

Division of Research
Graduate School of Business Administration
The University of Michigan

October 1985

PROCEDURES TO DETERMINE BOTH APPROPRIATE
PRODUCTION RATIOS AND MINIMUM INVENTORY
REQUIREMENTS TO MAINTAIN THESE RATIOS
IN FLEXIBLE MANUFACTURING SYSTEMS

Working Paper No. 448

Kathryn E. Stecke
The University of Michigan
Graduate School of Business Administration
Ann Arbor, Michigan

ABSTRACT

During the operation of a flexible manufacturing system (FMS), there are various set-up decisions that have to be made somewhat periodically, for example, if the part mix is changed or when the production requirements are finished for some part type or a machine breaks down. Five production planning problems have been defined to address these set-up decisions and several of the problems have been addressed previously.

This paper focuses on two of these planning problems that have not yet been sufficiently addressed either in practice or in the literature. In brief, the problems are to:

1. Determine production ratios at which a set of part types, selected to be machined concurrently over the next time period, should be produced.
2. Determine the minimum in-process inventory (numbers of pallets and fixtures) requirements to maintain those production ratios.

First, aggregate methods are used to solve these problems, i.e., to suggest appropriate production ratios for the individual part types. Initial computations are based on machine utilizations. Several objectives are considered, each applicable in different FMS situations. These aggregate production ratios are useful to help solve other planning and operating problems. For example, they provide guidelines to help in determining appropriate part input sequences. They can decrease the size of subsequent scheduling problems by decreasing the set of feasible alternatives.

The ratios also provide input into more detailed models that are used to determine the actual, operating, production ratios. Two complementary models are suggested. A stochastic model, a multiclass, closed queueing network, will provide slightly pessimistic, but relatively accurate, aggregate and steady state performance evaluation results. A deterministic model, a timed Petri

net and its associated algebraic representation, will provide slightly optimistic and accurate results. Either model could be adequate to use to determine the operating production ratios and minimum inventory requirements. Many future research areas are also noted.

1. INTRODUCTION

A flexible manufacturing system (FMS) consists of a set of computer numerically controlled machine tools connected via an automated material handling system. The high level of automation allows efficient and flexible simultaneous machining of a variety of part types in unit batch sizes. The operation of these systems is different from the traditional assembly line or job shop situations. The FMS planning, scheduling, and control problems have sometimes similar, but often different counterparts in the conventional manufacturing systems.

Five production planning problems were defined in Stecke [1983] to help an FMS manager to set up his/her system in an efficient and productive manner prior to the start of production. Several of these have been addressed previously at various levels of detail. This paper addresses a different two of the five planning problems.

The plan of the paper is as follows. §2 begins by briefly reviewing the planning problems and various solution approaches to date. We discuss how these problems and appropriate solution procedures are different for FMSs as well as how their solutions relate to the FMS scheduling problems that would need to be solved subsequently. §§3 and 4 suggest solution approaches to determine aggregate production ratios for several relevant operating objectives and associated problems. §5 takes the results of §§3 and 4 as input into two models to use to help determine operating solutions. Future research needs are provided in §6.

2. PROBLEM DEFINITION

The following five planning problems are addressed and implemented in an FMS, periodically and in advance of the start of production of a new or

different part mix. Suri and Whitney [1984] call problems like these second level decisions, to be addressed over a time horizon of several days or weeks.

1. The part types that are to be produced next, and simultaneously over the upcoming period of time, have to be selected.
2. Within each machine type, machines may be partitioned into identically-tooled machine groups. Then each group can perform the same operations. Grouping is useful in that it automatically provides redundancy for breakdown situations; it automatically provides for alternative part routings; it decomposes the tooling problems into smaller problems and makes them easier to solve. However, grouping is not essential and sometimes cannot be performed. The necessary planning functions can be addressed directly in (5) below.
3. The production ratios at which the selected part types should be produced over time are determined.
4. The minimum numbers of pallets and fixtures of different fixture types required to maintain the production ratios need to be determined.
5. The cutting tools of all operations of all of the selected part types have to be loaded into some machine's (one or more) limited capacity tool magazine in advance of production. This determines which machine tools each operation can be performed on during the real time production of parts.

There are production requirements that usually change over time for a variety of part types. These production requirements are derived either from some forecast of demand or actual customer orders. Depending on many factors, such as system capacity or due dates, for example, usually some subset of the required part types will be chosen to be produced next over the upcoming time

period. When the production requirements for some part type(s) are finished, space is freed up in the tool magazines and either one or more part types can be input into the system (if space for all cutting tools can be found) or just the reduced set of part types can be machined (perhaps more pooling can be done). An alternative heuristic to select the part types to be produced next is suggested by Whitney and Gaul [1984]. They partition all part types into batches and then machine one batch at a time.

The grouping and loading problems (problems (2) and (5)) have been treated at several levels of detail. Queueing networks have been used to characterize appropriate solutions to these and other FMS problems at an aggregate level of detail and to provide qualitative or operational insights in Buzacott and Shanthikumar [1980], Cavallé and Dubois [1982], Dallery and David [1983], Dubois [1983], Hildebrandt [1980], Solberg [1977, 1979], Shanthikumar and Buzacott [1980], Shanthikumar and Stecke [1986], Stecke [1985], Stecke and Morin [1985], Stecke and Solberg [1985], Suri [1983], Suri and Hildebrandt [1984], and Yao [1984], for example. At a detailed level, the various problems have been addressed using mathematical programming (Stecke [1983], Berrada and Stecke [1984]), heuristics (Stecke and Talbot [1983], Whitney and Gaul [1984]), and at a less detailed level by Kusiak [1983]. Many of these studies assume that part type mix problems like the third and fourth planning problems have already been solved.

This paper addresses the third and fourth planning problems. In particular, aggregate approaches to help determine appropriate, "optimal" input ratios in which a selected set of part types should be produced are suggested in §§3 and 4. The main, overall system objective that is considered in this paper is to maximize production or system utilization. The equipment is expensive and many FMS users admit that a high utilization and maximum production is of major concern (i.e., Vought AeroProducts, Caterpillar

Tractor, John Deere, Renault Machines Outiles, Olomouc and Celakovice in Czechoslovakia,...). If due dates are relevant, these would impact other problems, such as part type selection as well as the part input sequence and scheduling procedures. Makespan might also be important, but maximizing utilization could help to attain makespan objectives.

Two distinct and relevant objectives to determine the production ratios to follow that will help maximize production are considered here. Each would be applicable in a different type of FMS. These various FMS situations and those relevant for consideration here are now described.

Some FMSs produce parts that are required in certain relative ratios. For example, perhaps the system machines many parts or components for later assembly purposes. Then the parts are required in certain, perhaps equal, output ratios of each. These requirements can be translated into operating production ratios as we shall see in §3, where we provide other potential scenarios for this application. There are also interesting part type selection, grouping, loading, part input sequence, and scheduling problems in this case, which are not the subject of this paper. However, we show that the selection of appropriate production ratios can impact and simplify these other FMS operating problems.

The systems that are of more interest here are those that machine independent part types. There may be requirements for varying numbers of each, but we are free to determine the relative ratios in which they should be produced. There are several scenarios possible in this case. Some approaches to operate the system are better than others (with respect to maximum utilization, say).

We first examine, in §3, the situation in which a set of part types has been selected to be machined, having varying production requirements, with the operating objective of starting and finishing all of these parts at the same time. There are plausible reasons for such an operating decision. Before

production begins, all required cutting tools have to find their places in some tool magazine(s). When all requirements are finished, all magazines can then be emptied and the system set up for the next mix of part types. This approach tends to minimize the frequency of tool changes. We show in §3 how to determine aggregate ratios of part types so that they begin and end all requirements simultaneously. One can see that this approach defines the workload constraints and the bottleneck machine (type). In general, it will neither maximize production nor utilization. However, this may be an appropriate approach for some FMSs, for example, if demand for some parts is dependent and certain relative output ratios are required. These output ratios would be translated into different production ratios. We show how to implement this approach in §5.

The operation of the FMS is different in §4. Aggregate ratios of part types are determined so as to keep the workload per machine on each of the machine types relatively balanced (or unbalanced, if that is applicable) over a time horizon. In this case, the operation of the FMS is more flexible. Now, the requirements of some part type are finished first, and all of the planning problems can be addressed again, including the determination of production ratios. The planning for set up of the system prior to production is more complicated and still manageable, but system utilization and production increases. Groups of pooled machines are also considered.

Throughout §4, it is seen that by determining ratios to balance workloads, idle time tends to decrease, the amount of buffer space required is less than otherwise, less lead time is required, and inventory requirements can be minimized. Some theoretical justification of the latter observation is provided by Shanthikumar and Stecke [1986].

In some systems, parts require one or more refixturings so that cutting operations can be performed on a different surface of the part. In these situations, there are relative ratios predetermined for each fixturing of each part type, but if the types are independent, ratios can again be found for the final products. Another situation that can be handled similarly is that of different components being machined in fixed ratios but for different assemblies. The ratios of assemblies can then be determined. These situations are addressed in §4.5. In §4.6, we formulate the problems as parametric linear and integer programs to allow the generation of many of the possible optimal solutions. Examples demonstrate the usefulness and further research needs regarding the use of these ratios.

In §5, we address the problems of determining: (1) actual operating production ratios; and (2) minimum inventory requirements to operate at these ratios. The suggestions of §§3 and 4 serve as input into any of several models that can help determine both the best ratios at which to operate the FMS and the minimum numbers of pallets and fixture requirements. Both queueing networks and Petri nets are used to determine these. Simulation might also be used but often, not that much detailed modeling capability is required. More aggregate or simpler models might be desirable because of the frequency of solution that might be required.

3. SIMULTANEOUS COMPLETION OF ALL PART TYPES

We begin by defining in Table I the notation that will be used subsequently in this section. Additional notation will be defined when it is required in §4.

Given the part types that have been selected to be machined simultaneously over some upcoming production period, and each part type's total processing time and production requirements, the problem is to find a set of aggregate

TABLE I

Notation

i	part types, $i = 1, \dots, N$
j	machines, $j = 1, \dots, M$
k	machine types, $k = 1, \dots, K$
a_i	production ratio of part type i
r_i	production requirements for part type i
n_i	number of pallets required for part type i
p_{ij}	processing time of part type i on machine j
p_{ik}	average workload of part type i on a machine j of type $k = p_{ij}/m_k$
tp_i	total workload of one part of part type $i = \sum_{k=1}^K p_{ik}$
m_k	number of machines of type k
n	total number of pallets required = $\sum_{i=1}^I n_i$

production ratios to be followed that allow all part types to finish at the same time.

The aggregate production ratios are obtained simply by solving the following equations for a_i , $i = 1, \dots, N$.

$$r_1 (tp_1)/a_1 = r_2 (tp_2)/a_2 = \dots = r_i (tp_i)/a_i = \dots = r_N (tp_N)/a_N \quad (1)$$

This situation can be thought of as a static, deterministic, aggregate, minimum makespan-like problem. Travel time, waiting time, and the like are not considered here. The real-time control of production accounts for these. These aggregate production ratios, a_i , serve as guidelines for production. For example, we shall see that they impact and can be used to help determine an appropriate part input sequence. Applicable scheduling procedures will direct the flow of work through the system. These other considerations, i.e.,

waiting time,..., are accounted for shortly in §5 to determine actual production ratios.

The following simple example illustrates the concepts. Table II contains processing time information and production requirements for two part types on two machine types (mills and drills, say). The information indicates that output ratios of 1:2 are required for part types 1 and 2, respectively.

TABLE II
Processing Times for Two Part Types on
Two Machine Types with Production Requirements

	Mill	Drill	r_i
PT ₁	10'	40'	50
PT ₂	20'	10'	100

Substituting the appropriate information into equation (1) and solving, the aggregate production ratios are: $a_1 = 5$ and $a_2 = 6$.

It can be seen that maintaining these relative input ratios over time will help to allow the completion of all requirements of both part types at the same time. This might be a goal of Whitney and Gaul's [1984] batching procedure. Begin a batch of part types, complete its requirements over some time horizon, and then begin the next batch. The frequency of tool changes is minimized. However, production rate is lower using this approach, rather than following the different objective of §4. Hence the total number of tools changed will also be less. Also, starting and ending conditions are important. Is the system starting empty? Can batches overlap? There might be a significant amount of idle time during the beginning and ending of a batch of part types.

This procedure can be applicable in various situations. Suppose, for example, that two of PT_2 are needed for every PT_1 . This required output ratio can be related directly to the production requirements in Table II of 100 and 50 pieces, respectively. More generally, several part types may be required in predetermined ratios to be fed to downstream workstations, say, for assembly purposes.

As another example, suppose that several final products are for the same customer. Then the orders should be shipped together when completed to keep freight costs down. If there is only a small area for finished goods inventory, then this approach is appropriate also. Batching the requirements of a customer will minimize finished goods inventory.

This approach can also be used in an MRP environment (dependent demand), where predetermined quantities (obtained by exploding the bills of materials) of several components are specified to be manufactured within a particular time bucket. The r_i 's represent the requirements within the time bucket. The derived production ratios would facilitate overlapped production of the downstream workstations. This could reduce the effective lead time while ensuring that the total requirements for the time period are met.

Operating the system in this manner defines the workloads on each machine. In the particular example of Table II, the workload unbalance is not too bad. Over time, the drills would be the bottleneck machine type if the output ratios of 1:2 were followed. In §4, another approach that is applicable for different types of systems determines aggregate ratios to provide a better workload balance.

Recall that we have not yet accounted for travel time, waiting time, or congestion. Only the aggregate input ratios have been determined. These other considerations are accounted for in §5, where these ratios are input into other models that determine the actual operating production ratios as

well as the minimum numbers of pallets and fixtures to maintain these ratios. Rarely, these ratios are slightly revised at this next stage. Finally, we note that this procedure generalizes immediately to N part types, M machines, and K machine types.

4. BALANCING WORKLOAD PER MACHINE

Given the processing time requirements of each part type on each machine type, the problem is to determine relative ratios at which the part types should be maintained in production, so as to keep the workloads on the machine types balanced.

For different types and sizes of problems, the following suggested solution procedures differ. Also, for larger problems there are multiple "optimal" solutions with respect to balancing, so that other, secondary criteria can be used to determine the ratios to follow. For these reasons, the following presentation consists of cases, presented in order of increasing complexity. Examples are used to illustrate. The benefits to be obtained from following the suggested procedures are first demonstrated in §4.3, as the situations become sufficiently complex enough to be of interest. Parametric integer and linear programs are suggested in §4.6 to solve many of the problems presented here for larger problems.

Aggregate production ratios are determined that are to be followed over time. These ratios are input into the more detailed models in §5. They can be revised, if need be. Often, the minimum number of pallets and fixtures to maintain these ratios can be found via the procedures of §§3 and 4.

4.1 TWO PART TYPES, TWO MACHINE TYPES

The simplest situation consists of two part types and two machine types with one machine of each type. Using the notation of Table I, the aggregate

ratios to balance the workload on both machines over time can be found by solving:

$$p_{11} a_1 + p_{21} a_2 = p_{12} a_1 + p_{22} a_2. \quad (2)$$

The solution is:

$$\frac{a_1}{a_2} = \frac{p_{22} - p_{21}}{p_{11} - p_{12}}. \quad (3)$$

Note that the quantity on the left (right) hand side of equation (2) is an aggregate measure of workload over time on machine one (two). Equating the two quantities balances the workload.

If the solution (3) consists of positive a_1 and a_2 , these are then the ratios to maintain over time to balance the workload. However, a little care has to be taken to ensure a feasible (all a_i greater than zero) solution to equation (2). In selecting the two part types to be produced, one has to utilize one machine type more, while the other part type has to utilize the other machine type more. Otherwise, a workload balance between the two machines is impossible and a queue has to build and idle time results. This situation would surface in (3) as the ratios, a_1 and a_2 , would then relate negatively to each other. Obviously, the processing time requirements on each machine type j cannot be identical.

To illustrate with the data in Table II, the production ratios that balance the workload on the mill and drill are:

$$a_1 = 1 \text{ and } a_2 = 3.$$

We return to this situation in §5, where these aggregate operating ratios are to be input into more detailed models.

The procedure generalizes immediately to systems containing pools of machines. If there are several machines (m_k) of each type k , then the workload per machine is balanced by solving:

$$p_{11} m_2 a_1 + p_{21} m_2 a_2 = p_{12} m_1 a_1 + p_{22} m_1 a_2 . \quad (4)$$

The solution is:

$$\frac{a_1}{a_2} = \frac{p_{22} m_1 - p_{21} m_2}{p_{11} m_2 - p_{12} m_1} . \quad (5)$$

To illustrate with the same data of Table II, if there are 2 mills and 4 drills, the relative ratios that balance the workload per machine on each mill and drill are:

$$a_1 = 3 \text{ and } a_2 = 2 .$$

The next generalization is to include a third part type.

4.2 THREE PART TYPES, TWO MACHINE TYPES

By including a third part type 3 on the system, the equation to solve to find the production ratios of three part types on two machines is:

$$p_{11} a_1 + p_{21} a_2 + p_{31} a_3 = p_{12} a_1 + p_{22} a_2 + p_{32} a_3 . \quad (6)$$

The solution to equation (6) is of the form:

$$a_2 = \frac{(p_{12} - p_{11}) a_1 + (p_{32} - p_{31}) a_3}{p_{21} - p_{22}} . \quad (7)$$

In this case, the solution is not a set of ratios. The solution described by (7) is an equation, in particular, a plane. If the problem is well-defined (along the lines described in §4.1), then there is an infinite number of solutions that will balance each machine's workload. The problem of how to choose one of these solutions is addressed in §5.

To illustrate, consider the inclusion of a third part type in addition to those of Table II, requiring 10 minutes on the mill and 20 minutes on the drill. These requirements are in reverse to those of part type 2. In this case, the solution to equation (6) is:

$$a_2 = 3a_1 + a_3. \tag{8}$$

Table III contains some possible solutions, all of which balance the workload.

TABLE III

Production Ratios from Equation (8)

a ₁	1	1	2	2	1
a ₂	4	5	8	7	6
a ₃	1	2	2	1	3
n	6	8	12	10	10

If, for example, the number of pallets in the system is the sum of the ratios, we see that there can be several ways to distribute some n pallets among the three part types so as to balance workload. We return to this issue in §4.6. Notice that although part types 2 and 3 have asymmetric machine requirements, their ratios will never be equal unless part type 1 is not produced.

Finally, note that in the case of pooling machines, the equation to solve for the optimal production ratios is:

$$p_{11} m_2 a_1 + p_{21} m_2 a_2 + p_{31} m_2 a_3 = p_{12} m_1 a_1 + p_{22} m_1 a_2 + p_{32} m_1 a_3. \tag{9}$$

One solution to equation (9) is of the form:

$$a_2 = \frac{(p_{12} m_1 - p_{11} m_2) a_1 + (p_{32} m_1 - p_{31} m_2) a_3}{p_{21} m_2 - p_{22} m_1}. \tag{10}$$

4.3 N PART TYPES, N MACHINE TYPES

The procedure described now to find the production ratios is different from those provided in the previous sections. In the present case, N equations are solved for N unknowns. The production ratios are unique. In order to obtain a feasible and meaningful solution (i.e., all $a_i > 0$), an initial, simple check should be made to see that the maximum processing time of each of the N machine types is required by a different part type.

The ratios to balance the workload on the machine types over time can be found by solving the N equations:

$$W = \sum_{i=1}^N a_i p_{ij}, \quad j = 1, \dots, N. \quad (11)$$

W is a measure of the workload on the machines. Notice that these equations can be solved quickly to find the unique solution by using linear programming with an arbitrary objective function.

To illustrate the procedure, consider the aggregate processing time information in Table IV. We want the workload, W of equation (11), to be the same on each machine type:

$$W = 10a_1 + 20a_2 + 10a_3 = 20a_1 + 10a_2 + 30a_3 = 50a_1 + 5a_2 + 20a_3.$$

TABLE IV

Processing Times for Three Part Types on Three Machine Types

	Mill	Drill	VTL
PT ₁	10	20	50
PT ₂	20	10	5
PT ₃	10	30	20

Since the three equations are dependent, a value for W has to be chosen. The relative ratios remain the same, regardless of the value of W . The selected W merely scales the a_i . Setting $W = 100$ and solving the three equations simultaneously, we obtain: $a_1 = 1.083$, $a_2 = 3.783$, $a_3 = 1.35$. Doubling each a_i , we obtain: $a_1 = 2.106$, $a_2 = 7.566$, $a_3 = 2.7$. Rounding these values translates to ratios of about 1:4:1 or 2:8:3.

Simulating this situation quickly showed that maintaining the ratios of 1:4:1, and including a second PT_3 every other period, both kept the machines balanced and minimized the in-process inventory. Of course, an appropriate part input sequence has to be determined and the calculated production ratios help with this problem also. They provide guidelines to follow. By following the production ratios of 1:4:1.5, several input sequences provided: a balanced workload; minimum work-in-process required; minimum buffer space required; and minimum idle time. In fact, the minimum number of pallets and fixtures of each fixture type required to maintain the ratios was either:

(1, 4, 1) or (1, 3, 2).

Only 6 pallets in total were required. For all other sets of aggregate production ratios that were simulated, queues built up, there was inserted idle time, additional pallets (inventory) were required to keep machines busy, workload was unbalanced, and more buffer space at each machine was required.

Linear programming can be used to find the unique ratios. Using the formulation (P1) of §4.6 and LINDO, a first IP solution is 1:3:2. The IP optimum is 1:4:1 and took 0.194 seconds of CPU time to find. These matched the integer solutions that were suggested above by rounding the linear optimal solution.

This simpler situation illustrates the concepts and benefits. The procedure generalizes directly to N machines and N part types.

4.4 N PART TYPES, M < N MACHINE TYPES

A more usual situation is when there are more part types being machined concurrently than machine types. As in §4.2, the solution will most often no longer be unique. There could be an infinite number of solutions, but as we shall see, most of these can often be eliminated as either infeasible or undesirable. Other criteria in addition to balancing can be used.

One procedure to find the optimal aggregate production ratios is similar to that described in §4.3: The M equations (11) are solved for the a_i . However, there could be many solutions. The following example illustrates this situation.

Consider the three-machine system described in Table V. In order for the workload on each machine to be identical, equation (11) provides the three equations:

$$\begin{aligned} W &= 10a_1 + 20a_2 + 10a_3 + 15a_4 = 40a_1 + 10a_2 + 30a_3 + 20a_4 \\ &= 50a_1 + 5a_2 + 20a_3 + 40a_4. \end{aligned}$$

TABLE V

Processing Times for Four Part Types on Three Machine Types

	Mill	Drill	VTL
PT ₁	10	40	50
PT ₂	20	10	5
PT ₃	10	30	20
PT ₄	15	20	40

W is an aggregate measure of workload on each machine. It is an indication of machine capacity over some time horizon. Setting $W = 100$, we can solve for a_1 , a_2 , and a_3 , as functions of a_4 :

$$\begin{aligned} a_1 &= 1.7808 - 1.1959a_4 \\ a_2 &= 4.3477 - .739a_4 \\ a_3 &= -.4347 + 1.1738a_4. \end{aligned} \tag{12}$$

It appears as though there would be an infinite number of solutions. In reality, the feasible set of solutions is small. For any integer value of a_4 greater than one, a_1 is negative (i.e., infeasible). The equations (12) are graphed in Figure 1. For any values of a_4 outside of the interval (.37, 1.49), either a_1 or a_3 is negative. For $a_4 = 1$,

$$a_1 = .585 \quad a_2 = 3.6 \quad a_3 = .739.$$

Rounding these values up to integer values, suggests aggregate production ratios of:

$$(a_1, a_2, a_3, a_4) = (1, 4, 1, 1). \tag{13}$$

It can be seen that in this example, any other ratios that would tend to balance workload would be fractional, i.e., $a_4 = .5$.

Figure 2 is a Gantt chart of one possible scenario. It depicts a flow shop, where each part visits the mill first, the drill second, and the lathe last. Using the aggregate production ratios (13) as guidelines, a good part input sequence was determined to be:

1, 2, 2, 2, 3, 4, 2.

The subscript of each part number in Figure 2 is the pallet/fixture number that is assigned to that individual part. The sequence is periodic and repeats as is. The small boxes (■) indicate the completion of a cycle of the input sequence. Three cycles are shown here.

The production ratios of equation (13) were very useful in the following ways:

1. They were useful as guidelines to help find a good, periodic, input sequence of parts.

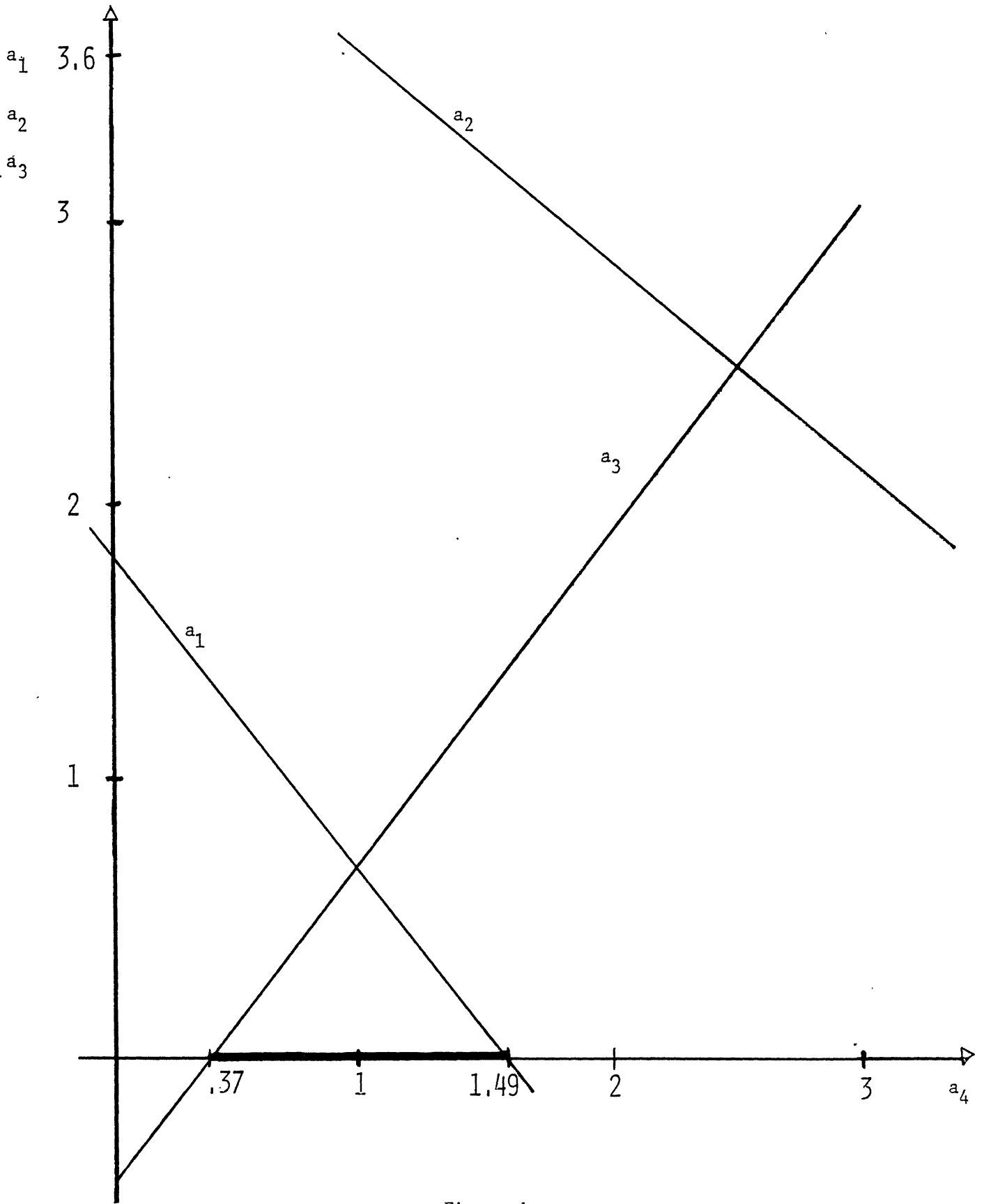


Figure 1

Graph of Feasible Solutions to Equations (12).

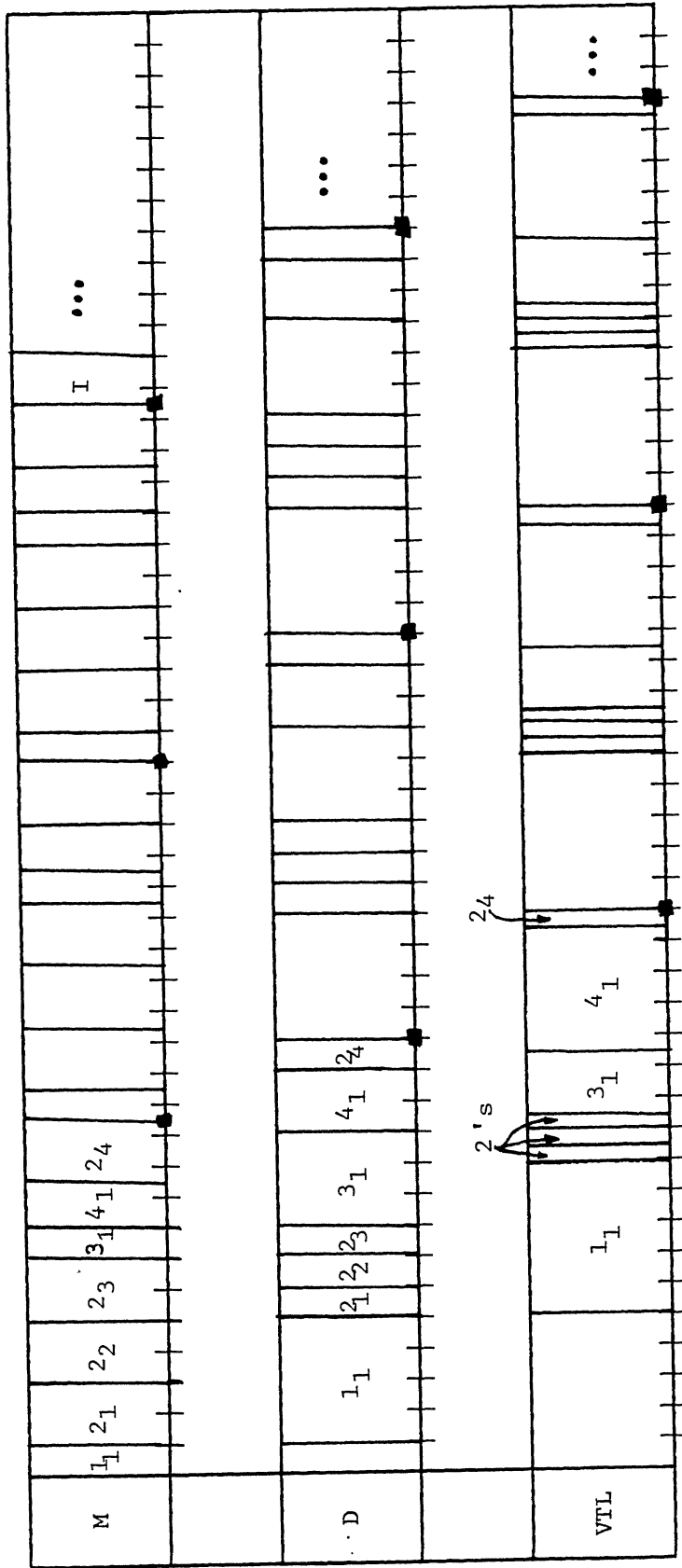


Figure 2

Gantt Chart for a Possible Scenario of the System of Table V.

2. They helped to find a schedule with very little idle time. Workload is nearly perfectly balanced.
3. The numbers of pallets/fixtures required to maintain these production ratios is exactly the values of the ratios: 1, 4, 1, 1. The total number of pallets required is 7.
4. The amount of buffer space at each machine to hold the WIP inventory is minimal.

The little idle time on the mill in Figure 2 can be decreased even further by following the ratios: (.5, 4, 1, 1). These ratios are closer to the optimal fractions that balance workloads. A part input sequence that provides an even better schedule (less idle time,...) while following these new ratios is:

1, 2, 2, 2, 3, 4, 2; 2, 3, 2, 2, 4, 2.

Again, the procedure generalizes immediately to $M < N$ machines and N part types. The usual situation is that there are more part types than machine types. This could result in several sets of "optimal" aggregate production ratios. In §4.6, parametric linear (integer) programming is used to provide many candidate, optimal and nonoptimal but good, ratios. In §5, queueing and idle time are considered to find actual operating ratios.

4.5 REFIXTURING

For most types of prismatic parts, after a series of operations are performed, they move off the system to be refixed. The part is clamped to a different fixture type on a different pallet. The part is then released to the system again and additional cutting and inspection operations are performed on a different surface of the part. Each refixturing in most respects can be treated as a new part type. However, for each part, the production ratios of the refixturings have to remain at one to one. If the end products are independent, aggregate ratios can be found for these that balance the workload.

Depending on the numbers of part types and machine types, the appropriate methods described earlier in §4 can be applied to find these ratios.

We again illustrate with an example. The system described in Table VI consists of two machine types processing two part types, each of which required a refixturing after passing through the mill and then the drill. In Table VI, PT_{ij} is part type i with pallet/fixture combination j .

TABLE VI
Processing Times for Two Part Types
Requiring Refixturings on Two Machine Types

	Mill	Drill
PT_{11}	10	40
PT_{12}	10	30
PT_{21}	20	10
PT_{22}	15	20

Aggregating the processing time information of Table VI and substituting into equation (2), the aggregate production ratios are: $(a_1, a_2) = (1, 10)$. Maintaining these ratios balances the workload. However, we shall see in §5 that in a flow shop situation, and ignoring for the moment the refixturing, set-up, and transportation times while considering the processing and queueing times, only three, rather than ten, fixtures for part type 2 are required to produce at the indicated production ratios. This determination includes waiting time and buffer requirements. When the delays due to transportation and fixturing times are accounted for, a few more pallets will sometimes be required.

4.6 DETERMINING AGGREGATE OPERATING PRODUCTION RATIOS

The previous methods of §4 provide aggregate production ratios to follow over time to balance the workload. In this section, more precise methods to determine the ratios are provided. The problem of determining ratios so as to balance workloads is formulated as both linear and integer programs.

The usual situation is that there are more part types simultaneously being machined than machine types. This ratio problem is similar to that of §§4.2 and 4.4, where there can be potentially many optimal sets of aggregate ratios to choose from. After the appropriate equations are solved, feasible intervals for the ratios could be found, as shown in §4.4. Because the ratios have to be greater than zero, the feasible intervals can be quite small. Of course if all part types in the current production plan dominate the same machine type, then no feasible ratios can balance the workload per machine.

In some situations, graphical methods are helpful to both determine the ratio intervals and to choose ratios, as in §4.4. In particular, a graph is useful for situations in which $N + 1$ part types are to be produced on N machine types. Otherwise, Tables of feasible combinations can be developed, as shown in §4.2, i.e., like Table III.

In any case, there could be a question concerning how to round fractional optimal aggregate ratios up or down to integer values. Fortunately, we shall see that performance does not appear to be sensitive to small variations in the ratios. In addition, considerations of transportation, fixturing, and queueing time may revise the ratios slightly (more inventory required), but in these situations also, experience has indicated that system performance does not appear to be very sensitive to variations in the ratios.

Both linear and integer programming can be used to find optimal and near-optimal sets of production ratios. The constraints are the M equations, one for each machine type:

$$\sum_{i=1}^N a_i p_{ij} = W, \quad j = 1, \dots, M. \quad (11)$$

Various sets of ratios can be generated by relaxing the equality constraints (11) by introducing slack (underload) and/or surplus (overload) variables for each machine type. These variables are:

X_{j1} = amount of time by which the workload on machine j is greater than W ;

X_{j2} = amount of time by which the workload on machine j is less than W .

Then the problem of finding aggregate production ratios so as to balance the workload per machine is:

(P1)

$$\text{Minimize} \quad \sum_{j=1}^M X_{j1} + \sum_{j=1}^M X_{j2}$$

subject to

$$\sum_{i=1}^N a_i p_{ij} - X_{j1} + X_{j2} = W, \quad j = 1, \dots, M$$

$$a_i \geq 1, \quad i = 1, \dots, N$$

$$X_{j1}, X_{j2} \geq 0, \quad j = 1, \dots, M.$$

Problem (P1) still provides a unique solution for a particular capacity workload W . However, multiple optima can be obtained by varying the parameter W .

If the objective function were changed to incorporate weights assigned to the overload (C_{j1}) and underload (C_{j2}) on each machine (type), then multiple optima can also be obtained by parametric ranging of these weights. The objective function becomes:

(P2)

$$\text{Minimize} \quad \sum_{j=1}^M C_{j1} X_{j1} + \sum_{j=1}^M C_{j2} X_{j2}.$$

The purpose of the following formulation is to combine both of the distinct objectives of §§3 and 4. The aim is to find ratios that balance the workload but finish all parts at around the same time, T . Let

T_{1i} = amount of time by which part type i finishes later than T ;

T_{2i} = amount of time by which part type i finishes earlier than T .

Then the problem is to find ratios so as to satisfy:

(P3)

$$\text{Minimize } \sum_{i=1}^N T_{1i} + \sum_{i=1}^N T_{2i}$$

subject to

$$\sum_{i=1}^N a_i p_{ij} = W, \quad j = 1, \dots, M$$

$$a_i - r_i (tp_i) X + T_{1i} - T_{2i} = 0, \quad i = 1, \dots, N \quad (14)$$

$$a_i \geq 1$$

$$T_{1i}, T_{2i} \geq 0, \quad i = 1, \dots, N.$$

Variations of (P1), (P2), and (P3) can occur by changing the objective function and the constraints. For example, the objective function could be:

$$\text{Minimize } B_1 \sum_{i=1}^N (T_{1i} + T_{2i}) + B_2 \sum_{j=1}^M (X_{j1} + X_{j2}).$$

Different sets of ratios could be obtained by varying the B_i 's, C_{ij} 's, X , and W . In any case, further research is required to determine how to set some of the parameters appropriately, such as W and X , so as to get meaningful ratios for the situation of trying to satisfy both objectives simultaneously.

We now describe how the IPs, LPs, and the methods of §§3 and 4 can be used to provide various optimal or near optimal sets of aggregate ratios. One way is to vary the parameters W , C_{j1} , C_{j2} , $j = 1, \dots, M$.

This is illustrated first with the results of Table VII. LINDO was used to provide both LP and IP optimum solutions to programs (P1) and (P2) for the four part type, three machine type example of §4.4, for varying parameters W , C_{j1} , C_{j2} . (Results for a ten part type problem are provided subsequently.) CPU time is given in seconds. Case 1 of Table VII is one solution obtained from manually solving the problem. Cases 2-13 and 17-19 all included the constraints that $a_i \geq 1$, for all i . We discuss this constraint shortly. Cases 14-16 required only that $a_i \geq 0$. Cases 2-10 and 14-16 have all C_{j1} and C_{j2} equal to 1, while Cases 11-13 have different C_{jk} 's. Finally, Cases 17-19 allow the workload W to be a variable of the problem.

We make the following observations and subsequently discuss the implications.

1. Perfect balance is obtained only with the LP optimal solutions of Cases 4, 11, 14, and 17. The remaining linear solutions and all of the integer solutions (except Case 19) are unbalanced, with either an overload or underload on one or more of the machine types.
2. In four of the 6 cases, the optimal integer solution is merely the scaling of the linear solution. Only Cases 7, 18, and 19 provide substantially different ratios. In three of the 5 cases, even the first IP solution is quite similar to, or a rounding of, the linear solution.
3. Varying W from 100 to 500 to 1000 provides quite different optimal sets of ratios.
4. If we for the moment assume that the sum of the ratios is suggesting the number of pallets, n , in a cycle, then from Table VIII and Figure 3 we see that n increases about linearly with W . Notice that the three integer solutions for $W = 100$ (Cases 3, 15, and 16) all sum to 6.

TABLE VII

Linear and Integer Optimum Solutions with Varying Parameters for the Problem of \$4.4

Case	W	C ₁₁	C ₁₂	C ₂₁	C ₂₂	C ₃₁	C ₃₂	Objective Function	a ₁	a ₂	a ₃	a ₄	CPU Time	
1	100							0.	.585	3.6	.739	1.		No Objective Function
2	100	1	1	1	1	1	1	48.75	1.	3.25	1.	1.	.188	LP Optimum
3	100							50	1	3	1	1	.350	IP Optimum
4	500							0.	1.	16.982	5.382	6.436	.186	LP Optimum
5	500							20	3	19	3	5		First IP Solution
6	500							15	3	18	3	5		Second IP Solution
7	500							10	6	20	1	2	.926	IP Optimum
8	1000							40.3125	1.	31.	12.625	13.563	.186	LP Optimum
9	1000							60	1	31	13	13		First IP Solution
10	1000							55	1	31	12	14	.993	IP Optimum
11	1000	50	0	50	0	1	1	0.	1.	31.	28.75	5.5		IP Optimum
12	1000							25	1	31	25	8		First IP Solution
13	1000							5	1	31	28	6		IP Optimum
14	100	1	1	1	1	1	1	0.	1.296	4.074	0.0	.37	.201	LP Optimum
15	100							40	2	4	0	0		First IP Solution
16	100							15	1	4	0	1	.606	IP Optimum
17	155.7							0.	1.	5.7	1.	1.43	.204	LP Optimum
18	240							5	2	9	1	2		First IP Solution
19	390							0	2	14	3	4	.616	IP Optimum

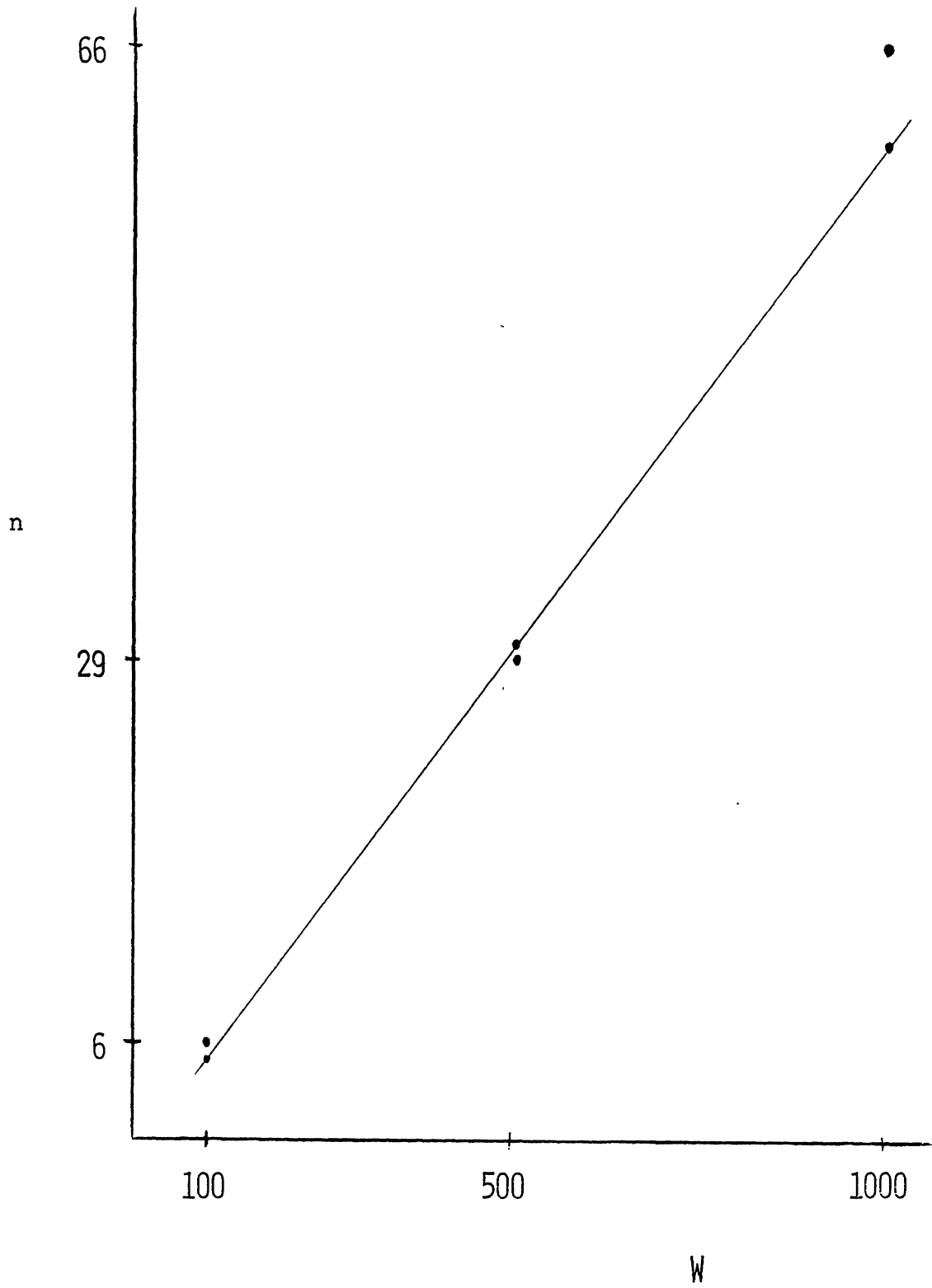


Figure 3

Number of Pallets as a Function of Workload.

TABLE VIII

Number of Pallets as a Function of Workload

W	100	500	1000
n	6	29	66

5. Varying C_{j1} and C_{j2} also provides substantially different sets of ratios.
6. Letting $a_i \geq 0$ in Cases 14-16, a_3 is always zero. This suggests that three is the "least compatible" part type, and that if we had a choice, it should not be produced with the other part types.
7. By letting the parameter W vary, even the IP optimum is balanced.
8. The CPU times (in seconds) are less than one second for all 19 cases. The times reported for the IP optimum all include the time to reach the LP optimum, since LINDO uses the LP optimum as a lower bound.

A larger example is now provided to continue to demonstrate possible uses of programs (P1) and (P3). Table IX provides processing time and demand information for a ten part type, three machine type, and five machine problem. Table X provides linear and integer solutions to problem (P1) (balancing aggregate workloads) for a variable W and for $W = 100, 500, \text{ and } 1000$.

We can make the following observations about the results of Table X.

1. The pairs of Cases 24 and 25, 28 and 29, and 35 and 36 all provide two different linear solutions. Rounding these does not cause a significant unbalance.
2. As W increases from 100 to 500 to 1000, $\sum a_i$ increases approximately from 8 to 40 to 80 as Figure 3 showed for the smaller problem.

TABLE IX

Processing Times for Ten Part Types on Three Machine Types
with Five Machines

	Mill (1)*	Drill (2)	VTL (3)	r_i
PT ₁	10	20	50	50
PT ₂	15	20	40	100
PT ₃	20	10	30	70
PT ₄	10	20	20	100
PT ₅	10	10	20	200
PT ₆	10	30	20	150
PT ₇	20	10	10	100
PT ₈	15	20	30	50
PT ₉	25	10	20	150
PT ₁₀	5	40	40	200

3. All of the IP optimums are balanced.
4. Many of the intermediate IP solutions provide nearly balanced workloads (Cases 21, 22, 26, 31, 32, 33, and 44). Any of these could be used to suggest compatible part types and appropriate ratios to follow.
5. Out of a candidate set of 10 part types, the solutions suggest various combinations of 3-7 part types that are compatible for possible simultaneous machining.
6. All integer optima are found within 3 CPU seconds, including the time to obtain the LP optimum.

*The number of machines of each type is in the parentheses.

TABLE X

Linear and Integer Optimum Solutions for a Ten Part Type, Three Machine Type, Five Machine Example for the Objective of Balancing Workloads

Case	W*	Objective Function	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	CPU Time	No. of a ₁ ≥ 0	∑ a ₁		
20	230	0.	1.	1.	1.	1.	1.	10.	1.	1.	1.	1.	.242	All	19	LP Optimum	a ₁ > 1
21	365	10	2	1	2	1	1	15	2	1	2	3			30	First IP Solution	a ₁ > 1
22	325	5	2	1	2	1	1	15	2	1	1	1			27	Second IP Solution	a ₁ > 1
23	350	0	2	1	2	1	2	15	1	1	2	2	1.174		29	IP Optimum	a ₁ > 1
24	100	0.	.625		1.875			5.625					.207	3	8	LP Optimum	a ₁ > 0
25	100	0.		2.632				4.737		.26				3	8	LP Optimum	
26		10		2			2	4						3	8	First IP Solution	
27		0		2			1	3		1	1	1	1.333	5	8	IP Optimum	
28	500	0.	3.125		9.375			28.125					.204	3	40	LP Optimum	
29	500	0.		3.	1.65	32.		9.3	.45					5	45	LP Optimum	
30		120				16			8	4	16			4	44	First IP Solution	
31		15	1	1	1				8	8	20			6	39	Second IP Solution	
32		10					1		8	2	8	20		5	39	Third IP Solution	
33		5	1	1					8	1	8	19		7	39	Fourth IP Solution	
34		0	2	1			1		8		8	19	1.456	6	39	IP Optimum	
35	1000	0.	6.25		18.75				56.25				.205	3	81	LP Optimum	
36	1000	0.		20.42				27.16		13.68	16.			4	78	LP Optimum	
37		125	7			1	2		3	32	36			6	81	First IP Solution	
38		125	7	4					3	32	36			5	80	Second IP Solution	
39		120	7						3	32	37			4	79	Third IP Solution	
40		120	11				1		32	32	36			4	80	Fourth IP Solution	
41		110	11			1			32	32	36			4	80	Fifth IP Solution	
42		110	9						1	32	37			4	79	Sixth IP Solution	
43		105	10							32	37			3	79	Seventh IP Solution	
44		30	16					28		18	16			5	80	Eighth IP Solution	
45		0	16				1	31		18	14		2.795	5	80	IP Optimum	

*Cases 20-23 specify that all a₁ ≥ 1 and allow W to be a variable. The remaining cases specify that a₁ ≥ 0 and fix W to be either 100, 500, or 1000.

Table XI provides linear and integer solutions for the same ten part type problem described in Table IX for the objective of finishing all requirements of the selected part types at about the same time. The program is that of (P3) without the balancing constraint. The constraints (14) are developed with the data of Table IX from equation (1) and the following:

$$\frac{45 (50)}{a_1} = \frac{45 (100)}{a_2} = \frac{40 (70)}{a_3} = \frac{30 (100)}{a_4} = \frac{25 (200)}{a_5} = \frac{35 (150)}{a_6} =$$
$$\frac{30 (100)}{a_7} = \frac{40 (50)}{a_8} = \frac{40 (150)}{a_9} = \frac{40 (200)}{a_{10}} = \frac{1}{T}$$

There is no significance to the parameter $1/T$.

Note that each of the 19 solutions of Table XI (18 are integer) provides a quite different set of ratios and all sets allow the completion of all requirements at nearly the same time. This suggests that secondary criteria (such as balancing or meeting due dates) could be incorporated to select the best of these (or even other) aggregate ratios to follow. It took 42.5 CPU seconds to reach the IP optimum, which is much higher than the solution time to the balancing problem. However, many good solutions are suggested.

4.7 BENEFITS AND USES OF AGGREGATE PRODUCTION RATIOS

We now outline many of the potential benefits and implications for various operating situations from using these ratios.

Secondary Criteria

Programs (P1), (P2), and (P3) can be run to provide alternative sets of ratios. Then secondary criteria, such as flow time or due date considerations, can be used to select the most appropriate set. We return to this issue later.

TABLE XI

Linear and Integer Optimum Solutions for a Ten Part Type,, Three Machine Type Example for the Objective of Finishing All Selected Part Types Simultaneously

Case	1/T	Objective Function	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	CPU Time	Σ a _i	
46	2000.	0.	1.12	2.25	1.4	1.5	2.5	2.62	1.5	1.	3.	5.	.282	22.	LP Optimum
47	562.3	8.47	4	8	5	6	8	7	5	3	7	16		68	First IP Solution
48	600.	7.5	4	8	5	6	8	7	5	3	7	15		67	Second IP Solution
49	642.7	7.36	4	8	4	6	8	7	5	3	7	14		65	Third IP Solution
50	700.	6.86	4	8	4	6	7	7	4	3	7	13		63	Fourth IP Solution
51	562.4	6.36	4	8	5	6	9	7	5	3	12	15		74	Fifth IP Solution
52	562.4	5.36	4	8	5	6	9	7	5	3	12	16		75	Sixth IP Solution
53	562.4	4.36	4	8	5	6	9	7	5	3	11	15		73	Seventh IP Solution
54	562.4	4.02	4	8	5	5	9	7	5	3	11	16		73	Eighth IP Solution
55	600.	3.83	4	8	5	5	9	7	5	3	10	15		71	Ninth IP Solution
56	600.	3.5	4	8	5	5	8	7	5	3	10	15		70	Tenth IP Solution
57	333.	2.9	5	13	8	9	15	16	9	6	18	27		126	Eleventh IP Solution
58	333.	2.9	6	13	8	9	14	16	9	6	18	27		126	Twelfth IP Solution
59	333.	1.9	6	13	8	9	15	16	9	6	18	27		127	Thirteenth IP Solution
60	333.	1.6	7	13	9	9	15	16	8	5	18	27		127	Fourteenth IP Solution
61	333.	1.4	7	13	8	8	14	16	8	6	17	26		123	Fifteenth IP Solution
62	1000.	1.2	2	4	3	3	5	5	2	2	6	9		41	Sixteenth IP Solution
63	750.	.93	3	6	4	4	7	7	3	3	7	12		56	Seventeenth IP Solution
64	250.	.2	9	18	11	12	20	20	11	7	23	36	42.552	167	IP Optimum

All Selected Part Types Have to be Produced

Cases 2-13 and 17-19 in Table VII, Cases 20-23 in Table X, and all cases of Table XI were run with $a_i \geq 1$. This can be appropriate since pallets come in discrete units. If all of the part types selected to be machined concurrently have to be selected (i.e., for assembly or due date purposes or the batching of a customer's orders), then all a_i should be forced to be greater than one. Otherwise, a fractional solution would be rounded up to one in any case. On the other hand, as we saw in the example of §4.4, an $a_i = .5$ could specify that one part i is introduced every other cycle.

Part Type Selection

The programs could also be used to help determine a compatible set of part types to be machined together as follows. All candidate a_i can be set to be greater than or equal to zero. Those i such that a_i is either equal to or close to zero in the optimal solution could be excluded from the mix of part types to be selected. For example, Cases 14-16 indicate that part type 3 is incompatible with the others and should be omitted from the current mix if that is possible. Cases 24-45 suggest various sets of compatible part types and different ratios for each. Other criteria could be used to select one of these sets.

Integer Versus Linear Solutions

In many cases, the integer optimum is close to the linear optimum. This suggests that it could be sufficient to round the linear solutions, especially if the problems are large. However, most often an FMS consists of only one, two, or three different machine types and the problems will not be very large. In any case, the integer solutions provide either the best way of rounding the linear solutions or suggest an alternative optimum. We conclude that the linear solution is often sufficient because:

- (a) rounding the fractions cannot change the aggregate balance by very much;
- (b) the linear solutions do provide the best balance;
- (c) these are aggregate ratios. Once secondary criteria, delay and travel times, and real-time congestion are considered, "optimum" matters even less. These ballpark, almost-balanced ratios appear to be sufficient from the computational experiments to date.

Given Production Ratios

Any given, relative ratios of two or more part types (in the case of assembly, for example), can be included in the objective function of (P1), (P2), or (P3).

Part Input Sequence

We suggest that the allocation ratios can also be used to help determine a part input sequence, as demonstrated in §4.4. Further research is required to specify precisely how this could be done. For a dedicated type of FMS, a periodic input sequence might be appropriate.

Other Constraints

Some remaining issues include a precise determination of how these ratios can best be used. Demand requirements, capacity constraints, and due dates could be incorporated. We use the "optimal" ratios of Case 13 to illustrate. The ratios of

1:31:28:6

tend to provide a nearly perfect workload balance. Yet because of demand requirements, it may be infeasible to produce only 1 of part type 1 out of

every 66 ($= \sum a_i$) parts. Demand and due date requirements might constrain the mix to be

1:10:10:5,

which is more unbalanced, but similar and somewhat proportional to the optimal ratios. Capacity may constrain the situation further. Suppose that a maximum of 10 pallets can be on the system and it is desired to construct a cyclic part input sequence of size 20. Then (assuming that demand and due dates are still met) this might change the above pallet distribution to

1:9:8:2,

and an appropriate input sequence is chosen from this mix.

This example demonstrates some of the other issues that have to be addressed when using these ratios. It also provides motivation for our conclusion that the workload parameter, W , should be kept small. Then n_i can be set equal to αa_i , where $\alpha = 1$ or 2 . In the example of §4.4 (Case 1), and letting $n_i = a_i$, the distribution of pallets is

1:4:1:1

and $n = 7$, which is manageable, reasonable, and was used to find a good input sequence.

Machine Breakdowns

The procedures that have been described here could also be used to analyze machine breakdown situations. For example, Hildebrant's [1980] hierarchical approach assumes that the production ratios are known for every possible failure state.

The ratios can also help determine the "hedging points" (optimal buffer levels) in the flexible assembly applications of Akella, Gershwin, and Choong [1984a, 1984b]. For each part type and for each failure state, the hedging

point is a function of: the part's value; the routing sequence (higher priority is given to parts that visit more unreliable machines); and the "extent of difficulty if the part is backlogged." For a particular failure state, the production ratios provide a measure of the "difficulty if the part is backlogged." Further research is required to suggest how the procedures developed here can be used to help specify the hedging points.

Due Date Criteria

Other criteria such as due date consideration, minimum flowtime, or minimum tool changing can be used to help choose the appropriate production ratios. More specifically, due dates may have been used to help solve the first planning problem to select the part types to be simultaneously machined next. From the feasible sets of ratios, those that best ensure that the due dates are met can be selected. To determine that the due dates can be met, processing time requirements, transportation, queueing, expected down time of the machine tools, for example, have to be considered. Also, real-time control has to occur to continuously monitor performance to ensure that no part type's due date is in jeopardy. If possible, appropriate action (or reaction) might be taken, in breakdown situations for example, to change the way the system is operated (to change ratios, for example), so as to meet the due dates. Some relevant due date-based criteria include tardiness, number of late jobs, and earliness (in a JIT environment).

In these types of situations, artificial intelligence might be useful, for the purpose of real-time, continuous monitoring. A rule based expert system could be developed to propose certain actions to take, if the system state changes drastically (i.e., a machine breakdown). Such a system could be used to choose, update, or change the production ratios as the system changes.

Different optimal ratios, all of which tend to balance workload, can be chosen by the expert system, as the system state indicates. The different set of ratios will specify different pallet distributions.

5. DETERMINING ACTUAL PRODUCTION RATIOS AND MINIMUM INVENTORY REQUIREMENTS

The aggregate ratios found by the methods of §§3 and 4 serve as guidelines to help provide input into the models now described to find actual operating ratios.

One model that is useful for these purposes is a multiclass closed queueing network (CQN), such as MVAQ. (See Hildebrant [1980], Cavailié and Dubois [1982], and Suri and Hildebrant [1984].) This aggregate stochastic model requires average input values and provides average output values. In particular, for each part type, the input required is the average visit frequency to each machine (group) and the average processing time of an operation at each machine (group). The outputs include the steady-state mean production rates of each part type, machine utilizations, and average queue lengths at each machine (group).

MVAQ can also model, at an average, aggregate level, load and unload times, refixturing times, queueing times, and transportation times. The production ratios found by the methods described in §§3 and 4 can be used to suggest numbers of pallets and fixtures of different types to maintain these calculated, optimal, aggregate ratios, subject to demand requirements, etcetera. These ratios can serve as input to the aggregate queueing network model. The output (machine utilizations) indicates how balanced the system is when the additional delay factors are included. In addition, the average production rates indicate if any due dates are in jeopardy.

Suri and Hildebrant [1984] indicate that this model is reasonably accurate and is about 10-20% pessimistic in its predictions, as compared to simulations of similar systems allowing more modeling detail. However, MVAQ is even more accurate in its relative predictions. For example, the ratios of the expected production rates of parts and machine utilizations matched those provided by simulation quite well. It is these relative values that would be of use here, to indicate ratios that provide a good balance.

The ratios could also be evaluated using the operational analysis based CQN models of Dallery and David [1983] to maximize the production rate of the various part types. These models require no sequencing assumptions (such as the usual FCFS at each machine (group)).

These models can't as yet provide information on the sequences that do provide the maximum production. The results might be used as a lower bound for an enumerative procedure.

Another useful model that could accept the production ratios found in §§3 and 4 as input to help find the minimum inventory requirements is a timed Petri net. (See Dubois and Stecke [1983].) This model could complement the queueing network model because it uses deterministic operation times. Also, it is not an aggregate model. The actual processing times and part routes are modeled. Set-up times, transportation times, and queueing times are modeled in all detail, unlike the queueing network models. Finally, it can also be used to help find an appropriate input sequence.

For a certain subclass of timed Petri nets (in particular, decision-free nets), the graphical model can be easily translated into linear state equations in a {max, +}-based algebra. (See Cohen, Dubois, Quadrot, and Viot [1983].) Decision-free means that no decisions are to be made. Everything needs to be specified in advance, such as the part routes, the input sequence,

and the like. A particular Petri net representation can be analyzed very quickly via some algorithms, based in part on Karp's [1978] efficient shortest path algorithm, to provide much information that is useful for performance evaluation.

Some of the output from the model includes the cycle time (hence the production rate), the bottleneck machine, its utilization, and the utilizations of all other machines. Some particularly useful information specifies that production can be increased by either:

- i) adding a machine of a particular type; or
- ii) inputting another pallet/fixture for a particular part type.

We indicate how this information can be used via an example.

Prior to this, recall the following. In §4, we indicated how to find aggregate production ratios that balance the workload on all machines. In most of the many examples that were examined, including all of those discussed in §4 (except the example of §4.5, which we return to shortly), the aggregate ratios found also provided the minimum numbers of pallets and fixtures required. When the aggregate processing time information indicates an unbalanced machine workload (as indicated by the total workload on each machine when only one part of each part type is considered), then the minimum number of pallets required per part type can be much less than the specified ratios. See the example of §4.5 that is defined in Table VI. We use this example to demonstrate how the Petri net model and the information provided can be easily used to determine the minimum inventory requirements.

The information that the Petri net program requires is:

- i) for each part type, the aggregate production ratios (the a_i);
- ii) also for each part type, the number of pallets/fixtures dedicated to that part type;
- iii) a part input sequence.

For the example described in §4.5, this information is:

- i) $(a_1, a_2) = (1, 10)$;
- ii) $(n_1, n_2) = (1, 4)$; (This is just to demonstrate. We know via simulation that the minimum number of pallets required is: $(1, 3)$);
- iii) $(1, 2, 2, 2, 2, 2, 2, 2, 2, 2)$.

The output from the Petri net program would indicate that we have too many pallets/fixtures for part type 2. Changing (n_1, n_2) to $(1, 3)$ in the next run provides the information that: production is maximized, the machines are balanced, there is no idle time, and these are minimum inventory and buffer requirements to maintain the optimal production ratios of $(1, 10)$.

Notice that the production ratios are also used to help find the optimal, perhaps periodic in between periods of breakdown, input sequence. It is very useful to know the relative numbers of each part type prior to the actual sequencing. Of course, most situations will not be as simple as this example of §4.5.

6. FUTURE RESEARCH NEEDS

Some considerations when using either of the two distinct objectives to determine aggregate ratios include the following. If requirements are finished simultaneously, tool magazines are changed less often. If the reason is to batch a customer's orders, then both finished goods inventory and freight costs are lower than otherwise. The objective of balancing workloads tends to decrease work-in-process inventory and increase both machine utilization and production rate. Which objective is more important is a function of many parameters. It depends on: where and how much value is added to the part; how much space is both available and required for in-process and finished goods inventory; the methods and costs of delivering the final products.

Since there are usually many good solutions for the balancing objective, it should be possible to determine ratios that satisfy both objectives "quite well." Further research is required to determine both how to do this as well as to specify a criteria that is satisfactory for both objectives. The tradeoffs involved in following one objective over the other also needs to be further investigated.

The ratios seem to impact many FMS operating problems and the methods suggest many areas in which further research is required. For example, we have seen that the ratios are useful to help determine the minimum inventory requirements in terms of numbers of pallets and fixtures. The ratios also provide guidelines for determining an appropriate part input sequence. However, further research is required to develop a more precise algorithm to find a good part input sequence. Some applicable work along these lines has been done by Hitz [1980] and Erschler, Lévêque, and Roubellat [1982]. These determine a periodic part input sequence, which could be appropriate for a relatively dedicated type of FMS in particular. However, these have been in flow shop situations, and did not allow several alternative routes. These studies assumed that there was only one pallet or fixture of each type. Both papers have also assumed that the production ratios of each part type have previously been determined. The ratios can be suggested by the procedures in §§3 and 4.

Hitz states that the efficiency of his branch and bound procedure for the permutation flow shop problem is partly a function of the "balance of work among machines." If the workload is balanced, the search is brief, since many descendent nodes can be fathomed immediately as infeasible. It appears that the procedures of both Hitz and Erschler et al. might benefit from using the approaches suggested here to calculate ratios in advance.

The ratios found here could be used to reduce the size of some other FMS combinatorial problems. The procedures might be used to reduce the size of the batching problems considered by Whitney and Gaul [1984]. The sizes of part input sequence problems could be decreased. The size of the binary matrix used by Dar-El and Sarin [1984] to list alternative routing combinations for the purpose of FMS scheduling could be reduced a priori through the use of these procedures. Whitney and Suri [1984] have formulated a large mixed integer programming problem to help select parts and machines. The objective is balancing workloads. The number of candidate part types impacts the problem complexity. The procedures reported here could be used to reduce the candidate set. In many of these cases, the ratios would impact a part type selection problem, which is hierarchically at a higher level. In all cases, further research is required to determine how these suggestions might be implemented and how these ratios could be used to simplify subsequent FMS planning and operating problems.

Easier and more automatic generation of the actual production ratios and minimum inventory requirements is needed. Also, implementation of secondary criteria for choosing ratios is needed for the situations in which multiple sets of ratios balance the workload. What are the best ways to automatically incorporate other criteria? Artificial intelligence techniques may be able to help here.

ACKNOWLEDGEMENTS

Part of this work was done while the author was visiting at Centre d'Etudes et de Recherches de Toulouse. The author would like to thank Eric Bensana, Gérard Bel, Didier Canals, and J.-B. Cavallé of C.E.R.T. and Ilyong Kim and Narayan Raman of The University of Michigan for helpful discussions. The author's work was supported in part by NSF Grant No. ECS 8406407.

REFERENCES

AKELLA, RAMAKRISHNA, CHOONG, YONG, and GERSHWIN, STANLEY B., "Performance of Hierarchical Production Scheduling Policy", Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, Ann Arbor, MI pp. 385-396 (August 15-17, 1984a).

AKELLA, RAMAKRISHNA, CHOONG, YONG, and GERSHWIN, STANLEY B., "Performance of Hierarchical Production Scheduling Policy", IEEE Transactions on Computers, Hybrids, and Manufacturing Technology, Vol. Chmt-7, No. 7, pp. 225-240 (September, 1984b).

BERRADA, MOHAMMED and STECKE, KATHRYN E., "A Branch and Bound Approach for Machine Loading in Flexible Manufacturing Systems", Working Paper No. 329, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor, MI (December 1984).

BUZACOTT, JOHN A. and SHANTHIKUMAR, J. GEORGE, "Models for Understanding Flexible Manufacturing Systems", AIIE Transactions, Vol. 12, No. 4, pp. 339-350 (December 1980).

CAVAILLÉ, JEAN-BERNARD and DUBOIS, DIDIER, "Heuristic Methods Based on Mean-Value Analysis for Flexible Manufacturing Systems Performance Evaluation", Proceedings of the 21st IEEE Conference on Decision and Control, Orlando, FL, pp. 1061-1065 (December 1982).

COHEN, GUY, DUBOIS, DIDIER, QUADRAT, JEAN P. and VIOT, M., "A Linear-System Theoretic View of Discrete-Event Systems", Proceedings of the 22nd IEEE Conference on Decision and Control, San Antonio, TX (December 14-16, 1983).

DALLERY, YVES and DAVID, R., "A New Approach Based on Operational Analysis for Flexible Manufacturing System Performance Evaluation", Proceedings of the 22nd IEEE Conference on Decision and Control, San Antonio, TX (December 14-16, 1983).

DAR-EL, E. M. and SARIN, SUBASH C., "Scheduling Parts in FMS to Achieve Maximum Machine Utilization", Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, Ann Arbor, MI, pp. 300-306 (August 15-17, 1984).

DUBOIS, DIDIER, "A Mathematical Model of a Flexible Manufacturing System with Limited In-Process Inventory", European Journal of Operational Research, Vol. 14, No. 1, pp. 66-78 (January 1983).

DUBOIS, DIDIER and STECKE, KATHRYN E., "Using Petri Nets to Represent Production Processes", Proceedings of the 22nd IEEE Conference on Decision and Control, San Antonio, TX (December 14-16, 1983).

ERSCHLER, JACQUES, LÉVÊQUE, DIDIER and ROUBELLAT, FRANCOIS, "Periodic Loading of Flexible Manufacturing Systems", Proceedings of the IFIP Congress, APMS, Bordeaux, France, pp. 327-339 (August 24-27, 1982).

HILDEBRANT, RICHARD R., "Scheduling and Control of Flexible Machining Systems When Machines Are Prone to Failure", Ph.D. Thesis, M.I.T., Cambridge, MA (August 1980).

HITZ, K. L., "Scheduling of Flexible Flow Shops-II", Report No. LIDS-R-1049, M.I.T., Cambridge, MA (October 1980).

KARP, RICHARD M., "A Characterization of the Minimum Cycle Mean in a Digraph", Discrete Mathematics, Vol. 23, pp. 309-311 (1978).

KUSIAK, ANDREW, "Loading Models in Flexible Manufacturing Systems", WP #05/83, Department of Industrial Engineering, Technical University of Nova Scotia, Halifax, Nova Scotia, Canada (March 1983).

SHANTHIKUMAR, J. GEORGE and BUZACOTT, JOHN A., "Open Queueing Network Models of Dynamic Job Shops", Working Paper No. 79-024, Department of Industrial Engineering, University of Toronto, Canada (August 1979).

SHANTHIKUMAR, J. GEORGE and STECKE, KATHRYN E., "Reducing Work-in-Process Inventory in Certain Classes of Flexible Manufacturing Systems", European Journal of Operational Research, 1986, forthcoming.

SOLBERG, JAMES J., "A Mathematical Model of Computerized Manufacturing Systems", Proceedings of the 4th International Conference on Production Research, Tokyo, Japan (August 1977).

SOLBERG, JAMES J., "Analytical Performance Evaluation of Flexible Manufacturing Systems", Proceedings of the 18th IEEE Conference on Decision and Control, San Diego, CA, pp. 640-644 (December 1979).

STECKE, KATHRYN E., "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems", Management Science, Vol. 29, No. 3, pp. 273-288 (March 1983).

STECKE, KATHRYN E., "A Hierarchical Approach to Solving Machine Grouping and Loading Problems of Flexible Manufacturing Systems", European Journal of Operational Research, 1985, forthcoming.

STECKE, KATHRYN E. and MORIN, THOMAS L., "The Optimality of Balancing Workloads in Certain Types of Flexible Manufacturing Systems", European Journal of Operational Research, Vol. 20, No. 1, pp. 68-82 (April 1985).

STECKE, KATHRYN E. and SOLBERG, JAMES J., "The Optimality of Unbalancing Both Workloads and Machine Group Sizes in Closed Queueing Networks of Multiserver Queues", Operations Research, Vol. 33, No. 4, pp. 882-910 (July-August 1985).

STECKE, KATHRYN E. and TALBOT, F. BRIAN, "Heuristic Loading Algorithms for Flexible Manufacturing Systems", Proceedings of the Seventh International Conference on Production Research, Windsor, Ontario, Canada (August 22-24, 1983).

SURI, RAJAN, "Robustness of Queueing Network Formulae", Journal of the Association for Computing Machinery, Vol. 30, No. 3, pp. 564-594 (July 1983).

SURI, RAJAN and HILDEBRANT, RICHARD R., "Modelling Flexible Manufacturing Systems Using Mean Value Analysis", Journal of Manufacturing Systems, Vol. 3, No. 1, pp. 27-38 (January 1984).

SURI, RAJAN and WHITNEY, CYNTHIA K., "Decision Support Requirements in Flexible Manufacturing", Journal of Manufacturing Systems, Vol. 3, No. 1, pp. 61-69 (January 1984).

WHITNEY, CYNTHIA K. and GAUL, THOMAS S., "Sequential Decision Procedures for Batching and Balancing in FMSs", Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, Ann Arbor, MI, pp. 243-248 (August 15-17, 1984).

WHITNEY, CYNTHIA K. and SURI, RAJAN, "Decision Aids for FMS Parts and Machine Selection", Proceedings of the First ORSA/TIMS Special Interest Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, Ann Arbor, MI, pp. 205-210 (August 15-17, 1984).

YAO, DAVID D. W., "Some Properties of Throughput Function of Closed Networks of Queues", Technical Report, Columbia University, New York, NY (1984).