April, 1986

LIDS-P-1555
VLSI 86-314

Stochastic Scheduling and Set-Ups in Flexible Manufacturing Systems

by

Stanley B. Gershwin

35-433
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

Abstract

This paper describes an approach for the incorporation of set-up times into a stochastic scheduling algorithm for a FMS. Set-up changes should not be performed too often because of the resulting reduction of capacity. They should not be performed too infrequently, because of the resulting increases of inventories and delays. The goal of the analysis reported here is to calculate the long-term average frequencies of set-ups and the fraction of time the system should be set up for each part type or family of part types.

To Appear in the Proceedings of the Second ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, edited by K. Stecke and R. Suri.

1. Introduction

In this paper, we describe an approach for the incorporation of set-up times into an earlier stochastic scheduling algorithm (Kimemia, 1982, and Kimemia and Gershwin, 1983) for a Flexible Manufacturing System (FMS). As before, the objective is to develop a feedback law which will respond to and mitigate the effects of random, potentially disruptive events.

This approach to scheduling differs from conventional mixed integer programming representations (Graves, 1981). In such formulations, there are a large number of binary or integer variables (for example, Afentakis, Gavish, and Karmarkar, 1984; Karmarkar and Schrage, 1985). These approaches model the system in detail, but their large computational requirements make them difficult to solve and interpret, and adding stochastic phenomena does not make them any easier. Maxwell and Muckstadt (1985) simplify the problem by dealing only with reorder intervals (i.e., set-up frequencies) and ignoring capacity questions. Kusiak, Vanelli, and Kumar (1985) treat only the grouping problem.

In the work reported here and in Kimemia (1982) and Kimemia and Gershwin (1983), the approach is to gain as much information as possible from a continuous representation of the flow of material in the system. Only after solving the continuous problem is the detailed discrete problem treated. The detailed scheduling problem is then much easier than it would be if it were treated without first solving the continuous problem.

In that earlier work, however, set-up times were not considered. That is, it was assumed that no time or cost was required for changing tools or other resources that are specific to a limited set of part types. In reality, however, even a FMS has limitations on its flexibility. The zero-set-up time assumption may be adequate for a small set of parts comprising a family, but not for selecting times to change configurations so that a new family can be produced. By <u>configuration</u> we mean a set of tooling that limits the FMS to doing operations on a specific part type or family of part types. To make parts that are in other families, the configuration must be changed.

In this paper, we assume that there is a non-zero set-up time. That is, some time is required to change the configuration of the system from making one part type (or family of part types) to making another. This is an issue of practical importance which arises in metal cutting systems, in which tool magazines can hold only a limited number of tools; printed circuit card assembly systems in which each insertion machine can hold only a limited number of types of electronic components; and VLSI fabrication, in which furnaces must be cleared of each kind of impurity before the next kind of impurity can be used.

It is not desirable to change the configuration too often because that reduces the amount of time available for productive work and thus the capacity. On the other hand, it is not desirable to change it too infrequently, because that will tend to increase inventories and delays before deliveries. The major innovation reported here is the explicit calculation of capacity for scheduling problems with non-zero set-up times.

The goal of the analysis is to calculate the long-term average frequencies of set-ups and the average fraction of time that the system is in each configuration (i.e., the fraction of time it is set up for each part type or family of part types). Further work at the detailed level is required to translate these quantities into times at which to perform set-ups.

An essential element of this analysis is the explicit recognition that different events take place at greatly different time scales, i.e., frequencies. At the higher level of the scheduling hierarchy, we calculate set-up frequencies, average production rates, and other long term quantities. The lower level has the responsibility of choosing times to perform set-ups and dispatch parts in a way that agrees with the high level rates. Because the high level analysis has no way of determining exactly when a set-up will take place, but instead only the rates of set-ups, it is reasonable to use the tools of probability.

The relationship of this work to longer and shorter term issues is described in Gershwin (1986).

In Section 2, we describe a formulation in which the set-up frequency is much less that that of the failures and repairs and production operations. Two examples are worked out in Section 3. Section 4 briefly describes the modifications to the analysis when set-up frequencies are comparable to repair and failure frequencies. We conclude in Section 5.

2. Infrequent Set-Ups

In this section, we consider the problem of devising a feedback law for the scheduling of a FMS. We wish to find the optimal production rate for each part type and the frequencies of set-ups. We also seek the fraction of time that the system spends in each configuration.

Time Scales

We assume that set-up time is not negligible compared to operation time. The horizon is long enough so that reliability behavior (i.e., repairs and failures of machines) may be replaced by average machine efficiency (or possibly by a random variable indicating the amount of material produced during an interval; parts are made often enough so that part flow may be approximated by a continuous function of time; and set-ups occur very many times during the problem horizon. The problem may be assumed to have at least two important time scales. The short time scale is that of machine failures and repairs which is treated by Kimemia and Gershwin (1983) and Gershwin, Akella, and Choong (1985). In that time scale, the system configuration never changes and so set-ups never occur. Changes in configuration take place over the long time scale.

Notation

S = continuous long time scale index variable.

s_{mk} = the time required to change configuration from m to k; this is specified.

w_m = the fraction of time the system is in configuration m (not including time during set-ups); this is a decision variable.

f_{mk} = the frequency that the configuration changes from m to k. That is, $f_{mk}\delta S$ is the probability that an observer sees the system change from configuration m to configuration k during the interval $(S, S + \delta S)$; this is a decision variable.

ϕ_{mk} = the frequency that the configuration changes from m to k while system is in configuration m. That is, $\phi_{mk}\delta S$ is the probability that an observer sees the system in configuration m change from configuration m to configuration k during the interval $(S, S + \delta S)$; this is a decision variable. Note that

$$f_{mk} = \phi_{mk}\, w_m \tag{2.1}$$

Basic Equations for Set-up Times in a Stochastic Environment

By accounting for the total time the system can be in any configuration or having its configuration changed, we have

$$\sum_m w_m + \sum_{m,k} s_{mk} f_{mk} = 1\ . \tag{2.2}$$

The fraction of time $w_k(S)$ that the system is in configuration k at time S satisfies

$$w_k(S+\delta S) = w_k(S) + \sum_{m\neq k} \phi_{mk}\, w_m(S)\delta S - \sum_{m\neq k} \phi_{km} w_k(S)\delta S\ .$$

Then, in steady state,

$$\sum_{m\neq k} \phi_{mk}\, w_m = \left(\sum_{m\neq k} \phi_{km} \right) w_k\ . \tag{2.3}$$

Interpretation

Equations (2.1)-(2.3) are similar to the equations for the steady-state probability distribution of a Markov process (in which w_m is analogous to the probability of finding the system in state m, and $\phi_{mk}\delta S$ is the probability of a transition from m to k in an interval of length δS.) The principal difference is the set-up time term in (2.2), which prevents the probabilities from summing to unity.

In the usual context, ϕ_{mk} is known. There are then enough equations to determine w_m (except in singular cases). Here, however, both ϕ_{mk} and w_m must be determined. Additional information is required.

Production Requirements, Capacity Constraints, and Inventory

Frequent set-ups reduce capacity and make production requirements infeasible. Infrequent set-ups increase cycle times (the time a part spends in the factory) and inventory. To describe these issues, we must introduce additional notation. The subscript L refers to long time scale quantities, to distinguish quantities here from similar short term quantities in Kimemia and Gershwin (1983), Gershwin, Akella, and Choong (1985), and Akella, Choong, and Gershwin (1984).

$\tau_{ij}(m)$ = the time that part type j requires at machine i while the system is in configuration m; this is specified.

u^m_{Lj} = the production rate of type j parts while the system is in configuration m; this is a decision variable.

$\bar{\Omega}^m_L$ = the long time scale capacity set. The long time scale production rate vector satisfies the long time scale capacity constraints:

$$u^m_L(S) \in \bar{\Omega}^m_L = \left\{ u \;\middle|\; u_j \geq 0,\ \sum_j \tau_{ij}(m)\, u_j \leq E\alpha_i \right\} \tag{2.4}$$

while the system is in configuration m. Here α_i refers to the repair state of machine i, and $E\alpha_i$ is the average availability of machine i over the time scale treated here. This issue and notation are discussed in Kimemia and Gershwin (1983) and Gershwin, Akella, and Choong (1985).

d_{Lj} = the long term average demand for part type j; this is specified. In order to satisfy demand, we must have

$$d_{Li} = \sum_m w_m\, u^m_{Li}\ . \tag{2.5}$$

Equation (2.2) indicates that as f_{mk} increases, w_m decreases. If w_m is too small (i.e., if the system does not spend enough time producing), then (2.5) may not be satisfied for any feasible u. Therefore, (2.5) imposes an upper limit on set-up frequency.

$x_{Lj}(S)$ = the difference between cumulative production and cumulative demand for type j at time S; this is a decision variable.

The dynamics of x are

$$\dot{x}_{Lj} = u_{Lj}^{m(S)} - d_{Lj} \ . \tag{2.6}$$

where m(S) is the configuration at time S.

An optimization problem is (2.1)-(2.4), (2.6), and

$$J(x_L(S)) = \int_S^T g_L^x(x_L(s))ds \tag{2.7}$$

where g is a measure of cost of surplus (x positive) or backlog (x negative). This is similar to the problem in Kimemia and Gershwin (1983) in which flow rates are chosen to respond to changes in machine repair states, rather than configurations.

Implementation: The solution is $u_L^m(S)$, w, and f or ϕ. ϕ indicates how frequently to change configurations. An algorithm must be found to determine the actual set-up instants so that the actual frequencies are as close to this as possible. One possible approach is to follow the earlier Gershwin, Akella, and Choong (1985) method for the lower level. Here, we let

$$W_m(S) = w_m S \ . \tag{2.8}$$

This is the calculated amount of time that the system should have been in configuration m in (0,S). Let $W^N{}_m(S)$ be the actual amount of time the system has been in configuration m between times 0 and S. Change the configuration to m whenever

$$W^N{}_m(S) \le W_m(S). \tag{2.8}$$

An alternative rule would be to change set-ups when all the x_{Li} of the current configuration have reached a specified value.

Relationship with earlier work: Figure 2.1 shows the relationship of the work described in this appendix with the earlier (Kimemia and Gershwin, 1983) algorithm. The long term average demand rate d_L is specified at a still higher level. The problem (2.7) is solved and the results are used in two ways: u_L^m is used as the demand rate d for the lower level algorithm; and ϕ and w are used to calculate set-up frequencies and actual set-up times. When the configuration (m) and the average demand rate d are specified, the lower level (k-g) algorithm can be run.

3. Examples with Infrequent Set-Ups

Example 1

Consider a flexible manufacturing system that makes three part types in two configurations. The system consists of one machine that has holders for two different tools. In configuration 1, the system can make types 1 and 2; in configuration 2, it can make types 1 and 3. It takes one hour on the average to do an operation on any part (after reliability is taken into account), so the capacity sets are:

For $m=1$,

$$u_1^1 + u_2^1 \leq 1 \tag{3.1}$$

where the superscript refers to the system configuration, and for $m=2$,

$$u_1^2 + u_3^2 \leq 1 \ . \tag{3.2}$$

The demands are $d_1 = .3$; $d_2 = .2$; $d_3 = .4$ parts per hour. Instead of considering the dynamic optimization problem, we treat the system in steady state. To satisfy demand, we must have

$$\left.\begin{aligned} w_1u_1^1 + w_2u_2^1 &= .3 \\ w_1u_2^1 \qquad\quad &= .2 \\ w_2u_3^2 &= .4 \end{aligned}\right\} \tag{3.3}$$

Because there are only two configurations, and because we assume that the system is in steady state, (2.3) becomes

$$\phi_{12}w_1 = \phi_{21}w_2 \ . \tag{3.4}$$

It takes s_{12} hours to change the tools from configuration 1 to 2, and s_{21} hours to change back. The normalization equation (2.2) is then

$$w_1 + w_2 + s_{12}\phi_{12}w_1 + s_{21}\phi_{21}w_2 = 1 \ . \tag{3.5}$$

The last two equations imply that

$$\begin{aligned} w_1 &= \frac{\dfrac{1}{\phi_{12}}}{\dfrac{1}{\phi_{12}} + \dfrac{1}{\phi_{21}} + s_{12} + s_{21}} \\ w_2 &= \frac{\dfrac{1}{\phi_{21}}}{\dfrac{1}{\phi_{12}} + \dfrac{1}{\phi_{21}} + s_{12} + s_{21}} \end{aligned} \tag{3.6}$$

Equation (3.3) now becomes

$$\frac{u_1^1}{\phi_{12}} + \frac{u_1^2}{\phi_{21}} = .3\left(\frac{1}{\phi_{12}} + \frac{1}{\phi_{21}} + s_{12} + s_{21} \right) \tag{3.7}$$

$$\frac{u_2^1}{\phi_{12}} = .2\left(\frac{1}{\phi_{12}} + \frac{1}{\phi_{21}} + s_{12} + s_{21} \right) \tag{3.8}$$

$$\frac{u_3^2}{\phi_{21}} = .4\left(\frac{1}{\phi_{12}} + \frac{1}{\phi_{21}} + s_{12} + s_{21}\right) \quad (3.9)$$

From (3.3),

$$w_1 + w_2 \geq .9 \quad (3.10)$$

or,

$$\frac{1}{\phi_{12}} + \frac{1}{\phi_{21}} \leq 9(s_{12} + s_{21}) \quad (3.11)$$

Instead of dealing with optimization problem (2.7), we assume that (3.1), (3.2), and (3.11) are satisfied with equality. We now have 6 equations ((3.1), (3.2), (3.7), (3.8), (3.9), (3.11)) in 6 unknowns (u_1^1, u_2^1, u_1^2, u_3^2, ϕ_{12}, ϕ_{21}). However, there is one redundant equation among them. The additional equation can be supplied by a simplified lower level model in which we minimize an estimate of the buffer space required due to batching. The results are

$$w_1 = w_2 = .45$$

$$\phi_{12} = \phi_{21} = \frac{2}{9(s_{12} + s_{21})}$$

$$u_1^1 = \frac{5}{9} \qquad u_2^1 = \frac{4}{9}$$

$$u_1^2 = \frac{1}{9} \qquad u_2^1 = \frac{8}{9}$$

The system's operation proceeds as follows: it may start in configuration 1, for example, producing type 1 and 2 pieces at rate 5/9 and 4/9 pieces per hour, respectively. These quantities are averages since we have assumed that the system actually suffers numerous repairs and failures while it is in each configuration. After about $2/(9(s_{12} + s_{21}))$ hours, we start changing over to configuration 2. This takes s_{12} hours. Production starts up again in configuration 2 in which type 1 and 3 pieces are made at rates 1/9 and 8/9 pieces per hour. The system stays in this configuration for about $2/(9(s_{12} + s_{21}))$ hours, and then the next changeover begins. It takes s_{21} hours and the system is in configuration 1 again.

Note that only the sum of the set-up times $s_{12} + s_{21}$ is important in this problem, and not the individual times. As the sum increases, set-ups are performed less often.

The decision of precisely when the changeover will take place is left up to the lower level. The quantities calculated here are requirements that the lower level must satisfy.

Example 2

Consider a machine that does operations on three classes of parts: finished (type 1 parts), semi-finished (type 2), and rough (type 3). The same tool can be used for any of the parts, but it is worn by use. Once it is used for semi-finished operations, it cannot be used for finished operations; and once it is used for rough operations, it cannot be used for finished or semi-finished operations. Thus, there is no changeover time in changing from finished to semi-finished or semi-finished to rough operations, since there is no tool change involved. Any other change does cost some time. Assume that the operation times and set-up times (when non-zero) are all 10 minutes. Thus

$$\tau_j(i) = \begin{cases} 10, \text{ if } i = j \\ \infty, \text{ otherwise} \end{cases} \tag{3.12}$$

and

$$\left.\begin{array}{l} s_{12} = s_{13} = s_{23} = 0 \\ s_{21} = s_{31} = s_{32} = 10 \end{array}\right\} . \tag{3.13}$$

Assume

$$\left.\begin{array}{l} d_1 = .02 \\ d_2 = .03 \\ d_3 = .04 \end{array}\right\} \text{ parts per minute.} \tag{3.14}$$

This implies that

$$\begin{array}{lcl} 0 \le u_1 \le .1 & & w_1u_1 = .02 \\ 0 \le u_2 \le .1 & \text{and} & w_2u_2 = .03 \\ 0 \le u_3 \le .1 & & w_3u_3 = .04 \end{array} \tag{3.15}$$

If we assume that u_i is always at its maximum value, then

$$\begin{array}{l} w_1 = .2 \\ w_2 = .3 \\ w_3 = .4 \end{array} \tag{3.16}$$

This implies that (2.2) can be written

$$2\phi_{21} + 3\phi_{31} + 3\phi_{32} = .1 . \tag{3.17}$$

Equation (2.3) becomes, in this example,

$$3\phi_{21} + 4\phi_{31} = 2\phi_{12} + 2\phi_{13} \tag{3.18}$$

$$2\phi_{12} + 4\phi_{32} = 3\phi_{21} + 3\phi_{23} . \tag{3.19}$$

There is a third equation in (2.3), but it is implied by the previous two.

Instead of solving the optimization problem (2.7), we make the following simplifying assumption: on the average, during each period that the system is in configuration j [i.e., capable of operating on type j parts], it produces Kd_j type j parts. Since the average length of time it is in configuration 1 is

$$\frac{1}{\phi_{12} + \phi_{13}}$$

and the average production rate in that state is .1, then

$$\frac{.1}{\phi_{12} + \phi_{13}} = .02K \ . \tag{3.20}$$

Similarly,

$$\frac{.1}{\phi_{21} + \phi_{23}} = .03K \tag{3.21}$$

$$\frac{.1}{\phi_{31} + \phi_{32}} = .04K \ . \tag{3.22}$$

The solution is:

$$\phi_{12} = \frac{10.625}{K} - .075 \tag{3.23}$$

$$\phi_{23} = \frac{2}{3}\phi_{12} \tag{3.24}$$

$$\phi_{31} = \frac{1}{2}\phi_{12} \tag{3.25}$$

$$\phi_{21} = .05 - \frac{3.725}{K} \tag{3.26}$$

$$\phi_{13} = \frac{3}{2}\phi_{21} \tag{3.27}$$

$$\phi_{32} = \frac{3}{4}\phi_{21} \tag{3.28}$$

in which

$$75 \leq K \leq \frac{425}{3}$$

so that all quantities above are non-negative. The solution is still not quite complete; the quantity K must be determined.

Interpretation

The solution has isolated two set-up cycles; 1 - 2 - 3 - 1 and 2 - 1 - 3 - 2. The first requires less set-up time, but the second may be preferred sometimes. The system's behavior will be a mixture of the two cycles. The cycle to be chosen at any time may depend on the amount of each part type that has been produced.

Other cycles are possible, such as 1 - 2 - 1. These partial cycles are not sufficient by themselves, but they may be selected if, for example, the production of type 3 parts is in excess compared with types 1 and 2.

4. More Frequent Set-Ups

In Sections 2 and 3, changes of configurations are assumed to be much less frequent that machine failures and repairs. In cases where these events occur at about the same frequencies, we can modify the definition of "configuration" to include failure state. That is, m now refers not only to the kind of tooling present at any instant, but also the failure state of the system. It includes α as defined by Kimemia and Gershwin (1983).

The equations of Section 2 are largely unchanged, but they must be interpreted differently since the time scale is shorter. Some components of ϕ_{mk} are specified quantities: the repair and failure rates of some machines. The rest are decision variables with the same meanings as in Sections 2 and 3.

Equation (2.4) must be modified as follows:

$$u^m \in \bar{\Omega}^m = \left\{ u \;\middle|\; u_j \geq 0,\ \sum_j \tau_{ij}(m)\, u_j \leq \alpha_i(m) \right\} \tag{4.1}$$

where $\alpha_i(m)$ is the repair state of machine i corresponding to configuration m.

5. Conclusion

This paper suggests an approach to incorporating set-up times into the earlier Kimemia-Gershwin scheduling method for unreliable FMS's. It follows the strategy of sidestepping the detailed combinatorial analysis that would be required if we represented each set-up event and each part explicitly. Instead, we deal with the rates of these events, and this greatly reduces the computational load.

Future work includes the development of the lower level algorithms. These algorithms must carry out the guidelines calculated here. It also includes the understanding of set-up cycles described in Example 2. Other related work is devising strategies for assigning parts to families; that is, deciding what will each configuration be able to do. Finally, the widely used concept of hierarchical decomposition of scheduling algorithms should be examined, so that there is a systematic method for assigning issues to hierarchical levels.

REFERENCES

P. Afentakis, B. Gavish, and U. S. Karmarkar (1984), "Computationally Efficient Optimal Solutions to the Lot-Sizing Problem in Multistage Assembly Systems," *Management Science*, Vol. 30, No. 2, February, 1984.

R. Akella, Y. F. Choong, and S. B. Gershwin (1984), "Performance of Hierarchical Production Scheduling Policy," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, Vol. CHMT-7, No. 3, September, 1984.

S. B. Gershwin, (1986), "An Approach to Hierarchical Production Planning and Scheduling," *Proceedings of the Symposium on Real-Time Optimization in Automated Manufacturing Facilities*, January, 1986.

S. B. Gershwin, R. Akella, and Y. F. Choong (1985), "Short-Term Production Scheduling of an Automated Manufacturing Facility," *IBM Journal of Research and Development*, Vol. 29, No. 4, pp 392-400, July, 1985.

S. B. Gershwin, M. Athans, E. R. Ducot, and J. G. Kimemia (1982), Systems Aspects of Flexible Automated Manufacturing Networks," MIT Laboratory for Information and Decision Systems Report LIDS-FR-1202, April, 1982.

S. B. Gershwin, S. C. Graves, Y. F. Choong, R. Akella (1985), "Control of Material Flow in Manufacturing Systems," Draft of MIT Laboratory for Information and Decision Systems Final Report.

S. C. Graves (1981), "A Review of Production Scheduling," *Operations Research*, Vol. 29, No. 4, pp 646-675, July-August, 1981.

J. G. Kimemia (1982), "Hierarchical Control of Production in Flexible Manufacturing Systems," MIT Laboratory for Information and Decision Systems Report LIDS-TH-1215.

J. G. Kimemia and S. B. Gershwin (1983), "An Algorithm for the Computer Control of a Flexible Manufacturing System," *IIE* Transactions Vol. 15, No. 4, pp 353-362 December, 1983.

U. S. Karmarkar and L. Schrage (1985), The Deterministic Dynamic Product Cycling Problem," *Operations Research*, Volume 33, No. 2, March-April 1985, pp 326-345.

A. Kusiak, A. Vanelli, and K. R. Kumar (1985), "Grouping Problem in Scheduling Flexible Manufacturing Systems," *Robotica*, Volume 3, pp 245-252.

W. Maxwell and J. Muckstadt (1985), "Establishing Consistent and Realistic Reorder Intervals in Production-Distribution Systems," *Operations Research*, Volume 33, No. 6, November-December 1985, pp 1316-1341.

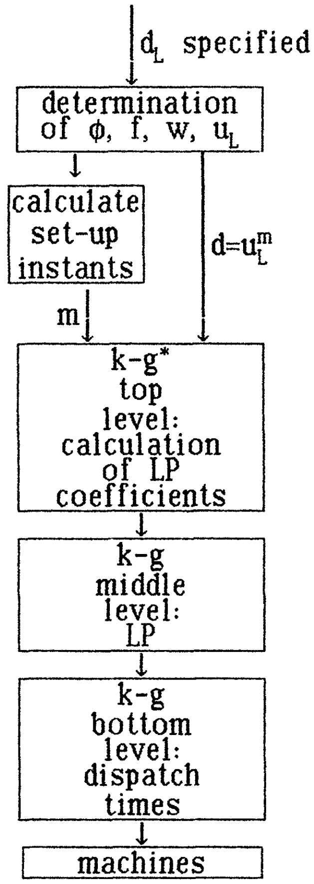

Figure 1. The Production Algorithm

* Kimemia and Gershwin [1983].

Acknowledgments

I am grateful for discussions with Yong F. Choong, Sheldon Lou, and Loren K. Platzman. This work was sponsored by the Defense Advanced Projects Agency and monitored by the Office of Naval Research under Contract N00014-85K-0213.