Revised August 1983

LIDS-P-1134 Revised

# AN ALGORITHM FOR THE COMPUTER CONTROL OF PRODUCTION IN A FLEXIBLE MANUFACTURING SYSTEM

by

Joseph G. Kimemia
Stanley B. Gershwin

This research was carried out in the M. I. T. Laboratory for Information and Decision Systems with support extended by the National Science Foundation Grant DAR78-17826.

# AN ALGORITHM FOR THE COMPUTER CONTROL OF A FLEXIBLE MANUFACTURING SYSTEM

by

Joseph Kimemia
AT&T Information Systems Laboratories, NP 2D-108
2220 Highway 66, Neptune, NJ 07753.

and

Stanley B. Gershwin
Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology,
35-310, 77 Mass Ave, Cambridge, MA 02139.

## ABSTRACT

The problem of production management for an automated manufacturing system is described. The system consists of machines that can perform a variety of tasks on a family of parts. The machines are unreliable, and the main difficulty the control system faces is to meet production requirements while the machines fail and are repaired at random times. A multi-level hierarchical control algorithm is proposed which involves a stochastic optimal control problem at the first level. Optimal production policies are characterized and a computational scheme is described.

**Acknowledgement**

This research was done in the Laboratory for Information and Decision Systems of the Massachusetts Institute of Technology, under NSF grant numbers DAR 78-17826 and ECS 79-20834

## 1. INTRODUCTION

A flexible manufacturing system (FMS) consists of a set of workstations, capable of performing a number of different operations, interconnected by a transportation mechanism. The FMS produces a family of parts related by similar operational requirements or by belonging to the same final assembly. All parts in the part family are produced simultaneously. Workpieces are introduced into the system at a loading station, and leave at an unloading station after undergoing a specified sequence of operations. The machines in an FMS are capable of performing operations on a random sequence of parts with negligible change over time from one part to the next. The flexibility of the system allows the parts the choice of one or more stations for each operation. This allows production to continue even when a workstation is out of service because of failure or maintenance.

The change over-time for parts in the same family is negligible in these systems because the machines are numerically controlled with a large number of tools used for each operation or because operations are performed by robots. In the first case, several tools must be selected and replaced for each operation so that no time is saved by working on two successive parts of the same type on the same machine. In both cases, software determines the operations, and changing the software can be done nearly instantaneously compared to operation times.

The FMS concept is not limited to metal cutting, such systems can also be used for the assembly of printed circuit boards, integrated circuit fabrication and automobile assembly lines. A survey of Flexible Manufacturing Systems appears in Dupont-Gatelmand [3].

The ability of an FMS to produce a family of parts simultaneously results in reduced finished and in-process parts inventories, and faster responses to changes in demand requirements when compared to traditional production methods. However, careful attention must be paid to production scheduling. The high capital cost of an FMS means that efficient use of system resources is very important.

In the majority of implementations, FMS's are part of a multi-stage manufacturing system. The input consists of parts that have undergone one or more processing stages and the finished family of parts are assembled into different final products.

Most manufacturing systems are large and complex. It is natural therefore, to divide the control or management into a hierarchy consisting of a number of different levels. Each level is characterized by the length of the planning horizon and the kind of data required for the decision making process. Higher levels of the hierarchy typically have long horizons and use highly aggregated data, while lower levels have shorter horizons and use more detailed information. The nature of uncertainties at each level of control also varies.

The managers of a manufacturing firm make production plans for finished products by considering forecasts of demand, sales, raw material availability, inventory levels and plant capacity. From the resulting master plan, the requirements for the components that go into the final products can be made. The various departments responsible for the manufacture of the components schedule their activities so as to meet the requirements dictated by the master production and the materials requirements plans [4][7].

In an FMS, the operations at the workstations and the material handling system are entirely under computer control. Decisions such as which parts should be loaded into the system and what workstations particular workpieces should visit next are taken by the FMS control computer. Human intervention is necessary only when unusual or unanticipated events take place. It is important therefore, to develop models and algorithms which allow the FMS controller to generate production schedules which satisfy demand requirements and to exercise control over the system so that the output conforms to the schedule. The task of the controller is complicated by random failures of the workstations. A good production policy should anticipate failures and demand

changes if it is to satisfy all of the objectives stated above.

It is important that this policy employ feedback so as to respond to failures and to allow human operators (who can deal with a wider range of situations than envisioned by system planners) to override control decisions on rare occasions.

In systems currently operating, a number of operating policies are employed. Typically, parts are loaded into the system whenever an opportunity arises, and planned production affects few decisions internal to the FMS [9]. Such policies do not typically consider capacity or reliability and can lead to congestion and under utilization.

Hutchinson [8] describes information and algorithm structures which have been implemented on an actual system. It is to be noted that a high degree of human intervention is necessary because of machine failures, human errors, maintenance and changes in operating environment. In this paper, an FMS control policy is described. Parts are loaded into the system in a way that will not overload the system or cause congestion and yet will meet long term production objectives. Because of the problem's complexity, the policy is organized hierarchically.

## 2. HIERARCHICAL SCHEME FOR THE OPERATIONAL CONTROL OF AN FMS

A four-level control structure specifically designed to compensate for workstation failures and changes in part requirements is proposed. The hierarchy is illustrated in Figure 1, in which the FMS controller is imbedded into the larger hierarchy of production management. The objective of the FMS controller is to satisfy a known, possibly time-varying demand for a family of items that is dictated by the Master Production Plan, subject to constraints imposed by the resources available.

The routing and scheduling policy described here is based on a set of assumptions on the time scales of various classes of events that occur in the operation of the flexible manufacturing system.

(i) The shortest time period is that of the set up when switching among the family of operations for which the machine is configured. It is assumed that these times are short compared to the following times and they may be ignored.

(ii) The next time period is that of the typical operation. If operation times are random, then we refer to the mean of the distributions. Operation times are assumed to be orders of magnitude larger than set-up times.

(iii) The next time period is that between machine failures or repairs. Again, mean times between failures and to repair (MTBF and MTTR) are considered.

(iv) The longest time period is the planning horizon for the problem under consideration. It is assumed that demand can be specified for a time period larger than the typical MTBF or MTTR. At this time period, the machines may be reconfigured for another part family.

### 2.1 The Flow Control Level (Calculates Short Term Production Rates)

The flow control level of FMS control determines the short term production rates of each member of the part family. The rates must be determined jointly because the parts share the time available at the workstations. In addition, the demand, the level of downstream buffer levels and the reliability of the workstations must be taken into account.

The mix of parts being produced must be adjusted continuously so as to take into account random failures of the workstations. If, for example, a part cannot be made because a certain workstation

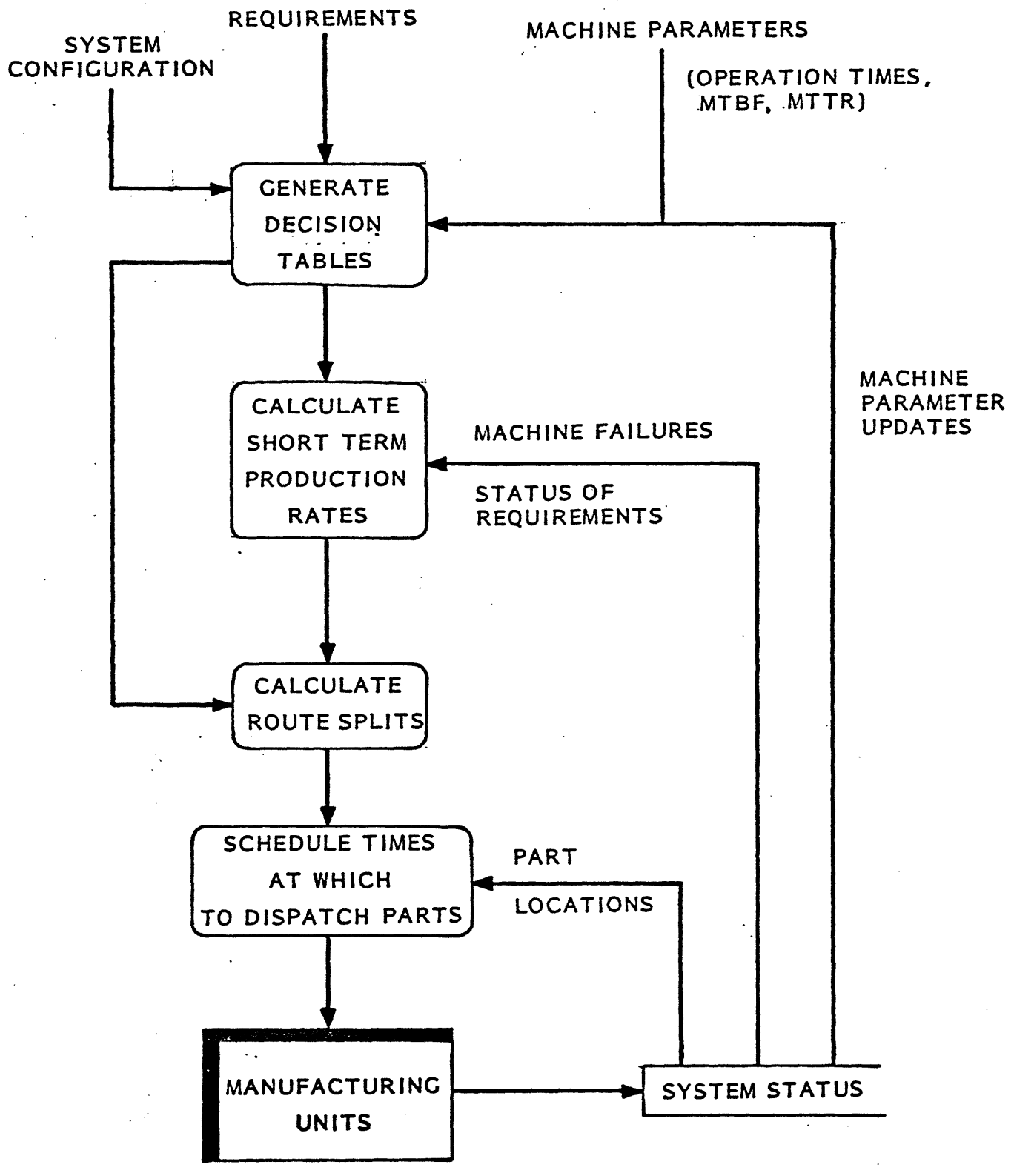



**Figure 1.** The Hierarchical Production Control Scheme

has failed, the lost production must be made up when the station is repaired. Using failure and repair statistics of the machines, the production rates should be chosen in a way that anticipates

station down-time. Adequate but not excessive downstream buffer levels should be maintained so as to satisfy downstream demand.

### 2.2 The Routing Control Level (Calculates Route Splits)

A part entering the FMS has one or more paths it can take through the system in order to complete its processing requirements. The proportion of parts that should follow each of the available paths is chosen by the route control level of the controller. The objective is to meet the production rate dictated by the flow controller while minimizing congestion and delay within the system.

The system can be modelled as a network of queues with the stations represented as the service nodes. The arrival rate of the parts is determined by the flow control level. The flow rate on each path can then be determined by a mathematical programming technique. Alternatively, it can be calculated together with total part flow as described below.

### 2.3 The Sequence Controller (Schedules Times at which to Dispatch Parts)

At the lowest level of control are scheduling algorithms that dispatch parts into the system and supervise the operations of the workstations. The objective is to maintain the flow rates chosen by the flow and route controllers.

We suggest a simple method which uses the flow rates calculated by the route controller to determine time intervals between loading parts on each path. Simulation results show that production rates and workstation utilizations determined by the flow and routing levels of the controller can be achieved provided they are feasible.

### 2.4 Generation of Decision Tables

At the highest level of the control scheme is the off-line calculation of the control policies to be used in the flow and routing levels. In principle, this is required only when a new schedule is established. In practice, it may be prudent to include a long term feedback loop to compile data on failure and repair rates as well as other parameters that may not be well known.

Whenever new estimates of parameters are found that are significantly different from earlier values, the calculation of control policies should be redone. While the calculation is being performed, production can continue using the previous control policy.

## 3. COMPARISON WITH OTHER WORK

The hierarchical approach to FMS planning and control has been suggested by a number of authors. Hildebrant [6] examines a three-level hierarchy that minimizes the time to produce a given number of parts. The top level calculates steady state production rates for each failure condition. Inventory levels are not considered and a change in production requirements means that the production rates must be re-computed. The second and third levels determine loading schedules for the parts. Olsder and Suri [11] use a dynamic programming formulation for the minimum time production problem. In this case, a feedback policy results which depends on the current failure state and production levels.

Hahne [5] and Tsitsiklis [13] study the problem of maximizing throughput in a system in which parts can be routed from an upstream machine to one of two unreliable downstream machines. They show that optimal policies are piece-wise constant functions of intermediate buffer levels. Calculation of exact optimal policies for the three machine system has large computational

requirements.

The hierarchical controller described here utilizes currently available capacity while anticipating workstation failures, repairs and changes in demand requirements. It differs from Hildebrant's [6] scheme in that flow rate decisions are made on the basis of the current inventory levels as well as the current set of working machines. Part types that are backlogged will tend to be favored over part types for which a surplus exists. Therefore, this control scheme satisfies requirements for a part family without the need for large finished and in-process parts inventories.

Buzacott and Yao [2] present a survey of other research in the modeling and control of flexible manufacturing systems. A survey of routing policies within the FMS is given by Buzacott [1].

## 4. THE FLOW CONTROL LEVEL OF THE FMS CONTROLLER

### 4.1 Problem Formulation

The flow control level of the FMS controller determines the production rates for the part family. The horizon is set by the FMS management and is of the order of one period of the master production plan. The routing and the sequencing levels ensure that the output of the system is the same as that set by the flow controller. For the lower levels of the hierarchy to be able to track the rates set by the flow controller, the rates must be at all times feasible for the current system configuration. It is important therefore for the flow controller to have complete and timely information of the operational status of all workstations. In addition, knowledge of the finished parts inventory is needed.

The FMS consists of M workstations. Workstation $m$ (m = 1, 2, ..., M) has $L_m$ identical machines. The concept of workstation is logical, not physical: the machines in a workstation need not be located closer to one another than other machines.

A family of N part types is produced. The material flow is modelled as a continuous process. This kind of model ignores combinatorial details which are treated at the lower levels of control. Its accuracy is adequate for the time horizon treated at the flow control level which is long compared to the time needed to produce individual parts. Let $u(t) \epsilon R^N$ be the production rate for the part family, the control variable. The downstream demand rate is $d(t) \epsilon R^N$ and is known in the interval $(0, t_f)$. Finished parts are stored in downstream buffers from where the downstream demand is satisfied. Define $x(t) \epsilon R^N$ by the following differential equation;

$$\frac{dx(t)}{dt} = u(t) - d(t) \tag{1}$$

The vector $x(t)$, termed the buffer state, measures the cumulative difference between production and demand for the parts. A negative value for a component of $x(t)$ gives the backlogged demand for the corresponding part. A positive value is the size of the inventory stored in the downstream buffers. Ideally, parts in an FMS are produced as they are required, keeping the buffer state close to zero.

The state of the workstations is called the machine state and is denoted by an M-tuple of integer variables $\alpha(t)$ with the component $\alpha_m(t)$ equal to the number of operational machines at station m.

Given that a machine at station m is operational, the probability of a failure in an interval of length $\delta t$ is $p_m \delta t$. The probability that a failed machine is repaired during the $\delta t$ time interval is given by $r_m \delta t$. The parameters $p_m$ and $r_m$ are the failure and repair rates for the machines at

station m. The dynamics of the machine state are therefore governed by

$$P(\alpha_m(t+\delta t)=l+1|\alpha(t)=l)=\begin{cases}(L_m-l)r_m\delta t & \text{for } 0\leqslant l<L_m\\ 0 & \textit{otherwise}\end{cases} \tag{2}$$

$$P(\alpha_m(t+\delta t)=l-1|\alpha(t)=l)=\begin{cases}L_m p_m\delta t & \text{for } 0<l\leqslant L_m\\ 0 & \textit{otherwise}\end{cases} \tag{3}$$

Note that

$$P(\alpha_m(t+\delta t)=l_1|\alpha_m(t)=l_2)=0 \text{ if } |l_1-l_2|>1$$

For two machine states i and j, it is convenient to define

$$\lambda_{ij}\delta t = P(\alpha(t+\delta t)=j|\alpha(t)=i) \quad \text{for} i\neq j$$

and

$$\lambda_{ii}=1-\sum_j\lambda_{ij}$$

The times between failures and to repair are thus modelled by exponentially distributed random variables with means $1/p_m$ and $1/r_m$ respectively. The machine state can be modelled by an irreducible Markov chain with a finite number of states. Each state communicates with M neighbors and transitions are due to the failure or repair of a single machine. The model assumes that machine failure rates do not depend on the part flow rate through the workstations. Where reliability does depends on the part flow rate, the failure rate then becomes a function of the production rate and part routing through the FMS.

Failure and repair rates are assumed to be independent of production rates and the number of operational machines for computational and expository convenience. It is easy to extend the model and the control method of this paper to include these effects.

Exponentially distributed times between failures are suitable where machine downtime is caused by the random failure of any one of a large number of components. The exponential model is also consistent with reported field data [12]. Non exponential distributions can be used to model failure rates that depend on the time since the last repair. However, in a practical implementation, the estimation of time-dependent failure and repair rates may prove to be difficult whereas the mean time between failures and to repair may be readily available.

The choice of the production rate is not arbitrary. The production rate at each instant is limited by the capacity of the currently operational machines. At time t, the production rate must lie in a set $\Omega(\alpha(t))$ which depends on the machine state and is thus subject to sudden changes.

To define $\Omega(\alpha)$, consider the machine state $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_M)$. Let $y^k_{mn}$ be the rate at which station m performs operation k on type n parts ( measured in parts per unit time interval). Let $\tau^k_{nm}$ be the time required to complete the operation. It then follows that

$$\sum_n\sum_k y^k_{nm}\tau^k_{nm}\leqslant\alpha_m \quad \text{for } \textit{all } m \tag{4}$$

The product $y^k_{nm}\tau^k_{nm}$ is the proportion of each unit time interval used by one or more operational machines at station m to perform operation k on type n parts. The left hand side of equation (4) is thus the total amount of work brought to station m per unit time by the part flow rate $y^k_{nm}$. The inequality follows because the amount of work brought to station m per unit time interval cannot exceed the time available at the operational machines.

Since no material is accumulated within the system, the total number of type n parts going through operation k per unit time interval is equal to the throughput of type n parts. This is expressed as

$$u_n = \sum_m^M y^k_{nm} \quad \text{for } \textit{all } k \text{ and } n \tag{5}$$

The set $\Omega(\alpha)$ is defined to be the set of all production rates $u = (u_1, u_2, \ldots, u_N)$ such that there exists feasible flow rates $y^k_{nm}$ satisfying (4) and (5). We note that $\Omega(\alpha)$ is the projection of a polyhedral set into a lower dimensional subspace. The feasible set is therefore convex and polyhedral.

This set is thus a representation of the capacity of the FMS. It would not be precise to denote capacity by a single number ( a production rate for all the parts flowing through the system) or even a vector ( a production rate for each part type). A set is required because of the sharing of resources among part types. The rate at which it is possible to manufacture one part is reduced by the production of other parts.

The flow control problem can now be stated. Given an FMS as described above, an initial buffer state $x(t_o)$ and machine state $\alpha(t_o)$, we wish to specify a production plan for $t_o \leqslant t \leqslant t_f$ that minimizes the performance index

$$J(x,\alpha,t_o) = E\left\{\int_{t_o}^{t_f} g(x(t))\,dt \,|\, x(t_o)=x, \alpha(t_o)=\alpha\right\} \tag{6}$$

Subject to (1), (2), (3) and $u(t)\epsilon\Omega(\alpha(t))$. The function g(x(t)) penalizes the controller for failing to meet demand and for keeping an inventory of parts in the downstream buffers. The performance index J(x, $\alpha$,$t_o$) is thus the expected total penalty incurred by the controller in the interval $(t_o,t_f)$. The function g(x) is given by $\sum_n g_n(x_n)$ where $g_n(x_n)$ are scalar convex functions satisfying

$$\lim_{|x|\to\infty} g_n(x) = \infty$$

and $g_n(0)=0$.

The cost function serves to enforce desired behavior on the controller. The ideal production policy would minimize the performance index by producing parts at exactly the demand rate thereby keeping the buffer state at zero. Such a policy is impossible because of the failures of the machines.

The class of production policies to be considered consists of functions u(x, $\alpha$, t) that satisfy for each x, $\alpha$ and t

$$u(x,\alpha,t)\epsilon\Omega(\alpha) \tag{7}$$

The production policies are therefore feedback control laws which give a feasible production rate for each buffer and machine state in the interval $(t_o,t_f)$.

**4.2 Characterization of Optimal Production Policies**

Define the cost-to-go, when the production policy u(x, $\alpha$, t), is applied as

$$J_u(x,\alpha,t)=E\left\{\int_t^{t_f} g(x(s))ds \,|\, x(t)=x,\alpha(t)=\alpha\right\} \tag{8}$$

The cost to go is thus the expected total penalty incurred by the controller for the remaining time given that the buffer and machine states are x and $\alpha$ at time t.

This function satisfies a partial differential equation. It can be derived informally by noting that, for any $\delta t>0$,

$$J_u(x(t),\alpha(t),t)=E\left\{\int_t^{t+\delta t} g(x(s))ds+J_u(x(t+\delta t),\alpha(t+\delta t),t+\delta t)\right\} \tag{9}$$

For small $\delta t$, this becomes, approximately,

$$J_u(x(t),\alpha(t),t)=g(x(t))\delta t+\sum_{\beta\neq\alpha(t)}\lambda_{\alpha\beta}\delta t J_u(x(t+\delta t),\beta,t+\delta t)+ \tag{10}$$

$$(1-\lambda_{\alpha(t)\alpha(t)}\delta t)[J_u(x(t),\alpha(t),t)+\frac{\partial J_u(x(t),\alpha(t),t)}{\partial x}\dot{x}\delta t+\frac{\partial J_u(x(t),\alpha(t),t)}{\partial t}\delta t]$$

where the derivatives of $J_u$ are evaluated at x(t), $\alpha(t)$ and t.

By letting $\delta t$ go to zero and re-arranging terms, this becomes

$$0=g(x(t))+\frac{\partial J_u}{\partial x}(u-d)+\frac{\partial J_u}{\partial t}+\sum_\beta\lambda_{\alpha\beta}J_u(x(t),\beta,t). \tag{11}$$

It is also possible to show that an optimal feedback control law $u^*(x,\alpha,t)$ and the optimal cost-to-go $J_{u^*}(x,\alpha,t)$ satisfy

$$0=\min_{u\in\Omega(\alpha)}\left\{g(x(t))+\frac{\partial J_{u^*}}{\partial x}(u-d)+\frac{\partial J_{u^*}}{\partial t}+\sum_\beta\lambda_{\alpha\beta}J_{u^*}(x(t),\beta,t)\right\} \tag{12}$$

Note that a control $u^*$ is determined, in this equation, by

$$\min_{u\in\Omega(\alpha)}\ \frac{\partial J_{u^*}}{\partial x}\,u \tag{13}$$

These results can be established formally by the techniques of Rishel [10] and Tsitsiklis [13].

We note that (13) is linear in u and that $\Omega(\alpha)$ is a convex polyhedral set. An optimal policy $u^*(x,\alpha,t)$ therefore takes values at extreme points of $\Omega(\alpha)$ whenever the gradient $\frac{\partial}{\partial x}J_{u^*}(x,\alpha,t)$ exists. For each machine state $\alpha$, an optimal policy divides the buffer state space into a set of regions in which the production rate is constant. Whenever the buffer state is in one of these regions, optimal production rates are constant. However, the regions do not cover the whole space. If the derivative $\frac{\partial}{\partial x}J_{u^*}(x,\alpha,t)$ does not exist, is orthogonal to a face of $\Omega(\alpha)$ or is zero, a unique minimizing value to (13) does not exist. The optimal production rate in that case depends on the extreme point policies in the neighboring regions.

In the following, we consider the time-invariant case, in which $d(t)=d$, a constant, and the final time t is infinite. In that case, the criterion is the average cost, rather than the total cost over the period $[t_o,t_f]$. As a result the cost-to-go function $J_u(x(t),\alpha(t),t)$ is time-invariant and is written $J_u(x,\alpha)$ and is called the value function.

There are two kinds of machine states: those for which the demand rate d is feasible, ie. those for which $d \epsilon \Omega(\alpha)$ and those for which it is not. The former machine states called feasible states, each has an associated fixed buffer. We call this buffer level $x_H^{\alpha}$, the hedging point. The hedging point is so designated because if $\alpha(t)$ remains constant for a long enough period and $\alpha$ is a feasible state then

$$\frac{d}{dt} J_{u^*}(x(t),\alpha(t)) = \frac{\partial J_{u^*}}{\partial x}(u^* - d) \tag{14}$$

since $u^*$ minimizes $(\frac{\partial J_{u^*}}{\partial x})u$ for all $u \epsilon \Omega(\alpha)$ and since the demand rate satisfies $d \epsilon \Omega(\alpha)$ then $(d/dt)J_{u^*}(x(t),\alpha(t))$ is negative and $J_{u^*}(x(t),\alpha(t))$ is a decreasing quantity. On the other hand, $J_{u^*}$ is the integral of a positive quantity, so it can not decrease without reaching a limit.

The limit is reached when x(t) is equal to $x_{\alpha}^{H}$ ( assuming that the machine state is constant for a sufficiently long time). After that time, $u^*$ is set equal to d and the buffer state x(t) stays constant at the hedging value.

The hedging point, which is the minimum of $J_{u^*}(x,\alpha,t)$ with respect to x, is the optimal buffer level with which to hedge against future failures. When demand is close to the capacity of the system, the hedging points are at high buffer levels because failures quickly result in deficits and recovery from a deficit is slow. The gradient $\frac{\partial}{\partial x} J_{u^*}(x,\alpha,t)$ can be regarded as a weighting on part production for an optimal control law defined by (13). The calculation of the optimal value function $J_{u^*}$ takes into account the relative costs of backlogs and inventory storage determined by the functions $g_n(x)$. Thus a part that has a high value index and is at the same time sensitive to machine failures would have correspondingly a large weighting. The exact solution to the flow control problem requires the solution of a coupled set of differential equations (12). This is only possible for small problems. Typical flexible manufacturing systems may have ten workstations and half a dozen different part types [9]. To solve a problem of that size, a practical computational method is required.

## 5. AN ESTIMATE-BASED (EB) CONTROL SCHEME

### 5.1 The Approach

The optimal policy in the flow control problem is determined from the optimal value function $J_{u^*}(x,\alpha,t)$ by the linear program (13). An optimal policy is a feedback law which for every machine state divides the buffer state space into regions within which the control is constant at an extreme point of the control constraint set.

The hedging point $x_{\alpha}^{H}$ ,which is the minimum of $J_{u^*}(x,\alpha,t)$ with respect to x, is the optimal buffer level with which to hedge against future failures.

Optimal policies cannot be computed in practice because of the large dimension of the flow control problem. We need a practical method for calculating sub-optimal control laws which produces good results when used in the flow control level of the hierarchy.

Given convex functions $\psi(x,\alpha,t)$ which are estimates of the optimal value function, consider a control policy $\hat{u}(x,\alpha,t)$ determined by

$$\hat{u}(x,\alpha,t) = \underset{z\in\Omega(\alpha)}{argmin}\left[\frac{\partial}{\partial x}\psi(x,\alpha,t)\right]z \tag{15}$$

The sub-optimal policy $\hat{u}(x,\alpha,t)$ like an optimal policy divides the buffer state space into a set of regions in each of which it takes values at an extreme point of $\Omega(\alpha)$.

The estimates $\psi(x,\alpha,t)$ should exhibit the properties of the optimal value function described above. The value of the estimate should be largest for machine states with the smallest production capacity. The relative magnitudes of the components of the gradient $\frac{\partial}{\partial x}\psi$ should reflect both the relative value of parts and their vulnerability to machine failures. The minimum with respect to x of $\psi(x,\alpha,t)$, which determines the hedging point for the sub-optimal policy, should be of a magnitude comparable to the optimal hedging buffer levels. If the estimates satisfy these criteria, we expect the sub-optimal policy to perform well and to meet demand requirements when they are close to system capacity. If the optimal value of the cost index is not sensitive to the location of the region boundaries, the cost $J_{\hat{u}}(x,\alpha,t)$ corresponding to the estimate based control policy should be close to the optimal cost.

### 5.2 Calculation of the Estimates

The control constraint set $\Omega(\alpha)$ is polyhedral, lies in the positive orthant and contains the origin. Define $\overline{H}(\alpha)$ and $\underline{H}(\alpha)$ to be sets such that

$$\overline{H}(\alpha) = \left\{u\in R^N \mid 0\leqslant u_n \leqslant \bar{q}_{\alpha n}\right\} \quad n = 1,2,\ldots,N \tag{16}$$

$$\underline{H}(\alpha) = \left\{u\in R^N \mid 0\leqslant u_n \leqslant \underline{q}_{\alpha n}\right\} \quad n = 1,2,\ldots,N \tag{17}$$

and

$$\underline{H}(\alpha) \subseteq \Omega(\alpha) \subseteq \overline{H}(\alpha) \tag{18}$$

$\underline{H}(\alpha)$ and $\overline{H}(\alpha)$ are hypercubes, the former contained in $\Omega(\alpha)$ and the latter containing the control constraint set. For example, Figure 2 shows the hypercubes for a sample control constraint set.

Define $\underline{\psi}(x,\alpha,t)$ and $\overline{\psi}(x,\alpha,t)$ by the following optimization problems.

$$\overline{\psi}(x,\alpha,t) = \min_{u(s)\in\underline{H}(\alpha(s))} E\left\{\int_t^{t_f} g(x(s))ds\right\} \tag{19}$$

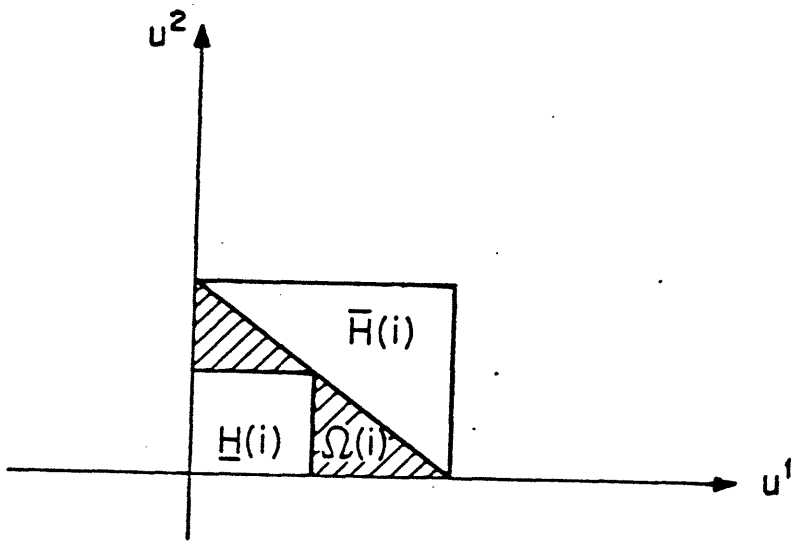



**Figure 2.** An Example of the Hypercubes for a Control Constraint Set

$$\underline{\psi}(x,\alpha,t) = \min_{u(t)\epsilon\underline{H}(\alpha(s))} E\left\{\int_t^{t_f} g(x(s))ds\right\} \tag{20}$$

both subject to (1), (2) and (3), and with initial conditions x(t)=x and $\alpha(t)=\alpha$.

From (13), the following holds,

$$\underline{\psi}(x,\alpha,t) \leqslant J_{u^*}(x,\alpha,t) \leqslant \bar{\psi}(x,\alpha,t) \tag{21}$$

Thus $\bar{\psi}$ and $\underline{\psi}$ are upper and lower bounds on the optimal value function. An estimate $\psi$ of $J_{u^*}$ can be obtained by taking a convex combination of the lower and upper bounds.

The cost function g(x) is separable. The constraints (16) and (17) affect each part separately. The optimization problems (19) and (20) are therefore decoupled and can be solved as a set of scalar problems one for each part.

The hypercubes $\underline{H}(\alpha)$ and $\bar{H}(\alpha)$ approximate the control constraint set. If the capacity for the production of part n is small in machine state $\alpha$, the corresponding limits $\underline{q}_{\alpha n}$ and $\bar{q}_{\alpha n}$ of the hypercubes are small. Likewise, if the capacity is large, the limits are also large. The calculation of the upper and lower bounds thus takes into account the relative productive capacity in all machine states, demand rates and the value of the parts as given by the cost function g(x). We expect therefore that the cost estimates satisfy the criteria above necessary for good performance by the estimate based controller.

### 5.3 Implementation of the Estimate Based Controller

There are two steps in the implementation of the EB-controller. Off-line (ie in the top box of Figure 1), upper and lower bounds to the optimal value function are computed by solving (19) and

(20). In practice this is done by discretizing the problems over discrete points in time and the buffer state. The estimate $\psi(x,\alpha,t)$ is computed from the bounds and stored. On-line (in the flow control level), whenever the system enters machine state i, the control u(t) is determined by the linear program (15) using the stored values of $\psi(x,\alpha,t)$.

Computational and storage costs of the EB controller grow exponentially with the number of workstations M and linearly with the number of parts N. However, the computation is done off-line and the estimates of the optimal value function can be stored in peripheral devices. The on-line computation consists of the linear program (15) and has N variables and M constraints. Typically, N is between 5 and 10, and M between 10 and 20. The program can thus be easily solved on a small computer.

The off-line computational cost can be reduced by pruning the machine state to exclude states with low probability. With the failure and repair rates typically found in manufacturing systems, a small number of states account for over 95% of the probability. A large number of states can therefore be eliminated without substantially altering the regions and hedging points corresponding to the estimate based control policy.

## 6. EXAMPLE

### 6.1 System and Analytic Results

To demonstrate the application of the hierarchical controller, consider the flexible transfer line of Figure 3. Each station has two identical machines. Two parts are produced. The first type requires two operations one at each station, while the second part requires a single operation which can only be performed at the first station.

The operation times and reliability data for the system are given in Tables 1 and 2. In this example, there are nine possible machine states. We will discuss only three of them, all machines operational ($\alpha$=(2,2)), one failed machine at station A ($\alpha$=(1,2)) and one station B machine failed ($\alpha$=(2,1)). The calculation, however, must include all nine states.

TABLE 1
PROCESSING TIME FOR THE PARTS IN MINUTES

| Part | Stage | |
|---|---|---|
| | A | B |
| 1 | 0.33 | 0.33 |
| 2 | 0.67 | not required |

TABLE 2
RELIABILITY DATA IN MINUTES

| Stage | MTBF | MTTR |
|---|---|---|
| A | 300 | 30 |
| B | 300 | 30 |

The cost function is given by

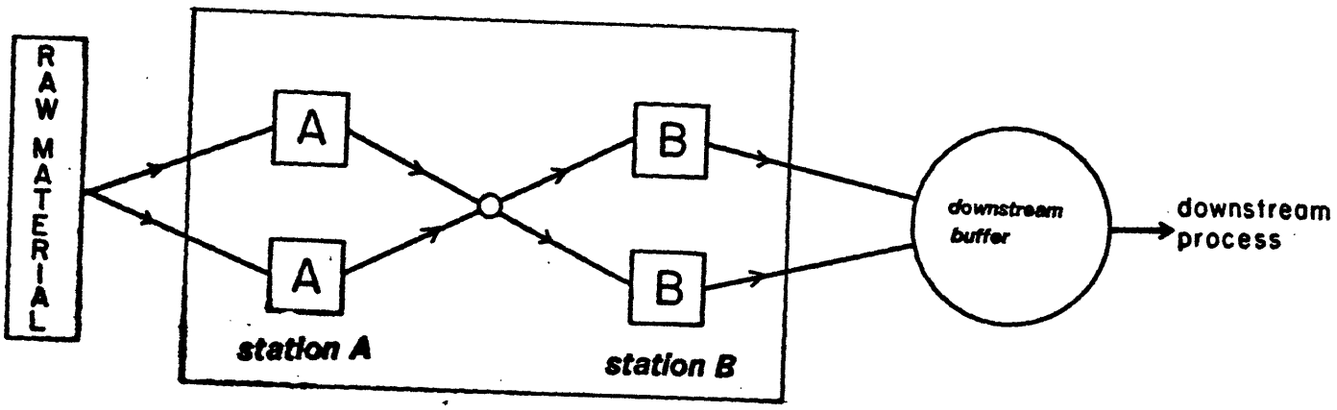


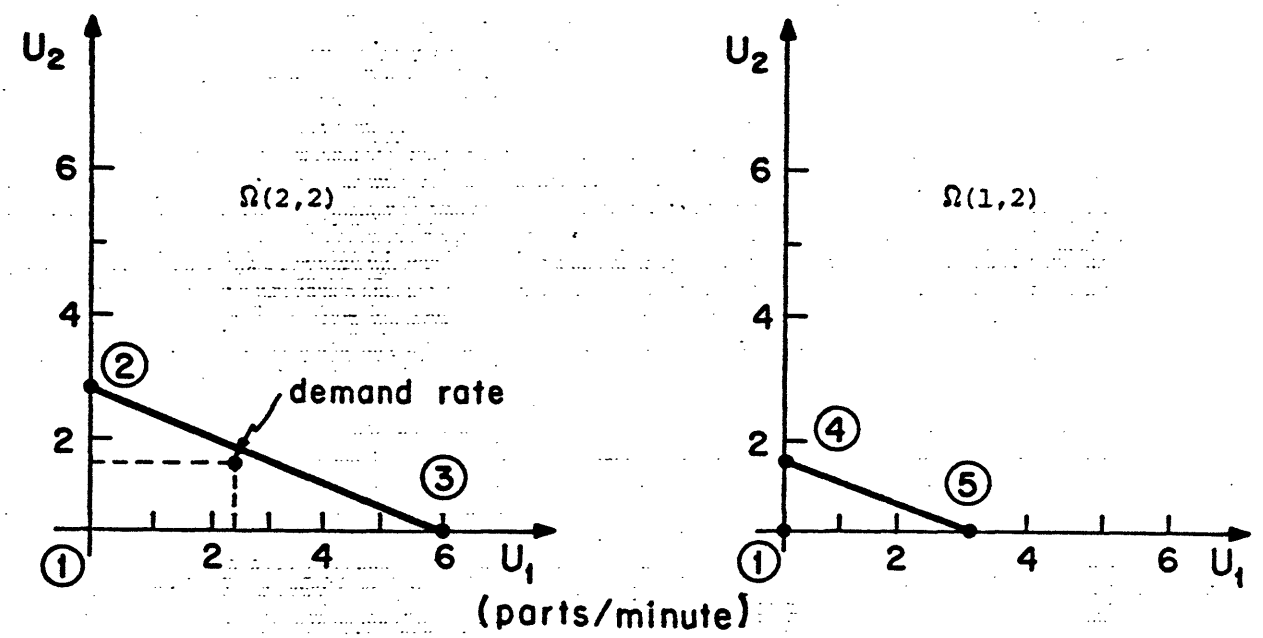

**Figure 3.** A Flexible Transfer Line

$$g(x)=\sum_{n=1}^{2}|x_n| \tag{22}$$

Thus the system is penalized equally for being ahead or behind demand requirements.

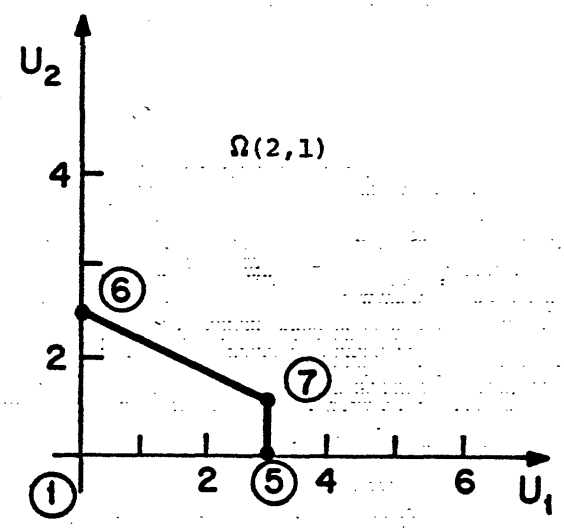
The production constraint sets for machine states (2,2),(2,1) and (1,2) are shown in Figure 4. The different effects of station A and station B failures are evident. The demand rate d(t) for the two part types are constant at 2.5 type 1 and 1.25 type two parts per minute. That is $d_1(t)=2.5$, and $d_2(t)=1.25$. Production can exceed demand only in machine states (2,2) and (2,1). In all other machine states, the demand rate is beyond the capacity of the system.

The control policy is characterized by the regions shown in Figure 5. In each region, the production vector is at an extreme point of a constraint set. It is indicated by the circled numbers in Figures 4 and 5. In finite horizon problems, the boundaries are time varying but maintain their structure. In infinite horizon problems, there is a solution only if the FMS can satisfy the long term demand requirements. In this case a steady production policy exists and is characterized by constant boundaries between the regions.



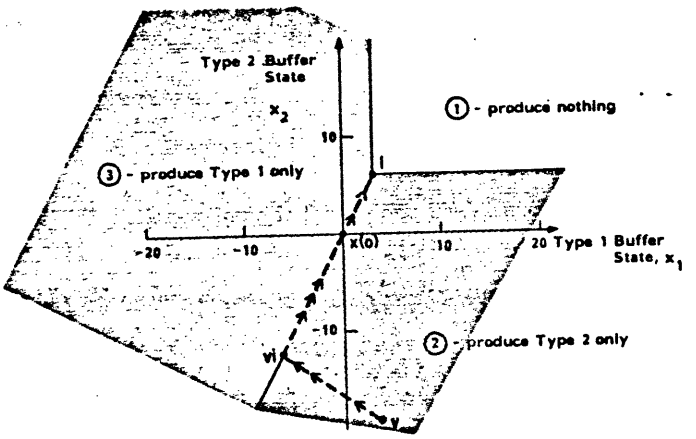

(a) All machines up
Machine state (2,2)

(b) One Type A down
Machine state (1,2)

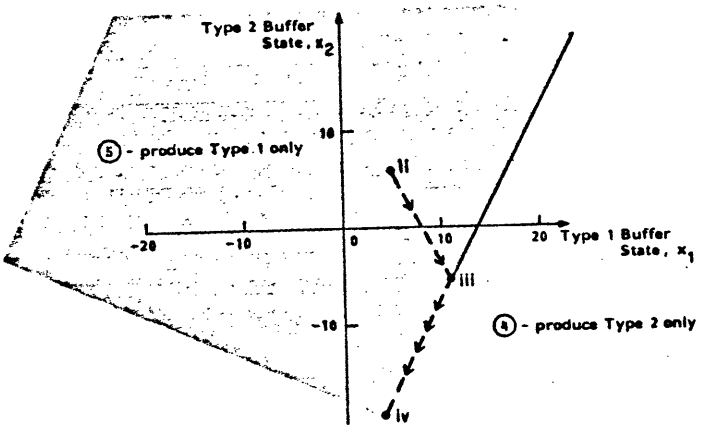



(c) One Type B down
Machine state (2,1)

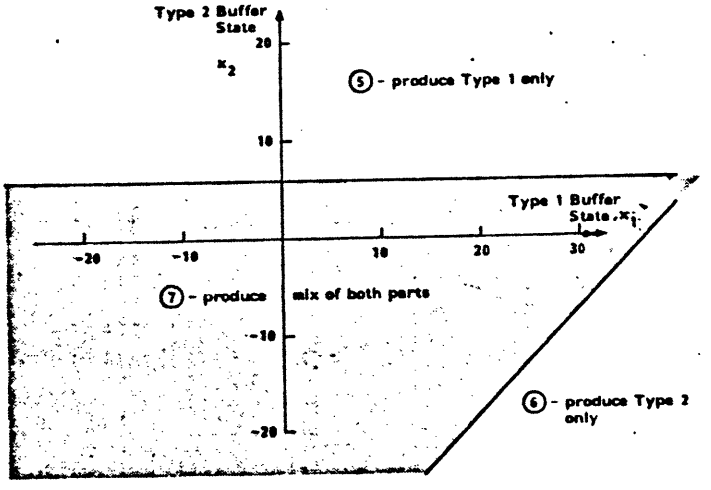
**Figure 4.** Control Constraint Sets




(a) All machines up (2,2)




(b) One Type A down (1,2)




(c) One Type B down (2,1)

Key:

——— Region boundaries

– – – → – – Sample trajectory

③ Production vector in control constraint set

**Figure 5.** Control Regions Corresponding to an Optimal Policy

The control policy was computed by evaluating an estimate of the optimal cost-to-go function and then solving equation (15) to obtain the regions.

Also shown in Figure 5 is the behavior of the buffer state trajectory. Initially, the system has all machines operating and the buffer state x(0) is 0. (The origin of Figure 5a). The point x(0) happens to lie on the boundary between two regions. (This is not always the case). The production vectors in the two neighboring regions both drive the trajectory towards the boundary. The trajectory moves in the positive direction as an inventory of parts is built up as a hedge against future failures. At point (i), production equals demand and the trajectory remains constant. That is $u(x,\alpha,t)=d(t)$ and $dx/dt=0$.

When a type A machine fails, the new production rate is found at point (ii) of Figure 5b. Initially only type 1 parts are produced resulting in an increase in the buffer level of type 1 parts. The buffer level of type 2 parts as a consequence, drops. At point (iii), the trajectory meets the boundary and a mix of both parts is produced, keeping the trajectory on the boundary. After approximately 25 minutes, the failed machine is repaired with the buffer levels at point (iv). The production rate is found at point (v) of Figure 5a. Type 2 parts are produced at the maximum rate to clear the backlog caused by the failure. Production of typel parts resumes at point (vi) and the trajectory follows the boundary to point (i) where once again production is at the demand rate. A similar set of events can be constructed for any other sequence of failures and repairs.

### 6.2 Simulation Results

The system of Figure 3 was simulated with the scheduling being performed by the hierarchical controller. Each station had an internal buffer with a capacity for 5 pieces and a last-in-first-out discipline. The simulation model was run for an equivalent of 14 hours. It should be pointed out that the buffer state x(t) refers to the difference between actual and desired production levels and can therefore take on negative values indicating backlogs. Internal storage buffers aid the sequence control level in generating schedules which maintain the production rates determined by the flow control level.

The availability and utilization of available time at each machine is given in Table 3. Station A is the system bottleneck. The controller is able to attain utilizations of 94% and 85% at the two station A machines. Station B on the other hand is lightly loaded with only 55% and 36% of the available time being used.

TABLE 3
UTILIZATION AND AVAILABILITY FOR THE SIMULATION

| Stage | Machine | Availability | Utilization |
|---|---|---|---|
| A | 1 | 0.95 | 0.94 |
| A | 2 | 0.91 | 0.85 |
| B | 1 | 0.92 | 0.55 |
| B | 2 | 0.92 | 0.36 |

Production statistics are shown in Table 4. On average, the production was 5.2 pieces behind demand for type 1 parts and 4.2 for type 2. The average in-process inventory in the system is small, 3 type 1 pieces and 1.2 type 2 pieces. At the end of the simulation, the system had produced the required number of type 2 parts and was two type 1 parts short of target. Thus the algorithm was able to track demand and at the same time keep the number of pieces inside the system small. It should be pointed out that the cost function (22) penalizes the controller equally for excess production and for backlogged demand. The preferred mode of operation is therefore to keep the

buffer trajectory close to zero when all machines are operational and to clear backlogs which result when failures occur, rather than to maintain a large inventory of parts as a hedge against future failures. The behavior can be modified by penalizing backlogged demand more than excess production and by weighting the parts differently in the cost function.

TABLE 4
PRODUCTION STATISTICS FOR THE SIMULATION

| Part | Average In-process Inventory | Mean Buffer State | Number of Parts Required | Number of Parts Produced |
|---|---|---|---|---|
| 1 | 3.0 | -5.2 | 2083 | 2081 |
| 2 | 1.2 | -4.2 | 1042 | 1042 |

A portion of the buffer state plotted at one minute intervals is depicted in Figure 6. The flow control level implemented in the simulation calculates the production vector at one minute intervals. This, in addition to the fact that production increases by integer amounts, accounts for the chatter of the trajectory in the vicinity of the region boundaries. However, the simulated buffer state behaves in a manner very close to that predicted by theory and shown in Figures 5a and 5b.

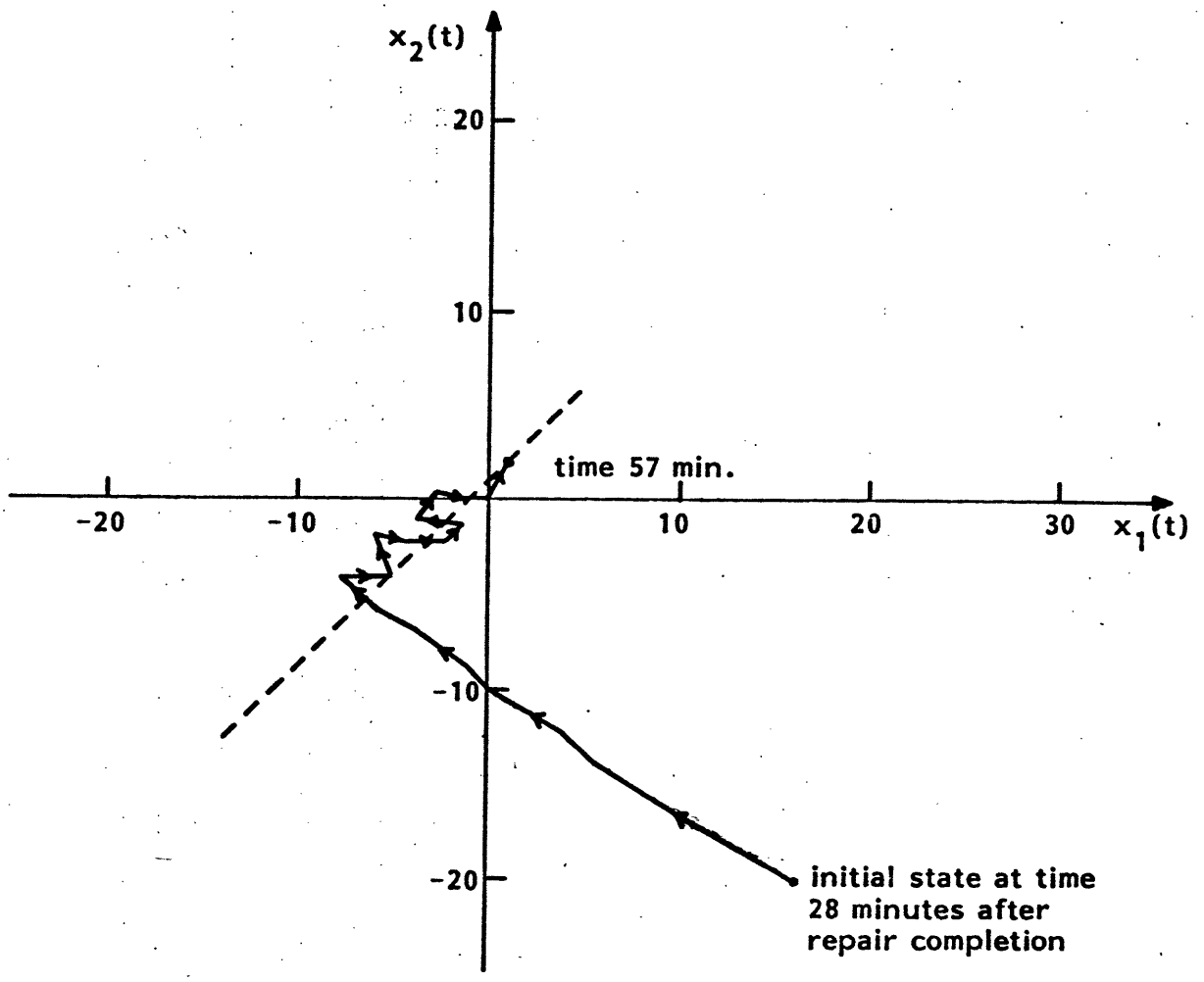



**Figure 6.** A Part of the Buffer State Trajectory for the Simulation Model

## 7. CONCLUSIONS

We have described a hierarchical algorithm for the control of production in an automated manufacturing system with unreliable machines. The algorithm is designed to fit into existing factory mangement structures. The example and the simulation results show that it is possible to accurately track demand requirements while maintaining a low in-process inventory, thereby realizing an important advantage of FMS's over traditional production methods.

The flow control level of the controller is responsible for regulating the input flow rate into the system so that production goals are met. It is important for the performance of the system that the production requirements should be feasible. The managers of the FMS should therefore have planning tools that ensure that the demand is within the capacity of the system.

This approach to short term production planning has several desirable features. Feedback is intrinsic to the approach, so that rational responses to random events are chosen. The control policy is adapted to the whole FMS, and not merely the first machine that a part encounters. this eliminates the buildup of material inside the system and, thus, congestion. It reduces the combinatorial complexity of the scheduling problem. It explicitly takes repair and failure information into account.

Our current research is aimed at implementing this approach and reducing the off-line computational effort. It is also aimed at improving algorithm performance by better maintaining the buffer state trajectory on region boundaries, when appropriate, and by modifying the sequence control level.

## 8. REFERENCES

[1] Buzacott, J. A. , "'Optimal' Operating Rules for Automated Manufacturing Systems", IEEE Transactions on Automatic Control, **27**, 2, (Feb, 1982).

[2] Buzacott, J. A. and Yao, D. , "Flexible Manufacturing Systems; A Review of Models" TIMS/ORSA Detroit Meeting (Apr, 1982).

[3] Dupont-Gatelmand, C., "A Survey of Flexible Manufacturing Systems", Journal of Manufacturing Systems, **1**, 1, 1-16, (1982).

[4] Halevi, G. , The Role of Computers in Manufacturing Processes, John Wiley and Sons (1980).

[5] Hahne, E. "Dynamic Routing in an Unreliable Manufacturing Network With Limited Resources ," M.I.T. Laboratory for Information and Decision Systems, Report No. LIDS-TH-1063 (1981).

[6] Hildebrant, R. "Scheduling Flexible Machining Centers When Machines are Prone to Failure," PhD Thesis, M.I.T Dept. of Aeronautics and Astronautics (May, 1980).

[7] Hitomi, K Manufacturing Systems Engineering, Taylor and Francis, London (1979).

[8] Hutchinson, G. "The Control of Flexible Manufacturing Systems: Required Information Structures," IFAC Symposium on Information-Control Problems in Manufacturing Technology, Japan, October (1977).

[9] Hutchinson, G. K, and Hughes, J.J. ,"A Generalized Model of Flexible Manufacturing Systems", Multi-Station Digitally Controlled Manufacturing Systems Workshop, Univ. of Wisconsin, Milwaukee (Jan 1977).

[10] Rishel, R. " Dynamic Programming and Minimum Principles for Systems with Jump Markov Disturbances," SIAM Journal on Control, **13**, 2 (February 1975).

[11] Olsder, G. J. and Suri, R. ,"Time Optimal Control of Parts-Routing in a Manufacturing System With Failure Prone Machines", Proc. of the 19th. IEEE Conference on Decision and Control, Albequerque, New Mexico (1980).

[12] Schick, I. and Gershwin, S. B. , "Modelling and Analysis of Unreliable Transfer Lines with Finite Interstage Buffers" M.I.T. Laboratory for Information and Decision Systems Report No. ESL-FR-834-6.

[13] Tsitsiklis, J. "Characterization of Optimal Policies in a Dynamic Routing Problem," M.I.T. Laboratory for Information and Decision Systems Report, No. LIDS-R-1178 (Feb, 1982).