opns. Res.* Vol.19 (1971), 349-370.
- [71] Yechiali, U.: Customers' Optimal Joining Rules for the  $GI/M/s$  Queue. *Mgmt. Sci.* Vol.18 (1972), 434-443.
- [72] Yechiali, U.: How Long Will You Wait for What You Really Want. Proceedings of the Conference on Stochastic Control and Optimization, Free University, Amsterdam, April 5-6, 1979 (1979), 51-56.
- [73] Zipkin, P. H.: Bounds for Row-Aggregation in Linear Programming. *Opns. Res.* Vol.28 (1980), 903-916.
- [74] Zipkin, P. H.: Models for Design and Control of Stochastic, Multi-Item Batch Production Systems. *Opns. Res.* Vol.34, No.1 (1986), 91-104.

Tadayoshi SHIOYAMA: Department of  
Mechanical Engineering,  
Kyoto Institute of Technology