

Research Support
University of Michigan Business School

April 1998

AGGREGATE PRODUCTION PLANNING OF
FLEXIBLE MANUFACTURING SYSTEMS
USING THE CONCEPT OF OPERATION TYPES

WORKING PAPER #98007

BY
TAMÁS KOLTAI
TECHNICAL UNIVERSITY OF BUDAPEST
ANDRÁS FARKAS
INTERNATIONAL MANAGEMENT CENTER-HUNGARY
KATHRYN E. STECKE
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

Aggregate Production Planning of Flexible Manufacturing Systems Using the Concept of Operation Types

Tamás Koltai

*Technical University of Budapest
Department of Industrial Management
1111 Budapest, Mûegyetem rkp. 9. T. Bld., Hungary
E-mail: koltai@vvg.bme.hu*

András Farkas

*Graduate School of Business Administration
International Management Center
1051 Budapest, Nádor utca 11, Hungary
E-mail: farkas@imc.hu*

Kathryn E. Stecke

*University of Michigan Business School
Department of Operations Management
Ann Arbor, Michigan 48109-1234, USA
E-mail: kstecke@umich.edu*

April 1998

Abstract

Many manufacturing systems often need to analyze whether there is sufficient capacity to perform several technologically-different operations for their current set of production orders. To perform such aggregate production planning, the concept of an "operation type" is introduced. A 0-1 mathematical programming model is developed to help determine how to configure a manufacturing system to avoid excessive capacity over- and/or under-utilization. Also, several management priorities, such as tooling cost, machine pooling, and flexibility issues, are considered. The model is illustrated through a sample problem and sensitivity analyses is performed to analyze the effects of unexpected changes in the capacity requirements and in the machine capacities.

Keywords: flexible manufacturing systems, production planning, mathematical programming, sensitivity analysis

1. Introduction

It is indigenous to a manufacturing system to check whether there is enough capacity from the system for the different operations demanded by the production requirements. Production planning for the equipment that performs different operations in conventional manufacturing systems is more straightforward than in flexible manufacturing systems. In some conventional systems, the capacity available for production can be determined directly from the available capacities of the different single-purpose machines, as they can usually perform only a small variety of operations. (Conventional and manual systems may also contain general-purpose machine tools.) A system with multi-purpose computer numerical control (CNC) machines provides additional opportunities to increase system utilization through the machine flexibility, since each machine can be used for a variety of operations. In this case, it is useful to know if there is enough capacity to perform all of the technologically different operations. If there is enough capacity, then next, the alternative uses of the machines for the different operations need to be analyzed. For example, in a combined system with mixed types (single- and general-purpose) of machines, decisions concerning whether certain operation types should be assigned to the single-purpose machines (or not) can be considered in order to leave the CNCs free for more complex or special operations.

The paper provides an aggregate production planning model. Aggregation means that the similar manufacturing operations, which require the same type of machines, are aggregated into operation types. Applying the concept of "operation types", the variety of the alternative uses of the production capacity can be reduced and a systematic evaluation of capacity utilization and workload balance is possible. Balancing the workload means that

- operation types should be assigned to the machines in such a way that there is enough capacity to perform all of the required operation types. That is, there is no overload of machines.
- operation types should be assigned to the machines in such a way that enough workload can be assigned to the machines to avoid idle time. That is, there is no underload of machines. If, for example, a machine is tooled just for drilling, and there is hardly any requirement for this operation type, then this machine is under-utilized, while perhaps there is a lack of capacity from other operation types.

In case of balanced workload, both capacity over- and under-utilization of machines are avoided.

An easy-to-use mathematical model and graphical illustrations are provided to analyze the capacity over- and under-utilization of the machines, as well as of the entire system. In addition to balancing workload, tooling costs, machine pooling, and other flexibility-related management considerations are part of the model as well.

Aggregation is a widely used tool in production planning. When some elements of a production system are treated in an aggregate manner, simpler models can be applied for capacity, inventory, and production planning. Obviously some information is lost by aggregation, but at a certain level of decision-making, an aggregate treatment can be sufficient. When more detailed information is required, then the aggregated concepts are disaggregated and/or a more detailed model is applied. In traditional production planning models, products and/or facilities are aggregated (Thomas and McClain, 1993). Products using the same setup of the production process are aggregated into product families and/or products with similar resource consumption are aggregated into product types. When a good production plan and capacity utilization is determined, a detailed production program can be prepared, in which product types and/or families are disaggregated into products (see, for example, Johnson and Montgomery, 1974 and Hax and Candea, 1984). Facility-level aggregation means that several different resources of the production, such as machines, workforce, and materials are considered as a single resource or facility (see, for example, Holt et al. 1960).

Special-purpose machines perform just a small set of technologically different operations. In this case, aggregation of machines is approximately equivalent to the aggregation of operations. In case of multi-purpose CNCs, machines and operations have to be treated separately, since the operation manager decides on the set of operations a machine has to perform. This fact was recognized by Niess (1980), who aggregated similar operations into "operation types". Niess developed an algorithm to determine a series of sets of orders for several production periods. The generated production program provides balanced capacity utilization, that is, there is no excess capacity over- and under-utilization. Niess applied this method for conventional production systems consisting of several single- and multi-purpose machines.

To solve the FMS production planning problems, Stecke (1983) proposed a nonlinear integer production planning model. The size of this model, however, can preclude frequent application,

especially for a large FMS. Stecke (1986) also proposed a hierarchical approach to the different planning issues. A hierarchical approach was also proposed by Van Looveren et al. (1986) and Jaikumar and Van Wassenhove (1989) to handle the increased complexity of FMS production planning. Stecke and Kim (1991) suggest an aggregate approach to sequence the production of orders for different types of parts to balance the workload for the various machine types.

Production planning of an FMS in an MRP environment was presented in Mazzola (1989). In this case, part type selection is determined by the MRP system in the rough cut capacity planning phase. For the selected part types, first grouping and loading of machines is solved, then detailed scheduling is made. Since part type selection is solved independently of the grouping and loading problem, the complexity of the model is somewhat reduced, but still the solution of real case size models requires efficient heuristics.

The complexity of the FMS production planning problems implies that aggregation could be an appropriate approach. The technological characteristics of these systems, however, require modification of traditional aggregation concepts. The capability to perform a large number of technologically different operations can cause difficulties in production planning as well as benefits for operation. The main contribution of this paper is to show how the increased complexity of FMS production planning problems can be addressed by aggregating operations into operation types. Then the aggregation concept applied for conventional systems can be extended for FMSs. The resulting models help to attain better capacity utilization, higher throughput, and ultimately more efficient operation.

The paper is organized as follows. In section 2, the concepts of operation type and ideal available capacity range are introduced and the basic definitions and notation used in this paper are explained and illustrated with the help of a sample problem. Section 3 presents a 0-1 mathematical programming model for allocating operation types to machines while considering workload balance and some typical management priorities. Section 4 illustrates the application of the model to a sample problem, and demonstrates that balanced workload can be achieved with different setup configurations. Section 5 demonstrates how sensitivity analyses can be performed to analyze the consequences of changes in capacity requirements and machine capacities. Finally, section 6 provides some general conclusions.

2. Basic Definitions and Concepts

Definitions and concepts required for the development of our production planning approach are described and illustrated in this section. The necessary subscripts and the parameters are summarized in Table 1.

A flexible manufacturing system (FMS) is a collection of machines, linked together by an automated materials handling system and directed by a central computer. An order (for a single part type) consists of r_i parts of type i . Each part type has a finite number of operations. An operation, o_j , is a machine visit, and is defined by its processing time on a given machine and by the set of cutting tools required. The number of cutting tools required for each operation j , for a particular order i , no_{ij} , are determined by the production requirements demanded by the orders, and by the tool lives of the required tools of the different types and by the processing times of each operation. A machine visit requires a pallet, a fixture, and whatever is on the fixture. This is usually a single part, but may also be multiple parts of the same type or multiple parts of several types, in one or more mounts. Each part type may have a (partial) precedence.

An operation type, ot_h , is a fundamental type of manufacturing process. For example, possible operation types can be milling, tapping, boring, drilling, or turning. The operation type(s) which can be assigned to a machine is (are) determined by the machine's processing capability. For example, a horizontal milling machine cannot perform vertical operations without refixturing the part. An operation type set, S_k , consists of a single or multiple operation type(s). For our purposes here, an operation (requiring a set of cutting tools) belonging to one operation type is indivisible. Consecutive operations may (or not) be assigned to the same machine. For each operation type (e.g., horizontal milling), there is a maximum number of tools (i.e., 210 tool types) that cover all operations that are members of that particular operation type. In any particular *period* over the short term planning horizon (i.e., one month, a couple of weeks), there is a set of required orders (i.e., 10-15 orders/shift). The operations of these orders of a particular operation type use a number of tools of different types (i.e., 84 tools). This is the tool set of the *period's* tooling requirement for those orders that require a particular operation type. We call this an operation type tool set, nt_h . The tool set of an operation type set is determined by the sum of the number of tools (including copies) required by the operation type(s) belonging to that operation type set. We call this operation type set tool set, ns_k . A limited

number of tool sets for each operation type, n_h , is available in a period for tooling up the machines. For example, three sets of tools of an operation type might be available from which two sets of tools can be used for tooling up two machines while one set of tools is being refurbished. Each operation type tool set must include any required multiple copies of tools needed for large orders.

The CNC machines can be single-purpose machines (a single operation type only can be performed) or multi-purpose machines (two or more operation types can be performed). Capacity of a machine, c_m , here is expressed in capacity units (CUs), over a period of, for example, a shift or two, a day, or a week. Processing time of operation j of order i , p_{ij} , is also expressed in terms of CUs, rather than in hours or minutes. The capacity requirement for an operation type set is the number of CUs per period demanded by the production requirements. The production requirements for the manufacturing system are given from a master production schedule (MPS) periodically, say each shift, day, or week. An upper capacity bound (UB) of a particular operation type set is the *maximum* amount of *available capacity* for an operation type set. It is calculated as the sum of the CUs of those machines which are capable of performing *any and all* operations belonging to that operation type set. A lower capacity bound (LB) of an operation type set is the *minimum* amount of planned *free* capacity for an operation type set that is available only to the operations that belong to that operation type set. It is calculated as the sum of the CUs of those machines which are capable of performing *only* those operations belonging to that particular operation type set. This is free capacity because if it isn't used, there is idle time on the machines. The ideal available capacity per period for an operation type set is a range defined by the difference between the upper and the lower capacity bounds of available capacity. The capacity is sufficient if the production requirements from all operation type sets are less than their corresponding upper bounds. When all operations have been assigned to machines and the workload is less than the lower capacity bound of any operation type set, then there is machine idle time.

Weights are given to the operation type sets to trigger several possible objectives. When higher weights are given to multiple operation type sets, the objective is to maximize "flexibility", i.e., to maximize the number of operation types assigned to a machine (to maximize the manufacturing capability of the machines). When higher weights are given to single operation type sets, the objective is to minimize "flexibility". In either case, an attempt is made to balance the workload per

machine over the planned periods. Intuitively, a better balance should be attained with higher flexibility.

The introduced notation and concepts are illustrated with the help of Figures 1, 2, 3, 4, and Table 2. There is a mixed type manufacturing system consisting of five metal cutting machines. This set of machines includes three multi-purpose CNC machines (M1,M2,M3) and two single-purpose machines (M4,M5) as shown in Figure 1. Single-purpose machines can either be conventional or CNC machines. The system was designed to perform three operation types (ot_1 , ot_2 , and ot_3), which are drilling, vertical milling, and horizontal milling, respectively. These main manufacturing capabilities, which are attainable through tooling up the machines, are indicated in Figure 1. For example, ot_1 and ot_2 at machine #3 indicates that M3 is capable of performing operation types 1 and 2, (i.e., M3 does drilling and vertical milling), whereas M5 is a single-purpose machine for ot_1 (drilling). Tooling is determined by all tools that are technically specified for the part types planned for production in a particular period. Hence, M5 requires the tool set of a single operation type set (S_1), while M3 requires all the tools of ot_1 and ot_2 which compose the tool set of a multiple operation type set (S_4). (Operation type sets S_k , $k=1,\dots,7$ are provided in Figure 2).

The production requirements determined by the MPS are given in Table 2 for one production period. One production period spans one shift. The different part types manufactured in the system are termed as orders. Identification of these orders is done by the code numbers #101 through #110. The required operation types are indicated in the heading of Table 2, while their corresponding operations are given by the calculated processing times expressed in CUs (i.e., order #101 requires 0.10 CUs for drilling, 0.50 CUs for vertical milling, and 0.06 CUs for horizontal milling). 0.10 CUs means that in a single eight hour shift, 48 minutes of drilling operation is required. In practice, for order #101, the 0.10 CUs of drilling may represent the drilling operations of 5 borings with different diameters and bore lengths, which can be processed with 5 different tools during 48 minutes. This can be done, for example, by one machine visit either on M5, or alternatively, on M3, or on M1.

The ideal available capacity range for each operation type set is displayed in Figure 2. The possible operation type sets are placed on the horizontal axis. In this case, there are three operation type sets (S_1 , S_2 , S_3) with single operation types, three operation type sets (S_4 , S_5 , S_6) with two operation types, and one operation type set (S_7) with three operation types. In general, the total

number of the operation type sets, K , can be calculated as

$$K = \left(\sum_{h=1}^{H-1} \frac{H!}{h!(H-h)!} \right) + 1,$$

where H is the total number of single operation types.

The vertical axes represent the lower capacity bounds (LBs) and the upper capacity bounds (UBs) of each operation type set. For example, for the operation type set S_1 , $LB=1$, because $M5$ is the only machine which is dedicated for ot_1 ; $UB=3$, since $M1$, $M3$, and $M5$ are all capable of performing ot_1 . For S_4 , which is an operation type set containing operation types ot_1 and ot_2 , $LB=3$ because $M3$ can perform ot_1 and ot_2 , $M4$ can perform ot_2 , and $M5$ can perform ot_1 . For S_4 , $UB=5$, as all of the machines can do either ot_1 or ot_2 . If the capacity requirement is smaller than the LB of any operation type set, then the system is underloaded, and there is free capacity on one or more machines. For example, if less than 1 CU from ot_1 is required, then the single-purpose drilling machine, $M5$, will have idle capacity. If the capacity requirement is greater than the UB of any operation type set, then the system is overloaded, i.e., there is not enough available capacity for the required operations. For instance, if more than 3 CUs are required from ot_1 , then $M1$, $M3$, and $M5$ will not have enough capacity to perform the drilling operations for the 10 orders in period 1. When the production requirements of each operation type set are within the lower and the upper capacity bounds (the shaded area in Figure 2), then there are neither unutilized nor excess capacity on any of the machines. In the most favorable case, the capacity requirements of a set of orders of a given period fall into the ranges defined by the lower and the upper bounds for each operation type set.

The capacity requirement for each operation type set is shown in Figure 3. It is calculated as the total processing time of all operations of the corresponding operation type set.

Figure 4 plots both the ideal available capacity ranges and the capacity requirements for each operation type set on one chart. The resulting graphical display is used to determine whether or not the manufacturing system is in complete technological balance, i.e., to check whether there is any excess capacity or lack of capacity from certain operation types. For example, in our case, Figure 4 illustrates that there are under-loadings at S_1 and S_4 . The under-loading at S_1 indicates that if all

drilling operations are processed on M5, then this machine still has idle capacity. The under-loading at S_4 indicates that idle capacity may exist on M3 and on M5. The under-loading of S_4 is higher than the under-loading of S_1 , since the total processing time of ot_2 does not require all of the available capacities of M3 and M4. Therefore the idle capacity of M5 is increased by the idle capacity of M3. The idle capacity of M3 can be utilized for a portion of ot_1 , but then the idle capacity of M5 increases the under-loading, as can be seen at S_4 .

Figure 4 also shows over-loadings in S_3 , S_6 , and S_7 . The over-loading at S_3 indicates that the total capacity requirements of ot_3 are higher than the total available capacity on M1 and on M2. The over-loading at S_6 indicates that the capacity requirements of ot_2 and ot_3 together are higher than the available capacities on machines M1, M2, M3, and M4. Finally, the over-loading at S_7 means that the capacity requirements of all of the operation types are higher than the available capacities of all of the machines.

In summary, on the one hand there could be idle capacity on certain machines, while there might not be enough capacity to fulfill all of the requirements for certain operation types. Figure 4 also displays the increased ideal capacity range when 20 percent capacity over- and under-utilization are acceptable for the management. The thin dotted line represents the increased upper bounds, and the thin dashed line shows the decreased lower bounds. The capacity requirements are within this extended range for all operation type sets.

3. Model Formulation

The objective of the model is the optimal allocation of the operation types to machines necessary to produce the orders of a given period. A 0-1 mixed integer programming model is developed, which ensures that the production requirements remain within the ideal capacity ranges, tool availability conditions are met, and some major managerial objectives are considered.

The formulation of the mathematical model is based on the notation defined in Section 2 and summarized in Table 1. The binary decision variable x_{km} is equal to one if operation type set k is assigned to machine m , and it is equal to zero, otherwise. The integrality constraint is as follows,

$$x_{km} = 0 \text{ or } 1, \quad k=1, \dots, K; \quad m=1, \dots, M. \quad (1)$$

Whenever an operation type set k is assigned to machine m , then this machine is capable of performing all of the operation types of that particular set.

Constraint (2) imposes that the total processing time of all operations of operation type set k should be higher than the lower capacity bound of operation type set k , that is,

$$(1-\alpha) \cdot \sum_{S'_{k'}} \sum_{m=1}^M c_m \cdot x_{km} \leq p_k, \quad \forall S'_{k'} \subseteq S_k; \quad k=1,\dots,K; \quad k'=1,\dots,K'. \quad (2)$$

Since the lower capacity bound is the capacity of those machines which are capable of performing only the operations of operation type set k , this constraint ensures that there will be no idle capacity on these machines. In real manufacturing systems, production managers must generally be resigned to a certain amount of idle capacity. Idle capacity is either planned and serves as buffer capacity to absorb the effect of unexpected events (i.e., machine breakdowns, tool breakages, quality problems), or it is the consequence of scheduling constraints. This fact is incorporated in the model with parameter α , which expresses the acceptable percentage of capacity under-utilization of the machines. For example, in Figure 4 the total processing time for all operations of S_1 is lower than the lower capacity bound of S_1 which would make constraint (2) infeasible even for 10 percent acceptable under-utilization ($\alpha=0.1$), but feasible for 20 percent acceptable under-utilization ($\alpha=0.2$). For S_2 , the production requirement is above the lower bound, therefore, constraint (2) is satisfied.

Note that the model helps to avoid capacity under-utilization resulting from improper assignment of operation types to machines. If machines are idle because of lack of production requirements, then there is no feasible solution for equation (2). In this case other operation management measures must be taken, such as decreasing the number of machines assigned to the current set of orders, increasing order number or order sizes, or ultimately accepting a high value for α .

Constraint (3) imposes that the total processing time for all operations of operation type set k should be lower than the upper capacity bound of operation type set k , that is,

$$(1+\beta) \cdot \sum_{\{\forall S_k | \text{ot}_h \in S_k\}} \sum_{m=1}^M c_m \cdot x_{km} \geq p_k, \quad k=1,\dots,K; \quad h=1,\dots,H. \quad (3)$$

As the upper capacity bound is the maximum amount of available capacity for the operation type set k , this constraint ensures that there will be enough capacity to perform all operations of operation type set k . In real manufacturing systems, production managers need to work around a certain amount of lack of capacity. In these cases, management may consider overtime, subcontracting, or other possible capacity adjustments. This fact is incorporated in the model with the parameter β , which expresses the acceptable percentage of capacity over-utilization. For example, in Figure 4 the total processing time for all operations of S_2 is lower than the upper capacity bound of S_2 , which makes constraint (3) feasible. For S_3 , the total processing time is higher than the upper bound and therefore, a certain amount of overtime is required. If 20 percent capacity over-utilization is acceptable, then constraint (3) for S_3 is feasible ($\beta=0.2$).

Constraint (4) ensures that only one operation type set is assigned to each machine, that is,

$$\sum_{k=1}^K x_{km} \leq 1, \quad m=1, \dots, M. \quad (4)$$

Constraint (5) describes the limitations on tool availability:

$$\sum_{(\forall S_k | ot_h \in S_k)} \sum_{m=1}^M x_{km} \leq n_h, \quad k=1, \dots, K; \quad h=1, \dots, H. \quad (5)$$

The left-hand side of constraint (5) calculates how many times each operation type appears in the allocated operation type sets. These calculated values should be less than the maximum number of available tool sets of each operation type.

There may be several operation type set allocations satisfying constraints (1)-(5), especially when relatively high values for α and β are chosen. However, they can *all* make workload balance possible. Stecke and Solberg (1985) and Stecke and Morin (1985) showed that balancing the workload of machines in FMSs may provide optimal throughput under certain conditions. Note, however, that workload is not allocated to the distinct machines by this model. Allocating operation types to machines determines how much workload can potentially be assigned to them in the next

phase, when production scheduling is done. For example, consider a system with two machines (M1, M2), and requirements for two operation types (for example drilling and milling). Let the required processing time be equal to 0.1 CU for drilling and 0.9 CU for milling. In *case a*, drilling is allocated to M1 and milling is allocated to M2. The maximum amount of production which can be performed is 0.1 CU on M1 and 0.9 CU on M2. In *case b*, both drilling and milling are assigned to M1 and only milling is assigned to M2. In this latter instance, 0.1 CU of drilling and 0.4 CU of milling on M1 and 0.5 CU of milling on M2 can provide a perfect workload balance. Allocation of the exact processing times to machines is not the objective of this model. However, if operation type allocation does not consider workload balancing, then the system may not be balanced when production scheduling is actually performed (see *case a*). Constraint (2) limits the possible capacity under-utilization of machines with the help of a specified α . This way the situation of *case a* of the previous example is avoided.

Besides workload balance, other management priorities can be considered with the introduction of weights assigned to the operation type sets. Applying such weights, the following objective function is formed:

$$\sum_{k=1}^K \sum_{m=1}^M w_{km} x_{km} \quad (6)$$

The resulting model can be either a minimization or a maximization problem.

One management priority is the *minimization of tooling costs*. After purchasing, tooling costs include the costs of preparation (sharpening, measuring, repairing) and loading (changing tools in the tool magazine, machine down time, adjustments) of the operation type set tool set. The tooling cost of an operation type set tool set can be expressed by the weights, w_{km} , which can be determined by the number of tools and by the technical complexity and geometrical characteristics of the parts. Applying these weights, an objective function is developed, which should be minimized.

Another management priority can be to utilize the benefits from *machine pooling*. It has been shown that machine pooling may improve system performance (Kleinrock, 1975, Stecke and Solberg, 1985). Pooling machines alleviates the effects of machine breakdowns, facilitates part routings, and

reduces material handling of the parts. The ideal pooling would be to load every machine with all tools for every operation type. This strategy, however, is usually neither feasible (tool and tool magazine limitations), nor necessary (there may be small workload requirements of certain operation types relative to others). Pooling should be done based on the production requirements. When machines are pooled, they are identically tooled to perform an operation type set S_k . That is, the workload of operation type set S_k , which is given by p_k , will be split among all of the machines pooled for S_k . The workload can be balanced among pooled machines when the average workload of the pooled machines is close to the average workload of all the machines. This can be expressed by the following weights.

$$w_{km} = \left| \frac{p_k}{\text{INT} \left[\frac{p_k}{p_{average}} \right]} - p_{average} \right|, \quad k=1, \dots, K; \quad m=1, \dots, M, \quad (7)$$

where

$$p_{average} = \frac{1}{M} \cdot \sum_{h=1}^H p_h. \quad (8)$$

A general belief of management, and a naive measure, is that flexibility can be defined by the number of operations that a machine can perform. More operations tends to mean more flexibility. In this case, if the same number of operations can be performed on two machines, then the flexibility is the same. Then an objective function that assigns higher weights to multiple operation type sets expresses the management priority of *maximizing flexibility*. Different managerial attitudes and business objectives, demand patterns, and manufacturing conditions may cause different evaluations of flexibility at different plants. The different evaluations of flexibility can be expressed with utility measures. In the literature, there are approaches to elicit the necessary information, and to quantify them (Keeney and Raiffa, 1976, Wakker, 1989). Depending on whether the advantages of flexibility

or the disadvantages of the lack of flexibility are expressed with the weights (w_{km}), either a maximizing or a minimizing objective function can be formulated.

An easy way to form weights for flexibility is to normalize the production requirements of each operation type set in a given period. In this case, flexibility will be required just for those operation type sets which have relatively high workloads. These weights can be calculated as

$$w_{km} = \frac{P_k}{\sum_{h=1}^H P_h}, \quad k=1, \dots, K; \quad m=1, \dots, M. \quad (9)$$

4. Numerical Illustration

The data used in Section 2 is applied to illustrate the formulation and solution of the aggregate production planning model. The manufacturing system consists of five identical CNC machines. These machines should be able to process the orders of one period, listed in Table 2. The operations of the orders are aggregated into three operation types, which form seven operation type sets. The total processing time of each operation type set, p_k , is given in the second row of Table 3. The acceptable levels of capacity under- and over-utilization of the machines are equal to 20 percent ($\alpha=\beta=0.2$). The available numbers of the operation type tool sets, n_h , are three, four, and two, respectively.

With this data, the mathematical model consists of seven lower bound constraints (2), seven upper bound constraints (3), five constraints (4), three tool availability constraints (5), and thirty five 0-1 decision variables. The model is solved for four different objective functions. The objective function coefficients (w_{km}) of the four objective functions are given in Table 3. Since all five machines are considered identical, the weight of an operation type set is identical for all the five machines ($w_{k,1} = \dots = w_{k,5}$). The corresponding optimal solutions are given in Table 4.

The weights, w_{km} , applied for the *tooling cost minimization* problem are given in the third row of Table 3. These data were estimated by experts and they reflect the higher costs of the use of multiple operation types. When these weights are applied, the use of dedicated machines is favored. This is reflected by the results, given in the second row of Table 4. It can be seen that only one operation type is assigned to each of the five machines. M2 is loaded for ot_1 , M1 and M5 are loaded

for ot_2 , and M3 and M4 are loaded for ot_3 . The results also show that the orders can be manufactured in a system consisting of dedicated machines only, with less than 20 percent capacity over- and under-utilization.

The weights, w_{km} , applied for the *machine pooling maximization* problem are given in the fourth row of Table 3. These coefficients are calculated by equations (7) and (8). In case of $w_{1,m}$, equation (7) yields infinity, which indicates that machines loaded just for ot_1 are not suitable for pooling at all. Since the total processing time of operation type set S_1 does not fill the available capacity of a single machine, it is obviously not logical to pool machines with this operation type set. This pooling would provide much excess capacity. In case of $w_{7,m}$, equation (7) yields zero, which means that a perfect workload balance can be gained with the total pooling of all five machines. This, however, is not possible because of the tool availability constraints. The results are given in the third row of Table 4. It shows that the first three machines are pooled and loaded for ot_1 and ot_2 , M4 is loaded for ot_2 and ot_3 , and M5 is loaded for ot_3 .

The first *flexibility maximization* problem uses utility measures in the objective function. The applied weights, w_{km} , are given in the fifth row of Table 3. These are the same coefficients as the weights for the tooling cost objective function, but now this function is maximized. It can be seen that the highest priority is given to the machines which are loaded for all of the operation types. The results, in the fourth row of Table 4, show that two machines are loaded with all of the operation types. This result may also suggest that two flexible machines and three dedicated machines can produce all operations of the orders in this period.

The second *flexibility maximization* problem uses the relative workload of operation type sets in the objective function. The applied weights, w_{km} , are given in the sixth row of Table 3. These coefficients are calculated by equation (9). In this case, one machine is loaded for all operation types, two machines for two operation types, and two machines are operated in a dedicated manner, as is shown in the fifth row of Table 4. This result was applied to the sample problem in section 2, and its implementation is shown in Figure 1.

The solutions of all four problems make it possible to balance the workload when the orders are assigned to the machines in the next planning phase, provided that 20 percent capacity over- and under-utilization is acceptable.

5. Sensitivity Analysis

Sensitivity analyses of the parameters of the model can help to analyze how certain changes affect the capacity over- and under-utilization of the manufacturing system. The production requirements (p_h) of an operation type and the machine capacity (c_m) are the most important parameters altered by unexpected changes. For these instances, sensitivity analysis can determine the *validity range* of a chosen parameter, within which the system remains balanced, that is, the capacity requirements remain within the ideal available capacity range. This sensitivity range can be determined by the *feasible increase* and the *feasible decrease* of a given parameter.

5.1. Sensitivity analysis of the total processing time of an operation type (p_h)

Every change concerning the orders may result in a change of their production requirements. A customer may request an increase of an order. Another customer may cancel an entire order. A customer may require a small modification of the product which may result in, for example, increased drilling time. In case of a rush order, the production requirements of all operation types may change. The sensitivity analysis of the total processing time of an operation type can help to analyze the consequences of these situations.

The sensitivity range of a particular processing time, p_h , can be determined by calculating the possible change of p_h for all operation type sets that contain operation type ot_h . That is, a range is computed for all S_k , when $ot_h \in S_k$. The feasible decrease of the total processing time of operation type h , Δp_h^- , is determined by the minimum of the algebraic differences between the processing times and the capacity lower bounds for all operation type sets which contain operation type ot_h , that is,

$$\Delta p_h^- = \underset{\{\forall S_k | ot_h \in S_k\}}{\text{Min}} \left[p_k - (1-\alpha) \cdot \sum_{S'_k} \sum_{m=1}^M c_m \cdot x_{km} \right] \quad \begin{array}{ll} k=1, \dots, K; & k'=1, \dots, K'; \\ h=1, \dots, H & \forall S'_k \subseteq S_k. \end{array} \quad (10)$$

The feasible increase of the total processing time of operation type h , Δp_h^+ , is determined by the minimum of the algebraic differences between the capacity upper bounds and the processing times for all operation type sets which contain operation type ot_h , that is,

$$\Delta p_h^+ = \underset{(\forall S_k | ot_h \in S_k)}{\text{Min}} \left[(1+\beta) \cdot \sum_{(\forall S_k | ot_h \in S_k)} \sum_{m=1}^M c_m \cdot x_{km} - p_k \right] \quad \begin{array}{l} k=1,\dots,K; \\ h=1,\dots,H. \end{array} \quad (11)$$

The results obtained from equations (10) and (11) when applied to the sample problem are given in Table 5. The sensitivity ranges are valid for 20 percent acceptable capacity over- and under-utilization ($\alpha=\beta=0.20$). The second row of Table 5 shows that the capacity requirement of ot_1 cannot be decreased without the violation of the lower capacity bound ($\Delta p_1^- = 0$). Figure 4 shows that for S_1 , the capacity requirements are equal to the 20 percent decreased value of the lower bound. Therefore the minimum in (10) is found at S_1 . The acceptable increase of the capacity requirement of ot_1 is equal to 0.9 ($\Delta p_1^+ = 0.9$). The change of the capacity requirement of ot_1 affects operation type sets S_1 , S_4 , S_5 , and S_7 . Checking the differences between the 20 percent increased value of the upper bound and the capacity requirements of these operation type sets, the minimum of equation (11) is found at S_7 .

For ot_2 , an acceptable decrease of p_2 is equal to 0.3 ($\Delta p_2^- = 0.3$). The change of the capacity requirement of ot_2 affects operation type sets S_2 , S_4 , S_6 , and S_7 and the minimum of the differences is found at S_4 , as is shown in Figure 4. The feasible increase of p_2 is equal to 0.5 ($\Delta p_2^+ = 0.5$) and this value is found at S_6 .

For ot_3 , $\Delta p_3^- = 1.1$, that is, a considerable decrease of the capacity requirements of ot_3 is possible without violating the lower bound. The value of zero for the feasible increase of the capacity requirement of ot_3 ($\Delta p_3^+ = 0$) shows that the production system cannot perform more of this operation type unless it exceeds the available capacity by more than 20 percent.

The results in Table 5 are independent validity ranges, that is, the feasible decrease and increase is valid only if the capacity requirements of a single operation type change. When the capacity requirements of more than one operation type set changes, a joint range for all of the simultaneously changing parameters can be determined.

5.2. Sensitivity analysis of the machine capacity (c_m)

The capacity of the machines may decrease because of machine breakdowns, unexpected production stops, or waiting for operators, repair-persons, or tools and materials. The capacity

increase can be regarded as scheduled overtime or as the organization of extra shifts. The sensitivity analysis of the machine capacity can help analyze these situations.

The sensitivity range of the available capacity of a particular machine can be determined by calculating the feasible change of the upper and lower capacity bounds of all of those operation type sets which are affected by the change of the operation type set assigned to that machine. For example, if a machine is tooled just for drilling, then the feasible change of the upper and lower capacity bounds of all of those operation type sets which contains drilling has to be examined.

The capacity decrease of a machine diminishes both the lower and upper capacity bounds. For our purposes, the decrease of the upper bound is relevant, because it may result in an infeasible capacity over-utilization. When the capacity of a particular machine changes, then all of the capacity upper bounds of those operation type sets, which contain any and all of the operation types assigned to this machine, are affected. The feasible decrease of the capacity of machine m , Δc_m^- , is determined by the minimum of the algebraic differences between the capacity upper bound and the processing time for all of the affected operation type sets, that is,

$$\Delta c_m^- = \underset{\{\forall S_k | x_{km}=1, \text{or } h \in S_k\}}{\text{Min}} \left[(1+\beta) \cdot \sum_{\{\forall S_k | \text{or } h \in S_k\}} \sum_{m=1}^M c_m \cdot x_{km} - p_k \right] \quad \begin{array}{l} k=1, \dots, K; \\ h=1, \dots, H; \\ m=1, \dots, M. \end{array} \quad (12)$$

The capacity increase of a machine augments both the lower and upper capacity bounds. For our purposes, the increase of the lower bound is relevant, because it may result in an infeasible capacity under-utilization. When the capacity of a particular machine changes, all of those capacity upper bounds of operation type sets, for which the operation type set assigned to the machine is a subset, are affected. The feasible increase of the capacity of machine m , Δc_m^+ , is determined by the minimum of the algebraic differences between the processing time and the capacity lower bound for all of the affected operation type sets, that is,

$$\Delta c_m^+ = \underset{\{\forall S_k | x_{k'm}=1, S_{k'} \subseteq S_k\}}{\text{Min}} \left[p_k - (1-\alpha) \cdot \sum_{S_{k'}} \sum_{m=1}^M c_m \cdot x_{km} \right] \quad \begin{array}{l} k=1, \dots, K; \quad k'=1, \dots, K'; \\ h=1, \dots, H \quad \forall S_{k'} \subseteq S_k; \\ m=1, \dots, M. \end{array} \quad (13)$$

The results computed for the sample problem using (12) and (13) are given in Table 6. The sensitivity ranges are valid for 20 percent acceptable capacity over- and under-utilization ($\alpha=\beta=0.20$). The second row of Table 6 shows that the capacity of M1 cannot be decreased without violating the upper capacity bounds ($\Delta c_1^- = 0$). All operation types are assigned to M1. Therefore, every operation type set contains at least one of the operation types of M1. Thus, the difference between the 20 percent increased value of the upper bounds and the capacity requirements for all operation type sets must be checked, and the minimum of equation (12) is found at S_3 . Figure 4 shows that for S_3 , the capacity requirement coincides with the acceptable capacity over-utilization. Therefore, the capacities of those machines which contain ot_3 cannot be decreased. This also explains the zero value found for the feasible decrease of the capacity of M2 ($\Delta c_2^- = 0$). The feasible increase of the capacity of M1 is equal to 1.1 ($\Delta c_1^+ = 1.1$). The operation type set assigned to M1 is S_7 . This is the largest set and it affects exclusively the lower capacity bound of operation type set S_7 . (S_7 is a subset of only itself.) Thus, the difference between the capacity requirements and the 20 percent decreased values of the lower bounds must be checked at S_7 and it is equal to 1.1. For M2, the assigned S_6 is a subset of S_7 and a subset of itself. The smallest difference between the capacity requirements and the 20 percent decreased value of the lower bounds is found at S_7 . Therefore, the feasible increase for M2 is also 1.1 ($\Delta c_2^+ = 1.1$).

Machine M4 contains a single operation type, ot_2 . In this case, the feasible decrease of the machine capacity is equivalent to the feasible increase of the production requirements of ot_2 , that is, $\Delta c_4^- = \Delta p_2^+ = 0.5$. Similarly, the feasible increase of the machine capacity is equivalent to the feasible decrease of the production requirements of ot_2 , that is, $\Delta c_4^+ = \Delta p_2^- = 0.3$. Machine M5 also contains a single operation type, ot_1 . The last row of Table 6 shows that the feasible capacity change of this machine is the same as the feasible change of the production requirements of ot_2 , except that the value of the feasible decrease must be changed by the value of the feasible increase.

6. Conclusions

FMSs are capable of performing a wide variety of technologically different operations. Therefore, any model trying to handle all of the operations has to cope with the problem of complexity. The aggregation of operations into *operation types* can help to address this problem. The paper provides an easy to use graphical and mathematical model to analyze the availability of

capacity. An operation type set assignment to machines is determined, which ensures that there will be no machines with excess idle time or too much overload of work. If machines are tooled in accordance with the result of the model, then workload balance can be achieved. Besides workload balance, other planning issues exist in an FMS. Several objective functions are also presented to determine alternatives for cheaper tooling, increased flexibility, and better pooling of machines.

The results of the model can be used as input to production programming. A production program has to be found, which provides the assignment of orders to the machines and includes the details of tooling. If the production program is consistent with the operation type set assignment, then workload balance will be achieved and the management priority expressed by the objective function will be ensured.

The number of decision variables is determined by the number of operation types. Therefore a careful aggregation strategy has to be chosen. If operations are aggregated into very few operation types, then the problem is easy to solve. However, then the model for disaggregation (i.e., for production programming) will be more complex. If the number of operation types is high, then the aggregate model is difficult to solve, but the subsequent production programming will be easier. This is, however, a general problem of all existing aggregate production planning models. The formulation of disaggregation models for detailed tooling and routing with various objective functions is a topic for further research.

References

- Hax, A.C. and Candea, C. (1984), *Production and Inventory Management*, Prentice-Hall, Englewood Cliffs, NJ.
- Holt, C.C., Modigliani, F., Muth, J.F., and Simon, H.A. (1960), *Planning Production, Inventories and Work Force*, Prentice-Hall, Englewood Cliffs, NJ.
- Jaikumar, R. and van Wassenhove, L.N. (1989), "A Production Planning Framework for Flexible Manufacturing Systems", *Journal of Manufacturing and Operations Management*, Vol. 2, pp. 52-79.
- Johnson, L.A. and Montgomery, D.C. (1974), *Operations Research in Production Planning, Scheduling and Inventory Control*, John Wiley and Sons, NY.

- Keeney, R.L. and Raiffa, H. (1976), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley and Sons, NY.
- Kleinrock, L. (1976), *Queueing Systems, Vol. 2: Computer Applications*, John Wiley and Sons, NY.
- Mazzola, J.B., Neebe, A.W. and Dunn, C.V.R. (1989), "Production Planning of a Flexible Manufacturing System in a Material Requirements Planning Environment", *The International Journal of Flexible Manufacturing Systems*, Vol.1, pp. 115-142.
- Mazzola, J.B., Neebe, A.W. and Dunn, C.V.R. (1998), "Multiproduct Production Planning in the Presence of Work-Force Learning", *European Journal of Operational Research*, (to appear).
- Niess, P.S. (1980), "Kapazität Abgleich bei Flexiblen Fertigungssystemen", *IPA Forschung und Praxis*, Springer Verlag, Berlin, Germany.
- Stecke, K.E. (1983), "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems", *Management Science*, Vol. 29, No. 3, pp. 273-288.
- Stecke, K.E. and Morin, T.L. (1985), "The Optimality of Balancing Workloads in Certain Types of Flexible Manufacturing Systems", *European Journal of Operational Research*, Vol. 20, pp. 68-82.
- Stecke, K.E. and Solberg, J.J. (1985), "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.
- Stecke, K.E. (1986), "A Hierarchical Approach to Solving Machine Grouping and Loading Problems of Flexible Manufacturing Systems", *European Journal of Operational Research*, Vol. 24, pp. 369-378.
- Stecke, K.E. and Kim, I. (1991), "A Flexible Approach to Part Type Selection in Flexible Flow Systems Using Part Mix Ratios", *International Journal of Production Research*, Vol. 29, No.1, pp. 53-75.
- Thomas, L.J. and McClain, J.O. (1993), "An Overview of Production Planning", In: *Logistics of Production and Inventory* (Editors, Graves, S.C., Rinnooy Kan, A.H.G., and Zipkin, P.H.), North-Holland, Amsterdam, The Netherlands.

- Van Looveren, A.J., Gelders, L.F., and van Wassenhove, L.N. (1986), "A Review of FMS Planning Models", In: *Modelling and Design of Flexible Manufacturing Systems*, (Editor, A. Kusiak), Elseviers Science Publishers B.V., Amsterdam, The Netherlands, pp. 3-31.
- Wakker, P.P. (1989), *Additive Representation of Preferences: A New Foundation of Decisison Analysis*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Captions and Table Headings

Figure 1. Allocation of operation types to the machines in the sample manufacturing system

Figure 2. Ideal available capacity ranges for the operation type sets

Figure 3. Capacity requirements for the operation type sets

Figure 4. Ideal available capacity ranges and capacity requirements for the operation type sets
($\alpha=\beta=0.2$)

Table 1. Notation

Table 2. Production requirements in CUs for one period

Table 3. Objective function coefficients

Table 4. The results of the operation type assignment

Table 5. Sensitivity analysis of the production requirements ($\alpha=\beta=0.2$)

Table 6: Sensitivity analysis of the machine capacity ($\alpha=\beta=0.2$)