

# A Decision-Making Approach to the Operation of Flexible Manufacturing Systems

GEORGE CHRYSOLOURIS

*Laboratory for Manufacturing and Productivity, Massachusetts Institute of Technology*

JAMES E. PIERCE

*Laboratory for Manufacturing and Productivity, Massachusetts Institute of Technology*

KRISTIAN DICKE

*Laboratory for Manufacturing and Productivity, Massachusetts Institute of Technology*

**Abstract.** This paper introduces a generic decision-making framework for assigning resources of a manufacturing system to production tasks. Resources are broadly defined production units, such as machines, human operators, or material handling vehicles; and tasks are activities performed by resources. In the specific context of FMS, resources correspond to individual machines; tasks correspond to operations to be performed on parts. The framework assumes a hierarchical structure of the system and calls for the execution of four consecutive steps to make a decision for the assignment of a resource to a task. These steps are 1) establishment of decision-making criteria, 2) formation of alternative assignments, 3) estimation of the consequences of the assignments, and 4) selection of the best alternative assignment. This framework has been applied to an existing FMS as an operational policy that decides what task will be executed on which resource of this FMS. Simulation runs provide some initial results of the application of this policy. It is shown that the policy provides flexibility in terms of system performance and computational effort.

**Key Words:** decision making, flexible manufacturing systems, scheduling

## 1. Introduction

Flexibility, defined as the sensitivity of cost to change, has been a major struggle for modern manufacturing systems. A tool that actually has achieved a considerable amount of flexibility in manufacturing systems is the Flexible Manufacturing System (FMS). There are approximately 800 FMSs installed throughout the world. In terms of hardware, FMSs have reached a relatively significant degree of flexibility; there are systems that can handle up to 1300 different parts. It appears, however, that the planning, scheduling, and control of such systems—the software side—has not yet reached the same level of flexibility. Most of the wealth of approaches, analytical treatments, and applications work done in this area by the scientific community (as indicated by the vast amount of academic literature on this subject) has yet to make its way into industrial practice. Although most of the scientific work is of high quality, the assumptions and approaches taken sometimes deviate significantly from the real-world problem; often they are not flexible enough to accommodate the unique situations that can occur in an FMS.

One of the most widely used modeling techniques for FMS is one that views this system as a network of queues and then makes different assumptions about the distribution or the

arrival times, processing times, due dates, etc. (i.e., see Solberg 1977; Stecke and Solberg 1985; Suri 1981). A comprehensive review of queueing models for the control of flexible manufacturing systems is provided in Buzacott and Yao (1986). The optimization strategies presented by various researchers include mathematical and dynamic programming. Heuristic procedures are also presented. Typically, these approaches are concerned with minimizing tardiness, maximizing throughput, or meeting specified demands. The conclusion drawn in Buzacott and Yao (1986) is that analytical methods are, for some problems, superior to simulation models, which are nonetheless of significant value for evaluating specific system designs. These conclusions were drawn because analytical models allow great insight into the performance of the system. However, it was also concluded that the development of pure analytical models is slow, and, due to the extent of abstraction from real systems, does not always appear to be directly useful. Similar conclusions are reached in Graves (1981), which presents a general, systematic classification of scheduling problems and reviews important developments for this class of problem. In O'Grady and Menon (1986), a large number of FMS articles are reviewed—without, however, any in-depth analysis of the various procedures. Indeed, the wealth of literature on the control and scheduling of FMS is so great that a good review is a formidable task. Consequently, no attempt is made in this paper to review, summarize, or comprehensively cite the body of literature on this subject.

In the last few years, a line seems to have been drawn between two broad categories of approaches to control and scheduling of FMS. The first attempts to address the problem using methods and techniques from operations research, control theory, simulation, or a combination thereof. The second category approaches the problem using artificial intelligence tools. Some of the work in the former category can be found in Kimemia and Gershwin (1983), Sarin and Dar-El (1986), Abidin (1986), Gangan, et al. (1987), and Ben-Arieh et al. (1988). There are fewer publications in the latter category; some of these are Ranky (1988), Suave and Collinot (1987), and Tabe and Salvendy (1988).

The approach described in this paper treats the control and scheduling problem of an FMS as a decision-making problem: *how manufacturing resources are assigned to production tasks*. A resource can be any production unit in a manufacturing facility. For example, a resource may be a machine, an AGV, a human operator, or a machining cell with associated material handling automation. A task can be any activity performed by a resource. In this paper, resources correspond to machines and tasks correspond to operations; however, the term *resources* and *tasks* are retained in order to describe the proposed decision-making approach in the fullest generality. The approach leads to a modular system that utilizes some aspects of operations research approaches as well as some artificial intelligence techniques such as rule-based systems and heuristics (see Chryssolouris et al. [1984, 1985, 1986a, 1986b, 1986c, 1988, 1990]). Particular flexibility is allowed in terms of the optimization criteria that can be considered and the system performance that can be weighed against execution time.

In section 2, the proposed decision-making framework will be described. Section 3 describes the modeling of an existing FMS for use with the framework. Section 4 describes the application of the framework via simulation. Section 5 discusses the results of the application. Finally, conclusions are presented in section 6.

## 2. A Decision-Making Concept for Manufacturing Systems as Applied to FMS

In the process of assigning tasks to resources, three major levels of decision making are used: strategic, tactical, and operational.

**Strategic** decision making is concerned with establishing the managerial policies and goals of the overall organization. Decisions at this level are extremely important since they determine to a great extent the success or failure of the organization. An important characteristic of strategic decisions is their long-lasting effects, resulting in long decision horizons in order for the proper information to be analyzed adequately. This information usually is processed in an aggregate form, with unnecessary details omitted.

The **tactical** level addresses the assignment of resources to production work; however, this allocation occurs in a somewhat aggregate manner in order to realize a relatively rough assignment of the production facilities. Information at this level is more detailed than at the strategic one, and the decision horizons are of medium length.

At the **operational** level, the detailed assignment of work to resources is accomplished by considering the overall goals, objectives and constraints of the organization as they filter from the strategic level down to the operational one.

Figure 1 shows how these decision-making levels reflect the structure of a manufacturing organization as they relate to the production function of the organization: the **factory** corresponds to the enterprise level, the **job shop**, which may be an FMS, to the production level, and the **work center** to the process level, respectively. A work center represents a group of resources that includes resources capable of performing similar manufacturing processes. For example, a turning work center will include all the machines of the FMS capable of performing turning work. Similarly, the milling work center will include all machines in the FMS capable of performing milling work, and so on. There is no need for these individual resources to be at the same location in the factory as long as they are controlled by the same center which receives work requests and assigns resources. The concept of the work center can also be extended to production functions like transportation, inspection, tooling, and maintenance. For example, a maintenance work center can

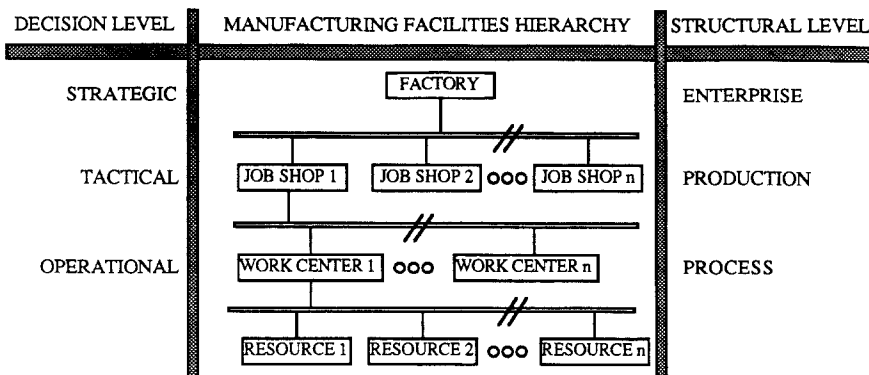


Figure 1. Manufacturing system hierarchical structure.

be viewed as a group of maintenance technicians who are dispatched occasionally within the factory to perform maintenance work on various pieces of equipment. The resources of a work center can be assumed to be either single machines or manufacturing cells: machines grouped together with auxiliary devices (e.g., robots) necessary to operate the machines automatically.

In light of this hierarchical structure, one can view an FMS as a job shop that consists of a number of work centers. Since the work centers are formed on the basis of a common manufacturing process or function and consist of similar resources, the decision problem of assigning a work center's resources to incoming production tasks can be viewed as one of single-stage production with parallel processors. Most of the significant work conducted in this area assumes that the processors are either identical or uniform and considers a single optimization criterion or attribute. The approach taken in this research does not place either of these restrictions on the problem; the work center resources therefore can be viewed as unrelated: the processing time of a given task will be different depending on the resource that performs it, and the decision-making process for assigning the task considers multiple criteria. The imposed decision-making framework is called MADEMA (MANufacturing DEcision MAKing). MADEMA assigns a work center's resources to production tasks following a number of steps that a human undertakes when making a choice.

These steps are

1. selection of decision-making criteria
2. selection of feasible decision alternatives
3. determination of the consequences of the selected alternatives with respect to the decision-making criteria
4. application of decision-making rules to determine the best of the selected alternatives

In the following sections (2.1-2.4), each of these steps will be described in detail.

### *2.1. Criteria*

The choice of decision-making criteria at the work center level must be guided by the need to reflect overall system objectives. Since work centers represent a relatively low level of control within a manufacturing system hierarchy, the connection between the criteria for decision making at the work center level and the overall system objectives may be difficult to ascertain. For such cases, like the ones modeled in this paper, a suggestion is made for the construction of a hierarchy of objectives that will help establish the correspondence between these objectives and the low-level decision-making criteria/attributes (Keeney and Raiffa 1976). This approach is adapted to the manufacturing environment and followed in the work presented in this paper.

In a broad sense, the objective of a manufacturing system is to produce products that will be profitable for the manufacturing organization. To maximize profit without specifying the time horizon within which this profit should be maximized, one must consider both revenue and cost; yet both revenue and cost in manufacturing systems can be affected by outside factors that are not always controllable by the manufacturing organization. It

appears, however, that between revenue and cost, manufacturing decisions are more directly related to cost, as manufacturing decisions determine levels of inventory, utilization of capacity, and a host of other factors that affect the cost of production. Based on a common cost structure, such as that presented in Niebel and Draper (1974), one can structure an objectives hierarchy that maps the relationships between organizational objectives and attributes/criteria that characterize work center performance. These attributes can be called proxy attributes/criteria, since their connection with overall organizational objectives is not necessarily a direct one. Certain standard attributes for characterizing work center performance have long been held to be important for short-term scheduling decisions (Conway, Maxwell, and Miller 1967). However, with few exceptions, most of the efforts in this area have been either to optimize one particular criterion or to incorporate all of the relevant attributes into an objective function and then to optimize this hard-wired objective function while scheduling decisions are being made (Graves 1986; Kao 1980). The approach taken in the present work is one in which the relative importance of attributes may vary and the consideration of attributes from decision to decision may change as well.

A key issue in the attribute/criteria selection process is the frequency with which the criteria should be considered. The process of selecting attributes each time a decision is required would be extremely time consuming; consequently, it is desirable to periodically select a set of criteria that will remain in effect for decision making for a certain period of time. Since the set of criteria would be selected so that it could be used for a sequence of decisions at the work center level, it is important that these criteria not be selected on the basis of transitory conditions, such as the properties of one or two tasks, but rather should be selected with respect to a typical mix of tasks over an extended period of time. On the other hand, to accommodate varying production needs, the decision-making framework should allow the relative importance of these criteria to be changed from a single, individual decision to the next. The definitions and the estimation procedures for the criteria used in this paper are presented in appendix A. Since the proposed criteria/attributes are not so extensive as to address every manufacturing situation, they should be viewed only as an illustration of how this approach can be applied under real conditions.

In the literature (Chryssolouris et al. 1986c), one can review how the attribute selection process can be conducted using a rule-based system. The rules used are extracted from an analysis of the characteristics of the various proposed attributes. In actual applications, the individuals responsible for entering rules would need to reflect the conditions peculiar to their respective situation. Actual implementation of the rule-based system consists of a backward chaining inference mechanism and separate rule bases associated with each attribute.

## 2.2. Alternatives

An alternative is defined as a set of possible assignments of available resources to pending tasks ( $R_{mi}, T_{nj}$ ). The times at which the decision-making activity occurs are referred to here as decision points, and the spacing between two subsequent decision points is referred to as a decision interval. The decision-making activity is triggered by a change of the status of the system, namely either by the completion of a new task. The decision horizon is a

time interval that begins at the decision point. The decision horizon may vary from zero—meaning no further considerations are included in the decision-making process other than the conditions present at the time of the decision—to any time that is considered critical for the decision-making process and that allows the efficient performance of this process. A long decision horizon may result in a better decision quality, but it will also require a longer time for the decision-making process to be performed, due to the larger number of factors that have to be taken into account.

There are a number of issues that have to be taken into account to address the problem of determining alternatives (Chryssolouris et al. 1985). These issues may be simplified when a number of assumptions are made. Some of the assumptions that keep the problem close to reality and, at the same time, allow the formulation of a flexible strategy for determining alternatives follow.

- There is no preemption. This is an assumption that at least for the manufacturing industry holds true due to the penalty that one would have to pay in terms of changing set ups for interrupting a task and continuing with the performance of another.
- No resource may remain idle in an alternative if there is an unassigned task which may be processed by the resource.
- Jobs/tasks arriving at the work center have process planning information: information concerning which resources are recommended for a particular task. One can further assume that there may be differences in processing time and operation cost for the different resources capable of performing a task.
- There is a stream of arrivals at the work center.
- Similar to the arrival rates, there is also a rate of failure of the different resources.
- Although decisions made at any given decision point may consider future assignments of tasks to resources, each resource is committed to only one task at a time so that the decision-making process has maximum flexibility to react to any unforeseen event (break-downs, new arrivals, etc.).

Using the previous assumptions, the procedure described next is followed to determine a set of feasible alternatives at any decision point.

### ***2.2.1. Algorithm Description***

1. Establish the maximum number of alternatives ( $X$ ) that the decision-making process can consider. This number is related directly to the computational burden and the quality of the decision-making process.
2. Determine the minimum processing time  $P_{\min}$  of all currently pending tasks  $n$  at the work center with resources  $r$ . These pending tasks are assigned to the work center based on process planning or MRP information.
3. Establish a decision horizon (DH) as an integer multiple of the  $P_{\min}$  ( $DH = h_{\text{total}} * P_{\min}$ , where  $h_{\text{total}}$  is an integer).

- 4a. Determine the number of resources ( $r_1$ ) which will become available within the time range  $[0, 1 * P_{min}]$ . The number of alternatives  $|\{AL^1\}|$  can be estimated (assuming  $n > R_1$ ) as:

$$|\{AL^1\}| = \frac{n!}{(n - r_1)!}$$

- 4b. If  $|\{AL^1\}| \geq X$ , then randomly select (without replacement)  $X$  alternatives out of  $\{AL^1\}$  and proceed with their evaluation (section 2.3); skip the remaining steps.
- 4c. If the condition in step 4b was not satisfied and if  $h_{total} = 1$ , then select all  $|\{AL^1\}|$  alternatives and proceed with their evaluation (section 2.3); skip the remaining steps. In this case, the decision horizon forces the decision-making process to consider less than the maximum allowable  $X$  alternatives—the process considers only  $|\{AL^1\}|$  alternatives.
5. Let  $h = 2$ .
- 6a. Determine the number of resources  $(r_{h_i})_{i = \{1, \dots, |\{AL^{h-1}\}|\}}$ , which will become available within the time range  $[(h - 1) * P_{min}, h * P_{min}]$  for each alternative in the set  $\{AL^{h-1}\}$ . The number of alternatives  $|\{AL^h\}|$  can be estimated as:

$$|\{AL^h\}| = \sum_{i=1}^{|\{AL^{h-1}\}|} \frac{(n - r_1 - r_{2_i}, \dots, - r_{(h-1)_i})!}{(n - r_1 - r_{2_i}, \dots, - r_{h_i})!}$$

- 6b. If  $|\{AL^h\}| \geq X$ , then, as in step 4b, randomly select (without replacement)  $X$  alternatives out of  $\{AL^h\}$  and proceed with their evaluation (section 2.3); skip the remaining step.
- 6c. If the condition in step 6b was not satisfied and if  $h_{total} = h$ , then select all  $|\{AL^h\}|$  alternatives and proceed with their evaluation (section 2.3); skip the remaining steps. In this case, as in step 4c, the decision horizon forces the decision-making process to consider less than the maximum allowable  $X$  alternatives—the process considers only  $|\{AL^h\}|$  alternatives.
- 6d. Increment  $h$  by 1. Proceed to step 6a.

Steps 6a–6c are merely repetitions of Steps 4a–4c applied to successive time ranges  $\{[1 * P_{min}, 2 * P_{min}), \dots, [(h - 1) * P_{min}, h * P_{min}]\}$ . The procedure continues until either the maximum of  $X$  alternatives is selected (as in step 4a) or the decision horizon forces an end to the selection of alternatives (as in step 4c).

The previous procedure includes, at any given decision point, an estimate of the number of possible alternatives. This estimation is based on the number of resources available within the decision horizon. This number, in turn, is influenced by the completion time of the tasks currently being processed on the various resources. Given the number of resources available within the decision horizon, one can apply combinatorial calculations to estimate the number of alternatives. These calculations, however, can be carried out only if the decision horizon is expressed in terms of number-of-tasks-per-resource and not in time units; yet the implementation of the above procedure requires the decision horizon to be expressed in time units. For this reason, the decision horizon is expressed

in time but as a multitude of the minimum task processing time (e.g.,  $3P_{\min}$ ) so that this multiplication factor (e.g., 3) can be used for the combinatorial calculations as an approximation of the number of tasks per resource that are allowed for the given decision horizon.

### 2.3. Consequences

The consequence of an alternative is defined as the values of the different criteria for this alternative. Since an alternative consists of an ordered set of resource/task pairs, the value of a criterion for an alternative is the aggregation of the values of this criterion for each individual pair ( $R_{mi}$ ,  $T_{nj}$ ). These values are estimated with the help of data such as the start/completion date of a task or parameters associated with resource performance. This data may be made available through adequate use of database concepts and data manipulation mechanisms (Chryssolouris et al. 1988).

In cases in which the decision horizon is not long enough to incorporate into an alternative the assignment of all pending tasks at the decision point, the estimation of the criteria values for an alternative can be erroneous as it may not include the consequence of leaving tasks unassigned until a later decision point. To address this issue, the following procedure has been established.

1. Order all resources of the work center according to their available times, as these times result from the particular alternative under consideration.
2. Randomly assign a pending task, from among the unassigned ones, to the first resource of the previously constructed list, and reorder the resources based on their available times so that the new assignment is reflected.
3. Repeat this process until no tasks remain unassigned for this alternative.

The procedure described above is essentially a sampling process that may be repeated a number of times. The criterion value for an alternative is the result, on one hand, of the calculation of the fixed, prespecified assignments of tasks to resources based on this alternative and, on the other hand, of an average value resulting from the randomly created samples of the unassigned tasks.

### 2.4. Decision-making Rules

Once the consequences of the different alternatives have been established, the problem of selecting the best alternative is reduced to evaluating a decision matrix. In general, the evaluation of decision matrices involves the execution of a procedure that specifies how attribute information is to be processed to arrive at a choice. Two major approaches exist for the processing of attribute information: noncompensatory and compensatory. Belonging to the category of noncompensatory models for multiple-attribute decision making are methods such as MAXIMIN and MAXIMAX, which do not permit trade-offs between attributes (Hwang and Yoon 1981). While these methods allow decision making without attribute preference information, their applicability is relatively limited due to the fact that only a small portion of the available information is used in making a decision. These methods, therefore, have not been considered in the context of this paper, although they are



simple and computationally advantageous. Compensatory models do permit trade-offs between attributes; a number is usually calculated and assigned to categorize and represent each alternative. A variety of approaches have been proposed within this category, with the most widely used method being Simple Additive Weighting (SAW). SAW has been well summarized in Hwang and Yoon (1981), and basic underlying considerations are given in MacCrimmon (1968). The SAW method offers simplicity, which is very important for the decision environment at the work center level, and, at the same time, utilizes all available information. In this paper, SAW will be used for the purpose of demonstrating the proposed framework.

A matrix created to solve a decision problem within a manufacturing environment will most likely include attribute values ( $A_{ij}$ ) with different units. To facilitate the computations required for the evaluation of the matrix, a normalization procedure, with the aim of obtaining comparable scales, is used as described in Hwang and Yoon (1981).

SAW and other multiple-attribute decision-making techniques require information about the relevant importance of each attribute. This information can be expressed in various ways, such as ordinal preference, cardinal preference, or marginal rate of substitution between attributes. Cardinal preference is usually given by a set of weights normalized to the sum of

one. In SAW, the weights assigned to indicate the relative importance of the various attributes become the coefficient of the normalized attribute values ( $a_{ij}$ ). The total score for each alternative is the sum of the products obtained by multiplying the normalized attribute values by the weight assigned to the attributes. The decision to be made is based on the total score for each alternative, with preference given to the alternative with the highest score.

In the literature (Srinivasan and Shocker 1973), a variety of techniques (e.g., the Eigenvector method) have been suggested for assessing the weights by collecting the judgement of the decision maker concerning the relative value of attributes. However, the use of any of these methods heavily involves the decision maker who, through a lengthy iteration process, assesses the relative importance of the attributes. Furthermore, an often elaborate computational procedure is applied in order to produce the final values of the weights. It therefore is questionable whether such an approach can be utilized for assessing the weights in the work center environment as information about attribute preference is normally not available within the decision-making time frame. Thus, assuming that the relative preferences (expressed in a set of normalized weights  $w$ ) among the different attributes are not known, the following procedure has been implemented.

The information required for an optimal decision is the values of the "correct" weights  $w$ . If one were to consider the maximum expected "gain" (or conversely the minimum expected "loss") as a criterion for selecting an alternative, it could be expressed as

$$\min_i \int_{\Delta} [U^*(w) - U_i(w)] f(w) dw \tag{1}$$

where  $\Delta$  is the feasible region of  $w$ -space,  $U^*$  is the utility of the most preferred alternative at a given  $w$ ,  $U_i$  is the utility of the alternative  $AL_i$  at  $w$ , and  $f(w)$  is the density function of the  $w$  probability distribution. Expression 1 can be further written as

$$\min_i \left[ \int_{\Delta} U^*(w) f(w) dw - \int U_i f(w) dw \right] \tag{2}$$

or, since the first term is constant over  $i$

$$\max_i \int_{\Delta} U_i(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} \quad (3)$$

The integral in expression 3 can be evaluated using numerical integration techniques (Hammersley and Handscomb 1964; Korobov 1957; Sag and Szekeres 1964) so that the decision-making process may proceed even if the attribute weights are unknown.

The MADEMA approach described in sections 2.1–2.4 has following major characteristics:

1. It allows a variety of criteria to be considered in a flexible fashion. In cases where simple decisions and a single criterion are most important, the concept can be condensed to the application of a simple dispatch rule.
2. The procedure for selecting alternatives follows a breadth-first logic. All alternatives that are feasible within the time range closest to the decision point—so-called close alternatives—are selected before any far alternatives—those in a subsequent time range—are selected. The emphasis on close as opposed to far alternatives is justified because unknown future task arrivals or production interruptions are more likely to render far alternatives infeasible. The evaluation at the decision point of a far alternative will not, in a dynamic scheduling situation, provide a good estimate of the alternative's eventual actual utility.
3. The performance of the approach, in terms of the quality of the generated schedule, can be tailored and optimized for any given computation time limit. This optimization can occur in two areas:
  - i. The number of alternatives selected for evaluation. This is controlled by the two parameters alternatives cap ( $X$ ) and decision horizon (DH) of the alternatives selection procedure of section 2.2.
  - ii. The accuracy of the evaluation of each selected alternative. This is controlled by the number of samples (SR, for Sampling Rate) taken in the consequences determination procedure of section 2.3. The larger the sampling rate, the more accurate but computationally expensive is the evaluation of each alternative.

This is in contrast to conventional hard-wired scheduling procedures that do not permit the explicit consideration of limits on computation time. The applicability of these procedures depends on the user's computation time limits and on problem size. This last factor, however, is difficult to determine a priori in a dynamic scheduling situation. MADEMA can be tailored to any computation time limit–problem size situation, and is thus always applicable.

In the following discussion, issues related to the implementation of this framework to an FMS scheduling and control problem will be addressed.

### 3. A Flexible Manufacturing System Model

Based upon the hierarchical approach previously defined, one can view an FMS as a job shop that consists of a number of work centers. As stated previously, a work center is defined

as a set of resources that perform similar processes. To illustrate practically how such a hierarchical structure applies to an FMS, one can consider the following example of an FMS installed in an aerospace company:

- 5 five-axis machining center, each machine having 90 tools and a tool exchange station
- 3 automated guided vehicles (AGVs) that follow a wire-guided path and are used for the delivery of both tools and workpieces to the machines.
- 1 automatically controlled washing station
- 1 CMM (Coordinate Measuring Machine)
- 2 material review stands for on-demand part inspection
- 1 tool preset and load area
- 2 ten-pallet carousels for setting up workpieces, tombstones, or any other appropriate fixture, with each carousel containing 2 load/unload stations

This example FMS is presented schematically in figure 2 and has been modeled as consisting of five work centers:

Work Center 1: Machining, contains the five machining centers

Work Center 2: Inspection, contains the CMM

Work Center 3: Review, contains the material review stands

Work Center 4: Washing, contains the automatic washing station

Work Center 5: Transportation, contains the three AGVs

The types of parts loaded in the example FMS are assumed to be independent of the FMS operation as they are determined from factors outside the system. Each part has different operations to be performed. Each part arrives at the FMS accompanied by a process plan (namely, a set of instructions that determine the sequence of the different operations as well as their technological constraints).

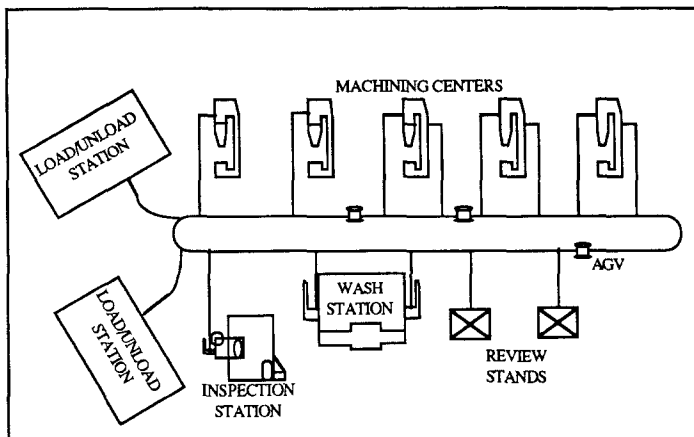


Figure 2. FMS schematic.

The operating scenario assumed is as follows:

1. Parts arrive at the loading station of the FMS; as they arrive and as a pallet is available, they are fixtured on the 2 ten-pallet carousels.
2. The parts are held on the carousels until they are picked up by an AGV and brought to any of the different work centers. Once a part leaves the loading station, it remains within the system until it is finished.
3. Upon completion of a part, an AGV will bring the part back to the loading station where it will be unloaded, removed from the station, and further processed through the rest of the manufacturing system.

The allocation of resources proceeds as follows:

1. Work Centers 1, 2, 3, and 4 (machining, CMM inspection, manual inspection, and washing) utilizing the MADEMA logic, decide which of the parts pending at the loading station or floating within the system should be assigned to these work centers.
2. Once these decisions are made, requests are issued to the transportation work center which, based on the FIFO rule, decides which AGV will move which part. These requests are modeled as transportation tasks.

The following assumptions, which correspond to the real operation of the system and which facilitate the decision-making process are made:

- Tools are available on every machine to perform any operation on any part. This assumption is made assuming that the presetting and loading of tools are made prior to the actual operation of the FMS.
- The processing times for transportation are assumed to be equal, irrespective of where the AGVs are and what part they carry.
- The parts arriving at the FMS consist of a number of operations. These parts are accompanied by a process plan which determines the operational characteristics of the different operations as well as the sequence of these operations. This process plan becomes the input to the MADEMA system. Based on these plans, MADEMA assigns operations to the different work centers.
- Resources within one work center are assumed to have the same technological characteristics; however, they may have different operating costs, and operations assigned to one resource or another may have different processing times. Nevertheless, all resources within a work center are capable of performing any operation assigned to the work center. (In parallel processing terminology, the resources within a work center are assumed to be unrelated.)

#### **4. Test Case Simulation**

Based on the hierarchical model of the FMS described above, a test case was devised that would allow the investigation of FMS behavior under a number of different operating

schemes. LISP was chosen for the implementation of the above model because of its flexibility as well as its ability to be interfaced effectively with artificial intelligence schemes. The simulation runs of the FMS were performed under the following conditions:

- There are ten different part types, distinguished from one another by a different sequence of operations and by different operations. The process plan of each part type has between one and five operations. The actual number of operations, and their sequence, used in the sample case is shown in table 1.
- Processing times for operations are determined individually using a uniform distribution and are based on the work center to which the particular operation will be assigned. However, depending upon the work center, the boundaries for processing times vary. Namely, for machining, the processing time will be a minimum of 20 minutes and a maximum of 100 minutes; for inspection, a minimum of 10 minutes and a maximum of 20 minutes; and for review and for washing, a minimum of 5 minutes and a maximum of 10 minutes (table 2). Transportation operations are assumed to have an average processing time of two minutes.
- The operating costs (per hour) for the resources in each work center were assumed to be between \$30 and \$45 for the machining work center, between \$60 and \$75 for review and inspection work centers, and \$6 for the washing and transportation work centers (table 2).
- Arrivals were specified for ten 8-hour shifts, based on a mean poisson inter-arrival rate of 0.0055 for part types 3, 5, 6, 8, and 10 and a rate of 0.00275 for part types 1, 2, 4, 7, and 9. The FMS was assumed to be loaded at the start of the simulation with a set of arrivals.
- The due dates of the different parts were assumed known and the operations due dates were based on these known part due dates.

Considering the previous assumptions, three issues were investigated with appropriate simulation runs.

1. The MADEMA concept and the resulting operational policy uses multiple criteria and, because of that it should be able to address a variety of performance measures simultaneously. Dispatch rules, on the other hand, usually are applied to focus on one particular aspect of the system and often ignore any other aspects. The point to be made here is that MADEMA, it would be expected, will perform relatively well with respect to a number of performance measures; however, a dispatch rule would be expected to perform very well with respect to one performance measure and other relevant performance measures will not be addressed. For the purpose of testing this hypothesis, three dispatch rules, the earliest part due date (EDD), the minimum part slack time (SLK), and the shortest processing time (SPT), were chosen. These three dispatch rules are representative of a large variety of dispatch rules that have been suggested in the literature and are supposed to address one particular performance measure of a system. MADEMA was tested as an operational policy using four criteria: mean tardiness, mean flowtime, mean operation cost, and capacity utilization. The definition of all the criteria in the MADEMA operational policy, as well as definitions of the different performance



measures as applied to this work, are included in appendix A. Whereas criteria are used for the decision-making process at any time a decision has to be made, performance measures are global measures that are calculated over time after a number of decisions have been made. In essence, performance measures provide a measure of the quality of the decision-making process and effectively are the parameters that one would consider from a management point of view to judge the efficiency of the system's operation.

2. The use of MADEMA as an operational policy for making individual decisions at any time an operation must be assigned to a manufacturing resource is based on the simultaneous consideration of a number of criteria which may be changed from time to time. In this way, the operational policy provides the often required flexibility in terms of addressing a particular aspect of the system for a specific period of time as needed. This flexibility can be accommodated by selecting the different decision-making criteria of the MADEMA operational policy. To investigate this point, simulation runs of the previously described FMS were performed using the criteria capacity utilization and mean operating cost and then using simply the criterion of mean tardiness (appendix A).
3. Also investigated is the issue that an operational policy for a flexible manufacturing system should not only have flexibility in terms of what criteria should be addressed, but also in terms of how much the solution should be improved. The suggested policy provides this second dimension of flexibility in terms of performance and computational effort. As mentioned in a previous section, the number of alternatives one considers when following this operational policy is related directly to the decision horizon: when an individual decision is made, how far ahead the decision-making process "looks" to determine the effect of the decision in the future. The longer the decision horizon, the better the decisions made will be and the longer they will take. This point was tested by using a flexible horizon which was varied among values of 0, 10, and 20 minutes when the MADEMA operational policy was used with the decision-making criteria flow-time and capacity utilization.

The performance measures used for the comparisons among the different operational policies were mean tardiness, mean flow time, mean wait time, mean operation cost, mean resource utilization, maximum tardiness, maximum flow time, and maximum wait time (appendix A).

## 5. Results and Discussion

Table 3 summarizes the results of the tests that compare the performance of the MADEMA operational policy using four criteria (mean tardiness, mean flowtime, mean operational cost, and capacity utilization) versus three dispatch rules (EDD, SLK, SPT). The performance measures selected are such that they provide a comprehensive view of the performance of the system under the different policies. The application of the dispatch rules, in this context, refers to the operation-selection procedure; the selected operations were randomly assigned to resources. With the MADEMA operational policy, because an alternative represents a comprehensive assignment of pending operations to available resources, there was no need to apply a resource selection rule. Each dispatch rule tends to optimize

*Table 3.* Performance measures comparing various decision-making rules.

Performance Measures upon Completion of Last Job	EDD	SLK	SPT	MADEMA
Mean T	69.00	69.00	68.00	65.00
Mean W	1.60	1.40	1.44	1.15
Mean F	110.50	110.20	110.00	106.00
Mean Util	41.20	41.00	40.90	40.20
Max T	139	137	137	129
Max W	11	17	17	17
Max F	187	187	187	177

a different aspect of the system. For example, table 3 shows that of the three dispatch rules, SPT performs best with respect to mean flow time while SLK performs best with respect to mean wait time (for the given test). It appears that MADEMA, on the other hand, performs somewhere in the middle and attempts to optimize simultaneously a number of performance measures. MADEMA may not achieve the best value that one could achieve with the application of a dispatch rule; however, it achieves a compromise value. This means that if one is certain about what aspect of the system one should optimize, one can apply a dispatch rule or, for that matter, apply MADEMA by limiting it to the application of a single criterion. However, if one is not certain about what aspect of the system is more important than another or others, or in the case that many aspects of the system in terms of performance are important simultaneously, the utilization of multiple criteria would appear to be a preferred solution.

Table 4 summarizes the results of a comparison between applying a MADEMA operational policy with two criteria—mean operation cost and capacity utilization—versus MADEMA with one criterion—tardiness. This table shows that the same decision-making framework, simply by changing the criteria one applies, can optimize entirely different aspects of the system's performance.

Table 5 summarizes the results of utilizing the MADEMA operational policy with two criteria—capacity utilization and flow time, and using different decision horizons as individual decisions are made regarding the assignment of operations to resources. An increase in the decision horizon provides a significant increase in the CPU time as the number of alternatives that are considered at any given decision-making point increases; on the other hand, an increase in the decision horizon improves the performance measures. Figure 3 schematically shows the effect of the decision horizon on the performance measure mean flow time and on the CPU time. It can be seen that it would be possible for one to tailor the decision-making framework, based on a computational price that one would like to pay and based on the performance one would like to achieve, to the needs of the particular system. This provides the often needed additional dimension of flexibility of the operational policy in terms of computational effort versus decision-making quality.

## 6. Conclusions

Based on the results presented in this paper, the proposed decision-making framework, MADEMA, seems to have three characteristics.



Table 4. Comparison of system performance.

Performance Measures After 6 Shifts	MADEMA-OC	MADEMAT
Mean T	44.08	17.69
Mean W	68.12	48.49
Mean F	170.68	151.92
Mean Cost	62.44	63.03
Mean Util	43.18	43.56
Max T	598	290
Max W	623	415
Max F	729	613

Performance Measures upon Completion of Last Job		
Mean T	29.92	12.01
Mean W	47.64	34.24
Mean F	150.89	138.61
Mean Cost	62.08	63.46
Mean Util	38.17	38.54
Max T	598	290
Max W	623	415
Max F	729	613

Table 5. The effect of changing the length of the decision horizon.

Performance Measures After 6 Shifts	DH = 0	DH = 10	DH = 20
Mean T	25.29	19.29	18.54
Mean W	52.88	46.47	45.47
Mean F	154.84	148.33	146.98
Mean Cost	61.92	61.73	61.54
Mean Util	42.92	42.88	42.72
Max T	356	393	478
Max W	504	547	600
Max F	792	743	758

Performance Measures upon Completion of Last Job			
Mean T	17.17	13.10	12.58
Mean W	37.32	33.22	32.63
Mean F	140.12	135.89	135.03
Mean Cost	61.69	61.39	61.27
Mean Util	38.00	37.95	37.84
Max T	356	393	478
Max W	504	547	600
Max F	702	743	758
Computational burden			
Mean number of alternatives	3.05	6.84	12.61
Cpu time (seconds)	6964.98	9043.4	12693.25

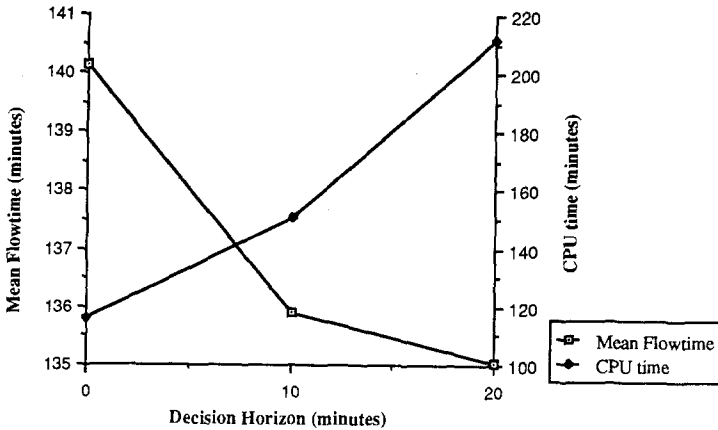


Figure 3. CPU time and mean flow time vs. decision horizon length.

1. It may provide an FMS with an operational policy that will not necessarily address one particular aspect of the system's performance, but a number of aspects. This can be attributed to the fact the MADEMA utilizes multiple criteria as it makes decisions as to the assignment of resources to operations.
2. It may provide flexibility in terms of selecting criteria so that if the aspects of the system that one would like to address are known, one may tailor the selection of criteria to the desired performance.
3. It may provide an additional dimension of flexibility in terms of the computational effort required to make assignment decisions versus the quality of these decisions.

### Acknowledgment

The research described in this paper was supported in part by Computer Aided Manufacturing-international (CAM-i).

### Note

The notation "time range [a, b]" denotes, in accordance with convention, " $a \leq \text{time} < b$ ."

### References

- Abdin, M. F., "Solution of Scheduling Problems of Job-Shop Type FMS with Alternative Machine Tools," *Computers and Industrial Engineering*, Vol. 11, Nos. 1-4, pp. 241-245 (1986).
- Ben-Arieh, D. H., Moodie, C. L., and Chi-Chung Chu, "Control Methodology for FMS," *IEEE Journal of Robotics and Automation*, Vol. 4, No. 1, pp. 53-59 (1988).
- Buzacott, J. A., and Yao, D. D., "Flexible Manufacturing Systems: A Review of Analytical Models," *Management Science*, Vol. 32, pp. 890-905 (1986).
- Chryssolouris, G., Chan, S., and Cobb, W., "Decision Making on the Factory Floor: An Integrated Approach to Process Planning and Scheduling," *Robotics and Computer-Integrated Manufacturing*, Vol. 1, Nos. 3-4, pp. 315-320 (1984).

- Chryssolouris, G., Wright, K., and Cobb, W., "Decision Making Strategy for Manufacturing Systems," *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, Tucson, Arizona, pp. 64-72 (1985).
- Chryssolouris, G., "On Decision Making in Manufacturing Systems," Japan-USA Symposium on Flexible Automation, Osaka, Japan, pp. 459-470 (1986a).
- Chryssolouris, G., Wright, K., and Cobb, W., "Manufacturing Systems Control with the Aid of Utility/Value Functions," ASME Paper 86-WA/DSC-27 (1986b).
- Chryssolouris, G., Wright, K., and Soo, P., "A Rule-Based System for Selection of Performance Criteria in Manufacturing," *Proceedings, 14th North American Manufacturing Research Conference*, Minneapolis, Minnesota, pp. 543-551 (1986c).
- Chryssolouris, G., Wright, K., Pierce, J., and Cobb, W., "Manufacturing Systems Operation: Dispatch Rules Versus Intelligent Control," *Robotics and Computer-Integrated Manufacturing*, Vol. 4, Nos. 3/4, pp. 531-544 (1988).
- Chryssolouris, G., Pierce, J., Lee, M., and Domroese, M., "Aspects of Decision Making in Resource Allocation to Manufacturing Tasks," *Working Paper*, Laboratory for Manufacturing and Productivity, M.I.T.
- Conway, R.L., Maxwell, W.L., and Miller, L.W., *Theory of Scheduling*, Addison-Wesley, Reading, MA (1967).
- Gangan, S., Khator, S.K., and Babu, A.J.C., "Multiobjective Decision Making Approach for Determining Alternate Routing in a Flexible Manufacturing System," *Computers and Industrial Engineering*, Vol. 13, Nos. 3-4, pp. 112-117 (1987).
- Graves, S.C., "A Review of Production Scheduling," *Operations Research*, vol. 29, No. 4, pp. 646-675 (1981).
- Graves, S.C., "A Tactical Planning Model for a Job Shop," *Operations Research*, Vol. 34, No. 4, pp. 522-533 (1986).
- Hammersley, J.M., and Handscomb, D.C., *Monte Carlo Methods*, Methuen, MA (1964).
- Hwang, Ching-Lai, and Yoon, Kwangson, *Multiple Attribute Decision Making*, Springer-Verlag, New York, NY (1981).
- Kao, E.P.C., "A Multiple Objective Decision Theoretic Approach to One-Machine Scheduling Problems," *Computers and Operations Research*, No. 7, pp. 251-259 (1980).
- Keeney, R., and Raiffa, H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley and Sons, New York (1976).
- Kimemia, J.G., and Gershwin, S.B., "An Algorithm for The Computer Control of Production in Flexible Manufacturing Systems," *IIE Transactions*, Vol. 15, No. 4, pp. 353-362 (1983).
- Korobov, N.M., "The Approximate Calculation of Multiple Integrals Using Number-Theoretic Methods," *Dokl. Akad. Nauk SSSR*, 115, pp. 1062-1065 (1957).
- MacCrimmon, K.R., "Decision Making Among Multiple-Attribute Alternatives: A Survey and Consolidated Approach," RAND Memorandum, RM-4823-ARPA (1968).
- Niebel, B.W., and Draper, A.B., *Product Design and Process Engineering*, McGraw-Hill, New York, NY (1974).
- O'Grady, P., and Menon, U., "Concise Review of Flexible Manufacturing Systems and FMS Literature," *Computers in Industry*, 7, pp.155-167 (1986).
- Ranky, P.G., "A Real-Time, Rule-Based FMS Operation Control Strategy in CIM Environment—Part II," *International Journal of Computer Integrated Manufacturing*, Vol. 1, No. 3, pp. 185-196 (1988).
- Sag, T.W., and Szekeres, G., "Numerical Evaluation of High-Dimensional Integrals," *Maths Comp*, 18, pp. 245-253 (1964).
- Sarin, S.C., and Dar-El, E.M., "Scheduling Parts in an FMS," *Large Scale Systems*, No. 11, pp. 83-93 (1986).
- Sauve, B., and Collinot, A., "An Expert System for Scheduling in a Flexible Manufacturing System," *Robotics and Computer-Integrated Manufacturing*, Vol. 3, No. 2, pp. 229-233 (1987).
- Solberg, J.J., "A Mathematical Model of Computerized Manufacturing Systems," *Proceedings, 4th International Conference on Production Research*, CIRP, Tokyo, Japan, pp. 1265-1275 (1977).
- Srinivasan, V., and Shoemaker, A.D., "Linear Programming Techniques for Multidimensional Analysis of Preference," *Psychometrika*, Vol. 38, No. 3, pp. 337-369 (1973).
- Stecke, K.E., and Solberg, J.J., "The Optimality of Unbalancing Both Workloads and Machine Group Sizes in Closed Queueing Networks of Multiserver Queues," *Operations Research*, Vol. 33, No. 4, pp. 882-910 (1985).
- Suri, R., "New Techniques for Modelling an Control of Flexible Automated Manufacturing Systems," *Proceedings of the 8th Triennial World Congress on Control Science and Technology*, IFAC, Kyoto, Japan, pp. 2027-2033 (1981).
- Tabe, T., and Salvendy, G., "Toward a Hybrid Intelligent System for Scheduling and Researching of FMS," *International Journal of Computer Integrated Manufacturing*, Vol. 1, No. 3, pp. 154-164 (1988).

## Appendix A. Definitions of Criteria and Performance Measures

### 1. MADEMA Criteria

a. To define the criteria and describe how the MADEMA approach selects an alternative, the following definitions and notation are used:

- $R_j, j \in \{1, \dots, J\}$  = the  $j$ th resource on the work center
- $R_j^{\text{rate}}$  = cost per unit time for operation  $R_j$
- $\text{Task}_i$ , i.e.,  $\{1, \dots, I\}$  = the  $i$ th task which is both present at the work center and unassigned to any resource at the beginning of the decision-making process
- $\text{alt}_q, q \in \{1, \dots, Q\}$  = the  $q$ th alternative generated when the work center is required to make a decision
- $T^{\text{cur}}$  = the current time
- $T_i^{\text{arr}}$  = the time at which  $\text{Task}_i$  arrived at the work center
- $T_i^{\text{dd}}$  = the due date of  $\text{Task}_i$
- $t_{ij}^{\text{proc}}$  = the processing time of  $\text{Task}_i$  of  $R_j$
- $t_i^{\text{proc}}(\text{alt}_q)$  = the estimated time required to process  $\text{Task}_i$  if  $\text{alt}_q$  is implemented  
     =  $t_{ij}^{\text{proc}}$  if  $\text{alt}_q$  assigns  $\text{Task}_i$  to  $R_j$ , otherwise  
     = the average of  $t_{ij}^{\text{proc}}$ , taken over all  $R_j$  capable of processing  $\text{Task}_i$
- $T_j^{\text{avail}}$  = the time at which  $R_j$  will become available, which will occur when it finishes processing all task orders assigned to it from past decisions
- $T_j^{\text{avail}}(\text{alt}_q)$  = the time at which  $R_j$  will finish processing any task orders assigned to it from both past decisions and from  $\text{alt}_q$
- $T_i^{\text{start}}(\text{alt}_q)$  = the expected "begin processing" time of  $\text{Task}_i$  if  $\text{alt}_q$  is selected
- $T_i^{\text{comp}}(\text{alt}_q)$  = the expected completion time of  $\text{Task}_i$  if  $\text{alt}_q$  is selected  
     =  $T_i^{\text{start}}(\text{alt}_q) + t_i^{\text{proc}}(\text{alt}_q)$

b. Given the above definitions, the MADEMA criteria are estimated as follows. ( $L$  = number of pending tasks in the work center at the decision time):

#### (1) Mean tardiness

$$\text{TARD}(\text{alt}_q) = \frac{\sum_{i=1}^L \text{Max}[0; (T_i^{\text{comp}}(\text{alt}_q) - T_i^{\text{dd}})]}{L}$$

#### (2) Mean flow time

$$\text{FLOW}(\text{alt}_q) = \frac{\sum_{i=1}^L [T_i^{\text{comp}} - T_i^{\text{arr}}]}{L}$$

**(3) Mean operation cost**

$$\text{COST}(\text{alt}_q) = \frac{\sum_{i=1}^L [t_i^{\text{proc}}(\text{alt}_q) * R_j^{\text{rate}}]}{L}$$

**(4) Capacity**

$$\text{CAP}(\text{alt}_q) = \sum_{i=1}^L t_i^{\text{proc}}(\text{alt}_q)$$

**2. MADEMA Performance Measures**

a. For performance measures calculated at time T, the previous definitions are modified/ extended as follows:

- Job<sub>k</sub>, k ∈ {1, . . . , K} = the kth job of the set of K jobs which have arrived at the job shop on or before time T.
- I<sub>k</sub> is the number of tasks in Job<sub>k</sub>.
- TJ<sub>k</sub><sup>arr</sup> = the time at which Job<sub>k</sub> arrived  
 TJ<sub>k</sub><sup>start</sup> = the time at which Job<sub>k</sub> was started  
 TJ<sub>k</sub><sup>comp</sup> = the time at which Job<sub>k</sub> was completed  
 TJ<sub>k</sub><sup>dd</sup> = the due date of Job<sub>k</sub>

b. Given the above definitions, the MADEMA performance measures are calculated as follows:

**(1) Mean tardiness**

$$\bar{T} = \frac{1}{K} \sum_{k=1}^K \begin{cases} \text{Max}[0; \text{TJ}_k^{\text{comp}} - \text{TJ}_k^{\text{dd}}], & \text{for completed jobs} \\ \text{Max}[0; \text{T}^{\text{cur}} - \text{TJ}_k^{\text{dd}}], & \text{for incomplete jobs} \end{cases}$$

**(2) Mean wait time**

$$\bar{W} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{I_k} \begin{cases} 0, & \text{for tasks that have not yet been released} \\ \text{T}^{\text{cur}} - \text{T}_i^{\text{arr}}, & \text{for tasks that are released but not started} \\ \text{T}_i^{\text{start}} - \text{T}_i^{\text{arr}}, & \text{for tasks that have been started} \end{cases}$$

**(3) Mean flow time**

$$\bar{F} = \frac{1}{K} \sum_{k=1}^K \begin{cases} \text{TJ}_k^{\text{comp}} - \text{TJ}_k^{\text{arr}}, & \text{for completed jobs} \\ \text{T}^{\text{cur}} - \text{TJ}_k^{\text{arr}}, & \text{for incomplete jobs} \end{cases}$$

**(4) Mean operation cost**

$$\bar{\$} = \frac{1}{K_{\text{comp}}} \sum_{k=1}^{K_{\text{comp}}} \sum_{i=1}^{I_k} (T_i^{\text{comp}} - T_i^{\text{start}}) * R_j^{\text{rate}}$$

where  $K_{\text{comp}}$  is the number of completed jobs in the job shop.

**(5) Mean utilization**

$$\bar{\%} = \frac{1}{J_{\text{tot}}} \sum_{j=1}^{J_{\text{tot}}} 100 * \left( \frac{R_j^{\text{busy}}}{R_j^{\text{busy}} + R_j^{\text{idle}}} \right)$$

where  $J_{\text{tot}}$  is the total number of resources in the job shop, and  $R_j^{\text{busy}}$  and  $R_j^{\text{idle}}$  are the total busy and idle times of  $R_j$ , respectively.

**(6) Maximum tardiness**

$$T_{\text{max}} = \text{Max} \left\{ \begin{array}{ll} \text{Max}[0; TJ_k^{\text{comp}} - TJ_k^{\text{dd}}], & \text{for completed jobs} \\ \text{Max}[0; T^{\text{cur}} - TJ_k^{\text{dd}}], & \text{for incomplete jobs} \end{array} \right\}$$

**(7) Maximum wait time**

$$W_{\text{max}} = \text{Max} \sum_{i=1}^{I_k} \left\{ \begin{array}{ll} 0, & \text{for tasks that have not been released} \\ T^{\text{cur}} - T_i^{\text{cur}} - T_i^{\text{arr}}, & \text{for tasks that have been released but not started} \\ T_i^{\text{start}} - T_i^{\text{arr}}, & \text{for tasks that have been started} \end{array} \right\}$$

**(8) Maximum flow time**

$$F_{\text{max}} = \text{Max} \left\{ \begin{array}{ll} TJ_k^{\text{comp}} - TJ_k^{\text{arr}}, & \text{for completed jobs} \\ T^{\text{cur}} - TJ_k^{\text{arr}}, & \text{for incomplete jobs} \end{array} \right\}$$