

Division of Research
Graduate School of Business Administration
The University of Michigan

January 1982
Revised, January 1984
Revised, May 1984

OPTIMALITY OF BALANCED WORKLOADS IN FLEXIBLE
MANUFACTURING SYSTEMS

Working Paper No. 289 -c

Kathryn Stecke
Graduate School of Business Administration
The University of Michigan, Ann Arbor Michigan

and

Thomas L. Morin
School of Industrial Engineering
Purdue University, West Lafayette Indiana

European Journal of Operational
Research, forthcoming

ABSTRACT

Generalized concavity and symmetric mathematical programming are used to analyze the optimality of balanced workloads to maximize the expected production of flexible manufacturing systems (FMSs) using a single-server closed queueing network model. In particular, we prove that balancing workloads maximizes expected production in certain types of m -machine FMSs with n parts in the system. Our results are compared and contrasted with previous models of production systems.

A flexible manufacturing system (FMS) is an automated alternative to the conventional means of batch manufacturing in the metal-cutting industry that consists of a number of numerically controlled machine tools which are linked together by an automated material handling system. Computers control most real-time activities such as the actual machining operations, part movements, and tool interchanges. An FMS can simultaneously and efficiently manufacture several part types. This combination of automation and increased flexibility offers the potential for vast improvements in productivity, but as noted by Graves [1981], also increases the complexity of the problems faced by production managers. For example, the operation of an FMS requires a careful system set-up prior to production to achieve a good system utilization during production, even though technological hardware developments eliminate machine set-up time. Several existing FMSs are described in Cavallé, Forestier, and Bel [1981], Stecke and Solberg [1981b], Dupont-Gatelmand [1982], and Barash [1982].

This paper studies an idealized version of the FMS loading problem, which is one of the set-up problems of an FMS (see Stecke [1983a]). The loading problem involves determining the best allocation of operations and associated cutting tools of a set of part types among the machine tools subject to technological and capacity constraints.

The most widely applied loading objective is to balance, or equalize, the total workload assigned to each machine in: job shops (Deane and Moodie [1972], Caie, Linden, and Maxwell [1980]); flow shops (Gutjahr and Nemhauser [1964], Ignall [1965], Magazine and Wee [1979]); and FMSs (Buzacott and Shanthikumar [1980], Berrada and Stecke [1983], Kusiak [1983], Stecke and Talbot [1983]). However, the applicability and optimality of balancing has recently come under scrutiny. For example, Stecke and Solberg's [1981b] simulation results demonstrated that balancing workloads is not necessarily

the best objective in an FMS. Other studies of finite-buffer stochastic flow lines also indicated that balancing the assigned workload is not always optimal (see Makino [1964], Hillier and Boling [1966, 1967], Payne, Slack, and Wild [1972], Rao [1976], Magazine and Silver [1978], and El-Rayah [1979]). In particular, the numerical studies (see Hillier and Boling [1966, 1967] and El-Rayah [1979]) discovered a "bowl phenomenon" in which the expected production of a finite-buffered, balanced flow line is increased by assigning proportionately lower average processing or service times to the middle machines on the line.

Queueing network models have recently been used to analyze design issues and planning problems of FMSs. (For example, see Buzacott and Shanthikumar [1980], Cavallé and Dubois [1982], Dubois [1983], Stecke and Solberg [1982], and Suri [1983].) Queueing networks have been shown to be robust models of FMSs even when the assumptions of the model are not satisfied (see Suri [1983] and §1).

In the context of a closed queueing network (CQN), the loading problem is that of allocating a total amount of work among a system of (possibly grouped) machines so as to maximize expected production. Using a CQN, it has been shown that (Stecke and Solberg [1982]):

- i) the best way to partition the machine tools of a particular type into machine groups is to unbalance as much as possible the number of machines in each group;
- ii) for these better system configurations, expected production is maximized by a particular unbalanced allocation of workload per machine.

However, in some practical situations, because of the discreteness of operation times, different machine tool requirements, and limited capacity tool magazines, balancing the workload per machine can be best. This paper

characterizes situations in which balancing is optimal: for those systems in which there is no grouping, or pooling machines of similar type into machine groups. Also, the fact remains that balancing is the almost universally applied loading objective, at least at present. Balancing is applicable to some FMSs.

In particular, in this paper we use a single-server CQN model to analyze the optimality of balancing for adequately buffered flexible manufacturing systems in which each operation is assigned to only one machine. We show that balancing maximizes the expected production in these systems. Specifically, we use generalized concavity and symmetric mathematical programming to establish the optimality of balanced workloads. An efficient means of implementing a balancing FMS loading objective is provided in Berrada and Stecke [1983].

There is a related Computer Science literature. Price [1974], Trivedi and Kinicki [1978], Trivedi, Wagner, and Sigmon [1980], and Trivedi and Sigmon [1981] maximize throughput in central server, single class, single-server CQN subject to various cost constraints. The studies optimize different parameters such as service rate (of a CPU, say), capacity of servers (I/O devices), device speeds, and main memory size, subject to budgetary limitations. The parameters relate cost considerations to performance.

In this paper, a different non-central server CQN of single-server queues is considered. Rather than the budgetary constraints of the previous studies, we impose a constraint on total system workload that appears as a result of our unique scaling of workload and throughput. Therefore, the objective function and constraints are somewhat different. The motivation of our particular scaling results from our studies of optimal machine allocation and optimal workload assignment in FMSs.

Even though the objective function (to maximize expected production, or throughput) is not concave (see Stecke [1983b]), the production function is still well-behaved. In the situations studied here, the local maximum (which we prove is a balanced workload) is a global maximum.

The plan of the paper is as follows. The closed queueing network model is described in §1. Notation and results from generalized concavity and symmetric mathematical programming that are required to characterize properties of optimal workloads are summarized in the Appendix. Properties of the production function and some preliminary results that are required to establish global optimality of balancing for this particular version of the FMS loading problem are provided in §2. The main result is given in §3. A discussion of the relationships between this CQN and other models of manufacturing systems is contained in §4.

1. CLOSED QUEUEING NETWORK MODEL OF AN FMS

A flexible manufacturing system can be modeled as a closed network of arbitrarily-connected queues. The particular case of a central server CQN is depicted in Figure 1. There are m machines and n parts in the system. The average processing time of an operation by machine i is t_i , $i=1, \dots, m$.

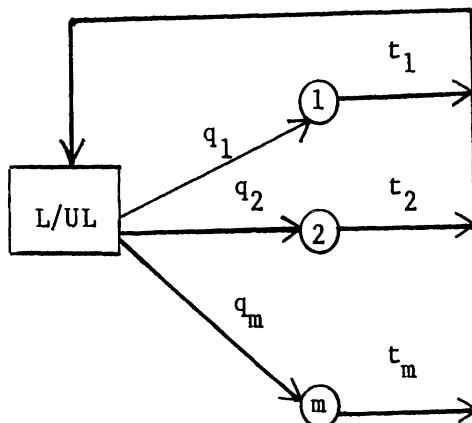


Figure 1

A Closed Queueing Network Model of a Flexible Manufacturing System.

Routing through the system is arbitrary, and can be described by the relative arrival rates (the q_i of Figure 1) to the machines. These can be obtained by any nonnegative solution to $q_i = \sum_j p_{ji} q_j$, where the p_{ij} 's are first-order Markovian probabilities. Our formulas permit any scaling of the q_i 's. For example, if the q_i 's are scaled to sum to one, q_i may be interpreted as the probability that a part leaving the load/unload station (L/UL) via a transporter goes next to machine i . Therefore, q_i is the expected number of visits to machine i per visit to the transporter (or L/UL). Other relevant routing possibilities are described in Stecké and Schmeiser [1983].

A measure of relative workload assigned to machine i is w_i (Buzen [1973], Reiser and Kobayashi [1975], Solberg [1977]), which is defined as the visit frequency times the average processing time, or $q_i t_i$, $i=1, \dots, m$. These workloads are relative since the q_i 's can be scaled in any manner.

For our purposes w_i was scaled, where $\sum_{j=1}^m q_j t_j / m$ is the average workload per machine, to provide:

$$X_i = q_i t_i / [(\sum_{j=1}^m q_j t_j) / m]. \quad (1)$$

X_i is a scaled measure of workload, whose values lie between 0 and m , for all i .

There are several reasons for choosing this particular scaling: the total amount of work to be allocated among the machines is a fixed constant that equals the total number of machines: $\sum_{i=1}^m X_i = m$; the scaled workload is now independent of any particular, chosen scaling of the q_i ; regardless of the number of machines, a balanced loading has a unit workload: $X_1 = X_2 = \dots = X_m = 1$; this particular scaling of workload results in the production function (defined in equation (3) below) being a dimensionless function, whose values are also normalized to lie between zero and one; and finally, the workload

scaling defined by equation (1) allows new alternative definitions of the production function (equation (3) below). See Stecké [1981] or Stecké and Schmeiser [1983] for these alternative definitions. These are useful for providing insight into what this production function, associated with a CQN model, is.

A state of the CQN model of an FMS is given by $\tilde{n} = (n_1, \dots, n_m)$, where n_i is the number of parts at machine tool i , both those waiting and those in process. For all i , $n_i \in \{0, 1, \dots, n\}$ and $\sum_{i=1}^m n_i = n$. The steady-state probability of being in state \tilde{n} is $p(\tilde{n}) = p(n_1, \dots, n_m)$, which for this CQN model has the product form solution: $p(\tilde{n}) = \frac{1}{G(m, n; X)} X_1^{n_1} X_2^{n_2} \dots X_m^{n_m}$, where:

$$G(m, n; X) = \sum_{\substack{n_1 + \dots + n_m = n \\ n_i \geq 0}} X_1^{n_1} X_2^{n_2} \dots X_m^{n_m}, \quad i=1, \dots, m. \quad (2)$$

It can be seen that the function $G(m, n; X)$ is the normalizing constant that is required for the probabilities, $p(\tilde{n})$, to sum to 1. For the FMS depicted in Figure 1 with n parts in the system, the expected production rate, which is the expected number of parts produced per unit of time, can be defined as a function of $G(m, n; X)$, which in turn is a function of assigned workload, X_i . In fact, for a particular scaling of q_i , the production function, $Pr(m, n; X)$, is given by Reiser and Kobayashi [1975] as:

$$Pr(m, n; X) = \frac{G(m, n-1; X)}{G(m, n; X)} = \frac{\sum_{n_1 + \dots + n_m = n-1} X_1^{n_1} X_2^{n_2} \dots X_m^{n_m}}{\sum_{n_1 + \dots + n_m = n} X_1^{n_1} X_2^{n_2} \dots X_m^{n_m}}. \quad (3)$$

The alternative definitions, referred to above, do provide additional intuition into just what $Pr(m, n; X)$ means. For these insights, we refer the reader to Stecké and Schmeiser [1983].

As an example, the production function for two single machines and any number of parts, n , is:

$$\begin{aligned}
 \Pr(2, n; X) &= \frac{\sum_{n_1+n_2=n-1} \binom{n_1}{X_1} \binom{n_2}{X_2}}{\sum_{n_1+n_2=n} \binom{n_1}{X_1} \binom{n_2}{X_2}} \\
 &= \frac{\sum_{n_1=0}^{n-1} \binom{n-1}{n_1} (2-X_1)^{n-1-n_1}}{\sum_{n_1=0}^n \binom{n}{n_1} (2-X_1)^{n-n_1}}, \quad \text{since } X_1+X_2 = m = 2 \\
 &= \frac{X_1^n - (2-X_1)^n}{X_1^{n+1} - (2-X_1)^{n+1}}, \quad \text{after dividing both numerator and denominator by } (2-X_1)-X_1 = 2(1-X_1). \quad (4)
 \end{aligned}$$

Many performance measures that can be obtained from CQN models, such as the expected production rate, are insensitive to the form of the service time distribution-see Helm and Schassberger [1982] and Dukhovny and Koenigsberg [1981]. In fact, for the performance measure of expected production, the service time distribution can be arbitrary.

The assumptions of our CQN model of a flexible manufacturing system are that:

1. There are n parts (or pallets) circulating through a system of m machines.
2. There is a buffer at each machine tool that has the capacity to hold all n parts, including the part being machined.
3. The queue discipline at each machine tool can be either FCFS, infinite server, LCFS preempt-resume, processor sharing (see Baskett et al. [1975]), random selection, or one developed by Kelly [1979] which allows an arbitrary distribution to be defined at each node.

The main restrictive assumption is the limited number of allowable queue disciplines, which is why product form queueing networks are not used to study scheduling problems.

Queueing network models have been shown to be accurate in qualitatively predicting steady-state behavior of FMSs. For example, Solberg [1977] compared results from his CQN computer program, CAN-Q, to those of a detailed simulation of the Sundstrand/Caterpillar FMS of Peoria, Illinois (Stecke [1977]) to find that the performance measures of all machine utilizations and expected production rate differed from those of the FMS by less than 3 percent. Similar results were observed by Kimemia and Gershwin [1978], Secco-Suardo [1978], and Dubois [1983]. Queueing network models have also been used to model other nonmanufacturing systems in which the service time distributions were not exponential, with encouraging results. For example, Hughes and Moe [1973], Giammo [1976], Lipsky and Church [1977], and Rose [1976, 1978] have verified in empirical studies that queueing network models reproduce observed quantities with reasonable accuracy. Attempts to explain the observed robustness through operational analysis can be found in Denning and Buzen [1978] and Suri [1983].

2. PRELIMINARY RESULTS

The production function given in equation (3) is difficult to characterize analytically. However, it can be evaluated numerically using Buzen's efficient algorithm [1973]. The function behaves so well empirically that some researchers (i.e., Secco-Suardo [1978] and Solberg [1979]) have conjectured that it must be concave. Concavity would be desirable because it would insure that a local maximum, if it exists, is a global maximum. However, Stecke [1983b] has shown that, contrary to conjecture, the production function is not concave in general, even though it is concave in a few restrictive cases. Fortunately, however, the function satisfies weaker

generalized concavity conditions, which are also sufficient to insure that a local maximum is a global maximum.

Using the Definitions (D) and Theorems (T) in the Appendix, we first establish two results, which will be used later to prove the main result.

Proposition 1. The set χ of feasible loadings is a closed S-convex set.

Proof. From equation (1), we have $\chi = \{(X_1, \dots, X_m) \mid \sum_{i=1}^m X_i = m, X_i \in \mathbb{R}, 0 \leq X_i \leq m\}$. Therefore, χ is clearly closed. It is also clearly convex and symmetric. Then by T11, χ is S-convex. ||

Proposition 2. The quotient of two symmetric functions in the same variables on the same symmetric set χ is a symmetric function.

Proof. Suppose that $f(x)$ and $g(x)$ are symmetric functions on the symmetric set χ . Then by D5,

$$f(xP) = f(x) \text{ and}$$

$$g(xP) = g(x)$$

for all $x \in \chi$ and for any permutation matrix P . Let

$$h(x) = f(x)/g(x)$$

for all $x \in \chi$ such that $g(x) \neq 0$. Then

$$h(xP) = f(xP)/g(xP)$$

$$= f(x)/g(x)$$

$$= h(x).$$

Therefore, $h(x)$ is a symmetric function on χ . ||

In Theorems 3 and 4 below, we prove directly the S-concavity of both the normalizing constant, $G(m,n;X)$, and the production function for two machines, $Pr(2,n;X)$.

Theorem 3. $G(m,n;X)$ is strictly S-concave.

Proof. By T8 it is sufficient to show that

$$(X_2 - X_1) \left(\frac{\partial G(m,n;X)}{\partial X_2} - \frac{\partial G(m,n;X)}{\partial X_1} \right) < 0,$$

for $X_1 \neq X_2 \neq 0$.

The quantity on the left-hand side of the inequality is

$$\begin{aligned} & (X_2 - X_1) \sum_{n_1 + \dots + n_m = n} \left(\frac{n_2}{X_2} - \frac{n_1}{X_1} \right) X_1^{n_1} X_2^{n_2} \dots X_m^{n_m} \\ &= \frac{X_2 - X_1}{X_1 X_2} \sum_{n_1 + \dots + n_m = n} (n_2 X_1 - n_1 X_2) X_1^{n_1} X_2^{n_2} \dots X_m^{n_m} \end{aligned}$$

Consider the two cases: either $n_1 = n_2$ or $n_1 \neq n_2$.

Since $X_i \geq 0$, $i = 1, \dots, m$, if $n_1 = n_2$, then the quantity is negative and we are done.

Suppose $n_1 \neq n_2$. Then for every n_1, n_2 such that $n_1 > n_2$ (say $n_1 - n_2 = \delta > 0$), by symmetry there exists a corresponding pair n'_1, n'_2 such that

$$n'_1 < n'_2 \text{ and } n'_1 - n'_2 = -\delta.$$

In fact, $n_1 = n'_2$ and $n_2 = n'_1$. Therefore, all terms vanish except those states such that $n_1 = n_2$, and we are left with:

$$\frac{-(X_2 - X_1)^2}{X_1 X_2} \sum_{n_1 + \dots + n_m = n} n_1 X_1^{n_1} X_2^{n_2} \dots X_m^{n_m} < 0,$$

for all $X > 0$ such that $X_1 \neq X_2$. Therefore, $G(m,n;X)$ is S-concave. ||

Theorem 4. $Pr(2,n;X)$ is S-concave.

Proof. From (4), we have:

$$\begin{aligned} \Pr(2, n; X) &= \frac{X_1^n - (2-X_1)^n}{X_1^{n+1} - (2-X_1)^{n+1}} \\ &= \frac{X_1^n - X_2^n}{X_1^{n+1} - X_2^{n+2}}, & \text{for } X_1 \neq X_2; \\ &= n/(n+m-1), & \text{for } X_1 = X_2. \end{aligned}$$

Differentiating yields:

$$\frac{\partial \Pr(2, n; X)}{\partial X_2} = \frac{-(X_1^{n+1} - X_2^{n+1})nX_2^{n-1} + (X_1^n - X_2^n)(n+1)X_2^n}{(X_1^{n+1} - X_2^{n+1})^2}$$

Therefore,

$$\begin{aligned} &(X_2 - X_1) \left(\frac{\partial \Pr(2, n; X)}{\partial X_2} - \frac{\partial \Pr(2, n; X)}{\partial X_1} \right) \\ &= \frac{(X_2 - X_1) \{ [(n+1)X_2^n (X_1^n - X_2^n) - nX_2^{n-1} (X_1^{n+1} - X_2^{n+1})] - [nX_1^{n-1} (X_1^{n+1} - X_2^{n+1}) - (n+1)X_1^n (X_1^n - X_2^n)] \}}{(X_1^{n+1} - X_2^{n+1})^2} \end{aligned}$$

Since the denominator is positive for $X_1 \neq X_2$, it may be dropped, yielding after simplification:

$$(X_2 - X_1) \{ X_2^{n-1} [(n+1)X_2 (X_1^n - X_2^n) - n(X_1^{n+1} - X_2^{n+1})] + X_1^{n-1} [(n+1)X_1 (X_1^n - X_2^n) - n(X_1^{n+1} - X_2^{n+1})] \},$$

which upon rearranging

$$= (X_2 - X_1) (X_1^2 - X_2^2) \left\{ \sum_{i=1}^{n-1} X_1^{2n-2-2i} X_2^{2i} - n X_1^{n-1} X_2^{n-1} \right\}. \quad (5)$$

In order to show that equation (5) is not positive, it suffices to show that

$$X_1^{2n-2-2i} X_2^{2i} + X_1^{2i} X_2^{2n-2-2i} \geq 2 X_1^{n-1} X_2^{n-1}, \quad (6)$$

since the summation of (5) can be separated into $n/2$ inequalities of the form

(6). Assume that $2i < 2n-2$.

Subtracting the RHS from the LHS of (6) yields:

$$\begin{aligned} & X_1^{2i} X_2^{2i} (X_1^{2n-2-4i} - 2X_1^{n-1-2i} X_2^{n-1-2i} + X_2^{2n-2-4i}) \\ &= X_1^{2i} X_2^{2i} (X_1^{2(n-1-2i)} - 2X_1^{n-1-2i} X_2^{n-1-2i} + X_2^{2(n-1-2i)}) \\ &= X_1^{2i} X_2^{2i} (X_1^{n-1-2i} - X_2^{n-1-2i})^2 \\ &\geq 0. \end{aligned}$$

Therefore, (6) holds for all $i < n-1$. The proof for $2i > 2n-2$ follows

mutatis mutandis and equality holds if $i=n-1$. ||

Next consider the m machine case.

Lemma 5. $\Pr(m,n;X)$ is symmetric.

Proof. By D5 and T9, $\Pr(m,n;X)$ is symmetric if the value of the function remains the same when the X_i are permuted.

$$G(m,n;X) = \sum_{n_1 + \dots + n_m = n} X_1^{n_1} X_2^{n_2} \dots X_m^{n_m}.$$

By Theorem 3, $G(m,n;X)$ is strictly S-concave. Then $G(m,n;X)$ is symmetric by

T9. Since the production function is the quotient of two symmetric functions, by Proposition 2, $\Pr(m,n;X)$ is a symmetric function of X . ||

If in addition $\Pr(m,n;X)$ is quasiconcave, then by T12 $\Pr(m,n;X)$ is S-concave and we can use T15 to prove that balancing is optimal.

Theorem 6. The production function, $\Pr(m,n;X)$, is strictly quasiconcave.

Proof. The result is proven in Stecké [1983b]. ||

3. CHARACTERIZING OPTIMAL WORKLOADS

We now state and prove the main result.

Theorem 7. A balanced allocation of workload maximizes expected production;
i.e.,

$$X^* = [X_1, X_2, \dots, X_m] = [1, 1, \dots, 1].$$

Proof. By Proposition 1, the set of feasible loadings, χ , is closed and
S-convex.

By Lemma 5, $\text{Pr}(m, n; X)$ is symmetric.

By T12, since $\text{Pr}(m, n; X)$ is quasiconcave by Theorem 6, then $\text{Pr}(m, n; X)$ is
S-concave.

By T15, the set χ^* of points maximizing $\text{Pr}(m, n; X)$ over the set χ is a
closed S-convex set.

χ^* is not empty since $\text{Pr}(m, n; X) \in [0, 1]$ for all m, n , and $X \in [0, m]$.

The symmetric point of χ is the point $[1, 1, \dots, 1]$.

By T14, $[1, \dots, 1] \in \chi^*$.

Therefore, a balanced allocation maximizes the expected production. ||

Balancing is now justified for the systems examined here, i.e., FMSs with
no pooling of similar machines.

The following computer-drawn graphs demonstrate the behavior of the
production function. First, Figure 2 is a graph of $\text{Pr}(2, n; X)$ as a function of
 X_1 for $n=4, 5, \dots, 14$ and infinity. For each curve, 400 points $(X, \text{Pr}(2, n; X))$
were plotted. These were calculated using a variation of Solberg's CAN-Q
program [1980]. The maximum functional value, also calculated and plotted, was
attained at $X = [1, 1]$. Table I displays the calculated optimal allocation
ratios and the maximum normalized production rate for each n . For $q_1 = q_2 = .5$
(each machine is visited half of the time on the average), the average pro-
cessing times, t_1 and t_2 , vary so that $t_1 + t_2 = 2$. The optimal allocation

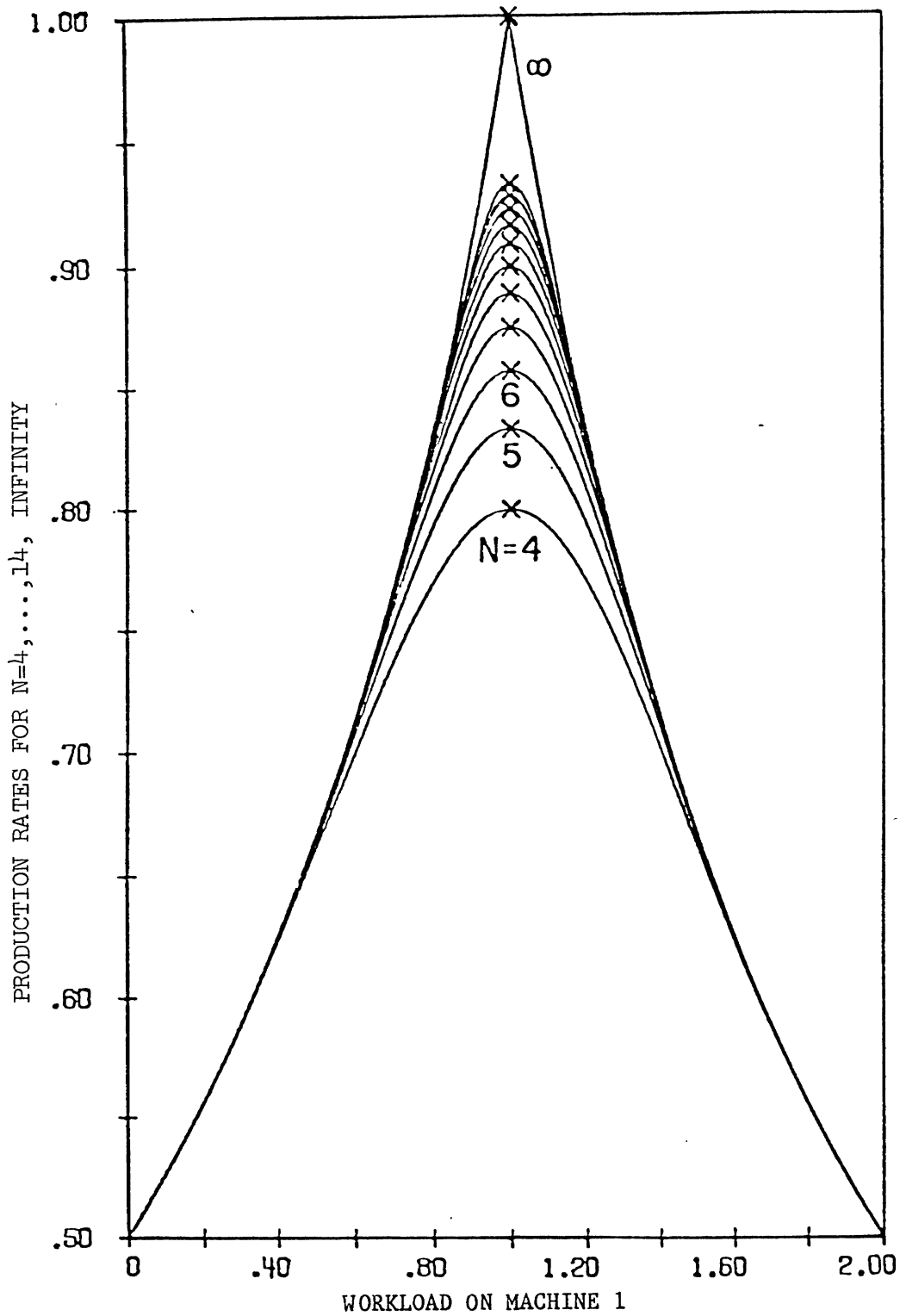


Figure 2

Production Rate as a Function of Workload Assigned to Machine 1 for 2-Machine Systems.

occurs when $t_1 = t_2 = 1$. Then $X_i = 2q_i t_i / (q_1 t_1 + q_2 t_2) = t_i$, where i is 1 or 2. The optimal allocation of workload in this system is balanced.

TABLE I

Maximum (Balanced) Production Rates and Corresponding Workloads for Two-Machine Systems

n	t_1	t_2	X_1	X_2	Max (Bal) Production Rate
4	1.0	1.0	1.0	1.0	.800
5	1.0	1.0	1.0	1.0	.833
6	1.0	1.0	1.0	1.0	.857
7	1.0	1.0	1.0	1.0	.875
8	1.0	1.0	1.0	1.0	.889
9	1.0	1.0	1.0	1.0	.900
10	1.0	1.0	1.0	1.0	.909
11	1.0	1.0	1.0	1.0	.917
12	1.0	1.0	1.0	1.0	.923
13	1.0	1.0	1.0	1.0	.929
14	1.0	1.0	1.0	1.0	.933
∞	1.0	1.0	1.0	1.0	1.000

Figure 3 is a graph of $Pr(3,n;X)$ as a function of X_1 for $n=4,5,\dots,14$ and infinity, along the plane $X_2 = X_3$. It is interesting to note that this two-dimensional slice of $Pr(3,n;X)$ is nonsymmetric even though the entire function is symmetric. The maximum is shown to be at $X_1=X_2=X_3=1$. The computer program generated both the balanced and the maximum normalized productions, as well as the optimal allocation ratios. These are shown for each n in Table II.

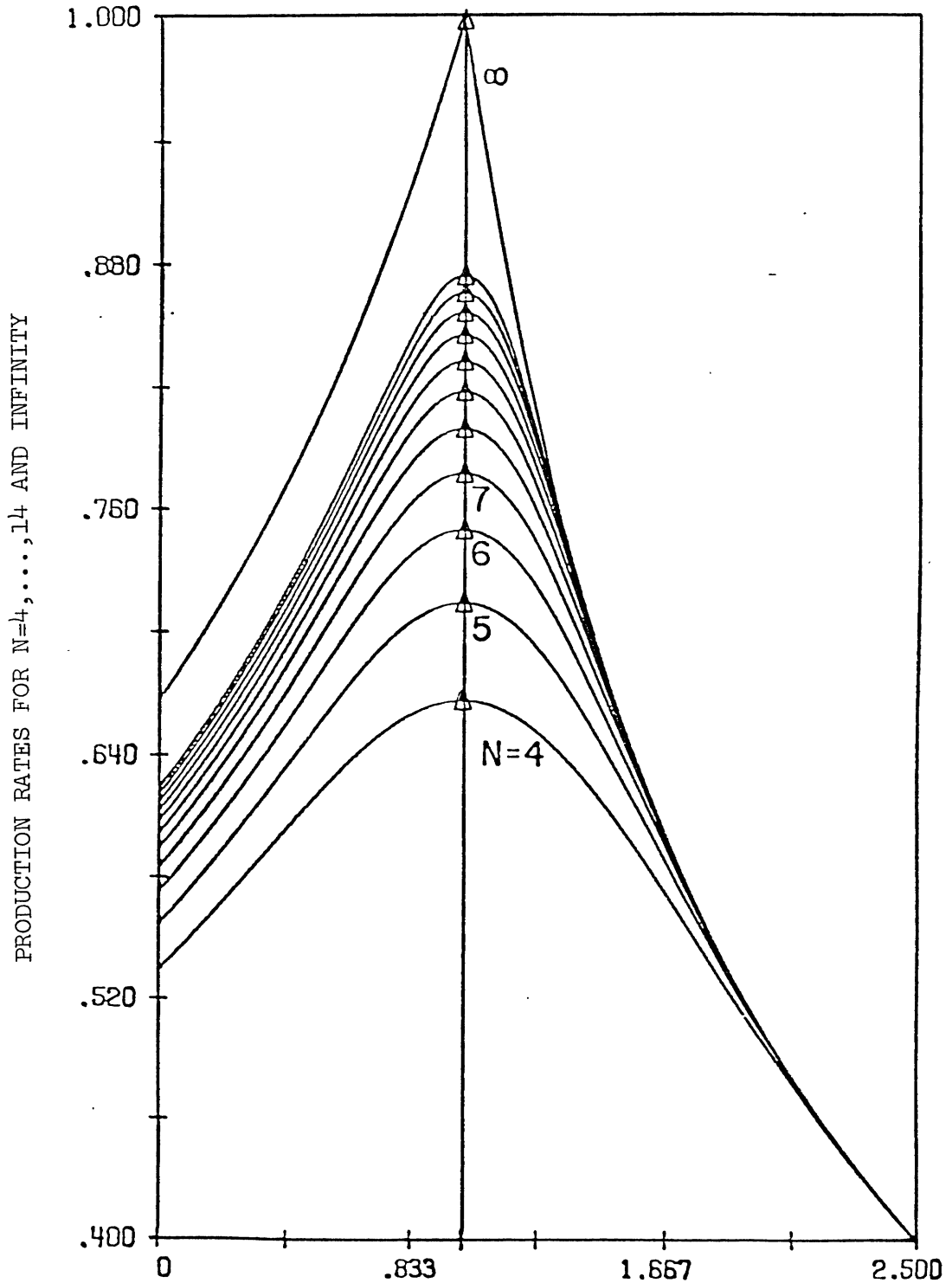


Figure 3

Production Rate as a Function of Workload Assigned to Machine 1 for 3-Machine Systems.

TABLE II

Maximum (Balanced) Production Rates and Corresponding Workloads for Three-Machine Systems

n	Balanced Production Rate	Maximum Production Rate	$X_1^*=t_1$	$X_2^*=t_2$
4	.667	.667	1.0	1.0
5	.714	.714	1.0	1.0
6	.750	.750	1.0	1.0
7	.778	.778	1.0	1.0
8	.800	.800	1.0	1.0
9	.818	.818	1.0	1.0
10	.833	.833	1.0	1.0
11	.846	.846	1.0	1.0
12	.857	.857	1.0	1.0
13	.867	.867	1.0	1.0
14	.875	.875	1.0	1.0
∞	1.000	1.000	1.0	1.0

Finally, Figure 4 displays $\text{Pr}(4,n;X)$ for $n=4, \dots, 14$ and infinity along the intersection of the planes $X_1=X_2$ and $X_3=X_4$. Table III gives values for $\text{Pr}(4,n;X)$. Notice that for all finite n , $\text{Pr}(3,n;X) > \text{Pr}(4,n+1;X)$. That is, as the number of machines increases, the actual expected production obviously

TABLE III

Maximum (Balanced) Production Rates and Corresponding Workloads for Four-Machine Systems

n	Balanced Production Rate	Maximum Production Rate	$X_1^*=t_1$	$X_2^*=t_2$
4	.571	.571	1.0	1.0
5	.625	.625	1.0	1.0
6	.667	.667	1.0	1.0
7	.700	.700	1.0	1.0
8	.727	.727	1.0	1.0
9	.750	.750	1.0	1.0
10	.769	.769	1.0	1.0
11	.786	.786	1.0	1.0
12	.800	.800	1.0	1.0
13	.812	.812	1.0	1.0
14	.824	.824	1.0	1.0
∞	1.000	1.000	1.0	1.0

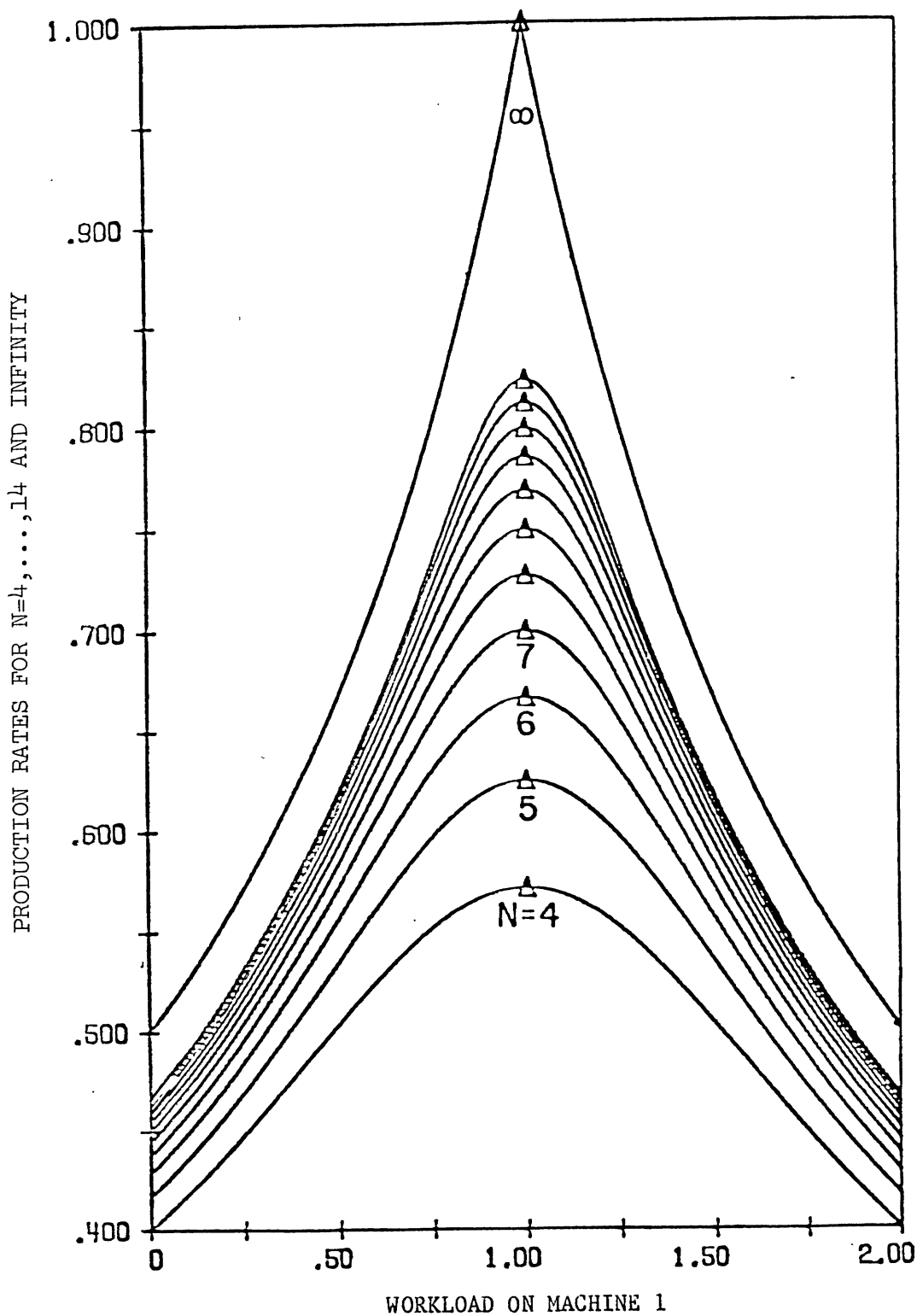


Figure 4

Production Rate as a Function of Workload Assigned to Machine 1 for 4-Machine Systems.

increases but the normalized expected production decreases. The apparent anomaly is the result of the normalization of production to the scaling between 0 and 1.

We conclude that even though $\text{Pr}(m,n;X)$ is not concave for any $m \geq 2$ and $n > 2$, balancing is optimal for all cases (fixed-route FMSs) considered here.

4. DISCUSSION

The results can be related to similar studies of workload allocation in manufacturing systems. Our results differ from the finite-buffer stochastic flow shop studies (Hillier and Boling [1966, 1967], Magazine and Silver [1978]) mainly because we assume an adequate buffer at each machine.

In fact, using our CQN model, the expected production is identical for both flow shops and job shops in which each operation is assigned to only one machine. To see this, let t_i (the average processing time of an operation by machine i) be identical for both systems. The routing mechanism, defined by Markovian probabilities p_{ij} , for a three-machine flow shop is given by the following transition matrix:

$$P_{\sim F} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix};$$

the routing for the job shop is given by:

$$P_{\sim J} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

Then solving balance equations:

$$q_i(k) = \sum_{j=1}^m p_{ij(k)} q_j(k), \quad \begin{array}{l} i=1, \dots, m \\ (k)=F \text{ or } J, \end{array}$$

for the two systems produces identical, steady state results. That is,

$$q_{i(F)} = q_{i(J)}, \quad i=1, \dots, m.$$

Since q_i , t_i , m , and n are identical for both the flow and job shops, the expected production rate is the same for both systems.

Some implications are as follows. Under the assumptions of our CQN model (in particular, random processing times and an adequate buffer at each machine), the two extreme system types (flow and job shops) are equally efficient. Intuition indicates that as variability decreases, the flow shop becomes more efficient than the job shop.

The Hillier and Boling [1966, 1967] result is basically the following:

For a stochastic flow shop with a finite buffer at each machine, the expected production is maximized by a specific unbalancing in the workload assigned to each machine.

Our analogous result is:

For a stochastic flow shop with an infinite buffer at each machine, the expected production is maximized by balancing the workload assigned to each machine.

In other words, as buffer size increases, the degree of unbalance in the optimal workload decreases, until in the limit, a balanced schedule is optimal.

APPENDIX:

RESULTS CONCERNING GENERALIZED CONCAVITY AND SYMMETRIC MATHEMATICAL PROGRAMMING

Definitions and previously published results which are required to prove the optimality of balanced workloads are reviewed below. The definitions and results concerning generalized concavity can be found in Mangaserian [1969] or Bazaraa and Shetty [1979], those concerning S-concavity can be found in Berge

[1963], and those on symmetric mathematical programming can be found in Greenberg and Pierskalla [1970].

Let f be a real-valued function mapping $\chi \rightarrow \mathbb{R}$, where χ is a closed subset of \mathbb{R}^m . We require the following Definitions (D) and Theorems (T):

D1 f is a quasiconcave function on the nonempty convex set $\chi \subseteq \mathbb{R}^m$ if and only if (iff) for any two points $x^1, x^2 \in \chi$, and for all $\lambda \in [0, 1]$,

$$f(\lambda x^1 + (1-\lambda)x^2) \geq \min\{f(x^1), f(x^2)\}.$$

D1 is not enough to insure that a local maximum is a global maximum. For this to be true we have:

D2 f is a strictly quasiconcave function on the convex set $\chi \subseteq \mathbb{R}^m$ iff for any two points $x^1, x^2 \in \chi$, and for all $\lambda \in (0, 1)$, with $f(x^1) \neq f(x^2)$,

$$f(\lambda x^1 + (1-\lambda)x^2) > \min\{f(x^1), f(x^2)\}.$$

In order to insure that a global maximum is unique, we require:

D3 f is a strongly quasiconcave function on the convex set $\chi \subseteq \mathbb{R}^m$ iff for any two points $x^1, x^2 \in \chi$ such that $x^1 \neq x^2$ and for all $\lambda \in (0, 1)$,

$$f(\lambda x^1 + (1-\lambda)x^2) > \min\{f(x^1), f(x^2)\}.$$

D4 χ is a symmetric set if $x \in \chi \implies x^P \in \chi$ for all permutation matrices P , where P is a permutation matrix if

- i) each row has only one entry equal to one;
- ii) each column has only one entry equal to one; and
- iii) all remaining entries are equal to zero.

D5 f is a symmetric function on a symmetric set χ if for any permutation matrix P

$$f(xP) = f(x) \text{ for all } x \in \chi.$$

D6 χ is S-convex if $x \in \chi \implies xS \in \chi$ for all doubly stochastic matrices S , where S is a double stochastic matrix of order m if all of its entries, p_{ij} , satisfy

$$i) \quad p_{ij} \geq 0, \quad \text{for all } i, j;$$

$$ii) \quad \sum_{i=1}^m p_{ij} = 1, \quad \text{for all } j; \text{ and}$$

$$iii) \quad \sum_{j=1}^m p_{ij} = 1, \quad \text{for all } i.$$

D7 f is a (strictly) S-concave function on an S-convex set χ if for any S

$$f(xS) (>) \geq f(x) \quad \text{for all } x \in \chi.$$

T8 Let D be an open interval in \mathbb{R} and let f be a symmetric differentiable function in $D^m \subseteq \mathbb{R}^m$. If for all $x = (x_1, \dots, x_m) \in D^m$ such that $x_1 \neq x_2$ we have

$$(x_2 - x_1) \left(\frac{\partial f}{\partial x_2} - \frac{\partial f}{\partial x_1} \right) (<) \leq 0,$$

then the function is (strictly) S-concave in D^m (Theorem 5, p. 221, Berge [1963]).

T9 An S-concave function f in \mathbb{R}^m is symmetric in the components x_1, \dots, x_m of $x \in \chi$; that is, the value of $f(x_1, \dots, x_m)$ remains the same when the x_i are permuted (Theorem 3, p. 220, Berge [1963]).

T10 If D is an open interval in \mathbb{R} , a necessary and sufficient condition for a differentiable and symmetric function f to be (strictly) S-concave in D^m is that for all $x_1, x_2 \in D$,

$$(x_2 - x_1) \left(\frac{\partial f}{\partial x_2} - \frac{\partial f}{\partial x_1} \right) (<) \leq 0$$

(Theorem 6, p. 224, Berge [1963]).

- T11 Symmetric convex sets are S-convex (but not necessarily conversely).
- T12 Symmetric (strictly) quasiconcave functions defined on a symmetric convex set χ are (strictly) S-concave (but not necessarily conversely).
- D13 A point $x = (x_1, x_2, \dots, x_m)$ is symmetric iff $x_i = y, \forall i=1, \dots, m$.
- T14 Every nonempty S-convex set contains a symmetric point.
- T15 If χ is a closed, S-convex set and f is S-concave on χ , then the set χ^* of points maximizing f over χ is a closed S-convex set (Greenberg and Pierskalla [1970]).

ACKNOWLEDGMENTS

We wish to acknowledge helpful discussions with Bruce W. Schmeiser, Leroy B. Schwarz, and James J. Solberg. The research was supported in part by National Science Foundation Grant No. APR 74 15256 to Purdue University.

REFERENCES

- BARASH, MOSHE M., "Computerized Manufacturing Systems for Discrete Products," Ch. VII-9 in The Handbook of Industrial Engineering, Salvendy, G. (Ed.), John Wiley & Sons, Inc., NY (1982).
- BASKETT, FOREST, K. MANI CHANDY, RICHARD R. MUNTZ and FERNANDO G. PALACIOS, "Open, Closed, and Mixed Networks with Different Classes of Customers," Journal of the Association for Computing Machinery, Vol. 22, No. 2, pp. 248-260 (April 1975).
- BAZARAA, MOKHTAR S. and C. M. SHETTY, Nonlinear Programming Theory and Algorithms, John Wiley & Sons, Inc., NY (1979).
- BERGE, CLAUDE, Topological Spaces, The Macmillan Company, NY (1963).
- BERRADA, MOHAMMED and KATHRYN E. STECKE, "A Branch and Bound Approach for Machine Loading in Flexible Manufacturing Systems," Working Paper No. 329, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI (April 1983).
- BUZACOTT, JOHN A. and J. GEORGE SHANTHIKUMAR, "Models for Understanding Flexible Manufacturing Systems," AIIE Transactions, Vol. 12, No. 4, pp. 339-350 (December 1980).
- BUZEN, JEFFREY P., "Computational Algorithms for Closed Queueing Networks with Exponential Servers," Communications of the Association for Computing Machinery, Vol. 16, No. 9, pp. 527-531 (September 1973).
- CAIE, JAMES, JANET LINDEN and WILLIAM L. MAXWELL, "Solution of a Single Stage Machine Load Planning Problem," OMEGA, Vol. 8, No. 3, pp. 355-360 (1980).
- CAVAILLÉ, JEAN-BERNARD and DIDIER DUBOIS, "Heuristic Methods Based on Mean-Value Analysis for Flexible Manufacturing Systems Performance Evaluation," Proceedings of the 21st IEEE Conference on Decision and Control, Orlando FL, pp. 1061-1065 (December 1982).
- CAVAILLÉ, JEAN-BERNARD, J. P. FORESTIER and GERARD BEL, "A Simulation Program for Analysis and Design of a Flexible Manufacturing System," in the Proceedings of the IEEE Conference on Cybernetics and Society, Atlanta GA, pp. 257-259 (October 1981).
- DEANE, RICHARD H. and COLIN L. MOODIE, "A Dispatching Methodology for Balancing Workload Assignments in a Job Shop Production Facility," AIIE Transactions, Vol. 4, pp. 277-283 (1972).
- DENNING, PETER J. and JEFFREY P. BUZEN, "The Operational Analysis of Queueing Network Models," Computing Surveys, Vol. 10, No. 3, pp. 225-261 (September 1978).
- DUBOIS, DIDIER, "A Mathematical Model of a Flexible Manufacturing System with Limited In-Process Inventory," European Journal of Operational Research, Vol. 14, No. 1, pp. 66-78 (January 1983).

- DUPONT-GATELMAND, CATHERINE, "A Survey of Flexible Manufacturing Systems," Journal of Manufacturing Systems, Vol. 1, No. 1, pp. 1-16 (1982).
- DUKHOVNY, I. M. and ERNEST KOENIGSBERG, "Invariance Properties of Queueing Networks and Their Application to Computer/Communications Systems," Information Systems and Operations Research, Vol. 19, No. 3, pp. 185-204, (August 1981).
- EL-RAYAH, T. E., "The Efficiency of Balanced and Unbalanced Production Lines," International Journal of Production Research, Vol. 17, No. 1, pp. 61-75 (January 1979).
- GIAMMO, T., "Validation of a Computer Performance Model of the Exponential Queueing Network Family," Proceedings of the International Symposium of Computer Performance Modeling, Measurement, and Evaluation, Harvard University. Also published in Acta Informatica, Vol. 17, No. 2, pp. 137-152 (1976).
- GRAVES, STEPHEN C., "A Review of Production Scheduling," Operations Research, Vol. 29, pp. 646-675 (1981).
- GREENBERG, HARVEY J. and WILLIAM P. PIERSKALLA, "Symmetric Mathematical Programs," Management Science, Vol. 16, No. 5, pp. 309-312 (January 1970).
- GUTJAHR, A. L. and GEORGE L. NEMHAUSER, "An Algorithm for the Line Balancing Problem," Management Science, Vol. 11, pp. 308-315 (1982).
- HEIM, W. E. and R. SCHAASBERGER, "Insensitive Generalized Semi-Markov Schemes with Point Processes," Mathematics of Operation Research, Vol. 7, pp. 129-138 (1982).
- HILLIER, FREDERICK S. and RONALD W. BOLING, "The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Operation Times," Journal of Industrial Engineering, Vol. 17, No. 5, pp. 657-658 (December 1966).
- HILLIER, FREDERICK S. and RONALD W. BOLING, "Finite Queues in Series with Exponential or Erlang Service Times: A Numerical Approach," Operations Research, Vol. 15, No. 2, pp. 286-303 (March-April 1967).
- HUGHES, P. H. and G. MOE, "A Structural Approach to Computer Performance Analysis," in Proceedings of the National Computer Conference, AFIPS Press, Montvale NJ, Vol. 42, pp. 109-119 (1973).
- IGNALL, EDWARD J., "A Review of Assembly Line Balancing," Journal of Industrial Engineering, Vol. 16, No. 4, pp. 43-52 (July-August 1965).
- KELLY, FRANK P., Reversibility and Stochastic Networks, John Wiley & Sons, Inc., NY (1979).
- KIMEMIA, JOSEPH and STANLEY B. GERSHWIN, "Multicommodity Network Flow Optimization in Flexible Manufacturing Systems," Report No. ESL-FR-834-2, M.I.T., Cambridge MA (September 1978).

- KUSIAK, ANDREW, "Loading Models in Flexible Manufacturing Systems," in the Proceedings of the Seventh International Conference on Production Research, Windsor Ontario Canada (August 22-24, 1983).
- LIPSKY, L. and J. D. CHURCH, "Applications of a Queueing Network Model for a Computer System," Computing Surveys, Vol. 9, pp. 205-221 (1977).
- MAGAZINE, MICHAEL J. and T. S. WEE, "The Generalization of Bin-Packing Heuristics to the Line Balancing Problem," Working Paper No. 128, Department of Management Sciences, University of Waterloo, Ontario, Canada (March 1979).
- MAGAZINE, MICHAEL J. and G. L. SILVER, "Heuristics for Determining Output and Work Allocations in Series Flow Lines," International Journal of Production Research, Vol. 16, No. 6, pp. 169-181 (May 1978).
- MAKINO, T., "On the Mean Passage Time Concerning Some Queueing Problems of the Tandem Type," Journal of the Operations Research Society of Japan, Vol. 7, pp. 17-47 (1964).
- MANGASERIAN, O.L., Nonlinear Programming, McGraw-Hill Company, NY (1969).
- PAYNE, S., N. SLACK and RAY WILD, "A Note on the Operating Characteristics of Balanced and Unbalanced Production Flow Lines," International Journal of Production Research, Vol. 10, No. 1, pp. 93-98 (1972).
- PRICE, THOMAS GORDAN, "Probability Models of Multiprogrammed Computer Systems," Ph.D. dissertation, Department of Electrical Engineering, Stanford University, Stanford CA (December 1974).
- RAO, NORI PRAKASA, "A Generalization of the 'Bowl Phenomenon' in Series Production Systems," International Journal of Production Research, Vol. 14, pp. 437-443 (1976).
- REISER, MARTIN and HISASHI KOBAYASHI, "Horner's Role for the Evaluation of General Closed Queueing Networks," Communications of the Association for Computing Machinery, Vol. 18, No. 10, pp. 592-593 (October 1975).
- ROSE, C. A., "Validation of a Queueing Model with Classes of Customers," Proceedings of the International Symposium on Computer Performance Modeling, Measurement, and Evaluation, Harvard University, Cambridge MA, pp. 318-325 (1976).
- ROSE, C. A., "A Measurement Procedure for Queueing Network Models of Computer Systems," Computing Surveys, Vol. 10, pp. 263-280 (1978).
- SECCO-SUARDO, G., "Optimization of a Closed Network of Queues," Report No. ESL-FR-834-3, Electronic Systems Laboratory, M.I.T., Cambridge MA (1978).
- SOLBERG, JAMES J., "A Mathematical Model of Computerized Manufacturing Systems," Proceedings of the International Conference on Production Research, Tokyo Japan (August 1977).

- SOLBERG, JAMES J., "Stochastic Modeling of Large Scale Transportation Networks," Report No. DOT-ATC-79-2, School of Industrial Engineering, Purdue University, West Lafayette IN (June 1979).
- SOLBERG, JAMES J., "CAN-Q User's Guide," Report No. 9 (Revised), NSF GRANT No. APR74 15256, School of Industrial Engineering, Purdue University, West Lafayette IN (July 1980).
- STECKE, KATHRYN E., "Experimental Investigation of a Computerized Manufacturing System," Master's Thesis, School of Industrial Engineering, Purdue University, W. Lafayette IN (December 1977).
- STECKE, KATHRYN E., "Production Planning Problems for Flexible Manufacturing Systems," Ph.D. dissertation, Purdue University, West Lafayette IN (August 1981).
- STECKE, KATHRYN E., "A Hierarchical Approach to Production Planning in Flexible Manufacturing Systems," in the Proceedings of the Twentieth Annual Allerton Conference on Communication, Control, and Computing, Monticello IL (October 6-8, 1982).
- STECKE, KATHRYN E., "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems," Management Science, Vol. 29, No. 3 pp. 273-288 (March 1983a).
- STECKE, KATHRYN E., "On the Nonconcavity of Throughput in Certain Closed Queueing Networks," Working Paper No. 356, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI (December 1983b).
- STECKE, KATHRYN E. and BRUCE W. SCHMEISER, "Alternative Representations of System Throughput in Closed Queueing Network Models of Multiserver Queues," Working Paper No. 324, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI (December 1983).
- STECKE, KATHRYN E. and JAMES J. SOLBERG, "The CMS Loading Problem," Report No. 20, NSF Grant No. APR 74 15256, School of Industrial Engineering, Purdue University, West Lafayette IN (February 1981a).
- STECKE, KATHRYN E. and JAMES J. SOLBERG, "Loading and Control Policies for a Flexible Manufacturing System," International Journal of Production Research, Vol. 19, No. 5, pp. 481-490 (September-October 1981b).
- STECKE, KATHRYN E. and JAMES J. SOLBERG, "The Optimality of Unbalanced Workloads and Machine Group Sizes for Flexible Manufacturing Systems," Working Paper No. 290, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI (January 1982).
- STECKE, KATHRYN E. and F. BRIAN TALBOT, "Heuristic Loading Algorithms for Flexible Manufacturing Systems," in the Proceedings of the Seventh International Conference on Production Research, Windsor Ontario Canada (August 22-24, 1983).

- SURI, RAJAN, "Robustness of Queueing Network Formulas," Journal of the Association for Computing Machinery, Vol. 30, No. 3, pp. 564-594 (July 1983).
- TRIVEDI, KISHOR S. and ROBERT E. KINICKI, "A Mathematical Model for Computer System Configuration and Planning," in Performance of Computer Installations, D. Ferrari, Editor, North-Holland, Amsterdam (1978).
- TRIVEDI, KISHOR S. and TIMOTHY M. SIGMON, "Optimal Design of Linear Storage Hierarchies," Journal of the Association for Computing Machinery, Vol. 28, No. 2, pp. 270-288 (April 1981).
- TRIVEDI, KISHOR S., ROBERT A. WAGNER and TIMOTHY M. SIGMON, "Optimal Selection of CPU Speed, Device Capabilities and File Assignments," Journal of the Association for Computing Machinery, Vol. 27, No. 3, pp. 457-473 (July 1980).