Journal of Intelligent Manufacturing (1992) **3**, 75–99

Design and planning problems in flexible manufacturing systems: a critical review

PANAGIOTIS KOUVELIS

Management Department, Graduate School of Business, The University of Texas at Austin, USA

Received November 1990 and accepted April 1991

This review paper describes the state-of-the-art research on flexible manufacturing systems (FMS) design and planning issues. The emphasis is on presenting research results coming out of the current FMS literature that help the FMS manager in setting up a highly efficient manufacturing system. In addition to that, it discusses relevant research contributions after 1986, that were not part of any of the previous survey papers on operations research models for FMSs. Also, applications of combinatorial optimization approaches to FMS planning problems are adequately exposed in the paper.

Keywords: Flexible manufacturing, design, planning, modeling

1. Introduction and problem classification

1.1. *Introduction*

Flexible manufacturing is an efficient alternative to conventional batch manufacturing. High work-in-process (WIP) levels and low machine utilization were always indicative of inefficiency in batch manufacturing. Even in the United States, the world's largest center of mass production, 75% of the parts produced by metal processing are in lots of less than 50 pieces (Merchant, 1983). The need for a highly productive alternative to batch manufacturing has always been the major driving force behind the flexible manufacturing system (FMS) development. Hard automation, that could efficiently turn thousands and millions of identical parts, was not the answer to efficient batch manufacturing. What was needed was flexible automation that could handle a large and constantly changing variety of produced items. The technological advancements of the 1970s such as on-line availability of computers and numerical control techniques made flexible automation possible, and the development of the first FMSs took place. FMSs have become widely used in a diverse set of industries. Application of FMSs can be found in the aerospace, agricultural, appliance, automobile, defense, electronic, machine tool, motor and engine components, and other industries throughout the United States, Japan and Europe. Technical descriptions of FMSs can be found in Dupont-Gatelmand (1982) and Hatvany (1983) among others.

A widely accepted definition of a FMS is the following (Browne *et al.*, 1984):

> 'A Flexible Manufacturing System is an integrated, computer-controlled complex of automated material handling devices and numerically controlled (NC) machine tools that can simultaneously process medium-sized volumes of a variety of part types. This new production technology has been designed to attain the efficiency of well-balanced machine paced transfer lines, while utilizing the flexibility that job shops have to simultaneously machine multiple part types.'

Sound analysis used to support the FMS design process is the key to achieving a good and cost-effective design. Typically, simulation models are used for the analysis performed during the FMS design. Simulation modelling is a powerful and flexible analysis technique, but it is not adequate in addressing design issues. Since simulation model development, debugging and analysis are time-consuming processes, and consequently extremely expensive activities, only a limited number of alternatives are explored for the design issues that are addressed. This point was emphasized by Solberg (1977):

'Long before the operating policies of a manufacturing system are considered many design decisions are made which affect the ultimate ability of production managers to control the performance of the system. Although some of the more advanced companies employ simulation methods to "fine tune" their system design, few make use of any formal methodologies at all in the critical early stages. Fundamental design issues, such as how large the system will be or the selection of processing and material handling equipment, are usually dealt by arbitrary choice or back of the envelope calculations. It is ironic that the most important decisions, those having the greatest long-term impact on system productivity are handled in the least careful manner.'

As one would expect, in many cases FMS managers struggle with scheduling problems under unrealistically tight capacity constraints due to inefficient designs and long-term plans.

A first needed tool for a FMS designer is an accurate performance model of the manufacturing system. Though queueing network models were introduced in the classical papers by Jackson (1957 and 1963) a long time ago, their extensive use as adequate performance models for automated manufacturing systems (in particular FMSs) appeared much later (Solberg, 1977; Buzacott and Shanthikumar, 1980; Stecke and Solberg, 1985; Yao and Buzacott, 1986 and 1987). Most of the previous survey papers on operations research approaches to FMSs (Buzacott and Yao, 1986; Kalkunte *et al.*, 1986 and Kusiak, 1985 and 1986) emphasize the modelling aspects to FMS operation. Also, they devote a significant effort in pointing out and classifying new and challenging problems in FMS production planning and control. This review paper describes the state-of-the-art research on FMS design and planning issues. Though in certain cases it is necessary for clarity of exposition of the research results to remind the reader of relevant modelling issues, the emphasis of our survey paper is in exposing important approaches and research results coming out of the current FMS literature that help the FMS designer and manager in setting up a highly efficient manufacturing system. In addition to that, it discusses relevant research contributions after 1986 that were not part of any of the previous reviews. The period 1986–91 is marked by the wide acceptance of approximate queueing network models as adequate performance models for automated manufacturing systems (research work initiated by Whitt (1983a and b), followed by Bitran and Tirupati (1988), Boxma *et al.* (1990 and 1991) and others). As expected, that motivated the development of new approaches to design and planning problems in manufacturing (Schweitzer and Seidmann, 1988a and b; Bitran and Tirupati, 1989a; Dallery and Stecke, 1990 and Boxma *et al.*, 1990 and 1991) that are presented, and appropriately interpreted in a FMS context for the first time, together with other recent optimization approaches in queueing network modelling frameworks (Shanthikumar and Yao, 1987 and 1988 and Kouvelis and Tirupati, 1991). Also, the application of combinatorial approaches (mostly graph theoretic and integer programming) to FMS planning problems has not been dealt with in detail in the previous surveys (with the exception of an exposition of results in loading and grouping problems (Kalkunte *et al.*, 1986; Stecke, 1986 and Stecke and Suri, 1989), and is discussed extensively in this review paper. A separate section of the paper is devoted to discussing research directions of potential interest to FMS researchers. Although the paper discusses a large set of research papers, it is not intended to be an exhaustive review of papers on the subject.

1.2. *Problem classification*

FMS is a large-scale system and, as such, it cannot be addressed in its entirety without devising appropriate decision hierarchies. Different hierarchical structures proposed in the FMS literature are briefly presented below.

Suri and Whitney (1984) propose a decision hierarchy consisting of three decision levels:

First decision level:	strategic decisions (i.e. part families selection, system capacity);
Second decision level:	batching and resource allocation decisions;
Third decision level:	scheduling, dispatching, tool management, and system monitoring decisions.

Stecke (1984) classifies FMS problems into four categories:(1) design; (2) planning; (3) scheduling and (4) control. In the same paper the term 'system set-up problem' is introduced for a set of FMS short-term planning problems that need to be simultaneously addressed (part type selection, machine grouping, production ratios, resource allocation, loading). A similar decision hierarchy to the one proposed by Stecke is also suggested by Kiran and Tansel (1985b).

An interesting structure of hierarchical production planning for FMSs from a control theorist's perspective can be found in Gershwin *et al.* (1984 and 1986). Their hierarchy partitions the production planning problem into sub-problems with successively shorter time scales, where the solution of each sub-problem imposes constraints on lower-level sub-problems. Their proposed hierarchy is:

Long term:	investment and initial design decisions;
Medium term:	design and planning decisions;
Short term:	real time control.

For purposes of exposition of our results we adopted a problem classification scheme that has the same classification criterion as that of Gershwin *et al.* (1986) (i.e. time scale). We treat initial design issues as part of our more general category of design problems. The higher level of the used hierarchy in the paper addresses the design problems (strategic decisions for Suri and Whitney (1984), first level of Stecke's (1984) hierarchy), and the next level deals with planning problems (second level decisions for all previously discussed hierarchies). We are dealing mostly with short-term planning problems. The third decision level for most hierarchies deals with real time control and scheduling problems of FMSs, and we are not reviewing research work for that decision level. For a review of real time control and scheduling problems of FMSs refer to Kalkunte *et al.* (1986) and Rachamadugu and Stecke (1987).

Our proposed problem classification is outlined below:

FMS design problems

(1) Optimal system configuration (i.e. determination of number and types of machines, level of WIP in the system);
(2) Specification of the FMS layout;
(3) Selection of a storage system (size of local buffers and/or central storage);
(4) Specification of the type and capacity of the material-handling system (MHS);
(5) Determination of other important system resources (i.e. number of pallets, number and types of fixtures, number and types of tools).

FMS planning problems

(1) Part type selection problem;
(2) Machine grouping problem;
(3) Loading problem (i.e. allocation of operations and associated cutting tools of the selected part types among the machine groups);
(4) Routing mix problem (i.e. route selection for each part type);
(5) Other planning problems (pallet/fixture allocation among selected part types, process planning, spare tool allocation, tool storage, machine processing rate optimization).

2. FMS design problems

2.1. *Optimal system configuration*

A frequently encountered design problem for a flexible manufacturing system is the determination of the optimal system configuration in order to minimize the cost of operation, subject to the constraint of achieving a minimum required system throughput. The optimal system configuration problem usually addresses the following two questions: (1) How many identical servers (machines) are actually required at each workstation? (2) What should be the level of the work in process inventory?

As a first step to tackle the problem, a central server closed queueing network is usually used as a performance model of a FMS to evaluate and predict the stochastic behavior of workflow in the system. The most widely used performance model is the CAN–Q model developed by Solberg (1977). The second step involves the development of an optimization procedure, either implicit enumeration or of a heuristic nature, that exploits some of the monotonicity properties of the performance measures of a closed queueing network (CQN), to address the above problems.

Vinod and Solberg (1985) developed an optimization algorithm using CAN–Q to determine the optimal system configuration. The basic formulation of the problem is as follows

$$\min z = \sum_{i=1}^{M} k_i c_i + k_N N \qquad (1)$$

$$\text{s.t. } \mathrm{TH}_M \geq P_0$$

where:

z = total operating and capital investment cost of the FMS;
k_i = operating and capital investment cost of a machine in workstation i per unit time ($i = 1, \ldots, M$);
c_i = number of machines (servers) at workstation i ($i = 1, \ldots, M$);
k_N = inventory holding cost per job in the system;
N = job population in the system per unit time;
TM_M = actual system throughput, which depends on c_i, $i = 1, \ldots, M$;
P_0 = desired system throughput;
M = total number of workstations.

The above integer program is very difficult to solve since the system throughput cannot be explicitly represented in terms of the decision variables, N and c_i, $i = 1, \ldots, M$. To obviate the problem the authors propose an approach that exploits the functional properties of the throughput function of the closed network. The proposed implicit enumeration algorithm requires a good starting solution in order to be efficient. Efficiency is also improved by use of appropriate bounds on the system throughput, which reduces the number of evaluations of the throughput function through CAN–Q.

Dallery and Frein (1986 and 1988) address the same optimization problem with a more general objective function. Their only requirement of the objective function is that it is monotonically increasing in the decision variables. Again, the FMS is modelled as a CQN. The proposed solution procedure goes through the following three steps. First, bounds on the system throughput are used in order to determine a good initial configuration. The above bounds are based on asymptotic bound analysis (refer to Kleinrock, 1975). Second, the marginal allocation scheme of Fox (1966) is used as a heuristic procedure to improve upon the starting configuration provided by the first step. Third, an implicit enumeration algorithm with starting solution given by the marginal allocation scheme is used to find the optimal solution. The basic difference between this method and that of Vinod and Solberg (1985) is the more sophisticated way of determining the initial solution that is used to start the implicit enumeration algorithm. The method shows computational efficiency superior to that of Vinod and Solberg.

In Kouvelis and Lee (1990a), a very computationally efficient algorithm for the optimal system configuration problem is presented. The algorithm exploits the concavity properties of the throughput function, and a monotonicity property characterizing the optimal server vector for special cases of the cost vector.

Shanthikumar and Yao (1987) considered the same problem where the total job population, which could also be considered as the total buffer capacity of the system, is given. Their formulation of the optimization problem is as follows

$$\max \sum_{i=1}^{M} \mathrm{f}_i(\mathrm{TH}_i(c_i)) - \mathrm{g}_i(c_i)$$

$$\text{s.t.} \sum_{i=1}^{M} c_i \leq C_{\max}$$

where:

c_i = number of servers at station i;
$\mathrm{TH}_i(c_i)$ = throughput of station i, given that c_i servers have been allocated to it;
$\mathrm{f}_i(\mathrm{TH}_i(c_i))$ = profit at station i as a function of its throughput;
$\mathrm{g}_i(c_i)$ = cost at station i as a function of number of servers allocated;
$C_{\max}$ = maximum total number of servers to be allocated.

They made the following assumptions with respect to the functions: $\mathrm{f}_i(\mathrm{TH}_i)$ is increasing concave in TH_i and $\mathrm{g}_i(c_i)$ is convex in c_i.

The major result of their paper is that $\mathrm{TH}_i(c_i)$ is proved to be an increasing concave function in c_i. Using this result, the optimization problem can be solved by the marginal allocation procedure of Fox (1966). The difficulty in applying the previous formulation is related to the meaningful definition of profit and cost functions of each individual workstation in a complicated manufacturing environment such as a FMS.

In a recent paper, Dallery and Stecke (1990) discussed a more restrictive version of the optimal configuration problem. They consider the CQN, which models the FMS, divided into a set of sub-networks. Each sub-network is allocated a particular workload. This workload is fixed and given. The problem the authors consider is to determine the best configuration of each sub-network that yields the highest throughput for the over-all original CQN, where the number of stations, the number of servers, and the workload allocated to each station defines a possible configuration for each sub-network. Some properties of the optimal solution of the above configuration problem for single class multiserver CQNs are discussed by Dallery and Stecke (1990). The simultaneous solution of the minimum cost configuration and the workload allocation among the machine groups of a FMS is considered also in Lee *et al.* (1989). For the case that the FMS consists of machines of a single machine type the authors present an implicit enumeration procedure for the problem.

For FMSs, that could be adequately modelled as open queueing networks (OQN) (Buzacott and Shanthikumar, 1980 and Buzacott and Yao, 1986a), the optimal configuration problem could be also addressed along the lines of the Bitran and Tirupati (1989b) model. The authors consider open network of single server queues with multiple classes, and permit general distributions to describe the arrival and service times. The only restriction is that inter-arrival times within each class, and the service times for all classes, have independent and identical distributions. For measuring the performance of such systems a decomposition approach is used (Whitt (1983a and b)), with an appropriate modification introduced by Bitran and Tirupati (1988), that incorporates specific properties of the automated manufacturing systems (i.e. squared coefficient of variation of inter-arrival and service times less than one, product interference in multi-product manufacturing environments). For optimization purposes the decomposition approximation equations can be adequately described with a system of equations $\phi(\lambda, \mu, \mathrm{ca}, \mathrm{cs}) = 0$, where:

$\lambda = (\lambda_i)$ = arrival rate vector, $i = 1, \ldots, M$;
$\mu = (\mu_i)$ = service rate vector, $i = 1, \ldots, M$;
$\mathrm{ca} = (\mathrm{ca}_i)$ = vector of squared coefficients of variation of inter-arrival times, $i = 1, \ldots, M$;
$\mathrm{cs} = (\mathrm{cs}_i)$ = vector of squared coefficients of variation of service times, $i = 1, \ldots, M$.

The important performance measure for such systems is the mean number of jobs at each station, that can be represented as

$$N_i = \phi(\lambda_i, \mu_i, \mathrm{ca}_i, \mathrm{cs}_i), \; i = 1, \ldots, M$$

according to the decomposition principle. If v_i is the average WIP value associated with each job at station i, then the formulation of the optimal configuration problem is

$$\text{Min} \sum_{i=1}^{M} F_i(\mu_i)$$

$$\text{s.t. } \phi(\lambda, \mu, \text{ca}, \text{cs}) = 0$$

$$N_i = \phi(\lambda_i, \mu_i, \text{ca}_i, \text{cs}_i), \; i = 1, \ldots, M$$

$$\sum_{i=1}^{M} \nu_i N_i \leq W_0$$

$$\mu_i \geq \mu_i^0, \quad i = 1, \ldots, M$$

where: W_0 is an upper bound on WIP investment. In the objective function of the formulation, the cost functions, F_i, represent cost of capacity additions to achieve a desired WIP value, W_0, and are assumed monotonically increasing in the service rate μ_i at station i. The service rate at each station is bounded below by the quantity μ_i^0. In the above formulation the decision variables are the continuous variables μ_is. For single server networks assignment of capacity can be easily translated into allocation of service rates. Treating capacity as a continuous variable precludes certain types of capacity changes (i.e. addition of machines) and thus is not appropriate for the initial design phase of the manufacturing system, but could be helpful for long-term capacity expansion issues (i.e. labor reallocation, tool acquisitions or machine modifications) for an existing FMS. In the Bitran and Tirupati (1989a) paper the following two simplifying assumptions are made: (1) the functions F_i are convex, and (2) the ca and cs are independent of capacity changes. The second assumption is justified by observations in Bitran and Tirupati (1989b). As the number of products increases the inter-departure process of a queue becomes independent of the queue and the service process, and resembles the inter-arrival process (for a more rigorous justification refer to Whitt (1987)). As a result of the two simplifying assumptions, the previous formulation is a convex program. Bitran and Tirupati (1988) propose a greedy heuristic solution approach to the problem, instead of using conventional convex programming methods, in order to facilitate parametric analysis of the problem. The parametric analysis leads to interesting trade-off curves between capacity, WIP and manufacturing lead times. The resulting curves are accurate enough for all practical purposes, due to the close-to-optimality performance of the heuristic. Suggestions for appropriate use of these curves for strategic manufacturing decisions (i.e. technology choice, capital investment, product mix) are reported in Bitran *et al.* (1987).

An extension of the Bitran and Tirupati (1988) formulation is presented in Boxma *et al.* (1990) for a multiserver OQN with exponential inter-arrival and service time processes. Under the same simplifying assumptions (1) and (2), the authors prove that a marginal allocation procedure, that adds a server (i.e. considers discrete options) at the workstation where the quotient of the increase of the objective function to the decrease of WIP is the smallest, generates nondominated solutions in the sense that you cannot decrease the investment cost without increasing WIP. The deviation from optimality of the heuristic is reported as satisfactory. Boxma *et al.* (1991) extend their formulation and the corresponding results for general OQNs, with nodes modelled as $GI/G/c_i$ queues. Decomposition approximations are used for performance evaluation of the OQN. Bitran and Tirupati (1989a) address optimal system configuration problems with discrete options for general OQNs. An option for a workstation (modelled as a $GI/G/c_i$ queue) includes the type of machine to be used, which affects the machine service rates, coupled with the number of such machines. They present a heuristic solution based on a linear relaxation of the proposed integer programming formulation for the problem. The heuristic reportedly performs well.

2.2. *FMS lay-out problem*

The design of the physical lay-out of a FMS is of tremendous importance for the effective utilization of the system. Tompkins and White (1984) emphasized the importance of lay-out decisions for effective material handling by pointing out that 20 to 50% of the total operating expenses in manufacturing are attributed to material-handling and lay-out-related costs. The lay-out decisions deal with the arrangement of the workstations (usually referred as the machine lay-out problem in FMS, Heragu and Kusiak (1988)) and the way the workstations are connected through the transportation lines of the material-handling system (MHS) (usually referred to as the MHS lay-out problem, Kouvelis and Lee (1990b)). The objective function usually considered in the literature is that of material-handling cost minimization.

In general, the plant lay-out problem is formulated as a quadratic assignment problem (QAP) (Francis and White, 1974). Gilmore (1963) and Lawler (1963) developed optimal procedures to minimize the total material handling costs. These optimal procedures are efficient for small-sized problems, because of the computational complexity of the QAP. Sahni and Gonzalez (1976) showed that the QAP is NP-complete. This led researchers to concentrate on heuristic algorithm development for solving the problem (Armour and Buffa, 1963; Hillier, 1963; Hillier and Connors, 1966; Drezner, 1980 and Picone and Wilhelm, 1984).

Heragu and Kusiak (1988) raise questions regarding the applicability of the QAP formulation for certain FMS implementations. In FMSs, the machine sizes are generally not equal, and as the clearance between machines tends to be constant, the distance between locations depends on the sequence of the machines, that violates the assumption of the QAP formulation for candidate location distances being independent of machine sequence. For such cases, the appropriate formulation of the FMS lay-out design is as a quadratic set-covering problem (QSP). The QSP problem is discussed in Bazaraa (1975), and a branch and bound approach for its solution is presented.

The QAP and QSP formulations ignore interactions between the lay-out decisions and the queueing performance measures of a FMS. The significance of such interactions has been demonstrated in Solberg and Nof (1980), where the CAN–Q model is used to explore important factors affecting lay-out configuration decisions. Four different lay-out configurations are considered: product lay-out; cart line; conveyor loop and process lay-out. The computational results showed that flow control issues, including interplay of processing requirements, travel times, part mix and process selections can yield circumstances favoring any of the four lay-outs considered.

Kouvelis and Kiran (1990) present a model that incorporates the queueing-related aspects of lay-out decisions. In a FMS, parts are loaded onto pallet-fixtures for automated handling and processing, and are transported between the stations *via* an automated MHS. FMSs are usually designed to accommodate a limited number of pallet-fixtures (i.e. job population in the system). This limited amount of work-in-process (WIP) tightly couples the workstations; hence, the throughput time (i.e. the average time that a part spends in the system) becomes more sensitive to material-handling delays and transportation times. Since the job population is fixed, the throughput rates are directly related to throughput times. Therefore, the effects of lay-out decisions on throughput rates (*via* transportation time) must be captured in a formulation of the FMS lay-out problem. Using a CQN model for measuring the FMS throughput, with one node of the network (let's call it node M) representing the MHS, this can be done by observing the effects of the lay-out decisions on the throughput function of the CQN. It is known from closed queueing network theory (e.g. Gross and Harris (1985)) that the throughput function of such a network is a function of the service rate at each station. The lay-out decisions by specifying the relative location of workstations affect the transit times between them, and hence determine the service rate of node M that models the MHS. Hence, the throughput function is a function of the lay-out decisions. Kouvelis and Kiran (1990) propose a modification of the regular QAP formulation by adding a lower bound production requirement constraint, a way to involve the throughput effects of the lay-out decisions in the optimization framework, and by including an additional term in the objective function, that reflects WIP considerations. The resulting formulation, referred as MQAP, is solved by an appropriate modification of the Lawler–Gilmore procedure (Francis and White, 1974), and is efficient for realistic FMS size (5–10 workstations) problems.

Rosenblatt (1986) used a dynamic programming approach to address the dynamic aspects of the plant lay-out problem, similar to the approach of Sweeney and Tatham (1976) for addressing the dynamic nature of the warehouse location problem. The approach solves numerous static plant lay-out problems, which are basically quadratic assignment problems. In the paper, it is reported that the Sweeney and Tatham state space reduction of the dynamic program does not work efficiently for the dynamic lay-out problem. The application of a heuristic procedure for state space reduction is suggested in the paper and it seems to be computationally promising.

The lay-out designer of a FMS faces the difficult task of developing a system that is capable of handling a variety of products with variable demands, alternate and probabilistic routings, at a reasonable operating cost. In particular, during the lay-out design phase, the operational product mix is highly uncertain. Usually at the time the lay-out decisions are made, the products and the process plans have not been completely determined. Furthermore, the product mix (given as part production rates) is subject to changes due to forecasting error and demand fluctuations. Kouvelis and Kiran (1991) developed single and multiple period lay-out models, that incorporate the queueing and product mix uncertainty aspects of the FMS lay-out decisions. Single period models are applicable for FMS implementations with very high lay-out changeover costs and uncontrollable changeover times. The product mix is uncertain during the lay-out design phase, but once realized during the system operation is expected to remain stable over the planning horizon. The multiperiod models are applicable to the majority of FMS implementations. The planning period for the FMS design is divided into smaller operational planning periods. The product mix for each planning period, though uncertain during the design phase, remains stable over that period. Most FMSs consist of cells of physically identical machines that can perform a variety of operations. At the beginning of each planning period (week, month) operations allocation decisions are made. These decisions cause a significant change in the operational product mix of the cell in each period, and in most cases imply sub-optimality of a static lay-out configuration. Kouvelis and Kiran (1991) developed a dynamic model that allows the designer to specify different lay-outs, one for each planning period based on the product mix characteristics (or the demand forecasts) for each period.

Afentakis (1986), using the modelling framework developed in Chatterjee *et al.* (1984), developed a graph theoretic formulation of the static FMS layout problem. The following two assumptions are made: (1) the MHS is unidirectional, where bypassing of certain machines is permitted; (2) An operation i can be performed only on a particular machine. The author models the MHS using the notion of a lay-out graph L(M, T), where the set of nodes M denotes the set of workstations, and the set of arcs T the material handling system links. If the part mix and the routing problem had been solved, and operations had been assigned to workstations, then one could proceed to the definition of a part transition graph $G_i(M, E_i)$, where: E_i = set of arcs, with $(j, k) \in E_i$ if part i must go from workstation j to workstation k.

There is a weight associated with each link $(j, k) \in E_i$, which represents the number of parts moving along link (j,k). Let's call G(M, E) the graph obtained by superposition of the part transition graphs for all parts, after removing for each link with more than one arc all but one arc. Then, the graph theoretic formulation of the FMS lay-out problem is:

Find the lay-out graph, L, with the following properties:

(i) Graph L has the same nodes as G;
(ii) If $(i, j) \in$ G, then there is exactly one path from i to j in L.
(iii) The sum of the weights associated with links (i, j) is minimized.

Afentakis (1986) classifies the problem as NP-complete, and proposes heuristic algorithms, based on switch-and-check techniques, for its solution. Afentakis *et al.* (1986) report simulation results on the dynamic nature of the FMS lay-out problem, and examine the influence of a number of system parameters, particularly the product mix, on several relay-out strategies.

Recently, some FMS researchers devoted their attention to analyzing, in detail, specific lay-out types that are implemented in FMSs. As reported in Heragu and Kusiak (1988) and Afentakis (1989) these are (for detailed description refer to the original references):

(1) Unidirectional loop network;
(2) Circular machine layout;
(3) Linear single-row machine layout;
(4) Linear double-row machine layout;
(5) Cluster machine layout.

Analysis of existing FMSs shows that the machine lay-out is determined by the type of material-handling devices used. In FMSs unidirectional loop network lay-outs are extensively implemented due to the wide use of efficient unicyclic material-handling networks. Such networks connect all workstations by a path passing through each workstation exactly once. A unicyclic material-handling network may represent a loop conveyor, tow line, overhead monorail system or wire paths of a unidirectional automated guided vehicle (AGV) system. Afentakis (1989) points out that such lay-outs are preferred to other configurations due to their relatively lower initial investment cost, since they contain the minimum number of required material-handling links to connect all workstations, and higher material-handling flexibility. Part of the flexibility aspects of the above configuration is its ability to satisfy all material-handling requirements for the part types scheduled for manufacturing in the system, and easily accommodate any future introduction of new part types or process changes. The reason for this is that the material-handling requirements of any part type processed at a workstation can be accommodated, as there is at least one directed path connecting any pair of workstations. Afentakis (1989) presents a mathematical programming formulation of the lay-out problem for such configurations under the objective of minimizing the average number of machines that parts cross per unit time, which is an indirect way to reduce congestion in the system by minimizing material-handling delays. He presents a sophisticated block interchange heuristic for its solution. Kouvelis and Kim (1991), using an appropriate workstation interchange argument, present dominance relationships for easily identifying local optimal solutions for the unidirectional loop network lay-out problem. Simple constructive heuristics, an optimal branch and bound procedure, and an appropriate decomposition principle for large size problems are also discussed in the paper. Kiran and Karabati (1990) address the same problem under a different objective, that of minimizing the total material-handling cost. An exact implicit enumeration procedure, enhanced with dominance relationships for identifying local optimal solutions, is developed.

Another lay-out type extensively implemented in flexible manufacturing environments is the linear single-row machine lay-out (LSRML). The machines are arranged along a straight track with a material-handling device moving jobs from one machine to another. In such environments, jobs enter the production line at one end of the track, which is usually close to the raw material storage point, and leave the line at the other end of it to enter a WIP or a finished goods storage area. Kouvelis and Chiang (1990) address the LSRML problem having as design objective the minimization of the total back-tracking distance of the material-handling device (i.e. the travel in reverse direction to the part flow movement). They proved that the problem is NP-complete. Special cases of the problem that are polynomially solvable are presented. The authors use simple local optimality conditions to develop effective heuristics for the problem. Kouvelis and Chiang (1991) discuss the use of a simulated annealing (SA) heuristic for the LSRML problem. Their extensive computational experimentation indicates that the SA heuristic, with its control parameters fine tuned to the best level for the specific application, provides, in reasonable computational time, near-optimal solutions. The heuristics of Kouvelis and Chiang (1990) are used to input the initial lay-out to the SA algorithm. For a more detailed discussion on algorithmic approaches to specially structured FMS lay-out problems, see Kouvelis and Kiran (1989).

The lay-out types (2)–(5) are usually implemented with the use of handling robots, AGVs, and gantry robots as material-handling devices. Heragu and Kusiak (1988) propose two heuristic procedures that exploit the special structure of the lay-out decisions for the above lay-out types. Mathematical programming models for analysis and design of lay-outs for robotic systems are presented in Sarin and Wilhelm (1984). Kiran and Tansel (1988) consider the storage location for WIP in flexible manufacturing cells with the material handling performed by robots. The operation of such cells is as follows: the robot picks up a part and places it on the machine. After completion of the operation the part may be taken to another machine for its next operation. If the next machine is currently occupied, the part is taken to the WIP storage area, where it waits till the required station becomes available. Kiran and Tansel (1988a) investigate the problem under discrete and continuous space assumptions. In the discrete case, the problem of WIP storage location for material-handling cost minimization is transformed to a generalized assignment problem. The continuous lay-out is discussed under different distance measures.

The MHS lay-out problem includes location of pick-up points and delivery stations, specification of track lay-outs and choice of unidirectional or bidirectional flows (flow path design). It is of extreme interest for FMS implementations with AGVs as the material-handling devices. Gaskins and Tanchoco (1987) propose an integer programming formulation for determining the optimal flow path that minimizes the total travel distance of loaded AGVs. It is assumed that the MHS network is given, and the only decision is to choose the direction of flow on each material-handling link, assuming that unidirectional flows are preferable. The formulation becomes inefficient for large-size MHS networks. Sharp and Liu (1987) proposed an analytical method for configuring the network of a fixed path, closed loop MHS. The purpose of the model is to make good initial decisions with respect to shortcuts, off-line spur construction and spur length. Kiran and Tansel (1988b) consider the optimal location of a pick-up point on a material-handling network. The pick-up point may connect the material-handling network to any one of the following: machining station, load/unload station, central or local storage. It could even serve as just a transfer point to/from another material-handling network. The authors extend Hakimi's result for node optimality of median location for directed networks. So, the pick-up point, if possible, should be located at a node of the material-handling network (i.e. a junction point or a workstation). For real FMS implementations locating the pick-up point at existing nodes may not be possible. Cases in which the pick-up point must be located on an arc of the MHN are analyzed, and it is demonstrated that the optimum location can be found in polynomial time. Kouvelis and Lee (1990b) present a different formulation of the MHS lay-out problem. A discrete directed network design problem on a plane formulation is given, and standard results from the graph theoretic literature as they apply to FMS implementations are discussed.

2.3. *Optimal storage capacities in a FMS*

As mentioned earlier, the most widely used model to evaluate the performance of a FMS is the CAN–Q model by Solberg (1977). The model makes an infinite buffer assumption, i.e. each workstation contains adequate buffer capacity to accommodate all parts that are queued for processing there. However, in most FMS implementations each machine has a very small buffer, usually of 2 parts per machine. Nevertheless, Suri (1983) showed that this assumption is quite robust for FMS performance evaluation. Also, Yao and Buzacott (1987) showed that, for FMS implementations with centralized MHSs (i.e. a loop conveyor) and zero buffer workstations, the product-form solution of the CQN model for the FMS performance remains valid for any service time distribution at the workstations. These robust results provide insights into the good performance obtained when Jackson CQNs are applied to FMSs with small local buffers. Simulation results demonstrating the above point (CAN–Q estimated the throughput of an actual manufacturing system with an error of less than 10%) are reported in Haider *et al.* (1986).

Performance modelling of FMSs with limited storage capacities has been extensively researched by Buzacott, Shanthikumar and Yao. Buzacott and Shanthikumar (1980) consider variations of a multiclass open queueing network model of a FMS with local storage only, with central storage only, and with some combination of the two. They point out that common storage is superior to local storage, and that for a system with M workstations ($M>2$), production capacity is conjectured to increase in the job population, N, where $B+2<N<B+M$, with B = central storage capacity. In current FMS implementations, central storage is augmented with local storage with one or two workpieces at each station. The supporting argument coming from practical experience is that such a scheme reduces machine idle time while maintaining efficient use of storage facilities.

Yao and Buzacott (1986) modelled a FMS as an open network of multiserver queues with general service times and limited storage capacity both at the system level and locally at each workstation. A rather interesting view of the MHS is expressed in the paper as they see it divided into two sections: (1) one section, the so-called MHS(I), handles the input flow to the workstations; (2) the other section, the so-called MHS(O), handles the output flow from the workstation.

The reason for that modelling peculiarity is their attempt to justify the assumption that machines are never blocked from finished jobs, since finished jobs are handled by the MHS(O), which is assumed to have the capacity to provide such a service level since it is basically a dedicated conveyor for finished jobs. Such a modelling approach does not agree with the physical implementation of a general FMS, but it is not restrictive, on the other hand, as one can ignore the MHS(O) and just keep the implied assumption of non-blocked machines from finished jobs. In the model the probability that input flow into a station is blocked due to limited local storage is explicitly calculated. The case of feed-back flow where part of the output flow from the workstation will be fed back to the central storage is also considered. In order to derive results for such a general model, a renewal approximation technique was used.

Yao and Buzacott (1986 and 1987) address the modelling of limited local buffers within the context of a CQN. In order to maintain the product form solution for such a network model, properties stronger than the Jackson networks are needed. The reversibility of the mean queue length process (Kelly, 1979), which is equivalent to the reversibillity of the Markov chain of the job routing, is required. Routings that can be shown to be reversible and of interest to FMS modelling are:

(1) Symmetric routing from station to station;
(2) Job routing in a central server network (CAN–Q situation);
(3) Probabilistic shortest queue routing, by which jobs are routed with the highest probability to the station which has the relatively shortest queue (more empty spaces in its buffer).

For those cases the calculation of the steady-state probabilities of the CQN is relatively easy and can be obtained from the product form solution for the Jackson CQN, with an appropriate adjustment of the normalizing constant, since the new state space results from truncation of the Jackson CQN model state space. In the paper, results are presented for a CQN with single server, exponential service times at stations with state-dependent processing rates, and under different operational strategies. An explanation of the need for specifying an operational strategy in the case of limited buffers follows.

One of the major attractive features of a closed queueing network model of a FMS like CAN–Q, which assumes infinite local buffers, is the ease of input specification. The basic inputs to the CAN–Q model are the visit frequencies of parts to the workstations (i.e. the number of visits of a part to the workstation per visit to the load/unload station) $e_i (i = 1, \ldots, M)$. Those visit frequencies can be easily determined once the product mix, the processed plan for the process parts and the operation assignment decisions are made. Actually, $e_i = \mathrm{TH}_i/\mathrm{TH}_M$ for the CAN–Q model, where TH_i = throughput of workstation i. But when we have a model with limited local buffers then $e_i \neq \mathrm{TH}_i/TH_M$, since it is possible that jobs attempting to enter the workstations are blocked. Now the e_i can be interpreted as the average proportion of operations in the central storage that need to be delivered to station i. So for such a model, a way for specifying the e_i from available input data is needed. That is achieved by specifying an operational strategy. The operational strategies discussed in the paper are:

(1) Fixed routing model: jobs are released into the system in such a way that e_i follow a set of fixed values;
(2) Fixed loading model: a set of values for $\mathrm{TH}_i/\mathrm{TH}_M$ are given;
(3) Dynamic routing model: the probabilistic shortest queue (PSQ) mechanism is used.

Yao and Buzacott (1986) generalize the results for a PSQ routing scheme for a multiclass, multiserver CQN with exponential service rates dependent on the part type processed at a particular workstation. Although in all of the above models the routing conditions are quite restrictive, still they are quite suitable for a FMS. The problem with them is the difficulty in specifying an appropriate operational strategy for a FMS.

A modelling idea used by Yao (1986) to develop an optimal central storage model for a FMS, is the controlled arrival single-stage queue model (CASQ). The model can be easily transformed to a CQN with the addition of another station indexed as station 0 with service rate $\mu_0(n_0) = \lambda(N - n_0)$, where

λ = arrival rate to the system
n_0 = number of jobs in the system
N = upper limit on the number of jobs in the system.

In an attempt to add more realistic control features to it, externally blocked jobs are not lost but queued outside of the FMS at the central common storage area. It has been proven by Shanthikumar and Sargent (1981) that the total number of jobs in a restricted open queueing network (OQN) has the same equilibrium distribution as the number of jobs in a birth and death queue with a state dependent birth rate

$$\lambda(n) = \lambda \quad n \leq N$$
$$\lambda(n) = 0 \quad n > N$$

and a state-dependent death rate equal to the throughput of a CQN, TH(n). Based on that fact, the CASQ model was developed by Shanthikumar (1979), where the system is transformed into a single-stage queue. Yao (1986) uses CASQ, with the FMS being the single-stage queue, to develop an inventory model for the system. The FMS is assumed to be equipped with N pallets and a central common storage area. Parts to be processed are delivered in batches from a central warehouse to the storage and then inserted into the FMS whenever a pallet is available. Since the production rate of FMS depends on the number of parts circulating in it, the following costs have to be considered: fixed ordering (material handling and transportation); inventory holding cost (associated with parts in storage) and costs associated with loss of production if the number of parts within the FMS drops below N. The model computes the optimal order point and optimal batch size under the above considerations.

2.4. *Material-handling system design*

Even though material handling is a critical factor in FMSs performance, with severe consequences due to its inefficient design (i.e. disorganized storage activities, excessive manual effort, idle machines, materials piled up on the floor, poor inventory control), little basic research has been devoted to issues related to the design of integrated material-handling systems in a FMS context. This section does not represent an exhaustive survey of operations research approaches to the MHS design problems, but intends to give a brief review of important work in the area, with particular emphasis on integrated MHS design applicable to FMS environments.

A brief description of some of the generic research areas of operations research that relate to material-handling is given by Maxwell (1981). A more detailed review of vehicle routing and scheduling problems related to MHS design is the paper by Kusiak (1985). A survey of operations research approaches to ten categories of material-handling issues (e.g. robotics, conveyor theory, transfer lines, FMS, equipment selection, storage alternatives, automated storage and retrieval systems (AS/RS), warehouse lay-out, palletizing, order picking and accumulation) is given in Matson and White (1982). Most of the research dealing with design issues of MHSs concentrates on specific systems, with conveyor systems receiving most of the attention (Muth and White, 1979). Lately, researchers concentrated on the design of automatic guided vehicle systems (AGVS), an efficient material-handling alternative for FMS implementations. An interesting model addressing the problem of determining the optimal number of vehicles in an AGVS was developed by Maxwell and Muckstadt (1982). An ingenious formulation of the problem into a transportation problem, with the objective to minimize the number of empty trips between workstations subject to flow constraints at each station, is presented.

For the MHS design engineer during the initial design phase of a FMS, the problem faced is that of selecting from a multitude of possible combinations of standard elements that can perform the required material-handling functions (i.e. robots, automatic guided vehicle, conveyors, AS/RS) a set that will do the job most effectively and economically. A system model for addressing the integrated MHS design of a FMS was developed by Kouvelis and Lee (1990b). The model is a discrete choice network design model for a FMS, that is developed within the framework of multicommodity network flows.

2.5. *Other design problems in FMS*

Other important resources of a FMS are pallets (devices holding the part during its transportation) and fixtures (devices holding the part during its machining operations). The determination of the number of available pallets and fixtures in the system is significant for an efficient operation of a FMS. It is well known that the modelling of a production system as a CQN, which implies that the number of jobs in the system is constant, is justified by using the following interpretation in a FMS context: the number of customers circulating in the network at any point in time is the same as the total number of fixtures available in the system. Consequently, when the total job population is determined in the system configuration problem, the total number of fixtures, with the pallets treated as special type of fixtures, is also determined. But still there are other design questions to be considered, and in particular, the selection between special-purpose (designed to carry a specific sub-set of the parts on a FMS) and general-purpose (almost all parts manufactured in the system could be mounted on) fixtures. Relevant trade-offs for such a decision (ease of scheduling, accommodation of product mix changes, fixture investment, WIP level) are discussed in Newman (1986).

It is surprising that the problem of determining the number of tools of the different tool types in a FMS has been consistently overlooked. In a FMS with versatile NC machines, each manufacturing operation can be performed by one or more cutting tools; and often those tools could be stored in the tool magazines of different machine types. The formulation of an optimization model, which will have the general constraint structure of the loading problem (see Section 3.3), is a suggested approach to the problem. Kouvelis (1991) presents an optimal tool selection procedure resulting from such a formulation. An extensive discussion of open research questions on tool management issues (including design and planning problems) for FMSs can be found in Gray *et al.* (1988 and 1989).

3. Planning problems in FMS

3.1. *Part type selection problem*

One of the fundamental decisions the FMS manager has to make is to determine the set of part types to be produced during the next short-term planning horizon. Although orders might exist for a large set of part types at the beginning of the planning horizon, it is important to specify a feasible sub-set of those part types to be produced over the planning horizon (i.e. allocation of tools and operations does not violate tool magazine capacity and time availability of the machines) that optimizes a specified system performance criterion (i.e. due dates, system utilization).

Whitney and Gaul (1985) suggest a sequential decision procedure for the part type selection problem, that partitions the part types into distinct and separate batches to be machined one at a time. The proposed method is of a heuristic nature, and considers various system performance measures and constraints such as machine utilization, due dates, degree of tool sharing, and the tool magazine capacity. Hwang (1986) presents an integer programming formulation of the part type selection problem, with the objective of minimizing the number of tool changeovers under a tool magazine capacity constraint. An optimization formulation with a different objective, that of minimizing makespan, is presented by Rajagopalan (1986). Heuristic procedures are suggested for its solution. Afentakis *et al.* (1989) present a new formulation of the part type selection problem for a FMS configured into a unidirectional loop network lay-out and operating with the use of cyclic scheduling policies. Heuristics are proposed for the solution of the problem.

The characteristic of the above part type selection approaches is that part types are partitioned into distinct batches, and each batch of part types is machined one at a time. In a particular batch, all part types are produced until all requirements are finished. A different approach to part type selection is presented in a series of papers (Stecke and Kim, 1986, 1988 and 1989 and Stecke, 1988b). After selecting a set of part types to be produced next, the authors' proposed approach determines the part mix ratios to be produced based on production requirements, processing times and system operational objectives (i.e. workload balancing, utilization, due dates). The part mix ratios are the relative number of parts of each part type that will be produced over a planning horizon. FMS production begins and continues according to the previously determined part mix ratios until the production requirements of a part type are completed. Then, either the reduced set of part types with their remaining production requirements are produced in new part mix ratios satisfying a specific operational objective or new part types are selected to supplement the others. Stecke and Kim (1988b) compared the different part type selection approaches. Their study indicates that their approach (termed as flexible approach) provides better machine utilization and minimizes makespan. The usefulness of the flexible approach, that is based on appropriate part mix ratio determination by solving an integer programming formulation with an appropriate operational objective, is also demonstrated in extensive simulation studies reported in Schriber and Stecke (1988).

A longer-term planning problem faced by FMS engineers is the identification of families of part types that could be produced by the FMS. This problem has been addressed to a large extent in the Group Technology literature. Group Technology (GT) is generally considered to be a manufacturing philosophy or concept that identifies and exploits the similarity of parts and operation processes in design and manufacturing. The GT principles, though applicable in general to most manufacturing systems, can be used as a vehicle for integrating the various elements of a computer-integrated manufacturing system like a FMS. One of the objectives of identifying part families is essentially to benefit from scale economies, that result from part processing similarities throughout the manufacturing cycle. This is of significant value in a medium volume production situation, as is the one faced in a FMS, where the economies of mass production are absent. The Group Technology literature (Kumar *et al.*, 1986; Vaneli and Kumar, 1986 and Wemmerlöv and Hyer, 1986) basically addresses the grouping of parts based on similarity of processing and/or design characteristics. Specific application of group technology techniques to the FMS part type selection were suggested by Kusiak (1984). To identify part families, clustering methods are proposed based on the use of appropriate coding systems to describe the characteristics of the part as related to its geometrical shape, type and sequence of operations required. The clustering algorithms require a measure of similarity between part type codes that, in most cases, is evaluated by some sort of distance measure defined on the part type coding scheme (i.e. average of differences between attribute values of two part type codes).

3.2. *Machine grouping problem*

Stecke, working within the modelling framework of CAN–Q, addressed certain issues relating to two of the FMS planning sub-problems, the machine grouping and the loading problems (Stecke and Solberg, 1985 and Stecke and Morin, 1985). The ideas and the results behind the machine grouping problem were not a surprise for the queueing theorists. It is well known that under stochastic conditions a pooled group of servers will perform more efficiently than the same number of servers working separately. The following interpretation of the pooling concept in a FMS context is useful in understanding the results that Stecke and Solberg (1985) obtained.

In examining the physical description of a FMS, one realizes that we could have machines able to perform exactly the same operations but not being of the same machine type. Machines are of the same machine type if they are physically identical, i.e. if they have the same axes of motion, dimension, horsepower, and capabilities. In queueing theory, the notion of pooling servers into service stations, particularly when we are dealing with product form queueing networks, requires the existence of identical servers in the same service station. The notion of an identical server in a FMS context is a machine of the same machine type and tooled identically. All of the above discussion is meant to show that our pooling alternatives in a FMS have certain limitations. Even if all the machines were of the same machine type, still maximum pooling of all of them into one machine group might not be feasible since it would require loading the tool magazine of each machine with all the tools required to perform all operations necessary for all parts that are going to be processed by the FMS. If we have more than one machine type then we should have at least as many different workstations as machine types, since limitations of the tool magazine capacity might force us to have more workstations. The FMS designer is then able to predict most of the time the minimum number of machine groups required.

Stecke and Solberg (1985) pursued an unconstrained optimization of the expected production rate of the manufacturing system as a function of the system configuration (machine grouping). Their results can be summarized as follows: (1) fewer groups are better, i.e. pool as much as possible; (2) if technological constraints define that g groups of m machines are required, then the more unbalanced configuration provides the larger expected production rate.

The above results were proved for small problems, because the expected production rate (throughput of the closed queueing network) as a function of the system configuration for a large number of parts in the system does not exhibit any nice concavity properties; and consequently the problem becomes cumbersome to solve analytically. A computer implementation of CAN–Q is used in Stecke and Solberg (1985) to evaluate the expected production rate for different system configurations for a large number of parts in the system and to justify the above conjectured results. The results of the machine grouping problem, though not proved rigorously, are widely accepted because of their intuitive interpretation.

But the machine grouping resulting from such an unconstrained optimization as the one applied above might not be even feasible for a real system, as important constraints such as the capacity of the tool magazine of each machine has been ignored. Many different approaches in formulating such a constraint have appeared in the literature. The formulation in Stecke (1986) has attracted the most attention and it is the one we discuss next.

Before formulating the tool magazine capacity constraint, the assignment of operations to machines has to be done. The operations assignment is formulated as:

$$1 \leq \sum_{l=1}^{M} x_{il} \leq q_i, \quad i = 1, \ldots, b \text{ operation index} \quad l = 1, \ldots, M \text{ workstation index}$$

x_{il} = 1, if operation i is assigned to each machine in group l; x_{il} = 0, otherwise

q_i = maximum number of times that operation i can be assigned (routing flexibility indicator).

The above formulation states the simple fact that each operation must be assigned to at least one machine of the machine type required by the operation. Of course, x_{il} should be forced to 0 if operation i cannot be performed by the machine type corresponding to machine group l.

A simple formulation of the tool magazine capacity constraint is the following:

$$\sum_{i=1}^{b} d_i x_{il} \leq t_l, \quad l = 1, \ldots, M$$

where:

d_i = number of tool slots required in a tool magazine by operation i,
t_l = capacity of the tool magazine for each machine in group l.

Stecke mentioned some other considerations such as: several operations may require some of the same tools and space in the tool magazine can be saved by eliminating tool duplication and considering overlap and weight balancing. If the first of the above considerations is included in the formulation, the resulting constraint is:

$$\sum_{i=1}^{b} d_i x_{il} + \sum_{p=2}^{b} (-1)^{p+1} \sum_{\substack{\forall \bar{B}/\bar{B} \subseteq B_k \\ \ni |\bar{B}|=p}} w_{\bar{B}} \prod_{i_k \in \bar{B}} x_{i_k l} \leq t_l, \quad l = 1, \ldots, M$$

where:

B_k = indexed set of sets of operations
$\bar{B}$ = index subset of B_k, such that $|\bar{B}| = p, p = 2, \ldots, b$
W_{B_k} = number of slots saved when the operations in B_k are assigned to the same machine.

The resulting formulation of the tool magazine capacity constraint is highly non-linear. But the non-linearity is not inherent in the problem. It is formulation specific. It can be easily avoided by selection of two sets of variables. One of them will represent the assignment of operations to machines and the other will assign tools to machines. Then other constraints of consistency between tool and operation assignment should be added. For more details on such a modified formulation, refer to Kouvelis and Lee (1991).

Stecke (1986) divided the machine grouping problem into two steps:

STEP 1: Find the minimum number of machines required to produce the set of part types considered for production.

That is accomplished by solving the following nonlinear integer programming problem:

$$\max \sum_{j=1}^{m} \gamma^j sl_j$$

$$sl_j = t_j - \left[\sum_{i=1}^{b} d_i x_{ij} + \sum_{p=2}^{b} (-1)^{p+1} \sum_{\substack{\forall \bar{B} \subseteq B_k \\ \ni |\bar{B}|=p}} w_{\bar{B}} \prod_{l_k \in \bar{B}} x_{ikj} \right]$$

$$\sum_{l=1}^{m} x_{il} = 1$$

where m is a decision variable along with x_{il}; and γ could be any number greater than 1 (preferably large). Of course the above problem is not easy to solve. As was mentioned before, the FMS engineer could use simple and effective reasoning to get a very good estimate of the minimum number of machines required. The above formulation, though sophisticated, still does not cover many aspects of a realistic problem. In particular, time availability constraints of the different machines over the planning horizon are ignored.

STEP 2: Use the optimal pooling results for the minimum number of workstations (determined in Step 1) as they are presented in Stecke and Solberg (1985).

A different way to look at the grouping problem as a long-term planning issue is the following: assume you have already ordered from a FMS vendor (or are currently existing in the system) a given number of physically identical machines. The grouping problem is that of allocating these machines to the different workstations, which subsequently determines the tools to be loaded on them, in order to optimize a performance measure of the manufacturing system. Since machines allocated to the same workstation in a FMS environment are tooled identically, we can treat them as identical servers in a queueing network modelling framework. Using the above interpretation of the grouping problem, we can apply the Shanthikumar and Yao (1988) model for its solution. For throughput modelling of the FMS a CQN can be used. Let TH(c) denote the FMS throughput as a function of the server vector $c = (c_i)$ (i.e. workstation configuration). Then the machine grouping formulation is

$$\text{Max TH}(c)$$

$$\text{s.t.} \sum_{i=1}^{M} c_i = C_0$$

$$c_i \geq 1, \quad i = 1, \ldots, M$$

where: C_0 = given number of physically identical machines.

The optimal solution to the above formulation is characterized by an intuitive monotonicity property, that roughly says that a workstation with a higher workload should be given more servers. Shanthikumar and Yao (1988) presented an efficient algorithm that solves the previous formulation, by appropriately using the characteristic property of the optimal solution to eliminate suboptimal allocations. It was also pointed out that a marginal allocation procedure generates the optimal solution for a two-workstation system. The authors conjecture its optimality for the general case.

In a CQN modelling framework, the machine grouping problem was addressed in Bitran and Tirupati (1989b). Since the authors treat capacity as a continuous variable, they address the problem of redistribution of homogeneous capacity (i.e. tools) among the workstations to minimize WIP. Using the decomposition approximation for performance modelling of a general single-server OQN (refer to Section 2.1), the formulation of the problem is

$$\text{Min} \sum_{i=1}^{M} \nu_i N_i$$

$$\text{s.t. } \phi(\lambda, \mu, \text{ca}, \text{cs}) = 0$$

$$N_i = \phi(\lambda_i, \mu_i, \text{ca}_i, \text{cs}_i), \quad i = 1, \ldots, M$$

$$\sum_{i=1}^{M} \mu_i = \mu_0$$

$$\mu_i \geq \lambda_i, \quad i = 1, \ldots, M$$

where μ_0 = total available capacity.

Under the simplifying assumption that ca and cs are invariant with capacity (refer to Section 2.1), the system of equations $\phi(\lambda, \mu, \text{ca}, \text{cs}) = 0$ can be solved initially (i.e. ca and cs can be treated as known parameters) and then $N_i = \phi(\mu_i)$, $i = 1, \ldots, M$. As under the assumed conditions, N_i is a convex function of μ_i, the reduced formulation is

$$\text{Min} \sum_{i=1}^{M} \nu_i \phi(\mu_i)$$

$$\text{s.t.} \sum_{i=1}^{M} \mu_i = \mu_0$$

$$\mu_i \geq \lambda_i, \quad i = 1, \ldots, M$$

and it fits the marginal allocation framework of Fox (1966). Consequently, a greedy heuristic will provide an optimal solution at the limit (as the allocated capacity increments at each step of the algorithm go to zero).

An extension of the Bitran and Tirupati (1989b) formulation is presented in the papers Boxma *et al.* (1990 and 1991). They treat capacity in a discrete manner (i.e. machine allocations), and use multiserver OQNs for performance modelling purposes. The resulting integer programming formulation is within the marginal allocation framework of Fox (1966), and can be easily solved to optimality with a greedy heuristic.

3.3. *Loading problem*

The results presented by Stecke and Solberg (1985) for the loading problem were surprising to industrial engineers. The transfer and assembly lines were the prevailing systems studied by industrial engineers for mass production. For those systems a philosophy of balancing the workload among service stations was widespread. Applying an unconstrained optimization of the throughput rate of a CQN with respect to the workload vector, Stecke and Solberg (1985) and Stecke and Morin (1985), in closely related papers obtained the results summarized below

(1) If all machine groups contain the same number of machines then a balanced workload per machine is optimal;

(2) For a system of unequally sized machine groups (a preferable system configuration), unbalancing the assigned workload per machine is optimal.

In addition, if a perfect unbalance (the theoretically optimum workload obtained in Stecke and Solberg (1985)) is not possible, then it is better to overload the larger group.

Stecke and Solberg (1985) proved rigorously the first of the above-stated results. The second result was conjectured and justified in many numerical examples where they used CAN-Q to evaluate the expected production rate of the manufacturing system. Stecke and Morin (1985) further examined the case of system configurations with equal number of machines in each machine group. They established rigorously the balanced allocation of workload for balanced system configurations (i.e. equal number of machines at all workstations) in order to maximize expected production rate, and also discussed certain properties of the expected production rate function (throughput of the CQN).

An elegant mathematical characterization of the loading problem results, and also many other interesting properties of the throughput function of a CQN are provided (Yao, 1985 and Yao and Kim, 1987). The major result presented is that the throughput function of a closed queueing network is a Schur concave function of ρ, where $\rho = (\rho_i)_{i=1}^m$ is the loading vector. This result was proven for single-server stations (Yao, 1985) and extended for multiserver stations with equal number of servers (Yao and Kim, 1987). Based on that, and using the fundamental property of Schur concavity,

$$\rho^1 \leq_m \rho^2 \Rightarrow \mathrm{TH}(\rho^1) \geq \mathrm{TH}(\rho^2)$$

(where ρ^1, ρ^2 are two loading vectors and the symbol $\leq_m$ denotes majorization ordering) one can easily see that the optimal allocation of a fixed amount of workload among the different workstations – always talking about the two special cases of CQNs, single-server stations or multiserver with equal servers – is achieved by a balanced workload, since the balanced loading vector majorizes all other loading vectors that have the same total workload.

Stecke (1983) addressing the loading problem as a constrained optimization problem identifies the following six candidate loading objectives:

(1) Balance the assigned machine processing times;

(2) Minimize the number of movements from machine to machine;

(3) Balance the workload per machine for a sub-system of groups of pooled machines of equal sizes;

(4) Unbalance the workload per machine for a system of groups of pooled machines of unequal sizes;

(5) Fill the tool magazine to as high a density as possible;

(6) Maximize the sum of operation priorities.

In Stecke (1983) the loading sub-problem was formulated as a mathematical program with the objective being one of the previously stated and the tool magazine capacity constraint. In the case that the loading problem is solved under the objective of balanced workload, Berrada and Stecke (1986) presented a branch and bound algorithm for solving the resulting non-linear mixed integer programming problem (MIP).

Kusiak (1983) formulated the FMS loading problem as a linear integer program, where the objective function is to minimize the total processing costs of the various operations. Kusiak was able to avoid the nonlinear tool magazine capacity constraint by ignoring the possibility that several operations may be performed by the same tool.

Chakravarty and Schtub (1984) approached the loading problem as a 0–1 mixed linear integer program by assuming that a single tool can handle all the machining operations of a particular part type. The efficiency of a tool depends on the machine to which the tool is allocated, and the objective is to minimize the total processing times. Co (1984) presented alternative 0–1 integer linear programming formulations for different objective functions, such as maximizing machine flexibility, minimizing number of consecutive operations in the same machine, etc. However, no solution procedure was given. Ammons *et al.* (1985) and Shanker and Tzen (1985) give formulations with bicriterion objectives. The objectives of the former are balancing workloads and minimizing workstation visits or crossings, while those of the latter are balancing workloads among machining centers and meeting the due dates of jobs. Greene and Sadowski (1986) use several objective functions such as minimizing makespan, minimizing mean flow time, and minimizing mean lateness in their formulations for the FMS loading and scheduling problems. Since the numbers of variables and constraints are very large, computational experiments were not performed. Finally, Sarin and Chen (1987) presented a very general 0–1 integer programming formulation of the problem. The problem is difficult to solve, as the number of integer variables can easily explode for even small size problems. A more efficient branch and bound algorithm for a similar formulation to the loading problem is presented in Kim and Yano (1991b). The loading problem has also been modelled as part of a general mathematical program for the FMS set-up problem in Kiran and Tansel (1985a), and was shown to be NP-complete by Kiran (1986).

It can thus be seen that the difficulties of the FMS loading problem, under the alternative formulations in the literature, lie in the nonlinear as well as the integer constraint of the problem. Kouvelis and Lee (1991) present a formulation of the FMS loading problem that is much more tractable to solve. By appropriately defining the operations and tool types of the system, the nonlinearity of the tool magazine capacity constraint can be avoided. Also additional constraints on the time availability of the machines for processing the parts are included. Such constraints are important because most FMSs are still subject to breakdowns, scheduled maintenances, and other disruptions, so that there may exist differential but finite time availabilities for different machines at the FMS. Moreover, the formulation exhibits a block angular structure, the exploitation of which gives rise to a very efficient branch and bound solution algorithm.

Heuristic algorithms are a viable alternative for the FMS loading problem. The major reason is that for large size problems the combinatorial complexity of the problem might prohibit its solution within reasonable time. In many cases, the FMS production manager is satisfied with a feasible operation and tool allocation solution, a necessary input for the lower decision level, i.e. the scheduling problem. The results of the scheduling problem directly define the quality of the solution of the short-term production planning problem. Because of the interdependence of the loading and scheduling sub-problems, optimality of the loading sub-problem loses significance as in many cases it might fail to provide feasible conditions for the scheduling problems (for further discussion see Hwang (1986)). In addition, the production environment in actual FMSs undergoes frequent changes (production order change, machines breakdowns, etc.) that cause the need for repeated solution of the loading problem. That favors the development of fast heuristics.

In Stecke and Talbot (1985) heuristic algorithms were developed for FMS, assuming that the grouping problem has already been solved. An assumption under which these heuristics work is that each operation can be performed by only one machine type. The first two of the algorithms are designed to minimize part movements in a FMS, an objective that is particularly important in a system having relatively high travel or pallet positioning times. The third heuristic is an attempt to meet the optimal allocation ratios that would have resulted under the unconstrained optimization formulation of the loading sub-problem in a CAN–Q modelling framework. Other heuristic approaches, viewing the loading problem as being of the bin-packing type, are discussed in Ammons *et al.* (1985); Shanker and Tzen (1985) and Rajagopalan (1986). Kim and Yano (1991a) view the loading problem as a two dimensional bin-packing problem. The two restricted dimensions are number of tool slots and processing times of operations. The width of a bin is the tool magazine capacity of the machine, and the bin's height is the available processing time on the machine. Heuristic algorithms, originally developed for the two-dimensional bin-packing problem, are tested for the loading problem, and the results are reported in Kim and Yano (1991a).

There seems to exist a strong connection between the loading and real time scheduling of parts in a FMS. In Stecke and Solberg (1981), an experimental investigation of operating strategies for a computer-controlled FMS is reported. The system under study was a real one, consisting of nine machines, an inspection station and a centralized storage area, all under computer control and interconnected by a MHS. A detailed simulation study was employed to test different alternatives of operating strategies, which basically involved policies for loading and real time flow control. The results are surprising. Minimizing the number of movements of parts can be much better than attempting to balance the workload, since travel time from machine to machine can be decreased significantly. Also, pooling of machines and duplication of operations assignments improved system performance.

The interaction of the loading problem and the scheduling of parts for certain FMS implementations are addressed in the papers Tang and Denardo (1988a and b). For some FMSs, the NC machines used require fine tuning during tool changes, with such fine-tuning operations taking significant time relative to the job processing time. Assuming that the allocation of operations to machines and the set of operations to be processed in the system for the next day are given, the authors address the problem of finding the optimum sequence of operations to be processed on a particular machine. The major constraint faced is the finite capacity of the tool magazine of the machine, i.e. for every job sequence a required number of tool switches is needed. In Tang and Denardo (1988a) the search for an optimum sequence is guided by the objective of minimizing the number of tool switches, that is applicable to FMS implementations with significant tool switching times. The fundamental observation in the paper is that for a given job sequence the optimal tooling policy of the machine (referred to as keep tool needed soonest (KTNS)) has the following properties:

(1) At any instant, no tool is inserted unless it is required by the next job;

(2) If a tool must be inserted, the tools that are not removed are those needed the soonest.

A heuristic procedure, based on the above observation, is given for the problem. In Tang and Denardo (1988b) the same problem under a different objective, that of minimizing the tool switching instances, is presented. Such an objective is applicable for machines having tool switching mechanisms able to perform multiple tool switches simultaneously, with switching times roughly constant and independent of the number of tool switches per switching instant. A specialized optimal branch and bound procedure is suggested for the solution of the problem.

3.4. *Routing mix problem*

The FMS routing mix problem is to determine which of the feasible routes of each part through the manufacturing system should be chosen and also the number of units of a particular product to be produced along the chosen routes. The term route of a part through the manufacturing system usually means a sequence of workstations the part has to visit in order to complete its processing requirements. The routing mix problem is very important since it impacts the routing flexibility of a FMS, a feature with tremendous importance for efficient real time scheduling of the automated system. Before going into a deeper discussion of the problem and its related concepts, modelling aspects of a FMS supportive to our discussion are presented.

Chatterjee *et al.* (1984) have developed a generic model for manufacturing systems to address, among others, the routing mix problem, using mathematical programming as the primary solution approach. They use the unit operation as the basic unit of analysis of parts operations. Kiran and Tansel (1985a) have given a precise definition of a unit operation as:

> 'elementary transformation of an input object into an output object essentially with no interruption. The occurrence of a unit operation requires the existence of an input object, a tool, a machine and possibly a fixture. The interruption at the end of a unit operation is caused by a need to change at least one of these four elements.'

Based on the above notion of unit operation, Chatterjee *et al.* (1984) describe parts operations as an acyclic graph $G_i(N_i, A_i)$ for a part i to capture the part specific operation sequencing rules, where N_i = set of operations of part i, anad A_i = set of arcs of the form (j, k) where j, k are operations which belong to N_i and operation j must precede operation k. Using that representation a successful definition of *routing of part i* as

> 'an ordering of the nodes from the part operation graph G_i that includes all the nodes and does not violate the precedence constraints'

can be given. An elegant algorithm for developing the set of all routings, call it R_i, is presented in the Chatterjee *et al.* (1984) paper. An appropriate routing flexibility measure is also presented: *routing flexibility part specific*: cardinality of the set R_i.

The authors would like to clarify an important point related to routing flexibility. The above flexibility measure captures one aspect of the routing flexibility issue, the one associated with the technology used in order to process a part. Using a given technology there are different sequences of operations we can perform in order to complete processing our part. But the most important aspect of routing flexibility, the one we build into our FMS, is the one associated with the appropriate allocation of operations and tools to the workstations (machine loading problem). Multiple assignment of the same operation to different workstations and appropriate tooling of the machines tremendously increases the routing flexibility of the system. Other aspects of routing flexibility are associated with the MHS design (for further discussion see Afentakis, 1986 and 1989).

Chatterjee *et al.* (1984) formulate the routing mix problem as a mixed integer program. Two sets of decision variables are used, the one having binary variables associated with the selection of routes of each part through the system and the other is a set of integer variables denoting the number of units of a part type produced according to a certain routing. The optimal decisions are based on the trade-offs between the set-up costs and the variable production costs at the workstations considering capacity constraints and the need to meet the demand for a particular part type. The objective function can be questioned for its general applicability, since the concept of set-up cost in most FMS implementations with automated tool-switching capabilities, that can easily interchange tools in seconds, seems to be of minor importance. Moreover, the proposed formulation doesn't capture the important linkage between the routing mix problem and the FMS loading problem, a significant routing flexibility aspect in a FMS.

The routing mix problem is also linked to the congestion phenomena in a FMS, an issue of significant importance due to the direct effect on the system performance. Appropriately addressing such considerations, Kinemia and Gershwin (1985) propose a network flow optimization procedure for solving the problem of optimal part routing. A detailed presentation of their model follows. The notation introduced by Kinemia and Gershwin is as follows:

M = number of workstations
N = number of different part types
k_i = number of operations for completion of a particular part type
τ_{ij}^k = time to complete operation k on a type i part at a workstation j
$\mathbf{y}_{ij}^k$ = flow rate of type i parts to station j for operation k
u_i = production ratio of type i parts
α_i = fraction of total production that is of type i
ρ_j = utilization of workstation j
$\mathbf{y}$ = a vector of flow rates
$f(\mathbf{y})$ = performance measure that is to be maximized.

Then the general flow optimization problem for FMS is the following:

max f(y)
s.t.

$$\sum_{j=1}^{M} y_{ij}^{k} = u_i \qquad k = 1, 2, \ldots, k_i \quad i = 1, 2, \ldots, N$$

$$\sum_{i=1}^{N} u_i = R$$

$$u_i = \alpha_i R$$

$$\rho_j(y) = \sum_{i=1}^{N} \sum_{k=1}^{k_i} y_{ij}^{k} \tau_{ij}^{k} \quad j = 1, \ldots, M$$

$$\rho_j(y) \leq 1$$

$$y_{ij}^{k} \geq 0$$

The authors propose an augmented Lagrangian method in combination with Dantzig–Wolfe decomposition to solve the problem. The solution gives optimal plans for routing parts to maximize the appropriate performance measure, which most of the time is the production output.

The flow optimization approach of Kinemia and Gershwin (1985) is an attempt to guarantee through the problem constraints that congestion phenomena at the workstations will be avoided. Thus, as a consequence, performance measures derived from an appropriate queueing model will be valid and can usefully be subjected to an optimization procedure. This formulation ignores the interaction of the loading and routing sub-problems.

Another paper addressing the routing problem is the paper by Cassandras (1984). He offers a high-level view of a manufacturing system as a network of stations where various operations are performed on resources carried to and from the stations according to specific requirements. Parts, tools and pallets are examples of such resources. His problem formulation idea is quite simple. The constraints of the problem are the classical ones found in a network flow optimization problem: conservation of flow, notions of a source and a sink and non-violation of capacity constraints. His objective: flow time minimization. The formulation follows the well-known network flow optimization problem pattern.

As we have mentioned repeatedly up to now, there is a strong interaction between the different FMS planning sub-problems. Clearly the FMS routing mix problem is linked to product mix and loading problems. Attempts to integrate the different FMS planning sub-problems into the same mathematical program have been reported. Kiran and Tansel (1985b) integrated five of the FMS planning problems into the so-called system set-up problem. A set of constraints included in their formulation addresses the operations assignment problem. The fundamental unit of analysis is the unit operation. Each unit operation has to be assigned to a machine that has the capability to perform it. If a unit operation is assigned to a machine then the tool which will perform that operation has to be assigned to the same machine. That simple idea is expressed by another set of constraints. Constraints for not violating the tool magazine capacity, in the non-linear form introduced by Stecke (1983), are also included. All the above set of constraints are similar to those of Stecke's loading problem. In addition, the constraint for the non-violation of time availability at each machine is used.

Kiran and Tansel (1985a) realized the importance of the congestion phenomena in a FMS and tried to introduce constraints to avoid them. Their modelling approach is indirect. In an attempt to avoid those phenomena they restrict the number of machine changes to an upper limit (external parameter) and also the number of fixture changes. The intuition behind it is to avoid delays and keep the traffic between workstations at a low level.

One of their decision variables is the number of parts to be produced from each part type, in other words the product mix problem is solved simultaneously with the other ones. A constraint that limits the number of parts which will be selected for simultaneous production is also introduced. The constraint formulation captures appropriately the linkages of the different sub-problems.

The proposed objective is the maximization of the total number of parts produced during the planning period. The model has a structure which can be exploited. For a sample problem with 100 tool types, 350 operations, 100 part types, and five machines, there are approximately 4000 decision variables and 6000 constraints. Relaxation based decomposition techniques are proposed.

3.5. *Other planning problems in a FMS environment*

Mazzola *et al.* (1989) propose a hierarchical model that integrates FMS production planning into a closed-loop material requirements planning (MRP) environment. Their framework involves a three-level hierarchy consisting of the FMS/MRP rough-cut capacity planning, grouping and loading, and detailed scheduling problems. The authors examine, in detail, the first two levels of the hierarchy. For the rough-cut capacity planning problem they present a generalized assignment problem formulation and suggest a heuristic procedure for its solution. For the grouping and loading problem a two-phase heuristic is presented. Mazzola (1989) presents batch-splitting heuristics for the FMS rough-cut capacity planning problem.

An interesting planning problem in a FMS environment is fixture allocation to the different part types. Given the product mix and the unit operations required by each part type, the decision faced is that of allocating the limited number of fixtures to parts in such a way that all required operations can be performed and an operational objective (usually throughput) is optimized. The fixture allocation problem was addressed as part of a more general model in Kiran and Tansel (1985a). The structure of the formulation bears significant resemblance to the operations assignment problem (refer Section 3.3).

Kusiak and Finke (1988) address a process planning problem applicable to FMS implementations. The flexibility of the machines and the material-handling system allows each part the opportunity to be manufactured by a variety of process plans (i.e. sequence of operations). Each of the process plans requires specific types of tools and auxiliary devices (i.e. fixtures, grippers, feeders) during the manufacturing process. The authors propose a model for the selection of a set of process plans for each part type with the minimum corresponding cost and minimal number of tools and auxiliary devices required. Both a graph theoretic and an equivalent integer programming formulation of the problem are proposed. Since optimization algorithms fail to address the computational complexity of real size problems, two quick and efficient heuristic procedures are presented for the solution of the problem.

Kiran and Krason (1988), Gray *et al.* (1989) pinpoint tool management as a critical issue for FMS performance. Stecke (1988a) reports as a case study the operating problems of a FMS that will become operational in 1989, and identifies as key planning issues the tool management problems. A classification of all relevant issues to tool management is presented in Gray *et al.* (1989). Such issues include spare tools management, tool storage policies, tool loading to machines and cutting/feed rate optimization.

Vinod and Sabbagh (1986) analyze the FMS performance subject to tool availability constraints. The aspects of tool unavailability are triggered by the need of tool replacement at a workstation due to any one or a combination of reasons: tool breakage, tool wear and tear, and generally poor tooling performance (i.e. poor quality finish, lack of dimensional accuracy). The approach for modelling the problem is simple and intuitive. The actual job processing time (i.e. service time at a node of a CQN for performance modelling purposes) is modified to account for the extended time required to process jobs due to interruptions caused by tool failures. In order to retain the product form solution of the results obtained by a CAN-Q type of model, the mean job processing time is assumed to include tool preparation and cutting time, and each machine at a specific workstation requires only one tool to process a job. Using the following additional notation: π_i = steady state distribution of availability of tools at workstation i, and $\mu_i(n)$ = load-dependent processing rate at workstation i when there are n jobs waiting to be processed, the modified job processing rate at workstation i, with c_i identical machines and processing rate of each machine, when appropriately tooled, μ_i^0, is given by

$$\mu_i(n) = \sum_{l=1}^{n-1} \pi_i(l)\, l \mu_i^0 + \sum_{l=n}^{c_i} \pi_i(l)\, n \mu_i^0 \qquad n = 1, \ldots, c_i$$

$$\mu_i(n) = \mu_i(n-1) \qquad n = c_i + 1, \ldots, N,$$

where N = the job population circulating in the system. In order for the load-dependent processing times to be calculated the steady state distribution π_i has to be derived. Vinod and Sabbagh (1986) suggest the use of a queueing model $M/M/c_i/c_i + y_i$ (where y_i = spares available for tools used at station i) to calculate the π_i, where the mean time between arrivals in the model is interpreted as the mean time to tool failure. The arrival rate and the service rate (i.e. μ_i^0) are assumed to be exponential random variables. For optimizing the levels of spares available for each tool a form of lexicographic partial enumeration is used. Kusiak (1986b) describes four basic tool storage policies, whereby spare tools are stored either in a tool magazine or in a remote tool storage area or both. The cutting speed/feed rate optimization for flexible machines is studied by Schweitzer and Seidmann (1988a and b). Having in mind a queueing network performance model of a FMS, specification of cutting speed/feed rate affects the processing rates of the flexible machines at a workstation. The authors present several nonlinear queueing network optimization methodologies, which determine the minimum cost processing rates given the desired FMS throughput, the WIP level, and tool cost functions. An intuitive and practically useful result coming out of the above work is that an optimal processing rate policy will use slight acceleration of processing rates at a few bottleneck machines, while allowing for significant tool cost reduction by lowering speeds at non-bottleneck machines.

4. Further research directions

4.1. *FMS design problems*

Although significant research effort has been spent in the development of analytical models for the design of FMS, very few have actually been used as major decision support tools during the design phase of such complex manufacturing systems. The gap between the mathematical design theory and the design practice is mostly due to the following factors.

4.1.1. *Absence of input data for the models*

Most of the design models assume that information on such items as the demand pattern for the different products, cost estimates for the system resources (machines, tools, pallets, fixtures, etc.), reliability of failure-prone system resources (machines, material-handling system), processing rates of flexible machines (i.e. machines that can process a wide range of operations, and consequently their processing rate depends on the operational product mix and the allocated set of operations and tools to them), operational product mix over the planning horizon (assumed rather stable), and even the length of the planning horizon, is available during the initial design phase of a FMS. It comes as a shock to most production researchers, particularly those that don't have a close contact with actual manufacturing environments, to find out that most manufacturing designers don't have at their disposal most of the above data, and in some cases they don't even have adequate ways to measure such inputs even for existing FMSs (Elmaghraby, 1987). The design engineer faces the challenge of designing a deterministic system capable of handling stochastic inputs, variable product mixes, changing processing technologies, non-deterministic processing rates, and probabilistic routings of a wide range of products traveling through the system over time.

4.1.2. *Data inaccuracy*

Even for the few cases where some of the above data is available, it is highly inaccurate. That might result either in the use of a historical analogy approach (i.e. use data from an already existing similar FMS), which might not be particularly useful for FMSs that respond to production requirements of different markets, or the expected forecasting inaccuracy due to the extended length of the planning horizons used for design purposes. To make things even worse, the short product life cycles of some products and the continuous changing nature of processing technologies in certain FMS related industries limits the applicability of highly accurate forecasting techniques (sophisticated time series, econometric models, etc.). In most cases the FMS designer will come up with specific ranges where the input parameters might lie, and in a few cases he might be able to describe probability distributions for the input parameter values over such a range. The majority of the previously described FMS design models, mostly due to the inherent combinatorial nature of the underlying problems and the use of an optimizing approach, are not amenable to easy sensitivity analysis where simultaneous variation of multiple input parameters is required.

4.1.3. *Inability of proposed algorithms to address real size problems*

This is mostly due to the inherent complex combinatorial nature of the design problems, and the limited processing capability of available computer technology. It doesn't provide any relief for the FMS designer to be informed that most of the FMS design problems are NP-complete. What the FMS designers need are robust designs that provide adequate operational performance for a wide range of input data.

4.1.4. *Lack of sub-problem integration*

In the FMS literature the different design sub-problems are addressed independently and only a few attempts have been made to link some of them. But the design of an integrated manufacturing system, which is the problem that a FMS engineer faces, requires the linkage of the different sub-problems. Optimality approaches to the solution of various design sub-problems, when significant interaction exists between them, does not imply optimality of the resulting FMS design. For example, there is significant interaction between the machine selection and the lay-out problem. The same set of machines under different lay-out configurations lead to different system throughputs (Solberg and Nof, 1980 and Kouvelis and Kiran, 1991), and for the same lay-out the determination of the number of machines of different machine types (even if the total number of available machines in the system has been predetermined) significantly affects the operational performance of the FMS (Kouvelis, 1988).

The existing FMS models provided a valuable first step in clearly identifying and structuring the relevant FMS design problems. But further intense research effort is needed in addressing the previously mentioned issues that limit the applicability of some of the existing models. There is a significant need for incorporating the inaccurate input parameter specification aspects of the design problem into appropriate models, and then attempt to develop solution approaches that lead to robust designs over a wide range of input instances. In many cases that might imply the abandoning of strict optimization approaches for the use of efficient (and rigorously bounded in terms of deviation from optimality) heuristics that facilitate the generation of sensitivity results for a wide range of input values (for an example of such an approach refer to Bitran and Tirupati (1989b)). In addition to traditional post-optimality and sensitivity analysis, it might be necessary to use non-traditional approaches that attempt to develop algorithms able to construct solutions that are close to optimality for all or most problem instances, where each instance of the problem is defined by a particular instance of the input string. An example of such an approach, though presented for a different class of problem, can be found in Tansel and Scheuenstuhl (1988).

Integration of the design sub-problems is a challenging research issue. Given the size and the complexity of the individual design sub-problems, the development and the solution of a single monolithic large-size combinatorial model does not seem to be a fruitful research avenue. There is a need for the development of a large-scale system model for a FMS that will consist of various subcomponents (submodels). Such a model should be able to describe the FMS both at a detailed and an aggregate information level, if it is to be used to satisfy the needs of the design phase. Of course, such a large-scale system model will be of significant mathematical complexity. The crucial research issue is the development of an appropriate hierarchical decision structure. Such a decomposition scheme should on the one hand recognize the sub-problem interaction, while on the other hand should generate sub-problems that are computationally tractable. One approach is to base the decomposition on the sub-problems already presented. The search for an appropriate objective for each sub-problem, one that is compatible with the over-all system design objectives, is in that case the immediate goal. The sequence in which the different sub-problems are to be solved in order to minimize deviations from an over-all optimal (or close to optimality) solution must also be determined.

The development of a generic modelling framework for a FMS can be used to provide other important side results. For example, it could be very useful to attempt to define rigorously some of the different flexibility concepts related to FMS (for references to flexibility concepts see Mandelbaum (1978), Buzacott (1982), Zelenovic (1982), Browne *et al.* (1984), Jaikumar (1984), Adler (1985), Kusiak (1985) and Lasserre and Roubellat (1985)). It is particularly interesting to explore the way that the various design decisions limit (or enhance) the different flexibility aspects of a FMS. A first step in that direction was presented in Afentakis (1986 and 1989).

The automated MHS is an important component in a FMS, since congestion at the MHS significantly limits the capacity of a highly interconnected system of workstations such as a FMS. The modelling of the MHS in the FMS literature has been done with the use of aggregate models (mostly CQNs). In our view, the design issues of the MHS have a strong linkage with the short-term (operational) material-handling problems in FMS. Consequently, an aggregate model like a CQN isn't appropriate in addressing these issues. Unfortunately, little work has been done in the direction of detailed model development for MHS design in FMS. Most of the existing knowledge for the design of those systems comes from the broader material-handling research literature, usually addressing MHS issues in production contexts different than the FMS one. The need for the development of new models, methods and techniques that concentrate on the unique material-handling requirements of a FMS, that are mostly due to the tremendous operational flexibility needs of such a system, is apparent.

4.2. *FMS planning problems*

Some of the factors previously discussed as limiting the applicability of mathematical design theory to actual practice are also present in the case of FMS planning problems. The shorter time horizon and the on-line availability of information, which as a side effect creates the challenge for efficient information management in the FMS environment, limit the effects of data unavailability and inaccuracy in the application of the developed models for FMS planning issues. But the computational complexity of the proposed algorithms, the lack of computing power and the lack of planning sub-problem integration are still present, and carry a heavier weight in limiting the success of operations research models in addressing the urgent operating problems of the FMS environment. We would like to point out that operations research-based planning models have been more widely implemented in practice than the respective design models .For an example of such applications refer to Stecke (1988a).

The FMS manager needs algorithms (or operating policies) that provide in a timely fashion good answers for pressing operating problems in a highly volatile manufacturing environment. In many cases even the knowledge of a feasible answer for an operating problem is adequate for practical purposes. Such an example is the extremely difficult FMS loading problem. Finding a feasible allocation of operations and tools to machines under tight tool magazine capacity constraints and/or tight machine time availabilities over a short horizon is a challenging problem by itself, and in most cases all that the FMS operator wants (Stecke, 1988a). This fact emphasizes the need for efficient heuristic algorithms for the planning problems, which are designed in a way that exploits the structure of the specific problem. As the planning horizon becomes shorter computational efficiency and feasible solution generation should gain weight over deviation from optimality as the guiding criteria for the design of the heuristic.

Research addressing different FMS planning sub-problems should continue, since such research provides valuable insight regarding a number of key planning issues. Although some of the FMS planning sub-problems have already been formulated into operations research models and solution approaches have been proposed, they cannot be considered as closed issues. Attempts to question their objectives and even the formulation of the problem, when sound reasons for doing so exist, can be proven very useful for the understanding of those complex problems. For example, for the FMS loading problem many different objective functions have been proposed for an appropriate integer programming formulation of the problem. The interesting research question is to identify the specific characteristics of a FMS implementation that makes one operational objective preferable to another, and even more clearly establish the cases for which the generation of a feasible solution is adequate for practical purposes for the short-term FMS production planning problem.

Pointing out and appropriately structuring new planning issues specific to the FMS environment or in certain cases specific to a particular FMS implementation is another valid research direction. In Buzacott and Yao (1986) the need for modelling, design and understanding of the operating constraints, that the existence of a centrally located tool delivery system might impose for certain FMS implementations, was pointed out. Still no significant results have appeared on the problem. We would like to point out further that the nature of certain planning problems, and we will use as an example the loading problem, is different for FMS implementations with a central tool delivery system. The tightness of the tool magazine capacity constraint is relaxed, since tools don't have to be on the machine all the time anymore. All that is required is that the necessary tools for a particular operation have been delivered to the machine before the operation starts. As it can be easily understood, congestion phenomena of the tool delivery system become of critical importance. Also, the issue of determining the tools that will be continuously placed on the various machines and the ones that are going to be stored in the central tool store area needs to be researched. The usefulness of tool-sharing policies for such FMS implementations has also to be addressed.

Models addressing the interaction of planning sub-problems have been developed (Kiran and Tansel, 1985a and Rajagopalan, 1986). The resulting mathematical programming formulations face the curse of combinatorial complexity. Testing the efficiency of different hierarchical decomposition schemes for the integrated FMS short-term planning problem (usually referred to as the 'FMS set-up problem') is needed. Development of heuristic procedures either for the whole problem or for each individual sub-problem should also attract more research effort. For the latter case, iterative schemes, that combine the heuristic results for the sub-problems and attempt to improve the performance value of the resulting planning solution, have to be developed.

Understanding of the FMS short-term planning problems is a crucial link to understanding FMSs. In order to design optimally FMSs we first need to understand how we efficiently allocate the system's resources in shorter planning horizons, and in order to have an efficiently operated and controlled FMS, we need to realize what are the flexibility limitations our short-term planning decisions impose on the real time operation of the system. Hopefully, future research effort will provide further insight on these issues.

References

Adler, P. S. (1985) Managing flexibility: a selective review of the challenges of managing the new production technologies potential for flexibility, Working Paper, Stanford University.

Afentakis, P. (1986) A model for layout design in FMS, in *Flexible Manufacturing Systems: Methods and Studies*, Kusiak, A. (ed.) North Holland, Amsterdam.

Afentakis, P. (1989) A loop layout design problem for flexible manufacturing systems. *International Journal of Flexible Manufacturing Systems*, **1**, 175–96.

Afentakis, P., Solomon, M. M. and Millen, R. A. (1989) The part-type selection problems, in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Mass.

Ammons, J. C., Lofgren, C. B. and McGinnis, L. F. (1985) A large scale machine loading problem in flexible assembly. *Annals of Operations Research*, **3**, 319–22.

Armour, G. C. and Buffa, E. S. (1963) A heuristic algorithm and simulation approach to the relative location of facilities. *Management Science*, **9**, 294–309.

Bazaraa, M. S. (1975) Computerized layout design: a branch and bound approach. *AIIE Transactions*, **7**, 432–7.

Berrada, M. and Stecke, K. E. (1986) A branch and bound approach for machine load balancing in flexible manufacturing systems. *Management Science*, **32**, 1316–35.

Bitran, G. R., Peterson, W. P. and Tirupati, D. (1987) Some recent advances in capacity planning for discrete manufacturing systems, in *Modern Production Management Systems*, Kusiak, A. (ed.), North Holland, Amsterdam.

Bitran, G. R. and Tirupati, D. (1988) Multiproduct queueing networks with deterministic routing: decomposition approach and the notion of interference. *Management Science*, **34**, 75–100.

Bitran, G. R. and Tirupati, D. (1989a) Capacity planning with discrete options in manufacturing networks. *Annals of Operations Research*, **17**, 119–36.

Bitran, G. R. and Tirupati, D. (1989b) Trade-off curves, targeting and balancing in manufacturing networks. *Operational Research*, **37**, 547–64.

Boxma, O. J., Rinnooy Kan, A. H. G. and van Vliet, M. (1990) Machine allocation problems in manufacturing networks. *European Journal of Operational Research*, **45**, 47–54.

Boxma, O. J., Rinnooy Kan, A. H. G. and van Vliet, M. (1991) Machine allocation algorithms for job shop manufacturing environments, Working Paper, Tilburg University, Netherlands.

Browne, J., Dubois, D., Rathmill, K., Seth, S. P. and Stecke, K. E. (1984) Classification of flexible manufacturing systems. *FMS Magazine*, **April**, 114–17.

Buzacott, J. A. (1982) The fundamental principles of flexibility in Manufacturing Systems, in *Proceedings of the 1st International Conference on Flexible Manufacturing Systems*, Brighton, UK.

Buzacott, J. A. and Shanthikumar, J. G. (1980) Models for understanding flexible manufacturing systems. *AIIE Transactions*, **12**, 339–50.

Buzacott, J. A. and Yao, D. D. (1986a) On queueing network models of flexible manufacturing systems. *Queueing Systems Theory and Applications*, **1**, 5–27.

Buzacott, J. A. and Yao, D. D. (1986b) Flexible manufacturing systems: a review of analytical models. *Management Science*, **32**, 890–95.

Cassandras, C. (1984) A hierarchical control scheme for material handling systems, in *Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems*, Ann Arbor.

Chakravarty, A. K. and Schtub, A. (1984) Selecting parts and loading FMS, in *Proceedings of the First ORSA/TIMS Conference on FMS*, Ann Arbor.

Chatterjee, A., Cohen, M. A., Maxwell, W. L. and Miller, L. W. (1984) Manufacturing flexibility: models and measurements, in *Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems*, Ann Arbor.

Co, H. C. (1984) Design and implementation of FMS, PhD Dissertation. Virginia Polytechnic Institute and State University.

Dallery, Y. and Frein, Y. (1986) An efficient method to determine the optimal configuration of a manufacturing system, in *Proceedings of the Second ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Ann Arbor.

Dallery, Y. and Frein, Y. (1988) An efficient method to determine the optimal configuration of a flexible manufacturing system. *Annals of Operations Research*, **15**, 207–25.

Dallery, Y. and Stecke, K. E. (1990) On the optimal allocation of servers and workloads in closed queueing networks. *Operational Research*, **37**, 694–703.

Drezner, Z. (1980) DISCON: a new method for the layout problem. *Operational Research*, **28**, 1375–84.

Dupont-Gatelmand, C. (1982) A survey of flexible manufacturing systems. *Journal of Manufacturing Systems*, **1**, 1–15.

Elmaghraby, S. E. (1987) On the measurement of capacity. Working Paper, NC State University, NC.

Fox, B. (1966) Discrete optimization via marginal analysis. *Management Science*, **13**, 210–16.

Francis, R. L. and White, J. A. (1974) *Facility Layout and Location: An Analytical Approach*, Prentice Hall, Englewood Cliffs, NJ.

Gaskins, R. J. and Tanchoco, J. M. A. (1987) Flow path design for automated guided vehicle systems. *International Journal of Production Research*, **25**, 667–76.

Gershwin, S., Hildebrandt, R., Suri, R. and Mitter, S. (1984) A control theorist's perspective on recent trends in manufacturing systems, in *Proceedings of the 23rd IEEE Conference on Decision and Control*, Las Vegas.

Gershwin, S. B., Hildebrandt, R. R., Suri, R. and Mitter, S. K. (1986) A control perspective on recent trends in manufacturing systems. *IEEE Control Systems*, **April**, 3–15.

Gilmore, P. C. (1963) Optimal and suboptimal algorithms for the quadratic assignment problem. *SIAM Journal*, **10**, 305–13.

Gray, A. E., Seidmann, A. and Stecke, K. E. (1988) Tool management in automated manufacturing: operational issues and decision problems, Working Paper, Ann Arbor.

Gray, A. E., Seidmann, A. and Stecke, K. E. (1989) Tool management in automated manufacturing: a tutorial, in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*. Mass.

Greene, T. J. and Sadowski, R. P. (1986) A mixed integer program for loading and scheduling multiple manufacturing cells. *European Journal of Operational Research*, **24**, 379–86.

Gross, D. and Harris, C. M. (1985) *Fundamentals of Queueing Theory*, John Wiley and Sons, New York.

Haider, S. W., Noller, D. G. and Robey, J. B. (1986) Experiences with analytical and simulation modelling for a factory of the future project at IBM, in *Proceedings of the Winter Simulation Conference*, pp. 641–8.

Hatvany, J. (1983) *World Survey on CAM*. Butterworth, UK.

Heragu, S. S. and Kusiak, A. (1988) Machine layout problem in flexible manufacturing systems. *Operational Research*, **36**, 258–68.

Hillier, F. S. (1963) Quantitative tools for plant layout analysis. *Journal of Industrial Engineering*, **14**, 33–40.

Hillier, F. S. and Boling, R. W. (1967) Finite queues in series with exponential and Erlang service time: a numerical approach. *Operational Research*, **15**, 286–303.

Hillier, F. S. and Connors, M. M. (1966) Quadratic assignment problem algorithms and the location of indivisible facilities. *Management Science*, **13**, 42–57.

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: Operations Research Models and Applications*, Ann Arbor.

Jackson, J. R. (1957) Networks of waiting lines. *Operational Research*, **5**, 518–21.

Jackson, J. R. (1963) Jobshop-like queueing systems. *Management Science*, **10**, 131–42.

Jaikumar, R. (1984) Flexible manufacturing systems: a managerial perspective, Working Paper, Harvard Business School.

Kalkunte, M. V., Sarin, S. C. and Wilhelm, W. E. (1986) Flexible manufacturing systems: a review of modelling approaches for design, justification and operation, in *Flexible Manufacturing Systems Methods and Studies*, Kusiak, A. (ed.), North Holland, Amsterdam.

Kelly, F. P. (1979) *Reversibility and Stochastic Networks*. Wiley, New York.

Kim, Y. D. and Yano, C. A. (1991a) A heuristic approach for loading problems of flexible manufacturing systems. *IIE Transactions*, in press.

Kim, Y. D. and Yano, C. A. (1991b) A branch and bound approach for the loading problem in flexible manufacturing systems: an unbalancing case. *International Journal of Flexible Manufacturing Systems*, in press.

Kinemia, J. G. and Gershwin, S. B. (1985) Flow optimization in flexible manufacturing systems. *International Journal of Production Research*, **23**, 81–96.

Kiran, A. S. (1986) Complexity of FMS loading and scheduling problems, Working Paper, University of Southern California.

Kiran, A. S. and Krason, R. J. (1988) Automatic tooling in a flexible manufacturing system. *Journal of Industrial Engineering*, **20**, 52–7.

Kiran, A. S. and Tansel, B. C. (1985a) A mathematical model for flexible manufacturing systems, Working Paper, University of Southern California.

Kiran, A. S. and Tansel, B. C. (1985b) A framework for flexible manufacturing systems, Working Paper, University of Southern California.

Kiran, A. S. and Karabati, S. (1990) The station location problem on unicyclic material handling networks, Working Paper, University of Southern California.

Kiran, A. S. and Tansel, B. C. (1988a) Optimum storage location in flexible manufacturing cells. *Journal of Manufacturing Systems*, **7**, 121-9.

Kiran, A. S. and Tansel, B. C. (1988b) Optimal pickup point location on material handling networks, *International Journal of Production Research*, **27** (9), 1475–86.

Kleinrock, L. (1975) *Queueing Systems I: Theory*. Wiley, New York.

Kouvelis, P. (1988) Design and planning problems in flexible manufacturing systems, PhD Dissertation, Stanford University.

Kouvelis, P. (1991) An optimal tool selection procedure for the initial design phase of a flexible manufacturing system. *European Journal of Operational Research*, in press.

Kouvelis, P. and Chiang, W. C. (1990) Linear single-row machine layout in automated manufacturing systems, Working Paper, UT Austin.

Kouvelis, P. and Chiang, W. C. (1991) A simulated annealing procedure for single row layout problems in automated manufacturing systems. *International Journal of Production Research*, in press.

Kouvelis, P. and Kim, M. W. (1991) Unidirectional loop network layout problem in automated manufacturing systems. *Operational Research*, in press.

Kouvelis, P. and Kiran, A. S. (1989) Layout problem in flexible manufacturing systems: recent research results and further research directions, in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Mass.

Kouvelis, P. and Kiran, A. S. (1990) The plant layout problem in automated manufacturing systems. *Annals of Operations Research*, **26**, 397–412.

Kouvelis, P. and Kiran, A. S. (1991) Single and multiple period layout models for manufacturing systems. *European Journal of Operational Research*, **52** (3), 300–14.

Kouvelis, P. and Lee, H. L. (1990a) The optimal system configuration of a flexible manufacturing system, Working Paper, The University of Texas, Austin.

Kouvelis, P. and Lee, H. L. (1990b) The material handling system design of automated manufacturing systems: a graph theoretic modelling framework. *Annals of Operations Research*, **26**, 379–96.

Kouvelis, P. and Lee, H. L. (1991) Block angular structures and the loading problem in flexible manufacturing systems. *Operations Research*, **39** (4), 666–76.

Kouvelis, P. and Tirupati, D. T. (1991) Approximate performance modelling and decision making for manufacturing systems: a queueing network optimization framework. *Journal of Intelligent Manufacturing*, **2**, 107–34.

Kumar, K. R., Kusiak, A. and Vannelli, A. (1986) Grouping of parts and components in flexible manufacturing systems. *European Journal of Operational Research*, **24**, 287–97.

Kusiak, A. (1983) Loading models in flexible manufacturing systems, recent developments in FMS and allied areas, in *The Design and Operation of FMS*. North Holland, New York.

Kusiak, A. (1984) Flexible manufacturing systems: a structural approach. *International Journal of Production Research*, **23**, 1057–73.

Kusiak, A. (1985) Material handling in flexible manufacturing systems. *Material Flow*, **2**, 79–95.

Kusiak, A. (1986a) Application of operational research models and techniques in flexible manufacturing systems. *European Journal of Operational Research*, **24**, 336–45.

Kusiak, A. (1986b) Parts and tools handling systems, in *Modelling and design of Flexible Manufacturing Systems*, Kusiak, A. (ed), Elsevier Science Publishers, Amsterdam, pp. 99–109.

Kusiak, A. (ed) (1986c) *Flexible Manufacturing Systems: Methods and Studies*. North Holland, Amsterdam.

Kusiak, A. (ed) (1986d) *Modelling and Design of Flexible Manufacturing Systems*. North Holland, Amsterdam.

Kusiak, A. and Finke, G. (1988) Selection of process plans in automated manufacturing systems. *IEEE Journal of Robotics and Automation*, **4**, 397–402.

Lasserre, J. B. and Roubellat, F. (1985) Measuring decision flexibility in production planning. *IEEE Transactions of Automatic Control*, **30**, 447–52.

Lawler, E. L. (1963) The quadratic assignment problem. *Management Science*, **9**, 586–99.

Lee, H. F., Srinivasan, M. M. and Yano, C. A. (1989) The optimal configuration and workload allocation problem in flexible manufacturing systems, in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Mass.

Mandelbaum, M. (1978) Flexibility in decision making: an exploration and unification, PhD Dissertation, University of Toronto, Canada.

Matson, J. O. and White, J. A. (1982) Operational research and material handling. *Journal of Operational Research*, **11**, 309–18.

Maxwell, W. L. (1981) Solving material handling design problems with OR. *Journal of Industrial Engineering*, **13**, 58–69.

Maxwell, W. L. and Muckstadt, J. A. (1982) Design of automatic guided vehicle systems. *IIE Transactions*, **14**, 114–24.

Mazzola, J. B. (1989) Heuristics for the FMS/MRP rough-cut capacity planning problem, in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Mass.

Mazzola, J. B., Neebe, A. W. and Dunn, C. V. R. (1989) Production planning of a flexible manufacturing system in a material requirements planning environment. *International Journal of Flexible Manufacturing Systems*, **1**, 115–42.

Merchant, M. E. (1983) Production: a dynamic challenge. *IEEE Spectrum*, **May**, 36–9.

Muth, E. J. and White, J. A. (1979) Conveyor theory: a survey. *AIIE Transactions*, **11**, 269–77.

Newman, W. E. (1986) Model to evaluate the benefits of FMS pallet flexibility, in *Proceedings of Second ORSA/TIMS Conference on FMS: Operations Research Models and Applications*, Ann Arbor.

Picone, C. J. and Wilhelm, W. E. (1984) Perturbation scheme to improve Hillier's solution to the facilities layout problem. *Management Science*, **30**, 1238–49.

Rachamadugu, R. and Stecke, K. E. (1987) Classification and review of FMS scheduling procedures, Working Paper, The University of Michigan, Ann Arbor.

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.

Rosenblatt, M. J. (1986) The dynamics of plant layout. *Management Science*, **32**, 76–86.

Sahni, S. and Gonzalez, T. (1976) *P*-complete approximation problem. *Journal of the Association of Computing Machinery*, **23**, 555–65.

Sarin, S. C. and Chen, S. C. (1987) The machine loading and tool allocation problem in an FMS. *International Journal of Production Research*, **25**, 1081–94.

Sarin, S. C. and Wilhelm, W. E. (1984) Prototype models for two dimensional layout design of robot systems. *IIE Transactions*, **16**, 106–25.

Schriber, T. J. and Stecke, K. E. (1988) Machine utilizations achieved using balanced FMS production rates in a simulated setting. *Annals of Operations Research*, **15**, 229–67.

Schweitzer, P. J. and Seidmann, A. (1988a) Optimizing processing rates for flexible manufacturing systems, Working Paper, University of Rochester, New York.

Schweitzer, P. J. and Seidmann, A. (1988b) Capacity range analysis and processing rate optimization for FMS's with distinct multiple job visit to work centers, Working Paper, University of Rochester, New York.

Shanker, K. and Tzen, Y. J. (1985) A loading and dispatching problem in a random flexible manufacturing system. *International Journal of Production Research*, **23**, 579–95.

Shanthikumar, J. G. (1979) Approximate queueing models of dynamic job shops, PhD Dissertation, University of Toronto, Canada.

Shanthikumar, J. G. and Sargent, R. G. (1981) A hybrid simulation/analytic model of a computerized manufacturing system, in *Proceedings of the 9th IFORS International Conference on Operational Research*, Hamburg, Germany.

Shanthikumar, J. G. and Yao, D. D. (1987) Optimal server allocation in a system of multiserver stations. *Management Science*, **33**, 1173–80.

Shanthikumar, J. G. and Yao, D. D. (1988) On server allocation in multiple center manufacturing systems. *Operational Research*, **36**, 333–42.

Sharp, G. P. and Liu, F. (1987) Analytical approaches to the design of material flow systems, Working Paper, Atlanta, Georgia.

Solberg, J. J. (1977) A mathematical model of computerized manufacturing systems, in *Proceedings of the 4th International Conference on Production Research*, Tokyo, Japan.

Solberg, J. J. and Nof, S. Y. (1980) Analysis of flow control in alternative manufacturing configurations. *Journal of Dynamic Systems, Measurement and Control*.

Stecke, K. E. (1983) Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. *Management Science*, **29**, 273–88.

Stecke, K. E. (1984) Design, planning, scheduling and control problems of FMS, in *Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems*, Ann Arbor.

Stecke, K. E. (1986) A hierarchical approach to solving machine grouping and loading problems of flexible manufacturing systems. *European Journal of Operational Research*, **24**, 369–78.

Stecke, K. E. (1988a) Algorithms for efficient planning and operation of a particular FMS, Working Paper, The University of Michigan, Ann Arbor.

Stecke, K. E. (1988b) Procedures to determine part mix ratios in flexible manufacturing systems, Working Paper, The University of Michigan, Ann Arbor.

Stecke, K. E. and Kim, I. (1986) A flexible approach to implementing the short-term FMS planning function, in *Proceedings of the Second ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, University of Michigan, Ann Arbor.

Stecke, K. E. and Kim, I. (1988) A study of FMS part type selection approaches for short-term production planning. *International Journal of Flexible Manufacturing Systems*, **1**, 7–29.

Stecke, K. E. and Kim, I. (1989) Performance evaluation for systems of pooled machines of unequal sizes: unbalancing versus balancing. *European Journal of Operational Research*, **42**, 22–38.

Stecke, K. E. and Morin, T. L. (1985) Optimality of balancing workloads in certain types of flexible manufacturing systems. *European Journal of Operational Research*, **20**, 68–82.

Stecke, K. E. and Solberg, J. J. (1985) The optimality of unbalancing both workloads and machine group sizes in closed queueing networks of multiserver queues. *Operational Research*, **33**, 822–910.

Stecke, K. E. and Talbot, B. F. (1985) Heuristics for loading flexible manufacturing systems, in *Flexible Manufacturing: Recent Developments in FMS, Robotics, CAD/CAM, CIM*. Raouf, A. and Ahmad, S. I. (eds), Elsevier Science Publishers, Amsterdam, pp. 73–84.

Stecke, K. E. and Suri, R. (ed.) (1989) in *Proceedings of the Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications*, Mass.

Suri, R. (1983) Robustness of queueing network formulae. *Journal of the Association of Computing Machinery*, **30**, 564–94.

Suri, R. and Whitney, C. K. (1984) Decision support requirements in flexible manufacturing systems. *Journal of Manufacturing Systems*, **1**, 61–9.

Sweeney, D. S. and Tatham, R. L. (1976) An improved long-run model for multiple warehouse location. *Management Science*, **22**, 748–58.

Tang, C. S. and Denardo, E. (1988a) Models arising from a flexible manufacturing machine – Part I: minimization of the number of tool switches. *Operations Research*, **36** (5), 767–77.

Tang, C. S. and Denardo, E. (1988b) Models arising from a flexible manufacturing machine – Part II: minimization of the number of switching instants. *Operations Research*, **36** (5), 778–84.

Tansel, B. C. and Scheuenstuhl, C. F. (1988) Facility location on tree networks with imprecise data, Working Paper, University of Southern California.

Tompkins, J. A. and White, J. A. (1984) *Facilities Planning and Design*, John Wiley and Sons, New York.

Vanneli, A. and Kumar, K. R. (1986) A method for finding minimal bottleneck cells for grouping part machine families. *International Journal of Production Research*, **24**, 387–400.

Vinod, B. and Sabbagh, M. (1986) Optimal performance analysis of manufacturing systems subject to tool availability. *European Journal of Operational Research*, **24**, 398–409.

Vinod, B. and Solberg, J. J. (1985) Optimal design of flexible manufacturing systems. *International Journal of Production Research*, **23**, 1141–51.

Wemmerlöv, U. and Hyer, N. L. (1986) Procedures for the part family machine group identification problem in cellular manufacturing. *Journal of Operations Management*, **6**, 125–47.

Whitney, C. K. and Gaul, T. S. (1985) Sequential decision procedures for batching and balancing in FMS. *Annals of Operations Research*, **3**, 301–16.

Whitt, W. (1983a) The queueing network analyzer. *Bell System Technology Journal*, **62**, 2779–815.

Whitt, W. (1983b) Performance of the queueing network analyzer. *Bell System Technology Journal*, **62**, 2817–43.

Whitt, W. (1987) Approximations for single class departure processes from multi-class queues, Working Paper, AT&T Bell Labs.

Yao, D. D. (1985) Some properties of the throughput function of closed networks of queues. *Operational Research Letters*, **3**, 313–18.

Yao, D. D. (1986) An optimal storage model for a flexible manufacturing system, in *Flexible Manufacturing Systems: Methods and Studies*, Kusiak, A. (ed.).

Yao, D. D. and Buzacott, J. A. (1986) Models of flexible manufacturing systems with limited local buffers. *International Journal of Production Research*, **24**, 107–18.

Yao, D. D. and Buzacott, J. A. (1987) Modelling a class of flexible manufacturing systems with reversible routing. *Operational Research*, **35**, 87–93.

Yao, D. D. and Kim, S. C. (1987) Some order relations in closed networks of queues with multiserver stations. *Naval Research Logistics Quarterly*, **34**, 53–66.

Zelenovic, D. M. (1982) Flexibility: a condition for effective production systems. *International Journal of Production Research*, **20**, 318–37.