BEBR
FACULTY WORKING
PAPER NO. 90-1646

Pre-Production Planning Decisions in Flexible Manufacturing Systems with Random Material Flows

Kathryn E. Stecke
N. Raman

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 90-1646

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 1990

•

Pre-Production Planning Decisions
in Flexible Manufacturing Systems
With Random Material Flows

Kathryn E. Stecke

School of Business Administration
University of Michigan, Ann Arbor, MI

and

N. Raman

Department of Business Administration
University of Illinois, Champaign, IL

ABSTRACT

In this paper, we consider the FMS planning problems of forming machine groups and assigning operations to these groups. An open queueing network representation of an FMS is used to show that, under total pooling, the grouping configuration which minimizes mean job flow time is one in which machine groups are maximally unbalanced and the larger groups are utilized more heavily.

Three grouping configurations - no pooling, partial pooling and total pooling, and three loading objectives are used for generating a variety of system configurations. A simulation experiment is used to study the impact of parameters such as system utilization, machine unreliability, batch size, and variation in operation processing times to compare the performance of these configurations. Experimental results show that the detrimental impact of these factors on mean job flow time can be reduced by partial pooling of machines, aggregating operations of a job to be performed at the same machine, and maximizing the sum of operation duplications. In particular, partial pooling of machines is both superior and robust across a wide range of values that these system parameters take.

1 Introduction

Greater product proliferation and market fragmentation, and shorter product life cycles have made firms increasingly aware of the importance of manufacturing flexibility. Unlike conventional manufacturing methods, programmable automation with computer-controlled and versatile machining and assembly capabilities promises an effective solution to the simultaneous requirement of manufacturing efficiency and process flexibility. Consequently, the design and operation of flexible manufacturing systems (FMSs) and the definition and classification of production flexibility are subjects of growing interest among researchers and practitioners alike.

The manufacturing issues faced in an FMS can be categorized into: i) Design problems, ii) planning problems, and iii) scheduling and control problems. FMS *design* problems address the long term issues relating to the system, and they include decisions regarding the selection of part types to be produced in the system, selection and layout of machine tools and the material handling system, design of buffers and the computer control architecture. FMS *planning* problems comprise resource allocation decisions during pre-production system setup. They include selecting the subset of part types for imminent manufacture from among the set of all part types that the FMS can produce, determining the ratio in which these part types will be manufactured concurrently, the assignment of pallets, fixtures, etc. to these part types, and the allocation of operations and cutting tools to individual machine tools. Because the operation of the machines in an FMS are tightly coordinated, the planning decisions play a critical role in determining the overall effectiveness of the system. FMS *scheduling and control* problems relate to the execution of orders and include the determination of part input sequence, the part processing sequence at each machine, and monitoring the actual system performance and taking the necessary corrective actions.

Much of the effectiveness of an FMS is derived from the versatility of its machines and the consequent part *routing flexibility*. Providing alternative routes for a part through the system renders it less susceptible to disruptions such as machine failures. The ability of a operation to select a machine in real time based upon the current system status also reduces part flow time relative to a conventional system in which each operation is typically assigned to only one machine. The number of such alternative part routes is determined by the objectives considered at the planning level and the resulting system configurations.

The objective of this paper is to evaluate alternative configurations, and to understand how their relative performance depend upon the underlying system characteristics. This is done in two stages. First, we consider several planning objectives, and use a mathematical programming model to determine machine configurations corresponding to these objectives under some simplifying assumptions. Next, we use an experimental study to extend the analytical results to a more general system. We also investigate the robustness of these configurations under varying degrees of system disruptions. The performance measure used in this study is average part flow time.

Previous research on dynamic FMSs is based primarily on queuing theoretic approaches. Buzacott and Yao (1986) present an excellent survey of this literature. Studies which address closely related issues include Stecke (1983), Stecke and Solberg (1985) and Stecke (1986). These investigations model an FMS as a closed network of queues, and they derive the optimal machine grouping configurations for the objective of maximizing system throughput. The underlying result of these studies is that under certain service disciplines and operation processing time distributions, the part production rate is maximized by pooling machines into unequal groups and assigning appropriately unbalanced workloads to these groups. Stecke (1983), in addition, considers various operation assignment objectives appropriate in an FMS. Stecke (1986) presents a hierarchical framework for implementing these objectives. Shanthikumar and Yao (1987, 1988) show that the throughput of a group is concave in the number of machines. This results in an efficient heuristic algorithm for assigning a given number of machines to individual groups.

A parallel body of research addresses the allocation of operations to machines in a *static* FMS environment. Ammons, Lofgren and McGinnis (1984), Chakravarty and Shtub (1984), Kusiak (1984), Rajagopalan (1986), Berrada and Stecke (1986), and Hwang (1986) present mathematical programming approaches to solve this problem for various objectives. However, because of the static nature of the problem considered, they do not explicitly address the impact of system utilization levels and unexpected disruptions.

The paper is organized as follows. In §2, we develop a general formulation of the minimum flow time (MMFT) problem. Given the complexity of this problem, we decompose it heuristically into two subproblems — the Machine Grouping Problem and the Machine Loading Problem. In §3, we model an FMS as an open network of $M/M/c$ queues to derive the characteristics of an optimal solution to the machine grouping problem. These results are used in §4 for formulating the machine loading problem and for generating various system configurations. In §5, we present an experimental investigation of these configurations which additionally addresses the impact of system uncertainties arising on account of machine failures, variation of processing times, and demand variations. We conclude in §6 with a summary of the main results obtained in this paper.

2 FMS Planning Problem

We consider an FMS consisting of M machines. Let N be the number of different part types produced in the system. These part types arrive randomly at the system; a given part type requires a series of operations to be performed in a specified sequence. Each operation can be assigned to one or more machines by ensuring that the tool required for processing that operation is available at the machine(s). Each machine has a tool magazine of limited capacity. For simplicity, we assume that each operation requires one tool, and tools are not shared among operations.

The objective of the pre-production setup problem is to assign operations to machines such that the expected part flow time is minimized.

We define the following notation:

n_j : Number of operations in part type j, $j = 1, 2, \ldots, N$,

p_{ij} : Processing time of operation i in part j, $j = 1, 2, \ldots, N; i = 1, 2, \ldots, n_j$,

ρ_m : Utilization of machine $m, m = 1, 2, \ldots, M$,

W_m : (Steady state) average time spent by a part type at $m, m = 1, 2, \ldots, M$,

T^m : Tool magazine capacity of machine m, $m = 1, 2, \ldots, M$,

λ^j : Arrival rate of job j, $j = 1, 2, \ldots, N$,

λ : Cumulative arrival rate at the system $= \sum_{j=1}^{N} \lambda^j$,

$x_{ijm} = 1$, if operation i in part type j is assigned to machine m, zero otherwise.

The minimum mean flow time problem can then be formulated as:

(MMFT1)

$$\min \frac{1}{\lambda} \sum_{m=1}^{M} W_m \sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} x_{ijm} \tag{1}$$

subject to

$$L_{ij} \leq \sum_{m=1}^{M} x_{ijm} \leq U_{ij}, \forall i, j \tag{2}$$

$$\sum_{j=1}^{N} \sum_{i=1}^{n_j} x_{ijm} \leq T^m, \forall m \tag{3}$$

$$\rho_m = \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{n_j} p_{ij} x_{ijm} \tag{4}$$

$$0 \leq \rho_m < 1, \forall m \tag{5}$$

$$x_{ijm} \in \{0, 1\}, \forall i, j, m \tag{6}$$

Equation (1) expresses the expected part flow time as the sum of the time spent by each operation at the machine to which it is assigned. Equation (2) ensures that the number of duplicate machine assignments for a given operation lies between the lower bound L_{ij} and the upper bound U_{ij}. [These bounds are prespecified.] Equation (3) relates to the constraints on the tool magazine capacity. Equation (4) measures the utilization of any machine in group g. Finally, Equations (5) and (6) specify the range of valid machine utilizations and the binary nature of the problem variables respectively.

This formulation makes three simplifying assumptions. First, we assume that all machines can process any of the $\sum_{j=1}^{N} n_j$ operations. Second, we assume that each operation requires one cutting tool which occupies one tool slot in the tool magazine, and these tools are not shared among operations. Third, all machines are considered to be equally efficient in terms of the operations processing times. The first assumption can be relaxed quite easily. We make this assumption primarily to simplify notation. It is also possible to relax the second assumption if each tool occupies a given number of tool slots, there is no tool commonality, and no saving results in the number of tool slots used resulting by assigning two or more tools of given types to the same tool magazine.

MMFT is a nonlinear 0-1 programming problem which is difficult to solve optimally. We decompose it heuristically into two subproblems — the Machine Grouping Problem and the Machine Loading Problem which are solved iteratively.

3 Machine Grouping Problem

It is possible in many FMSs to partition the available machines into *groups* of identically-tooled machines. All machines in each groups are, therefore, capable of performing the same set of operations. In addition, if each operation is assigned to only one group, the machines are said to be *totally pooled* into groups. In the machine grouping problem (MGP), we determine the number of groups G, the assignment of machines to individual groups, and the optimal group utilization levels.

Alternatives to total pooling include *no pooling* and *partial pooling*. No pooling refers to the configuration in which all machines are tooled differently, and each operation is assigned to only one machine. A conventional job shop is a typical example of no pooling. No two machines are tooled identically in partial pooling as well. However, in this configuration, an operation can be processed at more than one machine.

Clearly, in both no pooling and partial pooling configurations, MGP is solved trivially because $G = M$. We now discuss MGP for the total pooling case; first, however, we restate the problem appropriately.

(MMFT2)

$$\min \frac{1}{\lambda}\sum_{g=1}^{M} W_g \sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} x_{ijg} \tag{7}$$

subject to

$$\sum_{g=1}^{M} x_{ijg} = 1, \forall i, j \tag{8}$$

$$\sum_{g=1}^{M}\sum_{m=1}^{M} y_{mg} = M \tag{9}$$

$$\sum_{j=1}^{N}\sum_{i=1}^{n_j} x_{ijg} \leq T^m y_{mg}, \forall m, g \tag{10}$$

$$\rho_g = \frac{\sum_{j=1}^{N} \lambda_j \sum_{i=1}^{n_j} p_{ij} x_{ijg}}{\sum_{m=1}^{M} y_{mg}} \tag{11}$$

$$0 \le \rho_g < 1, \forall g \tag{12}$$

$$x_{ijg}, y_{mg} \in \{0, 1\}, \forall i, j, g, m \tag{13}$$

This formulation introduces the additional variable y_{mg} which equal 1 if machine m is assigned to group g, and is zero otherwise. Equation (8) ensures that each operation is assigned to only one machine. This constraint replaces Equation (2) in problem MMFT1. Equation (9) ensures that all M are assigned to groups. Equations (10)–(13) parallel Equations (3)–(6) in MMFT1 with subscript m replaced by g wherever appropriate.

MGP considers MMFT2 at an aggregate level by combining all part types into a single part type. We use a Jacksonian, open queueing network to model the FMS. Let the FMS comprise G groups such that group g consists of m_g machines. Clearly, $G \le M$, and $\sum_{g=1}^{G} m_g = M$. Let the (transition) probability that a part that has completed processing at group i will next visit group j be given by π_{ij}. The probability that the part will exit the system after finishing at group i is $1 - \sum_{j=1}^{G} \pi_{ij}$. Let λ be the external part arrival rate at the system and λ_g be the average part arrival rate at group g. From traffic balance, we have

$$\lambda_g = \gamma_g + \sum_{i=1}^{G} \lambda_i \pi_{ig}, g = 1, \ldots, G$$

where γ_g is the external arrival rate at group g. We define α_g, the visit ratio at group g, as the average number of times a part is processed at g. In addition, let $1/\mu_g$ be the average processing time of a part at g, $1/\mu$ the average part processing time, L_g the average number of parts at group $g$and T_g the maximum number of operations that can be assigned to groups g.

We have the following identities:

$$\rho_g = \frac{\lambda_g}{m_g \mu_g},$$

$$\alpha_g = \lambda_g / \lambda = m_g \rho_g \mu_g / \lambda \tag{14}$$

$$\frac{1}{\mu} = \sum_{g=1}^{G} \alpha_g / \mu_g = \frac{1}{\lambda} \sum_{g=1}^{G} m_g \rho_g \tag{15}$$

We also have the following relationships between the variables in MMFT and MGP:

$$\alpha_g = \frac{\sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} x_{ijg}}{\lambda}, \forall g$$

$$m_g = \sum_{m=1}^{M} y_{mg}, \forall g$$

$$\frac{1}{\mu_g} = \frac{\sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} p_{ij} x_{ijg}}{\sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} x_{ijg}}, \forall g$$

In addition, because all machines within a group are tooled identically, the maximum number of operations T_g that can be assigned to group g is given by

$$T_g = \min_m \{T^m \mid y_{mg} = 1\}$$

The expected part flow time can now be restated as

$$\begin{aligned} MFT &= \frac{1}{\lambda} \sum_{g=1}^{G} W_g \sum_{j=1}^{N} \lambda^j \sum_{i=1}^{n_j} x_{ijg} \\ &= \sum_{g=1}^{G} W_g \alpha_g \\ &= \frac{1}{\lambda} \sum_{g=1}^{G} L_g \end{aligned}$$

For a given G, the machine grouping problem MGP_G can then be formulated as:

$$\min \frac{1}{\lambda} \sum_{g=1}^{G} L_g(m_g, \rho_g) \tag{16}$$

subject to

$$\sum_{g=1}^{G} \alpha_g / \mu_g = 1/\mu \tag{17}$$

$$\alpha_g \leq T_g, \forall g \tag{18}$$

$$\sum_{g=1}^{G} m_g = M \tag{19}$$

$$0 \leq \rho_g < 1, \forall g. \tag{20}$$

Using Equation (15), we rewrite Equation (17) as

$$\sum_{g=1}^{G} m_g \rho_g = \lambda / \mu = M\rho. \tag{21}$$

where $\rho = \frac{\lambda}{M\mu}$ is the overall system utilization. Note that the right hand side of the above equation is a constant for the planning level decisions.

The solution approach to MGP consists of first finding the optimal number of groups G^*. At the next step, we solve MGP_{G^*} to obtain the optimal partition $\mathbf{m}^* = (m_1^*, m_2^*, \cdots, m_{G^*}^*)$ of machines into groups, as well as the group utilizations $\rho^* = (\rho_1^*, \rho_2^*, \cdots, \rho_{G^*}^*)$. To show that these two steps can be considered independently, we use the following result.

Lemma 1 *MFT does not decrease if any group g is decomposed into two subgroups $g1$ and $g2$, the other groups remaining unchanged.*

PROOF: See Appendix 1.

Lemma 1 directly leads to the following result.

Theorem 1 *MFT is minimized by minimizing the total number of groups.*

This result holds independent of how the machines are partitioned into groups, and how the workloads are allocated.

Even for known values of G, MGP remains quite difficult to solve primarily because of the cumbersome expression relating L_g to ρ_g and m_g. In order to develop some characteristics of the optimal solution, we first consider specific (feasible) values of $m_g, g = 1, 2, \cdots, G$ and determine the optimal utilization levels ρ^* corresponding to a given $\mathbf{m}$. For the case in which $m_1 = m_2 = \cdots = m_G$, we have the following result.

Theorem 2 *MFT is minimized for a system of machine groups of equal sizes by allocating balanced workloads to machines in each group.*

PROOF: The proof is straightforward once we note that L_g is convex in ρ_g, and therefore, for equal-sized groups, MFT is a sum of identical convex functions.

However, for unequal-sized groups, the optimal group utilizations will depend upon the number of machines in each group. In the following, we determine these utilizations for 3-, 4- and 5- machine systems ignoring tool magazine capacity restrictions, and extend these results to the general system through Conjecture 1. [We omit the trivial case in which $G = 1$.]

Three machines can be grouped in two ways: (1, 1, 1) and (1, 2). From Theorem 2 it follows that MFT is minimized in the (1, 1, 1) configuration by providing equal machine utilizations which are given by

$$\rho_g^* = \rho = \frac{\lambda}{3\mu}, g = 1,2,3$$

and the resulting minimum mean flow time is

$$MFT^*(1,1,1) = \frac{1}{\lambda}\sum_{g=1}^{3} L_g = \frac{1}{\lambda}\sum_{g=1}^{G} \frac{\rho_g^*}{1-\rho_g^*} = \frac{3\rho}{1-\rho}$$

For the (1, 2) configuration, the minimum MFT is given by

$$MFT^*(1,2) = \frac{1}{\lambda}[\frac{\rho_1^*}{(1-\rho_1^*)} + \frac{2\rho_2^*}{1-\rho_2^{*2}}]$$

In Appendix 2 we show that ρ_1^* and ρ_2^* are obtained by solving

$$(4-12\rho)\rho_2^* + (5-6\rho+9\rho^2)(\rho_2^*)^2 + (4-12\rho)(\rho_2^*)^3 + 3(\rho_2^*)^4 - 6\rho + 9\rho^2 = 0, and$$

$$\rho_1^* = 3\rho - 2\rho_2^*.$$

The resulting values of ρ_1^* and ρ_2^* are shown in Figure 1 for various values of ρ. Note that $\rho_2^* > \rho_1^*$ for all ρ. In addition, as $\rho \to 1$, $\rho_1^* \to \rho_2^* \to \rho$. MFT is minimized by providing higher utilization to the larger group. However, as the overall system utilization increases, the degree of imbalance reduces; in the limit, both groups have the same utilization.

INSERT FIGURE 1 HERE

Figure 2 compares the MFT values for (1, 1, 1) and (1, 2) configurations. Note that providing appropriately unbalanced utilizations to the two groups in the (1, 2) configuration decreases MFT. The difference between these configurations increases with an increase in overall system utilization.

INSERT FIGURE 2 HERE

Four machines can be grouped in 4 ways — (1, 1, 1, 1), (1, 1, 2), (2, 2) and (1, 3). Similarly, alternative configurations possible in a 5-machine system are (1, 1, 1, 1, 1), (1, 1, 1, 2), (1, 2, 2), (1, 1, 3), (2, 3) and (1, 4). Figures 3 and 4 depict the MFT values obtained under these configurations given optimal group utilization levels for 4- and 5-machine systems respectively. These figures extend the result obtained earlier for the 3- machine system. In addition, they bring out the relative impact of fewer groups and unequal group utilizations independently.

INSERT FIGURES 3 AND 4 HERE

Consider, for example, the 5-machine system. MFT decreases as the number of groups decreases from 5 to 2. For a given number of groups, MFT is minimized by maximally unbalancing the group sizes. For instance, (1,1,3) is superior to (1,2,2) when $G = 3$. Similarly, (1,4) is superior to (2,3) when $G = 2$. Figures 4 and 5 also show that reducing the number of groups is more effective than unbalancing groups sizes and allocating appropriately unbalanced workloads.

These results lead to the following conjecture.

Conjecture 1 *MFT is minimized by minimizing the number of machine groups, pooling machines into unequal groups, and by allocating appropriately unbalanced workloads to these groups.*

Because of the cumbersome nature of the MFT function, it is difficult to verify the generality of this assertion. However, it has been proved to be true for the several systems that we have examined. Theorem 1 and Conjecture 1 parallel the conjectures stated in Stecke and Solberg (1985) who studied the production rate function in closed queueing networks.

MGP then reduces to 1) determining the minimum number of groups required to process all operations, 2) allocating the available machines to these groups such that these group sizes are maximally unbalanced, and 3) determining the appropriate groups utilization levels. These steps are now discussed.

If all machines have the same tool magazine capacity T, then the minimum number of groups required is given by

$$G^* = \left\lceil \frac{\sum_{j=1}^{N} n_j}{T} \right\rceil$$

If all machines do not have the same tool magazine capacity, G^* can be found by the following procedure. Renumber all machines in the nondecreasing order of T^m. Then, G^* is the smallest integer K such that

$$\frac{\sum_{j=1}^{N} n_j}{\sum_{l=1}^{K} T^l} \leq 1.$$

The optimal grouping configuration is given by

$$\mathbf{m}^* = (1, 1, \cdots, M - G^* + 1)$$

The optimal group utilizations are obtained by solving the following problem.

$$\min \frac{1}{\lambda}\Big[\sum_{g=1}^{G^*-1} \frac{\rho_g}{1-\rho_g} + L(M - G^* + 1, \rho_{G^*})\Big]$$

subject to

$$(G^* - 1)\rho_g + (M - G^* + 1)\rho_{G^*} = \lambda/\mu$$

$$0 \leq \rho_g < 1, \forall g$$

Because G^* is known, $L(M - G^* + 1, \rho_{G^*})$ can be expressed as a function of only ρ_{G^*}. This problem can be solved in a manner similar to the 3-, 4- and 5-machine systems discussed earlier.

4 Machine Loading Problem

Given the solution to MGP, the machine loading problem deals with the allocation of operations to individual groups such that deviations of actual utilizations from their ideal values are minimized. This leads to the following formulation for the case of total pooling.

$$\min \sum_{g=1}^{G^*} \mid \rho_g - \rho_{g^*} \mid$$

subject to

$$\sum_{g=1}^{G^*} x_{ijg} = 1, \forall i, j \tag{22}$$

$$\sum_{j=1}^{N} \sum_{i=1}^{n_j} x_{ijg} \leq T_g, \forall g \tag{23}$$

$$\rho_g = \frac{\sum_{j=1}^{N} \lambda_j \sum_{i=1}^{n_j} p_{ij} x_{ijg}}{m_g} \tag{24}$$

$$0 \leq \rho_g < 1, \forall g \tag{25}$$

$$x_{ijg} \in \{0,1\}, \forall i, j, g \tag{26}$$

For the cases of no pooling and partial pooling, we replace subscript g with m. In addition, for partial pooling we substitute Equation (22) with

$$L_{ij} \leq \sum_{g=1}^{G^*} x_{ijg} \leq U_{ij}, \forall i, j \tag{27}$$

Because of the combinatorial nature of the problem, an efficient solution is unlikely to exist. We propose a heuristic solution approach which is a modification of the first fit decreasing heuristic for the bin packing problem. The algorithm consists of the following steps:

1. Initialization: a) Determine the target workloads $\theta_g, g = 1, 2, \ldots, G^*$.

   $$\theta_g = m_g \rho_g, g = 1, 2, \ldots, G^*.$$

   b) Initialize the counters for the current workload, the remaining assignable workload. and the remaining tool magazine capacity τ_g for each group.

   $$\begin{aligned} W_g &= 0, g = 1, 2, \ldots, G^* \\ \Delta_g &= \theta_g, g = 1, 2, \ldots, G^* \\ \tau_g &= \min_{m \in g} T^m, g = 1, 2, \ldots, G^*. \end{aligned}$$

   c) Form two lists of operations. For no pooling and total pooling, the primary list consists of one copy of each operation, and the second list is empty. For partial pooling, the primary list consists of L_{ij} copies, and the secondary list consists of $U_{ij} - L_{ij}$ copies of each operation. Arrange all operations in both lists in the decreasing order of $w_{ij} = \lambda_j p_{ij}$.

2. Select the group g^* with the largest Δ_g / m_g. Break ties in favor of the group with the largest m_g.

3. Assign the operation i^*j^* at the head of the list of unassigned operations in the primary list to g^*, and update workloads and available tool magazine capacity.

$$\begin{aligned} W_{g^*} &\leftarrow W_{g^*} + w_{i^*j^*}, \\ \Delta_{g^*} &\leftarrow \Delta_{g^*} - w_{i^*j^*}, \\ \tau_g &\leftarrow \tau_g - 1. \end{aligned}$$

If $\tau_g = 0$, eliminate group g from further consideration.

4. Repeat Steps 2–3 until all operations in the primary list are assigned.

5. Stop in the cases of no pooling and total pooling. For partial pooling, go to Step 6.

6. With respect to operation i^*j^* at the head of the unassigned operations in the secondary list, find group g^*, such that

$$\begin{aligned} x_{i^*j^*g^*} &= 0, \\ \frac{\Delta_{g^*} + w_{i^*j^*}}{m_{g^*}} &< 1, \\ \left| \frac{\Delta_{g^*} + w_{i^*j^*}}{m_{g^*}} - \rho_g^* \right| &< \left| \frac{\Delta_{g^*}}{m_{g^*}} - \rho_g^* \right| \\ g^* &= arg \max_g \{ \frac{\Delta_g}{m_g} \}. \end{aligned}$$

If these conditions satisfied for any group, assign operation i^*j^* to group g^* and update workload and remaining tool magazine capacity as shown in Step 3. Otherwise, discard all copies of operation i^*j^* from the secondary list.

7. Repeat Step 6 until the secondary list is empty.

In many systems, it may be appropriate considering other loading objectives in addition to the objective of ensuring appropriate group utilizations. Following Stecke (1983), we consider two such objectives. First is the objective of minimizing part movements. This objective is particularly useful for the case in which travel times are significant and/or the material transporters are heavily utilized. In addition, this objective leads to an aggregation of operations of a given job at any machine. Consequently, it has to join fewer machine queues.

The formulation of the machine loading problem corresponding to the objective of minimizing part travel (MLPMT) is given below.

$$\min \sum_{g=1}^{G^*} \sum_{j=1}^{N} \sum_{i=1}^{n_j-1} \mid x_{ijg} - x_{i+1,jg} \mid$$

subject to

$$\mid \rho_g - \rho_{g^*} \mid \leq \epsilon, \ \forall g \tag{28}$$

and (22)–(26)

In this formulation, ϵ denotes the maximum deviation from the ideal value permitted to the actual utilization of any group.

The second objective considered is the maximization of the weighted number of operation duplications; the assigned weights reflect the criticality of the individual operations. This objective attempts to reduce system congestion by providing more alternative routes to operations which have greater impact on the overall system performance.

Determining the criticality c_{ij} of a given operation i in job j is, however, difficult. If processing times are a measure of criticality, then longer operations will be assigned more often under this objective. On the other hand, if all operations are considered equally critical, then this objective will lead to more duplications of the shorter operations. We consider the relative merits of these two extreme scenarios in greater detail in §4.

The loading problem corresponding to the secondary objective of maximizing flexibility MLPMF is formulated below.

$$\min \sum_{g=1}^{G^*} c_{ij} x_{ijg}$$

subject to

(22)–(26), (28)

5 Experimental Study

In this section we use simulation experiments to extend our investigation to a general FMS. One objective of these experiments is to evaluate the robustness of the results obtained in the previous sections when the assumption regarding exponentially distributed operation processing times is relaxed. Specifically, we now consider parts with deterministic processing times. In addition, we measure the effectiveness of the various grouping and loading objectives under different values of the system parameters and in the face of system disruptions.

The two system parameters that we study are system utilization level ρ, and the coefficient of variation of the operation processing times $CVOPT$. While the impact of ρ on MFT is well known, different system configurations are likely to respond differently to a change in ρ. In addition, the disruptive impact of system uncertainties depends strongly upon the level at which the system is utilized.

Recent studies (see, for example, Monahan and Smunt 1989, Kochman 1989) show that mean part flow time is affected significantly by the variability in the operation processing times. While CVOPT can be considered as a surrogate for system disruptions, it merits independent consideration because it affects the coefficient of variation of service times at individual machines. An increase in CVOPT will result in larger MFT for any system. An important measure of the effectiveness of any configuration is its robustness against changes in CVOPT.

We consider two kinds of system disruptions. First is machine breakdowns. The degree of unreliability of a machine is usually expressed in terms of its mean time to failure (MTTF). [See, for example, Groff and Muth 1972.] Smaller the MTTF, greater the unreliability. The second type of disruption that we consider is the variation in the batch size of a given part. Such variations are caused, for example, by fluctuations in customer order quantity. In many multi- stage manufacturing systems, they follow also as a result of variations in yield and/or transfer batch sizes. From queueing theory (see, for example, Kleinrock 1975) that bulk arrivals, especially in varying batch sizes, result in larger MFT.

Conway et al. (1967) note that one of the major benefits of providing routing flexibility is that the system is less sensitive to the actual scheduling rule used. One of the major objectives at the FMS planning stage is to simplify the decisions that need to be made at the scheduling and control stage. Following this observation, we treat the insensitivity of a given configuration to the quality of the scheduling rule used as another measure of its effectiveness.

5.1 Experimental Design

We consider a dynamic FMS which produces ten part types to order; these orders arrive randomly at the system following a Poisson process. The FMS consists of six machines. All machines are identical in the sense that they can process any of the required operations, provided they are tooled accordingly, and the processing time of any operation is the same across all machines. Each part type requires six operations with fixed processing times. We ignore travel times between machines. In effect, we assume that parts are transported instantaneously from one machine to another.

Grouping Configurations

Three grouping configurations — no pooling, partial pooling and total pooling are considered. No pooling results in group sizes of one. We assume that under total pooling three groups are required. We consider the maximally unbalanced configuration (1, 1, 4), the balanced configuration (2, 2, 2), and an intermediate configuration (1, 2, 3). Solving MGP for (2, 2, 2), (1, 2, 3) and (1, 1, 4) configurations results in the optimal groups utilization levels shown in Table 1 corresponding to various values of the overall system utilization ρ. For the partial pooling configuration, we use $L_{ij} = U_{ij} = 2$ in order to make it comparable to the total pooling case with equal group sizes.

TABLE 1: Optimal Group Utilization Levels

ρ	*(1, 2, 3)*			*(1, 1, 4)*			*(2, 2, 2)*
	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	$\rho_1 = \rho_2 = \rho_3$
0.1	0.008	0.074	0.148	0.002	0.002	0.149	0.1
0.2	0.044	0.173	0.270	0.026	0.026	0.287	0.2
0.3	0.111	0.276	0.378	0.084	0.084	0.408	0.3
0.4	0.203	0.382	0.478	0.174	0.174	0.513	0.4
0.5	0.313	0.487	0.571	0.288	0.288	0.606	0.5
0.6	0.438	0.589	0.661	0.418	0.418	0.691	0.6
0.7	0.572	0.694	0.747	0.558	0.558	0.771	0.7
0.8	0.712	0.796	0.832	0.702	0.702	0.849	0.8
0.9	0.855	0.890	0.917	0.850	0.850	0.925	0.9

Loading Configurations

Alternative loading objectives discussed in §4 are used in conjunction with the grouping configurations mentioned above to generate the system configurations listed in Table 2. C1 through C5 are constructed by solving MLP for the three machine pooling cases. C1 is the base configuration which corresponds to a conventional job shop. It is used primarily as a benchmark to evaluate the relative performance of the other configurations.

TABLE 2: System Configurations Considered

System Configuration	*Loading Objective*	*Level of Pooling*	*Grouping Configuration*
C1	MLP	No Pooling	(1, 1, 1, 1, 1, 1)
C2	MLP	Partial Pooling	(1, 1, 1, 1, 1, 1)
C3	MLP	Total Pooling	(2, 2, 2)
C4	MLP	Total Pooling	(1, 2, 3)
C5	MLP	Total Pooling	(1, 1, 4)
C6	MLPMT	No Pooling	(1, 1, 1, 1, 1, 1)
C7	MLPMT	Partial Pooling	(1, 1, 1, 1, 1, 1)
C8	MLPMT	Total Pooling	(2, 2, 2)
C9	MLPMT	Total Pooling	(1, 2, 3)
C10	MLPMT	Total Pooling	(1, 1, 4)
C11	MLPMF	Total Pooling	(1, 2, 3)
C12	MLPMF	Total Pooling	(1, 1, 4)

Solving MLPMT results in configurations C6–C10 which parallel those obtained for the MLP objective under each machine pooling scenario. Configurations C11—C12 are obtained by solving MLPMF for (1, 2, 3) and (1, 1, 4) groupings. Each of these two configurations is further decomposed into two subconfigurations corresponding to the weights associated with the different operations. In C11A and C12A, all operations are given equal weights. This results in shorter operations being assigned to larger groups. Consequently, they are duplicated more often. In C11B and C12B, longer operations are assigned higher weights. This generates configurations in which the longer operations are assigned to larger groups.

Four system utilization levels — 60%, 70%, 80% and 90% are considered. Operation processing times are sampled from a uniform distribution to yield three levels — 0.0, 0.4, and 0.8 of CVOPT. We consider three levels of machine unreliability by varying the mean time to failure (MTTF) for each machine. An exponential distribution is used to represent the time to next failure for any machine. MTTF is selected to be 0.0, 10 $\overline{p}$ and 5 $\overline{p}$, where $\overline{p}$ is the average part processing time, corresponding to the three levels of increasing unreliability. In each case, the mean time to repair a machine is sampled from a uniform distribution with mean $0.3^*\overline{p}$.

Three levels of batch sizes are used in the study. In the first case, batch size is fixed at 1. In the second level batch size is sampled from the uniform distribution (3,7) while in the third level batch size is sampled from the uniform distribution (6,14) with mean 10. Note that in the two latter cases the ratio of the range to mean is the same.

Two scheduling rules — First-come-first-serve (FCFS) and Shortest Processing Time (SPT) are used to measure the impact of routing flexibility on scheduling rule performance. FCFS is used primarily to serve as a benchmark. SPT is widely regarded as an effective heuristic for the mane flow time problem. Thus, a large difference between the FCFS and SPT values for a given configuration implies that it is very sensitive to the quality of the scheduling rule.

The method of replications is used to obtain the summary statistics. Each scenario is replicated five times; within each replication, steady state statistics are obtained for over 4500 parts.

5.2 Experimental Results

Impact of Grouping Configuration

The reported values of MFT are normalized with respect to the average part processing time. The first set of results corresponds to the impact of configurations C1–C5. Figure 5 shows the impact of CVOPT on MFT for these 5 configurations. Two results follow from these graphs. First, the performance of partial pooling and total pooling relative to no pooling improves with an increase in CVOPT. Second, partial pooling performs the best across all values of CVOPT and at all utilization levels. Once again, the relative superiority of using partial pooling increases with CVOPT; it also increases with an increase in the utilization level. Among the total pooling configurations, C3 is superior at low CVOPT; however, as CVOPT increases, the unbalanced configurations C4 and C5 perform better, especially at high utilizations. In particular, C5 is the best configuration at 90% utilization and at CVOPT=0.8.

INSERT FIGURE 5 HERE

These configurations exhibit varying levels of sensitivity to the scheduling rule used as shown in Table 3 for 90% utilization level. C1 is most sensitive, and the impact of scheduling rule increases with an increase in CVOPT. This is expected because as difference among operation processing times increases, the impact of schedule quality increases. On the other hand, C2 is least sensitive to the scheduling rule used. Its insensitivity does not depend upon CVOPT. C3–C5 show varying degrees of sensitivity. In particular, the impact of using a better scheduling rule increases for the unbalanced configurations C4 and C5 with an increase in CVOPT. In particular, at CVOPT=0.8, the performance of C5 approaches that of C2.

TABLE 3: Impact of Scheduling Rules

CVOPT	*Configuration*	*MFT under*		*% Decrease under SPT*
		FCF	SPT	
0.0	1	3.46	3.46	0
	2	2.40	2.40	0
	3	2.45	2.45	0
	4	2.72	2.72	0
	5	3.11	3.11	0
0.4	1	4.50	3.49	22.6
	2	2.44	2.29	6.3
	3	2.90	2.57	11.4
	4	2.88	2.68	7.1
	5	2.97	2.89	2.7
0.8	1	8.25	4.75	42.5
	2	2.69	2.52	6.0
	3	4.38	3.33	23.9
	4	4.31	3.04	28.8
	5	4.14	2.77	33.2

Table 4 depicts the impact of machine unreliability. First, note that, in general, while increasing the level of unreliability increases MFT, the percentage increase comes down with an increase in CVOPT. This decrease is most prominent for the unbalanced configurations C4 and C5. Once again, C2 is the most robust configuration across all levels of unreliability, and its relative superiority improves with an increase in unreliability. Among the total pooling configurations, C3 is the most robust. As unreliability increases it results in increasingly better values of MFT than both C4 and C5. Note that all groups have 2 machines in C3. Therefore, if one machine fails, an alternative machine is available to process parts. At the other extreme, C5 has two groups with one machine each. Hence, if any one of these machine fail, the operations waiting at them are blocked.

TABLE 4: Impact of Machine Breakdowns

CVOPT	*Configuration*	*MFT under Level*			*% Increase over Level 0*	
		0	1	2	Level 1	Level 2
0.0	1	3.46	4.61	6.41	33.2	85.4
	2	2.40	2.80	3.37	16.5	40.3
	3	2.45	3.41	4.94	39.3	101.7
	4	2.72	4.94	6.86	81.9	152.4
	5	3.11	5.04	9.38	62.2	201.6
0.4	1	4.50	5.91	8.40	31.2	86.5
	2	2.44	2.86	3.40	16.9	38.8
	3	2.90	3.95	5.56	36.5	92.1
	4	2.88	4.10	6.52	42.2	126.6
	5	2.97	4.38	6.94	47.7	133.5
0.8	1	8.25	10.77	13.92	30.5	68.6
	2	2.69	3.08	3.58	14.9	33.2
	3	4.38	5.64	7.70	28.7	75.9
	4	4.31	5.92	8.31	37.5	92.8
	5	4.14	5.78	8.89	39.5	146.9

The impact of varying batch sizes is shown in Table 5. Once again we notice that as CVOPT increases the adverse impact of larger batch sizes decreases. C2 remains the most effective configuration; however, its performance is closely matched by C3 at low CVOPT. As CVOPT increases, the relative performance of C3 deteriorates. Interestingly, the unbalanced configurations exhibit greater sensitivity to batch size, and they perform poorly as the batch sizes increase. For example, while C5 is superior to C3 at batch size of 1 for CVOPT=0.8, it is much worse when batch size increases to 10.

TABLE 5: Impact of Batch Size

CVOPT	*Configuration*	*MFT under Batch Size*			*% Increase over BS = 1*	
		1	5	10	BS = 5	BS = 10
0.0	1	3.46	8.29	13.92	139.6	302.4
	2	2.40	7.54	13.50	214.1	462.5
	3	2.45	7.57	13.51	208.8	451.5
	4	2.72	9.20	15.86	238.6	483.9
	5	3.11	11.18	18.75	259.8	503.1
0.4	1	4.50	9.77	15.31	116.8	239.9
	2	2.44	7.55	13.46	208.8	450.4
	3	2.90	8.02	13.70	176.8	372.9
	4	2.88	8.88	15.19	208.5	427.6
	5	2.97	8.90	15.31	199.7	415.3
0.8	1	8.25	14.27	19.83	72.9	140.2
	2	2.69	7.79	13.80	189.9	413.3
	3	4.38	10.00	16.10	128.4	267.6
	4	4.31	10.87	17.72	152.3	311.4
	5	4.14	10.24	16.58	147.4	300.3

Impact of Secondary Loading Objectives

Table 6 compares the performance of the configurations generated by solving MLPMT with those obtained from MLP at 90% system utilization level. The results show that, in general for all configurations, operation aggregation leads to higher MFT at low CVOPT; however, it leads to superior performance at high CVOPT values. This is partly explained by the fact that with an increase in CVOPT, the coefficient of variation of service times (C_s) at each machine increases; this, in turn, leads to higher MFT. However, the aggregation of operations tends to reduce C_s, and therefore, MFT as well. Note, however, that in the case of partial pooling, operation aggregation is uniformly superior.

TABLE 6: Impact of Operation Aggregation

Configuration	*CVOPT*		
	0.0	0.4	0.8
No Pooling			
C1	3.46	4.50	8.25
C6	4.75	5.60	5.41
Partial Pooling			
C2	2.40	2.44	2.69
C7	2.25	2.31	2.33
Total Pooling			
C3	2.45	2.90	4.38
C8	2.85	3.10	3.79
C4	2.72	2.88	4.38
C9	2.70	2.75	2.77
C5	3.11	2.97	4.14
C10	2.97	3.02	3.11

Table 7 shows the impact of assigning operations based on MLPMF. Recall that C11A and C12A assign shorter operations to larger groups, and consequently, provide greater flexibility to them. On the other hand, C11B and C12B provide greater flexibility to the longer operations. For comparison purposes, we also show the MFT values obtained under MLP. The results indicate that at low values of CVOPT, the C11B and C12B are better, although they are comparable to C4 and C5 respectively. However, at higher CVOPT, C11A and C12A are significantly superior. This shows that at such values of CVOPT, it is preferable to provide more alternative routes to as large a number of operations as possible. More importantly, this result shows that the weights associated with each operation to indicate its criticality is likely to be dependent upon CVOPT.

TABLE 7: Impact of Operation Duplication

Configuration	*CVOPT*	
	0.4	0.8
C4	2.88	4.31
C11A	3.39	3.52
C11B	3.20	3.74
C5	2.97	4.14
C12A	3.28	3.46
C12B	2.91	4.62

6 Summary

This paper investigates the FMS setup problems of partitioning machines into groups, determining the appropriate group utilization levels, and assigning operations to these groups. An open queueing network representation of an FMS is used to show that, under total pooling, the mean job flow time is minimized when machine groups are maximally unbalanced, and the larger groups are utilized more heavily.

Three grouping configurations — no pooling, partial pooling and total pooling, and three loading objectives are used for generating a variety of system configurations. A simulation experiment is used to study the impact of various system parameters, and to compare the performance of these configurations. Experimental results show the superiority and robustness of partial pooling across a range of different values that these parameters can take. Its performance is improved further if overall part movement is reduced by performing several operations of a job at the same machine. The performance of various configurations under total pooling depends upon, among other factors, the overall system utilization level and CVOPT. In particular, configurations with unequal group sizes are superior under high system utilizations and high processing time variations. A reduction in either of these two factors tends to favor the use of equal-sized groups with balanced groups.

Among the loading objectives, greater operation aggregation leads to superior performance at high CVOPT for all grouping configurations. For partial pooling and total pooling with unbalanced configurations, it does so at low CVOPT as well. Experimental results also indicate that when CVOPT is low, longer operations should be provided duplicated more often. However, at high CVOPT, it is important to assign greater routing flexibility to a larger number of operations.

In summary, while high CVOPT, machine unreliability and large batch sizes are all detrimental to system performance, their impact can be reduced by partial pooling of machines, aggregating operations of a job to be performed at the same machine, and maximizing the sum of operation duplications.

Appendix 1

Proof of Lemma 1

First, note that group g, and after decomposition, subgroups $g1$ and $g2$ can be considered independent of other groups. Let α_{g1} (α_{g2}), m_{g1} (m_{g2}), and λ_{g1} (λ_{g2}) denote respectively the visit ratio, number of machines and arrival rate for $g1$ and $g2$ respectively. Then, we have

$$m_g = m_{g1} + m_{g2} \tag{29}$$

Also, from workload balance

$$\frac{\lambda_g}{\mu_g} = \frac{\lambda_{g1}}{\mu_{g1}} + \frac{\lambda_{g2}}{\mu_{g2}}$$

Hence,

$$m_g\rho_g = m_{g1}\rho_{g1} + m_{g2}\rho_{g2} \tag{30}$$

The increase in MFT because of decomposing g is

$$\begin{aligned} \Delta MFT &= [(\alpha_{g1}W_{g1} + \alpha_{g2}W_{g2}) - \alpha_g W_g] \\ &= \frac{1}{\lambda}[(L_{g1} + L_{g2}) - L_g] \end{aligned}$$

But

$$\begin{aligned} L_g &= \frac{\lambda_g}{\mu_g} + L_g^q \\ &= m_g\rho_g + L_g^q \end{aligned}$$

where L_g^q denotes the mean queue length at group g. Hence,

$$\begin{aligned} \Delta MFT &= \frac{1}{\lambda}[\{(m_{g1}\rho_{g1} + m_{g2}\rho_{g2}) - m_g\rho_g\} + \{(L_{g1}^q + L_{g2}^q) - L_g^q\}] \\ &= \frac{1}{\lambda}[(L_{g1}^q + L_{g2}^q) - L_g^q] \end{aligned}$$

From the theory of queues, we know that the mean queue length L_c^q in a single-channel system with c parallel servers is given by

$$L_q^c = p_0\frac{(c\rho)^c\rho}{c!(1-\rho)^2}$$

where ρ is the server utilization, and p_0 is the probability that an arriving part finds the system empty. Note that L_c^q is convex in ρ for a given c, and convex in c for a given ρ.

From Equations (29) and (30), it follows that ρ_g is a convex combination of ρ_{g1} and ρ_{g2}. Therefore, $L_g^q < L_{g1}^q + L_{g2}^q$, and consequently $\Delta MFT > 0$.

This proves the lemma.

Appendix 2

MFT under (1, 2) configuration

MGP for (1, 2) configuration can be written as

$$\min MFT = \frac{1}{\lambda}[\frac{\rho_1}{(1-\rho_1)} + \frac{2\rho_2}{1-\rho_2{}^2}]$$

subject to

$$\rho_1 + 2\rho_2 = 3\rho \tag{31}$$

$$0 \le \rho_g < 1, g = 1, 2.$$

Associating multipliers u with constraint (31) and using the Kuhn-Tucker conditions yields the following relationships at the optimal solution

$$u = \frac{1}{\lambda(1-\rho_1)^2} = \frac{(1+\rho_2^2)}{\lambda(1-\rho_2^2)^2}$$

or

$$\rho_1 = 1 - \frac{(1-\rho_2^2)}{\sqrt{(1+\rho_2^2)}} \tag{32}$$

From (31) and (32), we have

$$(4-12\rho)\rho_2 + (5-6\rho+9\rho^2)(\rho_2)^2 + (4-12\rho)(\rho_2)^3 + 3(\rho_2)^4 - 6\rho + 9\rho^2 = 0, and$$

$$\rho_1 = 3\rho - 2\rho_2$$

References

1. Ammons, J. C., C. B. Lofgren and L. F. McGinnis (1984), "A Large Scale Workstation Loading Problem," in *Proceedings of the First ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 249–255.

2. Berrada, M. and K. E. Stecke (1986), "A Branch and Bound Approach for Machine Load Balancing in Flexible Manufacturing Systems," *Management Science*, Vol. 32, 1316-1335.

3. Buzacott, J. A. and D. D. Yao, "Flexible Manufacturing Systems: Review of Analytical Models," *Management Science*, Vol. 32, 890–905.

4. Chakravarty, A. K. and A. Shtub (1984), "Selecting parts and Loading Flexible Manufacturing Systems," in *Proceedings of the First ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 284–289.

5. Conway, R. W. , W. L. Maxwell and L. W. Miller (1967), *Theory of Scheduling*, Addison-Wesley, Reading, MA.

6. Groff, G. K. and J. F. Muth (1972), *Operations Management: Analysis for Decisions*, Irwin, Homewood, IL.

7. Hwang, S. (1986), "A Constraint-Directed Method to Solve the Part Selection Problem in Flexible Manufacturing Systems Planning Stage", in *Proceedings of the Second ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 297–310.

8. Kleinrock, L. (1975), *Queueing Systems: Volume 1: Theory*, John Wiley, New York, NY.

9. Kochman, G. A. (1989), "The Performance of Manufacturing Systems in a Disruptive Environment," presented at the TIMS Meeting, Osaka, Japan.

10. Kusiak, A. (1984),"The Parts Families Problem in Flexible Manufacturing Systems," in *Proceedings of the First ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 237–242.

11. Monahan, G. E. and T. L. Smunt (1989), "Product-Process Relations in Batch Manufacturing," Working Paper # 89–1605, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.

12. Rajagopalan, S. (1986), "Formulation and Heuristic Solutions for Parts Grouping and Tool Loading in Flexible Manufacturing Systems", in *Proceedings of the Second ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 311-320.

13. Shanthikumar, J. G. and D. D. Yao (1987), "Optimal Server Allocation in a System of Multi-Server Stations," *Management Science*, Vol. 33, 1173–1191.

14. Shanthikumar, J. G. and D. D. Yao (1988), "On Server Allocation in Multiple Center Manufacturing Systems," *Operations Research*, Vol. 36, 333–342.

15. Stecke, K. E. (1983), "Formulation and Solution of Nonlinear Production Planning Problems for Flexible Manufacturing Systems," *Management Science*, Vol. 29, 273–288.

16. Stecke, K. E. (1986),"A Hierarchical Approach to Solving Machine Grouping and Loading Problems of Flexible Manufacturing Systems," *European Journal of Operational Research*, Vol. 24, 369–378.

17. Stecke, K. E. and J. J. Solberg (1985), "The Optimality of Unbalancing Both Workloads and Machine Group Sizes in Closed Queueing Networks of Multiserver Queues," *Operations Research*, Vol. 33, 882–910.

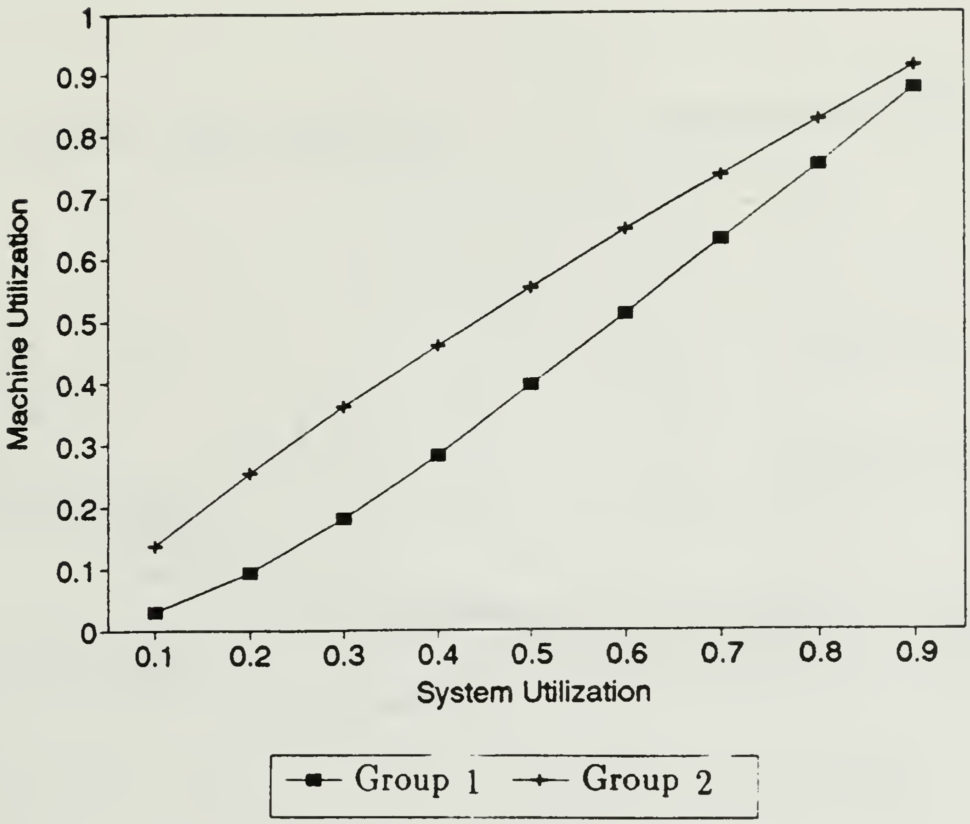

Figure 1: Optimal Machine Utilizations

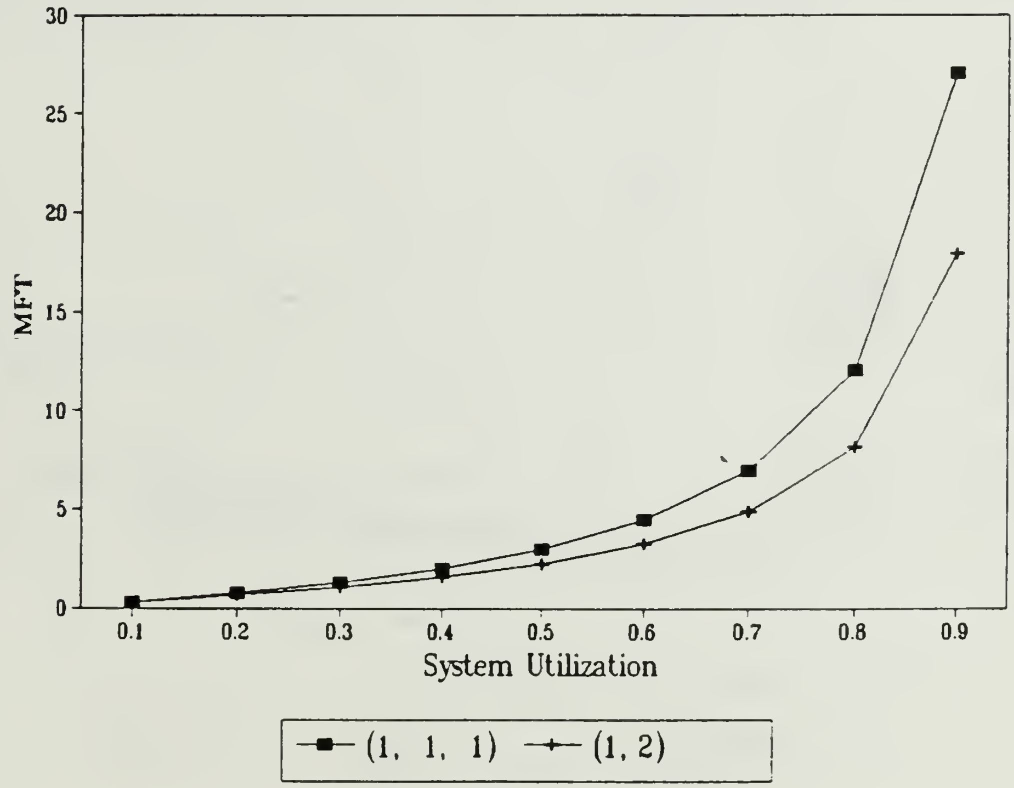

Figure 2: MFT under Alternative Grouping Configurations: 3- machines

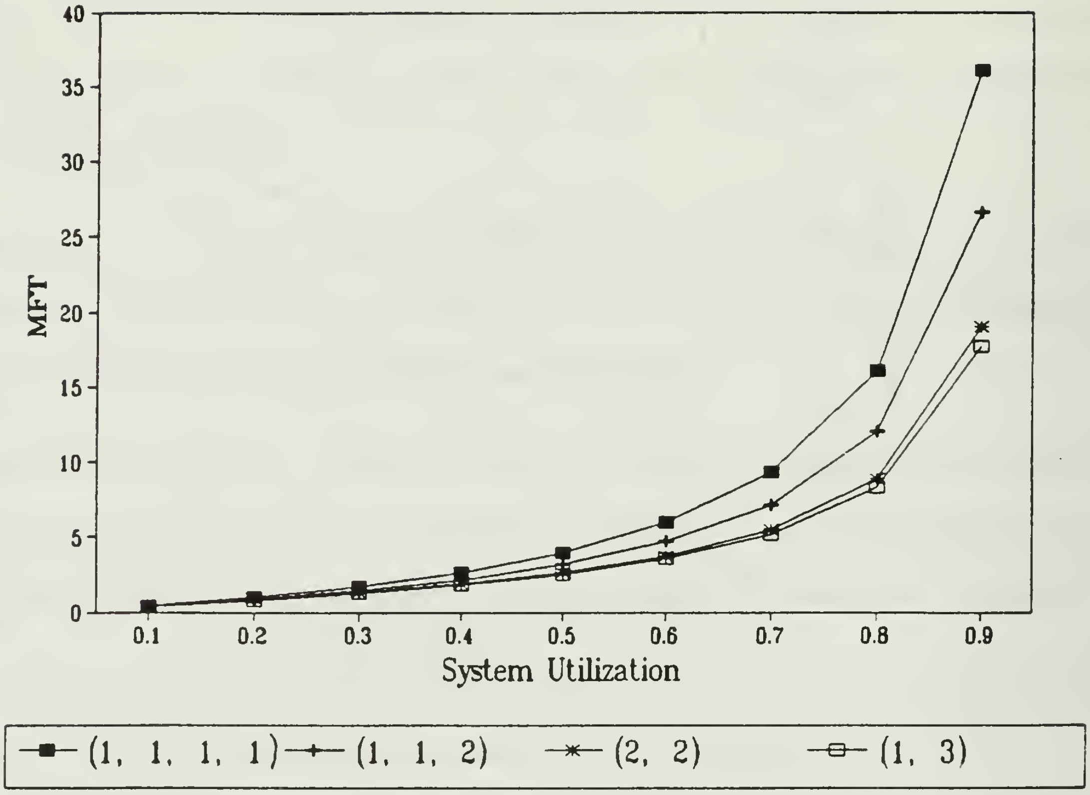

Figure 3: MFT under Alternative Grouping Configurations: 4- machines

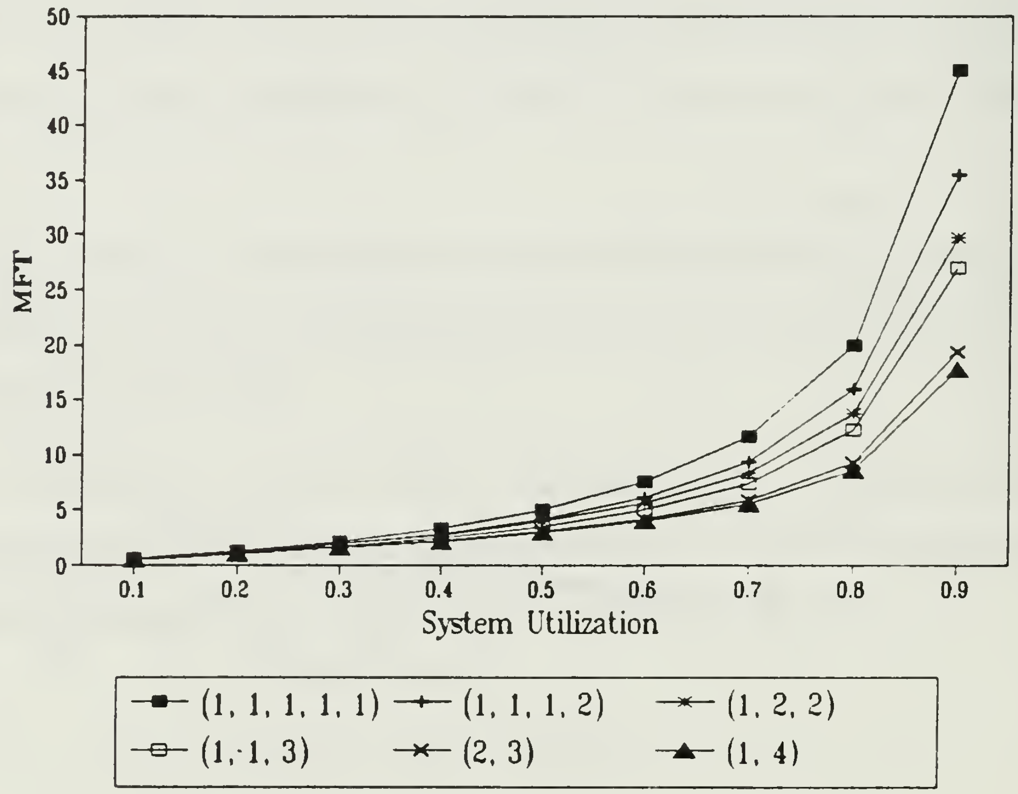

Figure 4: MFT under Alternative Grouping Configurations: 5- machines

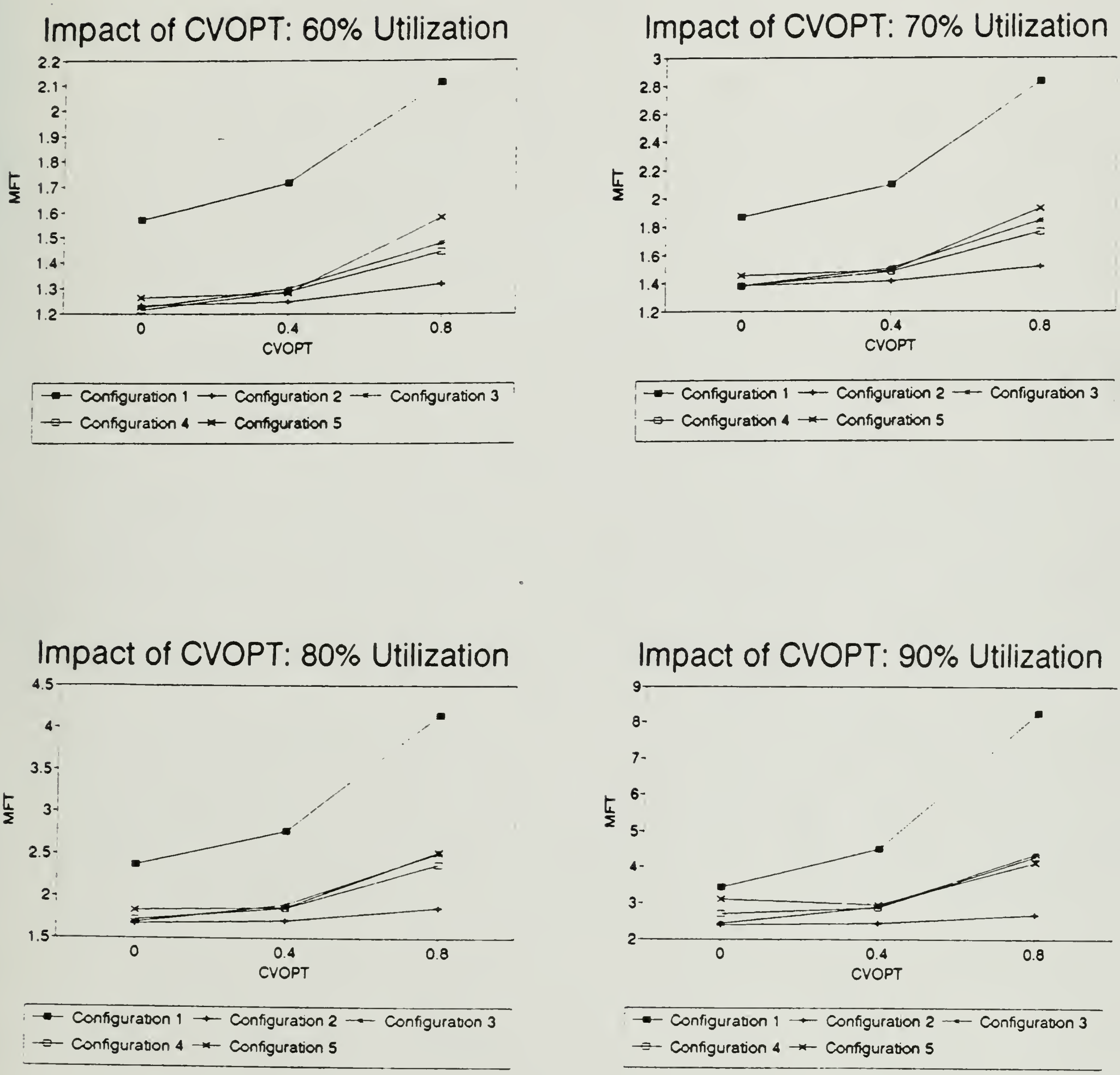

Figure 5: Impact of CVOPT