

Composite Performance-Dependability Analysis of Cellular Manufacturing Systems

N. Viswanadham, *Fellow, IEEE*, and R. Ram

Abstract— This paper deals with the transient analysis of manufacturing systems in the presence of failures and repairs. The exact model is decomposed into 1) a structure state process, modeling the failure and repair events which occur at a slower time scale, and 2) a performance model, modeling the part processing and material movement events occurring at a faster time scale. The solutions of these two models are combined to yield the mean and standard deviation of cumulative production over an interval. Such studies, called combined performance-dependability studies, are important in the manufacturing context.

We define combined performance and dependability (performability) measures in the manufacturing context, and provide a mathematical formulation. Detailed performability studies are carried out for cellular manufacturing systems, in particular for a flexible manufacturing cell and a two-cell manufacturing system featuring blocking and prioritized, centralized repair. We also present several numerical examples.

I. INTRODUCTION

AUTOMATED manufacturing systems (AMS's) are a complex interconnection of subsystems such as Numerically Controlled Machine Centers, Assembly stations, Automated Guided Vehicles (AGV's), Robots and Conveyors. These systems are highly capital intensive and have short process life cycles. To realize high pay-back ratio, the system has to operate at high throughput levels. One factor that has a major influence on system performance is the unscheduled down-time of the equipment due to failures. The structure of the interconnected manufacturing system changes with failures and repairs and consequently the performance levels deteriorate or improve. In this paper, we present an approach for the performance analysis of manufacturing systems in the presence of failures and repairs.

Throughput and Manufacturing Lead Time are two important performance measures in a manufacturing system. Throughput is the number parts produced per unit time and lead time is the amount of time the workpiece resides on the factory floor. Most previous performance studies are conducted under failure-free assumptions. Discrete event simulation, Petri Nets, Markov chains, and Queueing Networks are used to compute the performance measures. Here, we conduct composite performance-dependability studies to determine the moments

of cumulative production of a manufacturing system consisting of failure-prone subsystems.

This paper deals with the transient analysis of manufacturing systems in the presence of failures and repairs. The exact model is decomposed into 1) a structure state process, modeling the failure and repair events which typically occur at a slower time scale; and 2) a performance model, modeling the part processing and material movement events that typically occur at a faster time scale. The solutions of these two models are combined to yield the mean and standard deviation of cumulative production over a finite time horizon. Such studies are termed combined performance-dependability studies. The main contributions of this paper are the following:

- 1) It addresses the issue of assessing the performance of a cellular manufacturing system over a finite time horizon in the face of failures and repairs;
- 2) It quantitatively studies the variability induced in the system performance due to subsystem failures and repairs, by computing the moments (mean and standard deviation) of cumulative production of failure-prone manufacturing systems.

We define combined performance-dependability (performability) measures in the manufacturing context, and provide a mathematical formulation. This framework provides a unified approach to study the availability, throughput, or lead time-related performance of a failure-prone manufacturing system over a finite time. Detailed performability studies are carried out for cellular manufacturing systems, in particular for a flexible manufacturing cell and a two-cell manufacturing system featuring blocking and prioritized, centralized repair. We also present several numerical examples.

A. Previous Research

In the manufacturing literature, performance and reliability issues are dealt separately. Various analytical and simulation based modeling tools have been used for performance evaluation. Reliability analysis of manufacturing systems has received relatively little attention. Combined performance-reliability studies in the context of transfer lines with unreliable machines are surveyed in [6]. Dallery *et al.* [5] have developed a technique to compute the average production rate of a transfer line with unreliable machines. Albino *et al.* [2] present a method for obtaining the steady-state average throughput in the presence of failures and repairs. These studies focus on determining only average values of performance in the face of failures. However, recent studies point to a need to obtain more

Manuscript received November 20, 1992; revised June 23, 1993.

N. Viswanadham is with the Department of Computer Science & Automation, Indian Institute of Science, Bangalore 560 012, India.

R. Ram is with the Center for Development of Telematics, Bangalore 560 052, India.

IEEE Log Number 9214693.

complete information on system performance. For instance, in the context of transfer lines employing failure-prone machines, Gershwin [11] highlights the inadequacy of using only average values (such as mean throughput) in assessing the system performance, and argues that more information (such as the variance of the production) is necessary to give a more realistic assessment. Thus, transient analysis of an AMS for moments of performance over a finite time horizon is highly relevant. This paper considers techniques for computing the moments of certain cumulative performance measures over a specified finite time horizon.

Performability modeling is an important topic of research in the area of fault-tolerant computing systems. Beaudry [3] introduced performance-related reliability for gracefully degrading fault-tolerant systems. Meyer [14] introduced the term performability, and provided a formal definition of the performability measure. Efficient performability evaluation techniques have since been extensively investigated. Papers representative of the work in this area are [9], [10], [12], [18]. Performability analysis of manufacturing systems appears in Viswanadham *et al.* [19] and Ram [16].

B. Outline of the Paper

In Section II, we present the approach followed for computing the performability distribution and/or its moments and define the various terms used in this paper. In Section III we study a flexible manufacturing cell under various repair strategies and present numerical results. In Section IV, we consider a generic cellular manufacturing system with finite buffers and prioritized centralized repair, with multiple repairmen. We use stochastic Petri Nets to model the failure-repair behavior as well as to obtain the performance measures of the cellular manufacturing system.

II. COMPOSITE PERFORMANCE-DEPENDABILITY MODEL

It is possible to develop an exact monolithic performance-dependability model for manufacturing system taking to account setups, part processing, failures, repairs and reconfiguration. Single machine examples of such an analysis maybe found in [16]. However such an approach suffers from two problems: largeness and stiffness. Decomposition and aggregation methods are commonly followed to obtain near accurate solutions in these cases. Generally subsystem-wise decomposition and time-scale decomposition are two possible alternatives. In this paper we follow the latter.

The approach followed here is the following: we first obtain a structure state process (SSP), which describes the system evolution as influenced by failures and repairs. We define a performance model in each structure state and determine the reward (throughput or lead time), and use these values in the structure state process to determine the performability. The basis for such an approach is the large difference in the frequencies of failure/repair events and the part processing events. This leads to a natural hierarchical model:

- 1) A higher level (slow time-scale) structure-state process describing the failure-repair process.

- 2) A lower level (fast time-scale) performance model describing part processing and part handling activities.

In this paper, we assume that in each state of the structure state process, the performance model reaches steady state. This is a valid assumption because of the orders of magnitude difference in the failure rates and processing rates: a typical processing rate is 3 parts per hour whereas a typical failure rate is 1 in 200 hours. We may also mention that the time-scale decomposition proposed above leads to immense computational savings.

Failures can be classified into two types in the manufacturing systems context.

- 1) Operation dependent: Failures occur only when the machine is in operation.
- 2) State-independent (also termed *time-dependent*): Failures occur independent of the state of the machine, i.e., whether the machine is busy or idle.

Our approach basically deals with independent failures but could be adapted to operation dependent failures. We consider this point later in the paper.

Several situations can also arise with regard to repair. These include *no-repair*, *centralized repair*, and *decentralized repair*. The no-repair case is of interest in the case of night-shift operations. If there exists a separate repair station for each family of equipment, it is termed decentralized repair. If all equipment goes for repair to a common repair station, then we are considering a system with centralized repair.

A. Composite Performance-Dependability Measures

Manufacturing systems are composed of several subsystems. Each subsystem can be in a failed or operational state. The failure and repair events will result in deterioration or enhancement of system performance. Also each of these events change the structure of the system. We capture these interactions through the concept of structure state processes.

Definition: The *structure state* of an automated manufacturing system is a vector whose components describe the status of its constituent subsystems.

The structure state of the system changes due to failures and repairs as time progresses. The transitional dynamics from one state to another is captured via the structure state process defined below.

Definition: Let $Z(u)$ be the structure state of the manufacturing system at time $u \geq 0$. Then the family of random variables $\{Z(u), u \geq 0\}$ is called the structure state process (SSP).

The structure state process can be described by Markov chains, Queueing Networks, or Stochastic Petri Nets. We assume that the system reaches steady state in each of the structure states and use a performance model to compute the throughput or the manufacturing lead time, or, more generally the *reward* associated with that structure state; the reward is a quantitative index of the system performance in a specific structure state.

Let S be the state space of the SSP, which is partitioned into two disjoint sets, S_O and S_F ; S_O is the set of operational states and S_F is the set of non-operational states. Suppose the

structure state process $\{Z(u), u \geq 0\}$ is a homogeneous finite-state continuous-time Markov chain (CTMC) with state space S , and infinitesimal generator Q . Let $p_i(t)$ be the unconditional probability that the CTMC is in state i at time t , and let $P(t)$ represent the state probability vector of the CTMC. It is well-known that

$$\dot{P}(t) = P(t)Q$$

1) *Repairable Systems*: When the system is repairable in every state, the Markov chain is likely to be irreducible. If the system is not repairable in some states, then the Markov chain will have absorbing states. We deal with both the cases here. The steady-state probability vector $\pi = \lim_{t \rightarrow \infty} P(t)$ exists if the CTMC is irreducible and positive recurrent, and satisfies

$$\pi Q = 0; \sum_{i \in S} \pi_i = 1$$

Define

$$L(t) = \int_0^t P(u)du$$

Then $L(t)$ satisfies the differential equation

$$\dot{L}_i(t) = L(t)Q + P(0); L(0) = 0.$$

Note that $L_i(t)$ is the expected total time spent by the CTMC in state i during an observation period $[0, t]$. (Here, 0 represents a $|S|$ -dimensional vector of zeroes).

2) *Non-Repairable Systems*: For non-repairable systems, the Markov chain model of the structure state process will exhibit absorbing states that represent complete system failure. Here, the mean time to total system failure is of interest. To compute this, define

$$\eta_i = \lim_{t \rightarrow \infty} L_i(t) = \int_0^{\infty} P_i(u)du, i \in S$$

then η_i represents the mean time spent by the CTMC in state i until absorption. Also $\eta_i = 0$ for $i \in S_a$ and is finite for $i \in S_t$, where S_a and S_t are sets of absorbing and transient states. Note that $S = S_a \cup S_t$. To compute η_i , form a new matrix Q_t of size $|S_t| \times |S_t|$, where $|S_t|$ denotes the cardinality of the set S_t by restricting Q to only the set of transient states. Then the row vector $\eta = [\eta_1, \eta_2, \dots, \eta_{|S_t|}]$ satisfies the equation

$$\eta Q_t = -P_t(0)$$

where $P_t(0)$ is the initial probability vector restricted to the transient states. Then we have

$$\text{Mean time to absorption} = \sum_{i \in S_t} \eta_i$$

A survey of the numerical techniques for Markov chain transient analysis problems of the nature indicated above is presented by Reibman *et al.* [17].

2) *Performability Problem Formulation*: In a performability model, we combine the structure state model (or reliability model) which describes the system evolution as influenced only by failures and repairs, with the performance model that describes the system performance (throughput, lead time, etc.) in each state of the structure state process.

a) *Dependability measures*: We now define two important measures, reliability and availability. For this we define the indicator random variable

$$I(u) = \begin{cases} 1 & \text{if } Z(u) \in S_O \\ 0 & \text{if } Z(u) \in S_F \end{cases}$$

Definition: The *reliability* of a system over an observation period $[0, t]$ is the probability that the system is functioning properly throughout the observation period. Thus,

$$R(t) = P\{I(u) = 1, \forall u \in [0, t]\}$$

Definition: *Availability measures* consist of the following types.

- 1) The *point availability* or *instantaneous availability* $PAV(u)$ of a system, at time $u \geq 0$, is the probability that the system is properly functioning at time u .
- 2) The *cumulative operational time*, $O(t)$ of the system over $[0, t]$ is the amount of time the system is operational over $[0, t]$.
- 3) The *interval availability*, $IAV(t)$, is the fraction of time the system is operational over $[0, t]$.

$$IAV(t) = \frac{O(t)}{t}$$

While point availability $PAV(t)$ reflects the state of the system at a given point of time, $O(t)$ and $IAV(t)$ are more useful measures. Clearly, these measures are given by

$$PAV(u) = P\{I(u) = 1\};$$

$$O(t) = \int_0^t I(u)du;$$

$$IAV(t) = \frac{1}{t} \int_0^t I(u)du$$

The *steady-state availability*, A , of a system is the limiting value of availability $PAV(u)$ as $u \rightarrow \infty$. Thus,

$$A = \lim_{u \rightarrow \infty} PAV(u) = \lim_{u \rightarrow \infty} E[I(u)]$$

Note that $A = 0$ for non-repairable systems. For repairable systems, A is the fraction of time the system functions properly.

When the structure state process is described by a continuous-time Markov chain we have

$$PAV(t) = \sum_{i \in S_O} P_i(t);$$

$$A = \sum_{i \in S_O} \pi_i;$$

$$E[O(t)] = \sum_{i \in S_O} L_i(t)$$

b) Markov Reward Models: In this section, we present the composite measures that combine both performance and dependability aspects, using the notions of the structure state process and instantaneous and accumulated rewards. In each structure state, we define an index of performance such as throughput or lead time. This performance measure is in general called a reward.

Let $\{Z(u), u \geq 0\}$ be the SSP of the manufacturing system and S be its state space. Let γ_i be the (non-negative) reward in structure state i . Then we have:

- 1) Instantaneous reward at time t :

$$X(t) = \gamma_{Z(t)} = \gamma_i \quad \text{if } Z(t) = i$$

- 2) Accumulated reward over $[0, t]$

$$Y(t) = \int_0^t X(u)du = \int_0^t \gamma_{Z(u)}du$$

- 3) Distribution of accumulated reward

$$F_Y(y, t) = \text{Prob}(Y(t) \leq y)$$

Using the above definitions, we can compute the expected rewards $E[X(t)]$, $E[Y(t)]$ and their steady state values. These are indicators of the average performance of a manufacturing system in the face of failures.

Expected Instantaneous Reward (applicable to both repairable and non-repairable systems)

$$E[X(t)] = \sum_{i \in S} \gamma_i p_i(t)$$

- 1) Expected Accumulated Reward (applicable to both repairable and non-repairable systems)

$$E[Y(t)] = \sum_{i \in S} \gamma_i L_i(t)$$

- 2) Steady-State Expected Instantaneous Reward (of interest for repairable systems)

$$\lim_{t \rightarrow \infty} E[X(t)] = \sum_{i \in S} \gamma_i \pi_i$$

- 3) Expected accumulated reward until absorption (of interest for non-repairable systems)

$$\lim_{t \rightarrow \infty} E[Y(t)] = \sum_{i \in S} \gamma_i \eta_i$$

In general, for non-repairable systems, only transient analysis is of interest. For repairable systems, both transient analysis as well as steady-state analysis are of interest. In this paper, we are dealing with manufacturing system behavior over finite time-horizons, and hence we are primarily concerned with transient analysis.

The *distribution* of accumulated reward gives a complete picture of the performance of the manufacturing system in the face of failures. It is convenient to represent the distribution in the form of a two dimensional Laplace transform, $F_Y(y, t)$, given by [18]

$$F(s, \mu) = P(0)(\mu I + sR - Q)^{-1}e \quad (1)$$

where s and μ are Laplace variables associated with t and y , $R = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_{|S|})$ is a diagonal matrix of size $|S| \times |S|$ with the reward rates forming the diagonal, I is the identity matrix of size $|S| \times |S|$, and e is a column vector of dimension $|S|$ with unit entries. To obtain the performability distribution, one has to invert this two-dimensional transform. Efficient algorithms to carry out this inversion for the case where the structure state process is an *acyclic* Markov chain, which corresponds to non-repairable systems, have appeared in the literature [10]. However, the inversion is known to be computationally very expensive for non-acyclic Markov chains [9], [18] and one has to resort to numerical techniques. Thus, in the case of repairable systems, one is often interested in determining the *moments* of performability measures rather than its distribution. To obtain the moments, we use the following notation. Let the n th conditional moment of the accumulated reward be defined as

$$m_{n,i} = E[Y^n(t) | Z(0) = i], \quad i \in S, \quad n \geq 1$$

Let us define the $|S|$ -dimensional vector of rewards

$$\mathbf{m}_n(t) = [m_{n,1}(t), m_{n,2}(t), \dots, m_{n,|S|}(t)]$$

Then it is known that $\mathbf{m}_n(t)$ evolves according to the equation [13]

$$\frac{d\mathbf{m}_n(t)}{dt} = Q\mathbf{m}_n(t) + nR\mathbf{m}_{n-1}(t); \quad \mathbf{m}_0(t) = e^T \quad (2)$$

We obtain the unconditional n th moment from

$$E[Y^n(t)] = P(0)\mathbf{m}_n^T(t)$$

In the above, the superscript T indicates transposition. From the above equations, any moment can be obtained by solving the above ordinary differential equation (2).

c) Availability measures as special cases of performability measures: By setting the rewards as $\gamma_i = 1$ for $i \in S_O$, and $\gamma_i = 0$ for $i \in S_F$, we see that the performability measures specialize to availability measures. Then point performability $X(t)$ would correspond to point availability $PAV(t)$, and the accumulated reward $Y(t)$ to cumulative operational time $O(t)$. Thus, performability measures include availability/reliability measures as specific instances. Silva and Gail [8] have developed a uniformization based algorithm for numerical evaluation of the distribution of $O(t)$. Their methodology is applicable to Markov chains with hundreds of states. It is known that the computation of the distribution of $O(t)$ is far more difficult than the computation of the transient state probability vector $P(t)$.

d) Throughput-related performability: Let T_i be the throughput in structure state i . If we assign the reward $\gamma_i = T_i$, $i \in S$, then the accumulated reward $Y(t)$ gives the total production over the interval $[0, t]$. $E[Y(t)]$ is the average production of the system over $[0, t]$; $F_Y(y, t)$ gives the distribution of cumulative production.

e) *Lead time-related performability*: Let W_i be the average lead time in structure state i and ω be its target upper bound in the face of failures. Suppose we are interested in computing the probability of the event “average MLT over $[0, t] < \omega$.” This can be computed from the distribution function $F_Y(0, t)$ with rewards γ_i defined as $\gamma_i = (W_i - \omega)T_i$.

To see the above result, note that the average MLT over $[0, t]$ is the ratio of cumulative workpiece hours over $[0, t]$ to the cumulative production over $[0, t]$. More precisely,

$$\text{Average lead time over } [0, t] = \frac{\int_0^t W_{Z(u)} T_{Z(u)} du}{\int_0^t T_{Z(u)} du}.$$

From the above, it follows that

$$\begin{aligned} P(\text{Average MLT over } [0, t] \leq \omega) &= P\left[\frac{\int_0^t W_{Z(u)} T_{Z(u)} du}{\int_0^t T_{Z(u)} du} < \omega\right] \\ &= P\left(\int_0^t W_{Z(u)} T_{Z(u)} du < \omega \int_0^t T_{Z(u)} du\right) \\ &= P\left(\int_0^t (W_{Z(u)} - \omega) T_{Z(u)} du < 0\right) \\ &= F_Y(0, t) \text{ with } \gamma_i = (W_i - \omega)T_i \end{aligned}$$

Hence, we see that by appropriate assignment of rewards, one can study different aspects of manufacturing system performance in the face of failures. The concept of performability constitutes a powerful unifying framework for studying the availability, throughput, and lead time performance of a failure-prone manufacturing system. This paper focuses on availability and throughput-related performability.

III. PERFORMABILITY OF A FLEXIBLE MANUFACTURING CELL

Consider a flexible manufacturing cell (FMC), depicted in Fig. 1, with two identical machine centers M_1 and M_2 . An AGV transports workpieces between the pallet pool and the machine centers. The machine centers and the AGV are individually controlled by programmable controllers (labeled MC and AGVC in the figure) which are interconnected by a local area network (LAN) to a cell controller. The AGV and the machines are failure-prone, while the LAN and the associated computer control system are assumed to be extremely reliable. Assuming the failure times and the repair times to be exponentially distributed, we can formulate the structure state process as a continuous-time Markov chain (CTMC), and obtain the composite performance-dependability measures by associating suitable rewards with the states of the SSP. We consider the cases of no repair and decentralized repair.

A. Non-Repairable System

This study is important when the FMC is used in an unmanned night shift. The SSP is given by $\{Z(u), u \geq 0\}$ with state space

$$S = \{(i, j) \mid i \in \{0, 1, 2\}, j \in \{0, 1\}\}$$

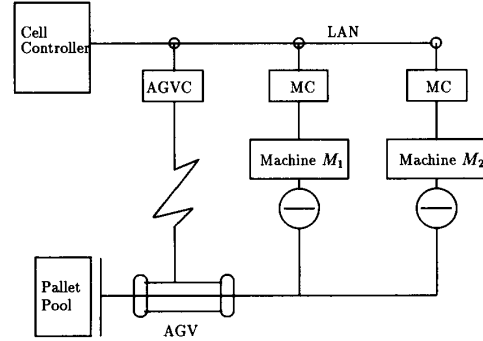


Fig. 1. A flexible manufacturing cell.

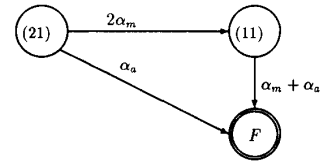


Fig. 2. SSP of a non-repairable FMC.

where i denotes the number of machines working, and j indicates the status of the AGV: up (1) or down (0). Further,

$$S_O = \{(21), (11)\}$$

$$S_F = \{(20), (10), (01), (00)\}.$$

For each structure state in S_F , the reward is zero since no production is possible when the AGV or both the machines are down. It is possible to lump all the states in S_F into one state and call it F . Then the Markov chain for the SSP is shown in Fig. 2.

Let α_m and α_a denote the failure rates of a machine and the AGV respectively. The state probability vector $P(t)$ and the rate matrix Q for this Markov chain are given by

$$P(t) = [p_{(21)}(t) \quad p_{(11)}(t) \quad p_F(t)]$$

and

$$Q = \begin{bmatrix} -(2\alpha_m + \alpha_a) & 2\alpha_m & \alpha_a \\ 0 & -(\alpha_m + \alpha_a) & \alpha_m + \alpha_a \\ 0 & 0 & 0 \end{bmatrix}$$

Let us assume that initially the system is in the state (21).

1) *Availability measures*: The instantaneous availability is given by

$$\begin{aligned} PAV(t) &= p_{(21)}(t) + p_{(11)}(t) \\ &= 2e^{-(\alpha_m + \alpha_a)t} - e^{-(2\alpha_m + \alpha_a)t} \end{aligned}$$

The expected operational time, $E[O(t)]$, is given by

$$\begin{aligned} E[O(t)] &= L_{(21)}(t) + L_{(11)}(t) \\ &= \frac{3\alpha_m + \alpha_a}{(2\alpha_m + \alpha_a)(\alpha_m + \alpha_a)} \\ &\quad - \frac{2e^{-(\alpha_m + \alpha_a)t}}{\alpha_m + \alpha_a} + \frac{e^{-(2\alpha_m + \alpha_a)t}}{2\alpha_m + \alpha_a} \end{aligned}$$

2) *Performability Measures*: Let the production rate of the system be $2T$ when both machines are working, and T when only one machine is working; thus, $\gamma_{(21)} = 2T$, and $\gamma_{(11)} = T$; of course, there is no production in the failed state, and $\gamma_F = 0$.

The instantaneous throughput at t is given by

$$E[X(t)] = 2Tp_{(21)}(t) + Tp_{(11)}(t) = 2Te^{-(\alpha_m + \alpha_a)t}$$

The expected cumulative production over $[0, t]$ is

$$\begin{aligned} E[Y(t)] &= \gamma_{(21)}L_{(21)}(t) + \gamma_{(11)}L_{(11)}(t) \\ &= 2T \left[\frac{1}{(2\alpha_m + \alpha_a)} \right. \\ &\quad \left. - \frac{2e^{-(\alpha_m + \alpha_a)t}}{\alpha_m + \alpha_a + \alpha_c} + \frac{2\alpha_m}{(2\alpha_m + \alpha_a)(\alpha_m + \alpha_a)} \right] \end{aligned}$$

B. System with Decentralized Repair

Assume that a dedicated repairman is provided for repairing machine and AGV failures. Unlike the non-repairable system, here one needs to distinguish among the failed states. We consider the state independent (or, time-dependent) failure case and the operation-dependent failure case separately.

1) *Decentralized Repair—Independent Failures*: In this case, the components fail irrespective of whether the system is operational or not. The SSP is the same as for the non-repairable system. All failure states are recoverable. Let β_a and β_m denote the repair rates of the AGV and a machine respectively. The SSP is shown in Fig. 3(a); since the failure/repair behavior of the machines and the AGV are independent, the SSP can be decomposed (exactly) into two CTMCs as shown in Fig. 3(b); in the latter figure, the failure/repair behavior of each resource type (machines or AGV) is described by a unique Markov chain. Thus, the transient state probabilities of the SSP, the $p_{(ij)}(t)$ s, can be obtained from

$$p_{(ij)}(t) = p_i(t)p_j(t)$$

where $p_i(t)$ is the probability that i machines are working at time t (where $i = 0, 1, 2$), and is obtained by solving (separately) the failure/repair model of the machines; and, $p_j(t)$ is the probability that j AGV's are working at instant t (where $j = 0, 1$), obtained from the corresponding model for the AGV.

2) *Decentralized Repair—Operation-Dependent Failures*: Assume that when the system is functional, the resources are all fully utilized. Here, since failures occur only when the system is operational, the state space of the SSP is

$$S = \{(21), (11), (20), (10), (01)\}$$

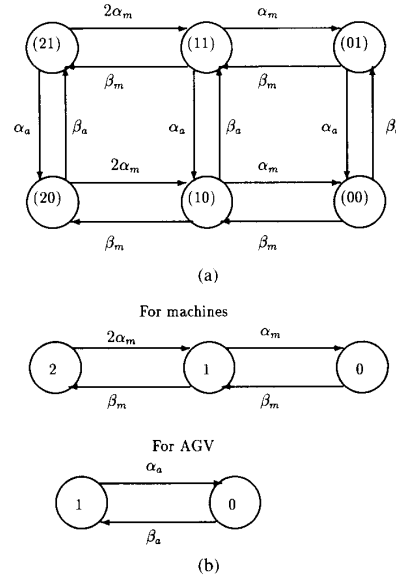


Fig. 3. SSP of a FMC with decentralized repair: state independent (time-dependent) failures. (a) SSP for a state-independent failure model. (b) Decomposed failure/repair models.

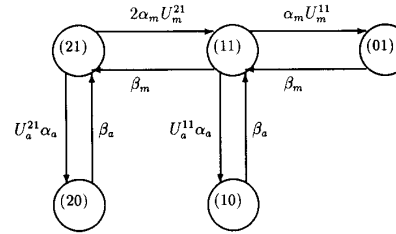


Fig. 4. SSP of a FMC with decentralized repair: state-dependent failures.

with

$$S_O = \{(21), (11)\}; \quad S_F = \{(20), (10), (01)\}$$

The Markov chain model of the SSP is shown in Fig. 4. Transitions representing failure will be allowed only when the resource is busy. Transition rates can however be computed as the product of the failure rates and percentage utilization of the resource. The average utilization can be obtained from the performance model. If U_k^{ij} represents the average utilization of the k th resource in the state (i, j) , the transition rates are as given in Fig. 4.

C. Numerical Study

As noted earlier, the expected availability/performability measures for the non-repairable system may be obtained analytically. However, for the repairable system, the SSP is a positive recurrent Markov Chain, and analytical solutions for performance-dependability measures are very hard to obtain. One has to resort to numerical techniques. In this section, we study the throughput-oriented performability of the FMC.

Consider the system operation over an interval of 24 hours (three consecutive shifts). Let the mean time to failure and

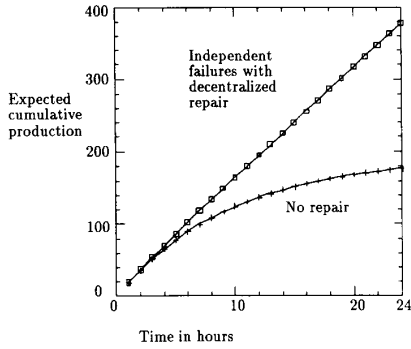


Fig. 5. Expected cumulative production of a failure-prone FMC.

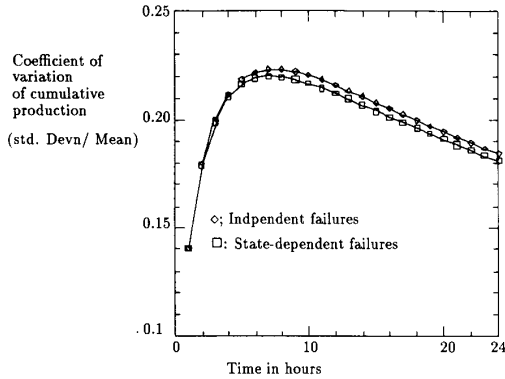


Fig. 6. Flexible Manufacturing Cell: Coefficient of variation of cumulative production.

the mean time to repair be respectively $1/\alpha_a = 24$ hours and $1/\beta_a = 2$ hours for the AGV, and $1/\alpha_m = 16$ hours and $1/\beta_m = 3$ hours for a machine. We study the cumulative production over the interval $[0, t]$ for the non-repairable system, and the system with decentralized repair. Assume that the production rate of each machine is $T = 10$ parts per hour. The reward assignment is: $\gamma_{(21)} = 2T = 20$, and $\gamma_{(11)} = T = 10$; and, the other states have reward 0, as they denote system failure.

We consider both independent and state dependent failures. Since the utilization in each operational state is 100%, $U_k^{i,j} = 1$ for all i, j, k . The graph in Fig. 5 compares the expected cumulative production of the repairable and non-repairable systems. For the non-repairable case, the cumulative production begins to saturate, since the system reaches a failure state and becomes non-productive. However, in the repairable case, the cumulative production keeps increasing since the repair facility maintains it in an operational condition. The curves for independent and state dependent failure cases are identical in both repairable and non-repairable situations. The results for independent failure case are only shown in Fig. 5.

Figure 6 shows the coefficient of variation (ratio of standard deviation to mean) of cumulative production for the two repairable case variants. The coefficient of variation is an index of the variability (or, uncertainty) in production induced due to resource failure and repair.

The coefficient of variation is lower for the state-dependent failure case relative to the independent failure case, because in the former no further failures are possible once the system reaches a failed state, whereas they can still occur in the latter. The coefficient of variation shows an interesting behavior: it initially increases with the shift duration, peaks, and then decreases. This behavior has implications for production scheduling, especially in the context of Flexible Manufacturing, where short production runs are the rule. Consider the case where the FM cell has an order for 375 parts. From Fig. 5, we see that the expected shift duration is 24 hours. One may consider the following alternatives:

- 1) Three separate shifts of 8 hours each (expected production of 125 parts for each shift) on separate days, starting the system in the fully operational state in each shift.
- 2) A single continuous run of 24 hours.

In general, the first alternative has the advantage of lower work in progress, while the second has a shorter delivery time. The expected cumulative production values for each of the two alternatives are very nearly identical. However, the variance of the cumulative production is higher in the case of the first alternative. Consider the state-dependent failures case. For $t = 8$ hours, the coefficient of variation is 0.22, and for $t = 24$ hours, the coefficient of variation is 0.18. Thus, the standard deviation of the production is $3 \times 0.22 \times 125 \approx 83$ parts for the first alternative, and $0.18 \times 375 \approx 68$ parts for the second alternative. Thus, the second alternative is preferable from the point of view of reduced variability. \square

Besides studying the performance of a FMC in the face of failures, this example also illustrated how the contribution of resource unreliability to variability in production systems can be quantified.

IV. PERFORMABILITY OF A CELLULAR MANUFACTURING SYSTEM WITH FINITE BUFFERS AND PRIORITIZED REPAIR

In this section, we deal with the performability evaluation of a cellular manufacturing system. First, we describe the architectural features of the system.

A. Architecture of the Cellular Manufacturing System

The cellular manufacturing system (see Fig. 7) consists of a number of cells which are linked together by a Material Handling System (MHS). Each cell consists of a number of machines, and a robot to handle within-cell part movement. Raw workpieces enter the system at the Load/Unload station, where they are fixtured, and are then routed to one of the cells. The MHS moves parts between the Load/Unload station and the various cells. The MHS is typically a conveyor system (which also serves as a centralized storage), or a pool of AGV's with a centralized storage facility. After completion of processing at one of the cells, a finished workpiece is moved by the MHS back to the Load/Unload station, where it is defixtured. The finished workpiece leaves the system, and a raw workpiece takes its place.

The maximum number of workpieces permitted inside a cell at any given time is limited, owing to the finite storage capacity available within the cell. On the other hand, the MHS

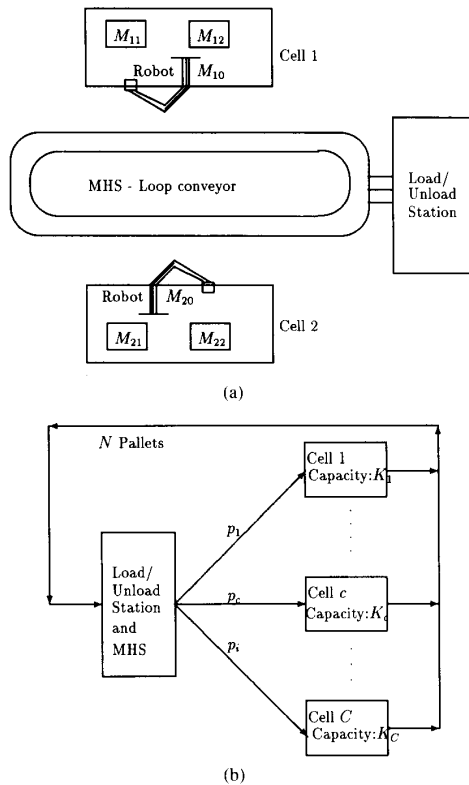


Fig. 7. Architecture of a cellular AMS, and its logical model. (a) A typical cellular AMS. (b) Logical model of a cellular AMS.

along with the central storage incorporates a sufficiently large buffer space so that it can be thought of as possessing infinite storage capacity. Thus, if a workpiece routed to a particular cell finds that the cell is full, it is refused entry (*blocked*), and is routed back to the centralized storage area. For instance, if the MHS is of loop conveyor type, the workpiece will remain in circulation on the conveyor. If the workpiece is blocked from entry to a cell, the MHS treats it as a new job entering the system, and tries to route it to another cell (possibly the same one). This mode of blocking is termed *rejection blocking* [15]. Here, once a workpiece is blocked from entry to a cell, the MHS does *not* stop service; instead it proceeds with its operation on the other workpieces waiting for transport. We also assume that within a cell no further blocking is caused once a workpiece is admitted.

Since each machine in the system is versatile, a particular machine can carry out all the required operations on a workpiece admitted to its cell. Thus, a given cell can maintain production (with a reduced level of throughput) if its material handling robot and at least one of the machines are working; we thus have cell-level fault-tolerance. At the system level, we assume that the cells are functionally equivalent, so that each cell can provide the necessary processing for a workpiece. Hence, one cell is sufficient to maintain production (at a reduced throughput). Thus, the cellular AMS exhibits a high degree of fault-tolerance. We say the manufacturing system is *available* (or, operational) if the MHS and at least one of the

cells are available; in turn, a cell is *available* if its robot and at least one of its machines are available.

Over a specified period of operation, owing to the randomly occurring subsystem failures and subsequent repairs, the cellular AMS will function in different configurations and exhibit varying levels of performance over the (random) residence times in these configurations. In the following, we study the *availability and performability* of the system when it uses centralized repair, with different number of repairmen,

B. Performance model of the Cellular AMS

To carry out a performance analysis for the cellular AMS (Fig. 7(a)), we view it as a central server model, as shown in Fig. 7(b). The combination of the MHS, Load/Unload station, and the central storage is modeled as the central server having infinite buffer capacity, and the cells are modeled as subnetworks with blocking, owing to their finite capacities. We shall refer to the combination of the MHS, the Load/Unload station, and the central storage, as the "MHS". The system description is based on the paper by Dallery and Yao [7].

1) *Modeling Issues*: Let the number of cells be denoted C , and the MHS be denoted cell 0. The constant number of workpieces circulating in the system is denoted N . The number of machines in cell c (including the material handling robot) is denoted $m_c + 1$. The robot in cell c is denoted $M_{c,0}$, and the other machines are denoted $M_{c,1}, \dots, M_{c,m_c}$. The maximum number of workpieces permitted inside cell c is denoted K_c . The number of workpieces inside cell c , denoted k_c , satisfies $0 \leq k_c \leq K_c$.

The system operates as follows: on fixturing, a workpiece is probabilistically routed to cell c , $1 \leq c \leq C$, with probability p_c ; these probabilities satisfy $\sum_{c=1}^C p_c = 1$. Depending on the number of workpieces currently present in cell c , there are two possibilities:

- 1) If $k_c < K_c$, then the workpiece is admitted to the cell, with the material handling robot in the cell removing it from the conveyor. Part routing within the cell follows a classical central server pattern. With probability $p_{c,i}$, the robot routes the workpiece to machine $M_{c,i}$ in cell c (here, $1 \leq i \leq m_c$); or, the workpiece exits the cell (with the robot placing it on the conveyor) with probability $p_{c,0}$. These routing probabilities satisfy $\sum_{i=0}^{m_c} p_{c,i} = 1$. The processing time (part move time in the case of the robot) at $M_{c,i}$ is exponentially distributed with mean $1/\mu_{c,i}$.
- 2) If $k_c = K_c$, the cell capacity is fully occupied; the workpiece is blocked, and remains on the conveyor. The MHS again chooses a destination cell (probabilistically) for the workpiece and tries to route it to that cell.

In all cases, the part move time on the MHS is assumed to be exponentially distributed with mean $1/\mu_0$.

Dallery and Yao [7] have shown that a queuing model of the system is product form under the First Come First Served (FCFS) queuing assumption, in spite of the blocking phenomena present in the model. They have derived an algorithm for the exact performance evaluation of the queuing

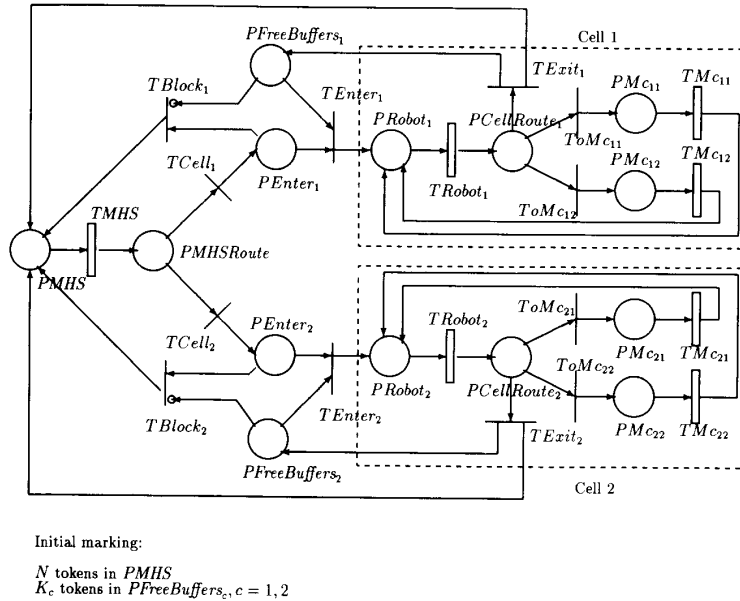


Fig. 8. GSPN model for the performance analysis of the cellular AMS.

model. In this paper, we use a GSPN based model for carrying out a system performance evaluation.

2) *GSPN Model—A Two-Cell System*: We consider a typical system, consisting of two cells ($C = 2$) and a load/unload station, linked together by a centralized loop conveyor type of MHS. Each cell has two machines ($m_c = 2; c = 1, 2$), and a material handling robot. The GSPN model of the system appears in Fig. 8. Table I gives the interpretation of the places and transitions in the GSPN model of Fig. 8. In the initial marking, we have N tokens in $PMHS$, and K_c tokens in each $PFreeBuffers_c$.

The transition $TMHS$ that represents the conveyor operation is modeled as an infinite server with processing rate μ_0 per workpiece, since it is simultaneously moving all the workpieces. Similarly, the number of tokens in the place $PFreeBuffers_c$ indicates the free buffers available in cell c (the marking of this place is the cell buffer capacity K_c minus the number of workpieces already present in cell c). Thus, when the MHS tries to route a part to cell c , it is permitted to enter that cell only if buffer space is available. Otherwise, the workpiece is routed to the conveyor, by firing the immediate transition $TBlock_c$; the inhibitor arc from place $PFreeBuffers_c$ to the $TBlock_c$ models the rejection blocking.

The operation of the robots and the machines in each cell are modeled using standard GSPN techniques. All these act as single server queues, with exponential processing times. When a workpiece leaves a given cell c , by firing transition $TExit_c$, a token is deposited in $PFreeBuffers_c$, indicating the freeing of one buffer in cell c .

C. Structure State Process

We assume an *independent failure model*. Let $\alpha_{c,i}$ ($1 \leq c \leq 2; 0 \leq i \leq m_c$) denote the mean failure rate of machine $M_{c,i}$,

TABLE I
 LEGEND FOR THE GSPN PERFORMANCE MODEL OF THE CELLULAR AMS

Places	
$PMHS$	Workpieces being handled by the MHS
$PHMS\ Route$	Workpiece being routed by MHS
$PFree\ Buffers_c$	Free buffer spaces in cell c
$PEnter_c$	Workpiece attempting to enter cell c
$PRobot_c$	Workpieces waiting for/ moved by the robot of cell c
$PCellRoute_c$	A workpiece in cell c is being routed
PM_c	Workpieces waiting for/ served by machine $M_{c,i}$ in cell c
Transitions	
	Immediate; probability of firing p_c
$TCell_c$	A workpiece is routed to cell c by the MHS
$TBlock_c$	Immediate; a workpiece trying to enter cell c is blocked
$TEnter_c$	Immediate; a workpiece is allowed into cell c
$TExit_c$	Immediate; a finished workpiece leaves cell c
$ToM_{c,i}$	Immediate; probability $p_{c,i}$
	In cell c , a workpiece is routed to $M_{c,i}$
$TMHS$	Timed; firing rate (number of tokens in $PHMS$) $\times \mu_0$; Models the infinite server operation of the MHS
$TM_{c,i}$	Times; firing rate $\mu_{c,i}$ Models the processing (or part move) time in $M_{c,i}$

and let its mean repair time be $1/\beta_{c,i}$; and, let α_0 and β_0 denote respectively the failure and repair rate of the MHS. All failure and repair times are assumed to be exponentially distributed. A pool of identical repairmen is available to work on the failed subsystems. A repairman can work on only one failed subsystem at a time. A failed subsystem can use the services of only one repairman. For example, if only one machine has failed, and there are two idle repairmen, one repairman starts working on the failed machine, and the other

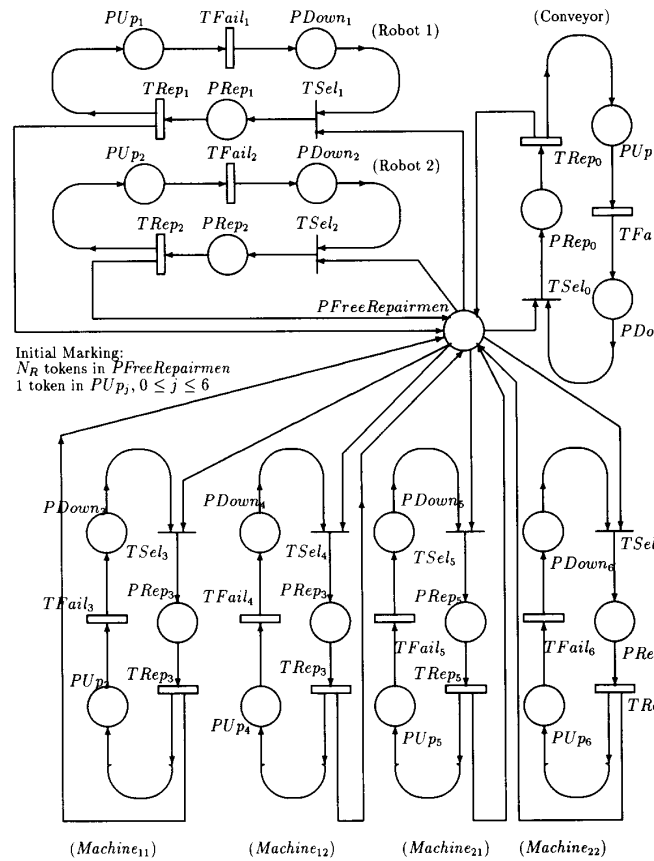


Fig. 9. GSPN model for the Structure state process of the cellular AMS.

repairman continues to be idle. Let N_R denote the number of repairmen in the system.

The following *non-preemptive priority* rule is applied for selecting the item for next repair:

- 1) MHS: priority level 3 (highest priority)
- 2) Robots, M_{10} and M_{20} : priority level 2
- 3) Other machines, M_{11} , M_{12} , M_{21} , M_{22} : priority level 1 (lowest priority)
- 4) Among failed units with the same priority level, each failed unit is chosen with equal probability, provided no higher priority items are waiting for repair.

The reason for the above priority allocation is obvious when one notes that the conveyor is a single point failure for the entire system, the robot for the cell.

Figure 9 shows the GSPN model of the structure state process; it models the failure and repair of the different subsystems. Table II gives the interpretation of the places and transitions of this GSPN, and also indicates the assignment of the numerical suffix to each place/transition. Place PUP_0 indicates that the MHS is working; timed transition $TFai_0$ models the random failure of the MHS; place $PDown_0$ indicates that the MHS has failed, and is awaiting repair; the immediate transition $TSel_0$ represents choosing the MHS for repair; place $PRep_0$ indicates that the MHS is under repair;

and, the timed transition $TRep_0$ models the repair on the MHS. Similarly named places and transitions with indices $i = 1, \dots, 6$ represent the corresponding status/activity with respect to the other subsystems. Initially, all the seven subsystems are assumed to be functioning, and all repairmen are idling. Thus, in the initial marking, a token is placed in each of the places PUP_i , $i = 0, \dots, 6$; the number of tokens in the place $PFReeRepairmen$ equals N_R .

In order to carry out the availability and performability analysis on this GSPN, we use the notion of a *GSPN reward model* developed by Ciardo *et al.* [4]. In Section II, we modeled the SSP as a Markov process, and associated rewards with the states of the Markov process. Here, we are using a GSPN model of the SSP. Since GSPNs are isomorphic to continuous-time Markov chains, a tangible marking s of the GSPN is a typical structure-state of the SSP, and the state space of the SSP of the performability model of the cellular AMS is the set S of the tangible markings [1] of the GSPN of Fig. 9. One can assign rewards with the tangible markings of the GSPN and obtain a stochastic (Markovian) reward model.

Modeling the SSP with a GSPN rather than a Markov chain directly is advantageous in the same fashion as a GSPN performance model is easier to use than a direct Markovian performance model. A GSPN model with reward

TABLE II
LEGEND FOR THE GSPN MODEL OF THE SSP OF THE CELLULAR AMS

Index assigned for each subsystem in the GSPN							
Index	0	1	2	3	4	5	6
Subsystem	MHS	Robot, M_{10}	Robot, M_{20}	M_{11}	M_{11}	M_{11}	M_{11}
Places							
PU_{p_i}	Subsystem i is working						
$PDown_i$	Subsystem i has failed, and is awaiting repair						
$PREP_i$	Subsystem i under repair						
$PFreeReparations$	Idle repairmen						
Transitions							
$TFail_i$	Timed transition; models breakdown of subsystem i						
$TRep_i$	Timed transition; models repair of subsystem i						
$TSel_i$	Immediate; failed subsystem i taken up for repair						
Priorities of transitions							
Transition	Priority						
$TTSel_0$	3						
$TTSel_1, TTSel_2$	2						
$TTSel_3, TTSel_4, TTSel_5, TTSel_6$	1						
$All\ timed\ transitions$	0						

specification may be converted to an underlying Markovian reward model, using (automated) techniques very similar to those used to convert a GSPN performance model into an underlying continuous-time Markov chain [4]. Once we obtain the underlying Markovian reward model, it can be analysed for various performability measures by well-established techniques, such as those referred to in Section II.

D. Availability Analysis

For computing availability measures, a reward rate of either 0 or 1 is to be associated with each tangible marking of the GSPN model of the SSP. Let $\#(PLACE, s)$ denote the number of tokens in place "PLACE" in tangible marking s . Cell c is available in this marking if the robot and at least one of the machines in the cell are functioning. Hence, we have

$$AVAIL(\text{Cell}_1, s) = \begin{cases} 1 & \text{if } (\#(PU_{p_1}, 0) = 1) \text{ AND } ((\#(PU_{p_3}, s) = 1) \\ & \text{OR } (\#(PU_{p_4}, s) = 1)) \\ 0 & \text{Otherwise} \end{cases}$$

and

$$AVAIL(\text{Cell}_2, s) = \begin{cases} 1 & \text{if } (\#(PU_{p_2}, 0) = 1) \text{ AND } ((\#(PU_{p_5}, s) = 1) \\ & \text{OR } (\#(PU_{p_6}, s) = 1)) \\ 0 & \text{Otherwise} \end{cases}$$

TABLE III
FAILURE AND REPAIR TIMES (IN HOURS) OF THE SUBSYSTEMS

Equipment	MHS	Cell 1		Cell 2			
	Conveyor	Robot, M_{10}	M_{11}	M_{12}	Robot, M_{20}	M_{21}	M_{22}
Failure time, $1/\alpha$	480	96	72	72	64	72	72
Repair time, $1/\beta$	40	15	12	12	15	12	12

In marking s , the entire system is available (functioning) if the MHS and at least one of the two cells are available:

$$AVAIL(\text{AMS}, s) = \begin{cases} 1 & (\#(PU_{p_0}, s) = 1) \text{ AND } ((AVAIL(\text{Cell}_1, s) = 1) \\ & \text{OR } (AVAIL(\text{Cell}_2, s) = 1)) \\ 0 & \text{Otherwise} \end{cases} \quad (3)$$

For availability analysis, set the reward associated with state $s \in S$ to be

$$\gamma_s = AVAIL(\text{AMS}, s)$$

With this reward assignment, one can evaluate various availability measures. The following numerical example considers cumulative availability measures, computed using the randomization algorithm [8] on the underlying Markovian chain of the GSPN of Fig. 9.

1) *Numerical Example 4.1—Distribution of Cumulative Operational Time:* We assume the values of the mean failure times and mean repair times of the various equipment given in Table III. We consider the system operation over a six-day week, working two eight-hour shifts a day, i.e., $t = 96$ hours.

We study the impact of the number of repairmen, N_R , on the attainable levels of system availability. All the subsystems are assumed to be properly functioning at the start. We study the (complementary) distribution of the cumulative operational time over the operation period of 96 hours; Fig. 10 gives this figure of merit for $N_R = 1, 2, 3$ repairmen. This graph shows the probability achieving a system uptime of u hours over the $t = 96$ hour duration, where u takes values in $[0, t]$. From this graph, we see that at least two repairmen are necessary to give an uptime of $u = 72$ hours with a probability of 0.8; a single repairman is unable to satisfy this requirement.

E. Performability Evaluation

In this section, we consider the evaluation of the mean and variance of performability measures. A potential difficulty in the performability evaluation of the cellular AMS is the requirement to solve a large number of performance models of the system, owing to the large state-space of the SSP. For instance, even for the non-repairable version of the system, there are 128 structure states. However, a closer look at the problem reveals that 1) several of the structure states are "failed" states, so the reward assignments are obtained trivially; and, 2) among the "working" structure states, several are equivalent in the sense that they give rise to the same performance model. More specifically, to obtain the throughput in a given marking s , it is necessary to consider only the equipment that are in working condition. Hence, we have to consider the different

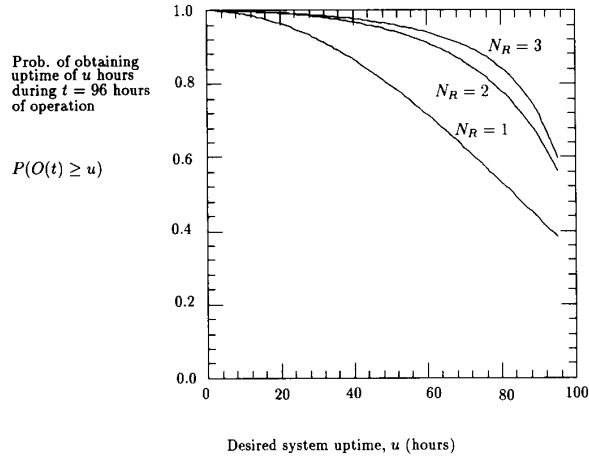


Fig. 10. Complementary distribution of the cumulative operational time.

TABLE IV
SIZE OF THE STATE-SPACE OF THE SSP FOR DIFFERENT NUMBER OF REPAIRMEN

No. of repairmen	0	1	2	3
No. of states in SSP	128	449	680	589

markings of only a subset of the places in the GSPN i.e., the markings of only the seven places “PUp0,” through “PUp6.” As an illustration of the first observation, for the non-repairable system, 64 possible markings with “PUp0” empty, denoting the MHS has failed, are failure states, and the throughput in those states is 0. To illustrate the second observation, assume that the MHS is functioning, all the equipment in cell 1 is working, and that the robot M_{20} of cell 2 has failed; there are four possible markings for the status of the machines M_{21} and M_{22} of cell 2, but these four states are equivalent as far as the performance model is concerned.

There are three possible working configurations for each cell: 1) Robot and both machines operational; 2) robot and first machine operational, but second machine failed; and 3) robot and second machine operational, but first machine failed. The number of distinct working configurations of the AMS is easily determined as follows:

- 1) Both cell 1 and cell 2 working: $3 \times 3 = 9$ configurations.
- 2) Cell 1 up, cell 2 down: 3 configurations.
- 3) Cell 2 up, cell 1 down: 3 configurations.

Hence, the number of separate runs of the performance model is only 15, one for each “working configuration” of the manufacturing system; this figure is much smaller than the state space of the SSP. For comparison, we show the size of the state space of the SSP for various values of N_R in Table IV.

A key feature to be considered is the choice of routing parameters dependent on the structure state of the system; for instance, if all the equipment in cell 1 are working, but one of the machines in cell 2 has failed, we would like to direct more workpieces to cell 1 than to cell 2. The following numerical example illustrates a specific probabilistic routing policy that is changed adaptively with structure state changes.

TABLE V
THE AVERAGE PROCESSING TIMES USED IN THE PERFORMANCE MODEL

Equipment	MHS	Robot			Robot		
		M_{10}	M_{11}	M_{12}	M_{20}	M_{21}	M_{22}
Processing time (min)	8	1	2	4	1	3	5

1) *Numerical Example 4.2—Performability Analysis:* In this example, we compute the performability measures for the system with different number of repairmen.

2) *System Parameters:* Table V summarizes the average processing times (move time for MHS and robots) used in the performance model.

The constant number of fixtures circulating in the system is $N = 10$. We assume that cell 1 can permit at most five workpieces to be present at a time ($K_1 = 5$), and cell 2 allows a maximum of 6 workpieces ($K_2 = 6$).

The routing probabilities under *failure-free conditions* are:

- 1) To the cells, from the MHS: $p_1 = 0.5$, $p_2 = 0.5$
- 2) In cell 1: $p_{1,0} = 0.2$; $p_{1,1} = 0.5$; and, $p_{1,2} = 0.3$
- 3) In cell 2: $p_{2,0} = 0.25$; $p_{2,1} = 0.55$; and, $p_{2,2} = 0.2$

These routing probabilities satisfy the condition $\sum_{i=0}^{m_c} p_{c,i} = 1$.

The following scheme is used for routing when some of the resources have failed:

- 1) Routing within the cells: Assume cell c is working but exactly one $M_{c,i}$, ($i = 1, 2$) has failed. In this case the working machine carries all processing load; e.g., if $M_{c,1}$ is working and $M_{c,2}$ has failed, set $p_{c,1} = 1 - p_{c,0}$.
- 2) Routing from the MHS to the cells:
 - a) If only one cell is working, it carries all workpieces are routed to it; e.g., if cell 1 is up and cell 2 is down, $p_1 = 1$, and $p_2 = 0$.
 - b) If both cells are working, and both have the same number of working machines, the routing probabilities to either cell are equal. But, if one cell has 2 working machines and the other has only one, the MHS routes a workpiece to the former with probability 2/3, and to the latter with probability 1/3.

As mentioned earlier, this choice of routing probabilities tries to allocate more workpieces to the cell with a higher number of operational equipment. One can also consider alternative routing strategies under failure conditions.

3) *Throughput-Related Performability—Moments of Cumulative Production:* Figure 11 shows the expected cumulative production over a 96 hour period, with one, two, and three repairmen. As expected, the average cumulative production increases with the number of repairmen employed. Figure 12 shows the coefficient of variation of cumulative production (standard deviation divided by mean). The use of more repairmen has not only increased the cumulative production, but also reduced the degree of variability induced by equipment failures.

Another important feature is the time-dependence of the coefficient of variation during the 96 hour period considered; it is lower for shorter intervals, and increases as time progresses. This has implications for maintenance planning. We mentioned

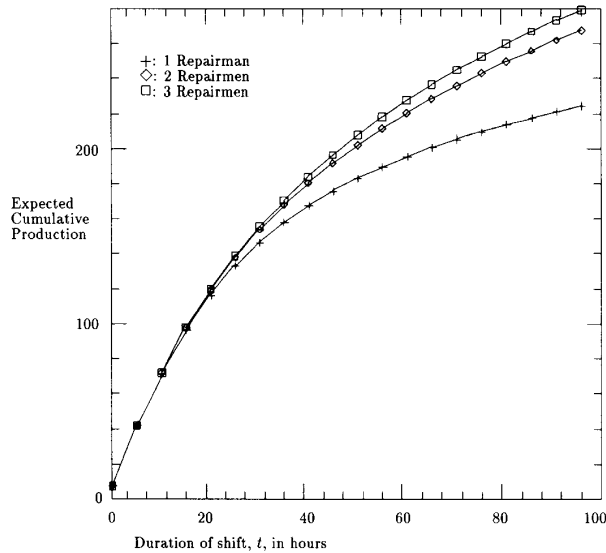


Fig. 11. Expected cumulative production of the cellular AMS over various subintervals.

that this system works two eight-hour shifts per day. Consider the three repairmen case. If preventive maintenance is carried out in the third shift every day (that is after 16 hours of operation), so that all resources are restored to fully operational condition, the coefficient of variation can be kept under 0.28. If preventive maintenance is carried out every alternate day (that is after 32 hours of operation), the coefficient of variation can be maintained at around 0.36. If the preventive maintenance is carried out after every 3 days (48 working hours), the coefficient of variation can be maintained at 0.45. The numerical results shown here how maintenance can be planned to control the degree of randomness introduced by resource failures.

Owing to the relatively large dimension of the state space of the SSP, and the *repairable* nature of the system (leading to an *irreducible* Markov chain model for the SSP) we do not attempt to compute *distributions* of performability measures. For instance, the randomization-based algorithm of Silva and Gail [9] has a computational complexity that is *exponential* in the number of reward rates. For the cellular AMS, we have 16 different reward rates (15 different working configurations, and a failure configuration), and a software implementation will be prohibitively expensive (computationally). Although other numerical techniques with polynomial-time complexity have been reported [18] for performability distribution evaluation for repairable systems, the size of the state space in our problem is much higher than the maximum state space size considered in the literature.

V. CONCLUDING REMARKS

This paper presented the performability framework for analysing the performance of AMS's in the face of subsystem failures. We showed how this framework provides for unified, quantitative analysis of the availability, throughput, and manufacturing lead-time related performance of a failure-

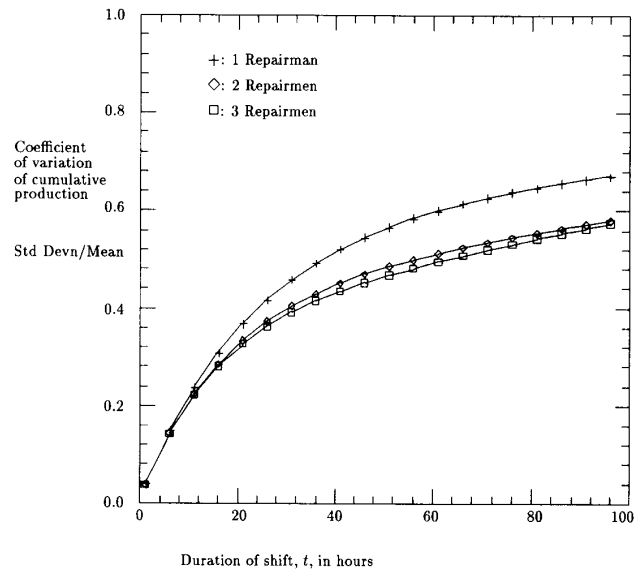


Fig. 12. Coefficient of variation of cumulative production of the cellular AMS.

prone manufacturing system, through the concept of structure state processes and appropriately chosen rewards.

We presented performability results for a flexible manufacturing cell, and a two cell system with blocking and prioritized, centralized repair, along with several numerical examples. By computing the first and second moments of the throughput-oriented performability measures, we also quantitatively studied the variability induced by equipment failures on system production.

The results reported here can form the basis of several enhancements, such as conducting performability studies with multiple part types, and developing efficient techniques for

computing the distribution of performability measures, thereby giving a complete quantitative study of the performance of failure-prone manufacturing systems.

ACKNOWLEDGMENT

We thank Prof. K. S. Trivedi, Duke University, for providing us the SPNP and SHARPE software tools which were utilized to obtain several of the numerical results presented here. Special thanks are due to Mr. V. Gopalakrishna, Dept. of Computer Science and Automation, Indian Institute of Science, for his help in the computation of the moments of performability measures. Thanks are also due to Dr. Y. Narahari, Indian Institute of Science, and Prof. Krishna Pattipati, University of Connecticut, for several useful discussions. We also thank the Department of Science and Technology, Government of India, for supporting this research under the project "Reliability Analysis of Automated Manufacturing Systems."

REFERENCES

- [1] M. Ajmone Marsan, G. Conte, and G. Balbo, "A Class of Generalised Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems," *ACM Transactions on Computer Systems*, vol. 2, no. 2, 1984.
- [2] V. Albino, G. O. Okogbaa, and R. L. Shell, "Computerized Integrated Performance—Reliability Measure of a Flexible Automated Production System," *Computers and Industrial Engineering*, vol. 18, no. 4, pp. 547–558, 1990.
- [3] M. Beaudry, "Performance Related Reliability Measures for Computer Systems," *IEEE Transactions on Computers*, vol. 27, no. 6, pp. 248–255, 1978.
- [4] G. Ciardo, J. Muppala, and K. S. Trivedi, "On the Solution of GSPN Reward Models," *Performance Evaluation*, vol. 12, pp. 237–253, 1991.
- [5] Y. Dallery, R. David, and X.-L. Lie, "An Efficient Algorithm for Analysis of Transfer Lines with Unreliable Machines and Finite Buffers," *IIE Transactions*, vol. 20, pp. 280–283, 1988.
- [6] Y. Dallery and S.B. Gershwin, "Manufacturing Flow Line Systems: A Survey of Models and Analytical Results," LMP Report, Cambridge, MA, MIT, 1992.
- [7] Y. Dallery and D. D. Yao, "Modeling a system of Flexible Manufacturing Cells," in *Modelling and Design of Flexible Manufacturing Systems*, A. Kusiak, Ed. Elsevier Science Publishers, B.V., 1986, pp. 289–300.
- [8] E. de Souza e Silva and H. R. Gail, "Calculating Cumulative Operational Time Distributions of Repairable Computer Systems," *IEEE Transactions on Computers*, vol. 35, no. 4, pp. 322–332, 1986.
- [9] E. de Souza e Silva and H. R. Gail, "Calculating Availability and Performability Measures of Repairable Computer Systems Using Randomisation," *Journal of the Association for Computing Machinery*, vol. 36, no. 1, pp. 171–193.
- [10] L. Donatiello and B. R. Iyer, "Analysis of a Composite Performance Reliability Measure for Fault-Tolerant Systems," *Journal of the Association for Computing Machinery*, vol. 34, no. 1, pp. 179–199, 1987.
- [11] S.B. Gershwin, "Variance of Output of a Tandem Production System," Technical Report LMP-92-003, Laboratory for Manufacturing and Productivity, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1992.
- [12] S. M. R. Islam and H. H. Ammar, "Performability Analysis of Distributed Real-Time Systems," *IEEE Transactions on Computers*, vol. 40, no. 11, pp. 1239–1251, 1991.
- [13] B. R. Iyer, L. Donatiello, and P. Heidelberger, "Analysis of Performability for Stochastic Models of Fault-Tolerant Systems," *IEEE Transactions on Computers*, vol. 35, no. 10, pp. 902–907, 1986.
- [14] J. F. Meyer, "On Evaluating the Performability of Degradable Computing Systems," *IEEE Transactions on Computers*, vol. 29, no. 8, pp. 720–731, 1980.
- [15] R. O. Onvural, "Survey of Closed Queuing Networks with Blocking," *ACM Computing Surveys*, vol. 22, no. 2, pp. 83–121.
- [16] R. Ram, "Performance and Performability Analysis of Automated Manufacturing Systems," Ph.D. Thesis, Dept. of Computer Science and Automation, Indian Institute of Science, Bangalore, 1992.
- [17] A. Reibman, R. Smith, and K. S. Trivedi, "Markov and Markov reward model transient analysis: An overview of numerical approaches," *European Journal of Operational Research*, vol. 40, pp. 257–267, 1989.
- [18] R. M. Smith, K. S. Trivedi, and A. V. Ramesh, "Performability Analysis: Measures, an Algorithm and a Case Study," *IEEE Transactions on Computers*, vol. 37, no. 4, pp. 406–417, 1988.
- [19] N. Viswanadham, Y. Narahari, and R. Ram, *Advances in Manufacturing and Automation Systems*, vol. 47, part 3, of *Advances in Control and Dynamic Systems*. Academic Press, chap. "Performability of Automated Manufacturing Systems," 1991, pp. 77–120.



N. Viswanadham, (SM'86-F'93), received the Ph.D. degree in 1970 from the Indian Institute of Science, Bangalore, India. N. Viswanadham is a TataChem Professor and Chair of the Department of Computer Science and Automation, of the Indian Institute of Science.

Dr. Viswanadham has held visiting appointments at several North American universities. He was a GE Research Fellow at the corporate research center in 1989. His current research interests are in the areas of fault tolerant control system design and modeling and control of flexible manufacturing systems. He is the author of several journal articles and conference papers. He is a joint author of two textbooks: *Reliability in Computer and Control Systems* (North-Holland, 1987) and *Performance Modeling of Automated Manufacturing Systems* (Prentice-Hall, 1992). He is co-editor of four other books. He is currently editor of *Sadhana: Academy Proceedings in Engineering Sciences*. He is also Associate Editor of the journals *Control Theory and Advanced Technology*, *Journal of Manufacturing Systems*, *Information and Decision Technologies*, *Intelligent and Robotic Systems*, *Asia-Pacific Engineering Journal*, and *Systems Automation: Research and Applications*. He was Associate Editor-at-Large for the *IEEE Transactions on Automatic Control*, in 1990 and 1991. He is a Fellow of the IEEE, The Indian National Science Academy, the Indian Academy of Sciences, and the Indian National Academy of Engineering.



R. Ram was born in Madras, India, in May, 1965. He obtained the B.E. (Honours) degree in Electronics and Communication Engineering, with Distinction, from the Madurai Kamaraj University, India, in 1986, and was awarded the University Gold Medal. Subsequently, he obtained the M.E. and Ph.D. degrees in Computer Science and Engineering, from the Indian Institute of Science, Bangalore, in January 1988 and February 1992, respectively. He served as the system administrator at the Intelligent Systems Laboratory, Department of Computer

Science and Automation, Indian Institute of Science, during 1989–1992. He is now employed as Program Manager for Performance Engineering at the Centre for Development of Telematics at Bangalore. Currently, he is leading a team of engineers investigating performance and dependability issues in large switching systems. He has several publications in prestigious international journals and refereed international conferences. His research interests are in the areas of performance modeling and evaluation, flexible manufacturing systems, and object-oriented techniques for discrete event simulation.