

CHAPTER I

PRODUCTION PLANNING AND SCHEDULING OF A HYBRID FLOWSHOP

TABLE OF CONTENTS

1. INTRODUCTION	3
2. PRODUCTION DECISIONS	4
3. PRODUCTION PLANNING AND SCHEDULING.....	6
3.1 PLANNING AND SCHEDULING	6
3.2 GENERIC SOLUTION APPROACHES FOR PRODUCTION PLANNING.....	8
3.3 DRAWBACKS OF CLASSICAL SOLUTIONS	9
4. PLANNING AND SCHEDULING OF A HYBRID FLOWSHOP	11
4.1 HYBRID FLOWSHOP.....	11
4.2 A SOLUTION FROM A CARPET FACILITY	12
4.3 SOLUTIONS FROM A JOB SHOP ENVIRONMENT	14
4.4 A SOLUTION FROM A TEXTILE FACILITY	15
4.5 OUR APPROACH	16
5. AN INTEGRATED PLANNING AND SCHEDULING MODEL	19
5.1 MATHEMATICAL FORMULATION OF THE PLANNING PROBLEM	19
5.2 GLOBAL PLANNING HEURISTIC	21
5.3 GLOBAL SCHEDULING HEURISTIC	23
6. CONCLUSION	24
REFERENCES	25

TABLE OF FIGURES

Figure 1.1: Production planning and scheduling systems	8
Figure 1.2: A three-stage HFS	12
Figure 1.3: (a) A production plan over 4-weeks horizon at the stage level in a 3-stage HFS, Ideal production plan	22
Figure 1.3: (b) A production plan over 4-weeks horizon at the stage level in a 3-stage HFS, Infeasible production plan	22
Figure 1.4: Global Planning Heuristic	23
Figure 1.5: An overview of the decomposition-based heuristic for the HFS planning and scheduling.....	24

LIST OF ACRONYMS

FPS	First Planned Stage
GPH	Global Planning Heuristic
GSH	Global Scheduling Heuristic
HCPL	Heuristic for Checking the Planned Load
HFS	Hybrid Flow Shop
JIT	Just-In-Time
MPS	Master Production Schedule
MRP	Manufacturing Resources Planning
SSH	Single Stage Heuristic

CHAPTER I

PRODUCTION PLANNING AND SCHEDULING OF A HYBRID FLOWSHOP

Abstract

This chapter studies the production planning and scheduling systems of a hybrid flowshop (HFS) manufacturing facility. In general, two main approaches are used to deal with production management decisions. (i) The hierarchical approach, where the decision system can be regarded as made up of different levels of cooperative subsystems or submodels and (ii) the monolithic approach, where all decisions are handled in an all-encompassing model. Each of these approaches has its advantages and drawbacks. Monolithic models, for instance, are hard to solve when used in industrial context. The hierarchical models are very sensitive in the sense that their performance depends strongly upon the aggregate information used, like aggregate available capacity. Indeed, shop floor aggregate information can not be easily estimated and hence, these models are unlikely to produce accurate and feasible plans at the shop floor level.

The main objective here is to introduce a decomposition-based approach for solving the HFS planning and scheduling problem. The main issue in this approach is to overcome non-efficient capacity utilization while performing a good production plan with the lowest production and holding costs. This is realized by a two-level hierarchical and iterative procedure. The size of the planning problem is first reduced by considering the production and the capacity of the bottleneck stage only, as representing the capacity of the overall production facility. The iterative process of the procedure then fine-tunes the available capacity estimation at the planning level by checking the production plan feasibility at the scheduling level.

1. INTRODUCTION

This chapter studies the production planning and scheduling systems of a Hybrid Flow Shop (HFS) manufacturing facility. In general, two main approaches are used to deal with production planning and scheduling decisions. The hierarchical approach where the decision system can be regarded as made up of different levels of submodels, each of which has its main objectives and scopes. The monolithic approach uses a global model encompassing all planning and scheduling decisions. Each approach has its advantages and drawbacks. Monolithic models, for instance, used in industrial context, are hard to solve. The hierarchical models are very sensitive in the sense that their performances depend strongly on the aggregate information used, like aggregate available capacity.

The objective here is to introduce a decomposition-based heuristic for solving the HFS planning and scheduling problem as well as a new integration approach of the two subsystems so as to overcome classical solutions drawbacks as, for example, non-efficient capacity utilization. The remainder of this chapter is organized as follows:

Section 2 recalls some basic definitions of production decisions so that the role of planning and scheduling decisions can be identified within the overall company decision system framework. Section 3 tackles in more details some of the key aspects of production planning and scheduling. We first define what is meant by planning and scheduling and how the integration of these two systems is carried out in classical approaches. We then give an overview of generic planning solutions and emphasize some of their critical aspects, like the aggregate capacity estimation and the evaluation and use of lead-time. Through an example from the carpet industry, section 4 introduces the production planning and scheduling of a hybrid flowshop manufacturing facility. We present there the typical process of this industry, the day-to-day methods used for solving the planning and scheduling problems, as well as some of their drawbacks. As solution approaches for tackling the production planning and scheduling problem, we recall an integrated planning and scheduling approach used in a job shop environment proposed by Lasserre (1992). We also recall an integrated solution approach in a hybrid flowshop facility proposed by Elmaghraby and Karnoub (1997). Eventually we summarize our theoretical approach, which is compared to the two first ones. Section 5 states the production planning subproblem of the HFS and presents in details our theoretical decomposition-based approach.

The remainder of this thesis (Chapter 2 through Chapter 5) is dedicated to the study of the scheduling subproblem of the HFS, which will be carried out through the design of Branch and Bound algorithms.

2. PRODUCTION DECISIONS

PRODUCTION

Bitran and Tirupati (1993) define “*production as the process of converting raw materials into finished products. Manufacturing systems are typically composed of large numbers of components, which have to be managed effectively in order to deliver the final products in right quantities, on time and at an appropriate cost. In systems characterized by multiple products, several plants and warehouses, a wide variety of equipment and operations, production management encompasses a large number of decisions that affect several organization echelons*”.

Decision systems in a manufacturing company are often very large and complex to tackle and comprehend. Two different approaches are generally pursued to deal with production management. The first approach, by using aggregate models, attempts at reducing the size, or at decomposing the production decision system so as to deal with a set of small modules, instead of one global problem. The second approach uses a frontal attack, through an all-encompassing model or a monolithic model. We hereafter summarize the former approach and emphasize difficulties and drawbacks of the latter.

HIERARCHICAL DECISION SYSTEM

Aggregate models have largely been used to deal with production management problems. See examples of aggregate models in Hax (1978), Hax & Candea (1984), Gelders and Van Wassenhove (1981), and Leisten (1998). The production decision system can be regarded as a hierarchical model made up of different levels with different scopes and

objectives. Production decisions are being handled separately at each level, instead of being dealt with simultaneously. A strong integration of the different levels is required since a company is first and foremost concerned with a global optimization problem. See Hax and Meal (1975), Bitran and Hax (1977, 1982) and Bitran, Haas and Hax (1981, 1982) for examples of hierarchical models.

In order to better understand production decisions in a manufacturing company and the way these decisions can be viewed from an aggregate viewpoint, we simply recall the taxonomy proposed by Anthony (1965) so as to classify them. Three categories of decisions have been defined, i.e. strategic planning, tactical planning and operations control. Bitran and Tirupati (1993) have proposed the following definitions:

- **“Strategic planning** decisions are mostly concerned with the establishment of managerial policies and the development of resources to satisfy external requirements in a manner that is consistent with the organizational goals. In the area of production management these decisions relate to the design of production facilities and include the following: (i) location and sizing of new plants, (ii) acquisition of new equipment, (iii) selection of new products lines, and (iv) design of logistic systems”.
- **“Tactical planning** focus on the resource utilization process. At this stage, after decisions have been made regarding physical facilities, the basic problem to be solved is the allocation of resources such as capacity, workforce availability, storage and distribution resources. Typical decisions in this category include utilization of regular and overtime labor, allocation of capacity to product families, accumulation of seasonal inventories, definition of distribution channels, and selection of transportation activities. These decisions involve a medium-range planning horizon, and the aggregation of items into product families. Models addressing these issues are classified as aggregate planning models”.
- **“Operations control.** Decisions in this category deal with day-to-day operational and scheduling problems, which require complete disaggregation of the information generated at higher levels. Typical decisions at this level include the following: (i) production sequencing and lot sizing at the item level, (ii) assignment of customer orders to individual machines, (iii) inventory accounting and inventory control activities, (iv) dispatching, expediting and processing orders, and (v) vehicle scheduling”.

Although the aggregate approach might present these three classes of decisions as independent, their interdependence and interactions are very strong because each class of decisions feed one another with information and constraints. A good integrated approach is necessary and vital to the company so as to minimize suboptimization.

MONOLITHIC MODEL

An alternative to the aggregate approach is to regard production decision systems as a global system and use only one integrated and detailed decision model instead. See Bitran and Tirupati (1993) and Ouenniche and Boctor (1998) and the references cited therein for examples of monolithic approaches. Bitran and Tirupati (1993) summarize two major drawbacks of such approach:

“The development of integrated decision models that deal with all the decisions simultaneously, while attractive in principal, has several drawbacks. These models tend to be very large, and in most practical situations, it would be very difficult, if not impossible, to obtain optimal solutions with reasonable effort. Even if computational power and methodological capabilities would permit solution of a large detailed model, the approach is inappropriate because it would not be responsive to the management needs at each level of the organization, and would prevent interaction between models and managers at each organization echelon”.

The research carried out in this thesis is mainly concerned with the production planning and scheduling of hybrid flow shop facility (see Section 4). We also advocate the separation between planning and scheduling and use the aggregate approach with a close or tight interaction and integration of the two subsystems, i.e. planning and scheduling. The reason of this separation is to reduce the size of the global problem. Indeed, a monolithic model, which is used to carry out production planning and scheduling decisions simultaneously over the planning horizon and at the detailed item level, is very complex and remains very hard to solve in general. Moreover, this separation can be carried out in such a way that suboptimization is minimized, i.e. by carrying out a good integration procedure (see our approach in Section 5).

In the next section we define the concepts of production planning and scheduling and how these two systems can be integrated. We then present an overview of some generic methods and approaches used to solve production planning subproblems. The scheduling subproblem will be dealt with more systematically in subsequent chapters. We finally highlight why existing solutions for production planning and scheduling are not satisfactory to obtain a good integration level.

3. PRODUCTION PLANNING AND SCHEDULING

In the production management literature production planning and scheduling decisions are viewed as two levels of decisions, see for example Thomas and McClain (1993). While aiming at managing production so that the right items in the right quantities are produced and delivered on time, these decisions are mainly concerned with costs reduction and effective resources utilization. We recall, hereafter, the role of production planning and scheduling when considered separately.

3.1 PLANNING AND SCHEDULING

DEFINITIONS

Thomas and McClain (1993) give the following definitions of production planning and scheduling:

*“**Production planning** is the process of determining a tentative plan for how much production will occur in the next several time periods, during an interval of time called the planning horizon. Production planning also determines expected inventory levels, as well as the workforce and other resources necessary to implement the production plans. Production planning is done using an aggregate view of the production facility, the demand for products, even of time”.*

“Production scheduling is more detailed than production planning, coupling individual products with individual production resources using smaller time units or even continuous time. Despite its greater detail, even a production schedule typically cannot be implemented exactly. Machine breakdowns, worker absences, new rush orders, and other disruptions impose last minute changes. In real-time, implementation of a production plan or schedule is often called dispatching. These decisions are an important part of the operation of a productive facility. They must keep the production plan on track, and feedback of the actual current situation helps the next plan to be (approximately) achievable”.

PLANNING AND SCHEDULING INTEGRATION

The two above-mentioned definitions present the production planning and scheduling system as a two-level production decision system. Figure 1.1 depicts a hierarchical view of these two levels. The planning system uses aggregate production information, like holding and setup costs, capacity, lead-time, horizon and orders, and generates a production plan which meets aggregate demand due dates by performing an overall but aggregate production cost minimization. See Hax and Meal (1975), Bitran and Hax (1977, 1982) and Bitran, Haas and Hax (1981, 1982) for examples of integrated production planning and scheduling.

Once the upper level planning system has produced a medium-term production plan, it feeds the lower level scheduling system with information and constraints, like a set of lots or jobs to be produced within their fixed due dates. In order to build a short-term plan, the lower level scheduling system uses detailed production information, like processing times and setup times of individual jobs and the availability time of individual machines. While respecting the imposed higher level planning constraints, the scheduling system produces a detailed production plan for the shop floor, which is devoted to implementation and, therefore, tackles only the first period of the planning horizon. The information gathered on the shop-floor state during the implementation of the schedule, like inventory levels and capacity availability, is sent back to the scheduling and planning systems. The scheduling plan might be rebuilt using feedback information from the shop floor like machines breakdown, arrival of urgent orders and shortage of raw materials. The planning system is rerun regularly in a rolling horizon basis after being fed back by new incoming demands, information concerning the implementation of previous plans, estimation of available resource capacities and inventory levels.

The border between the planning and the scheduling levels is not well defined. For instance, some lot sizing methods, which are generally referred to as planning methods, can be considered belonging to the operation control level when neither aggregation, (i.e. considering the items instead of product families), nor scheduling are involved. The output of the lot sizing method is then directly implemented at the shop floor level.

In our approach proposed in Section 5, the planning method builds a production plan at the item level and the output of such planning procedure still needs scheduling. Although it does not use aggregated product families, our method can be regarded as belonging to the planning level because it uses aggregate costs and capacity, and its output is a production plan over T periods.

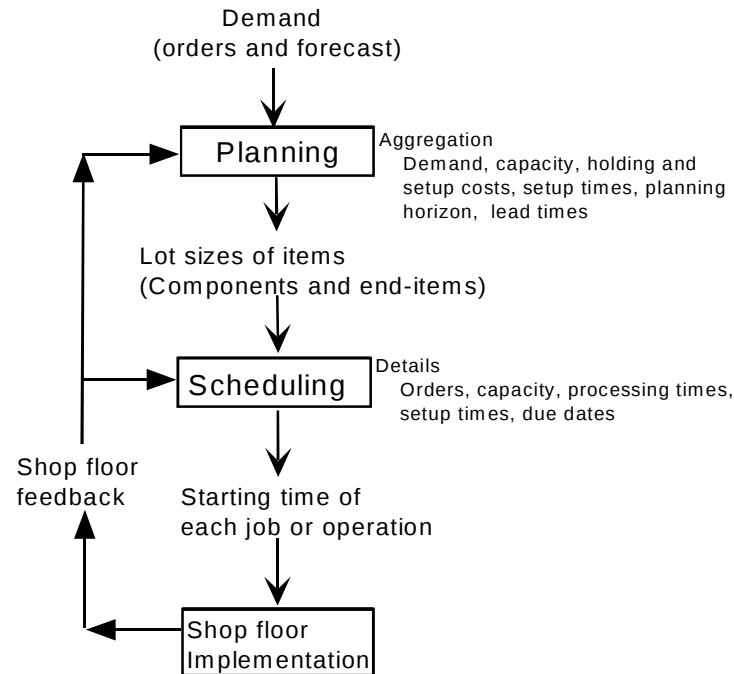


Figure 1.1: Production planning and scheduling systems

We hereafter present an overview of some generic methods and approaches used to solve production planning subproblems and highlight some of their drawbacks.

3.2 GENERIC SOLUTION APPROACHES FOR PRODUCTION PLANNING

Production process

Different approaches and solutions have been designed so as to deal with production planning. To a great extent, these solutions depend on the type of production process, the kind of manufactured items, nature of demand, company-customer relationships, as well as the organizational goals pursued by the company. In broad terms, a production system is a make-to-stock, make-to-order or an assemble-to-order system depending on the accuracy of demand and on the complexity of the bill of materials structure.

Two generic approaches and methods are generally used to deal with production planning: Push system (as Manufacturing Resources Planning, MRP), or Pull system (as Just-In-Time systems, JIT). Another approach uses lot sizing methods which are mainly based on mathematical models. The main objective of lot sizing is to build a production plan by fixing the size of item lots to be produced over a planning horizon, and where overall costs, i.e. production and holding costs, are minimized. For an extensive study on MRP, JIT and lot sizing techniques the reader can refer to Graves et al., (1993).

Push and Pull systems

Push systems, or Pull systems or a combination of both, are often used to handle most, or part, of the production planning system. This is done by planning in advance, or by settling a system that manages the release of orders to the shop floor, so that the items can meet their respective deadlines.

Production planning systems (like MRP or JIT) are essential to production facilities with complex item structures where the final product or the end-item is composed of sub-items. Each sub-item might also be in turn composed of sub-items, and the end-item can be seen as a complex structure of sub-items or as a multi-level hierarchy of sub-items. Companies manufacturing such elaborated or multi-level items need a systematic and elaborated system that helps managing inventories and the releases of production orders of end-items as well as sub-items at all levels of any product hierarchy structure. See Baker (1993) and Groenevelt (1993) and the references cited therein.

Lot sizing systems

Lot sizing procedures are used to group production items into item lots over a planning horizon. The main objective of lot sizing is to find a good equilibrium or tradeoff between the holding and production costs while respecting the deadlines set by the MRP system for dependent demand, and demand forecasts and/or customer orders for independent demand. Lot sizing can be used whenever a set of item orders, having deadlines or due dates, needs to be produced within a set of periods, i.e. the planning horizon. While respecting deadlines or reducing backlogging, the grouping of lots or items might save on holding costs and on production or setup costs.

Optimized lot sizing systems might be used to generate the Master Production Schedule MPS (MPS is the main input into the MRP system). They can also be used at different levels of the MRP system, i.e. at different levels of product structures in the so-called *requirements computation* used to build production plans for all components or items managed by the planning system. In less complex production systems, or for less elaborated product structures, this lot sizing procedure suffices as a planning system with demand forecast and/or customer orders as input.

3.3 DRAWBACKS OF CLASSICAL SOLUTIONS

Whatever the chosen production planning solution may be, the latter relies, in most cases, on aggregate information such as capacity, costs and lead times. The effectiveness of optimized planning system heavily relies on the accuracy of these parameters and how the lower level scheduling system deals with the production load and the time constraints imposed by the upper level planning. (See above Subsection 3.1).

Classical production planning methods use aggregate information like available capacity and lead-time as parameters. Actually, available capacity and lead-time are variables or decisions to be taken, and their values depend upon the planned load, the scheduling methods and the shop floor implementation of the schedule. The purpose of the remainder of this section is to emphasize drawbacks resulting from the use of capacity and lead times as parameters in production planning models, i.e. having fixed and aggregate values throughout the planning process.

IMPACT OF THE AGGREGATE CAPACITY

Despite a few attempts, all proposed production planning or lot sizing and scheduling solutions tackle production planning and scheduling processes separately, see Fontan and Imbert (1985), or Potts and Van Wassenhove (1992). This is certainly due to the fact that

the monolithic approach remains very hard to solve. Indeed, even the well-known economic lot scheduling problem is difficult to solve, see Elmaghraby (1978) or Zipkin (1991).

The production planning system sets the quantities of the items to be manufactured for each period of the planned horizon or due dates for individual orders of items, so that overall aggregate costs are as low as possible. These decisions are mainly based upon aggregate information on the available resources and products, i.e. capacity, items, lead times, holding and setup costs, etc. The scheduling system manages the assignment of the resources to the lots or jobs in such a way that their effective due dates established by the planning system are met.

One can observe that the aggregate available capacity considered at the planning level could either be overestimated or underestimated. Capacity overestimation results in an infeasible plan or unpredictable delays at the scheduling level while capacity underestimated results in waste of capacity, leading to constitution of larger inventories or to backlogging. These two possibilities (i.e. infeasible plan at the scheduling level and capacity waste) may occur although the higher-level production plan is feasible and optimal. They are mainly due to the degree of accuracy of the aggregate information used as capacity, lead times and costs. In order to overcome these difficulties, fine tuning of the aggregate capacity is carried out as follows:

- Overestimation of available capacity is easily detected afterwards, because of the unfeasibility of the proposed plan at the scheduling level. At the planning level, reduction of the estimated aggregate capacity might lead to a more realistic estimation of the available capacity. The danger of this strategy is to fall down into underestimation of the capacity case, which will lead to the capacity waste, since the higher level planning is suggesting a production load inferior to what the plant could deliver.
- Underestimation of capacity might not be easy to discover, and if suspected an augmentation in the estimated aggregate capacity might lead to a better estimation of available capacity. The drawback may be to fall in the overestimation case.

Therefore, fine-tuning the estimated aggregate capacity at the planning level so that it reflects the accurate available capacity at the shop floor level is far from being an easy task. Since backlogging costs can be very high, the easiest strategy leads to the constitution of a safety stock of end-items as well as an increase in work in process inventories. This high inventory level leads to a better response to customer demand, but generates higher costs.

A more realistic estimation of the load or capacity usage (to avoid waste or overloading) might only be known when a detailed schedule of the set of the lots or jobs at the shop floor level is established. As a matter of fact, the available capacity is not a fixed parameter, but depends strongly upon the way the available resources are utilized at the shop floor level. A good estimation of capacity utilization by a planned load at the floor shop level might then be determined by the scheduling of this load. Therefore, a good integration and interaction of the two levels (planning and scheduling) are required for an effective capacity utilization. This integration is better carried out when the scheduling

system feeds back the planning system with accurate information about the capacity utilization at its level.

LEAD TIME DILEMMA

The manufacturing lead-time of a work order (or an order lead-time), regarding a lot or a job, in a production plant, is the total amount of time needed to complete the order. This starts from the moment when production is authorized (i.e. the instant an order is released to the shop floor) until the moment when complete processing of the work order ends, and the corresponding material is ready for shipping or available for the next production stage. The components of an order lead-time in a production facility are queue times, processing times and moving or transportation times. An order lead-time is used, for several purposes, as an important information in production management. Lead-time might be used as a parameter in a decision system, as a decision variable and/or as a criterion for measuring the performance of a production facility and to be optimized. Karmarkar (1993) highlights the lead-time dilemma:

“Order release is generally taken to mean the level of control between planning and scheduling or execution, ...”. The lead-time of an order is used at the planning level to compute the corresponding order release time to the shop floor, so that this order is produced on time. “Traditional release models of both push and pull vary, and take lead times to be given and independent of order release parameters. In fact, order releases load production facilities, and the nature of this loading process relative to available capacity is the primary determinant of lead times in the facility. The fallacy here is in thinking of lead-time as an attribute of a part. Rather it is a property of a facility or shop that depends on total load and capacity, and thus varies over time”.

Lead-time and capacity utilization are then strongly correlated. Because of the queue time, essentially, lead-time is then related to the overall shop loading and, hence, to the available capacity - since a planned load is usually fixed upon the estimated available capacity. The lead-time also depends upon the effectiveness of the methods used to optimize production planning and scheduling. As for the capacity evaluation, lead-times estimation might affect the performance of production management. For the study of lead-time concept and related research, see Karmarkar (1993).

As an attempt to minimize these drawback effects, Section 5 presents an approach to integrate planning and scheduling decisions in a hybrid flowshop (HFS) production facility. The latter is first presented in the next section where we exemplify the HFS production planning and scheduling problem with a production process from a carpet facility.

4. PLANNING AND SCHEDULING OF A HYBRID FLOWSHOP

4.1 HYBRID FLOWSHOP

In this project, we are mainly concerned with process industries like cosmetics, pharmaceuticals, textile and food industries. In these industries, production facilities are

organized as multistage production flowshop facilities where a production stage may be made up of parallel production lines, machines or any other production facilities. Such production processes are called multiprocessor flowshop or Hybrid Flow Shop (HFS) processes (Figure 1.2). At some stages, the facilities (machines, lines, etc) are duplicated in parallel to increase the overall capacity of the shop floor, or to balance the capacities of the stages, or to eliminate or reduce the impact of bottleneck stages on the overall shop floor capacity.

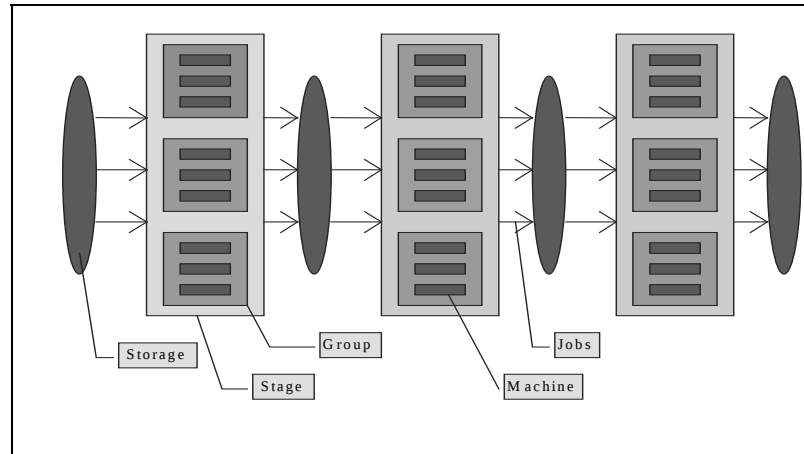


Figure 1.2: A three-stage HFS

In order to produce an item, one operation at each stage needs to be performed. Those operations have to be realized sequentially from the first to the last stage. A machine might process, at most, one job at a time or a set of jobs simultaneously. Some jobs might not need processing at some stages and simply skip that stage and go directly to the subsequent one. Work in process inventories or intermediate storage might be constituted between the different stages. For more background on the context of the hybrid flowshop problem see Pinedo (1995), sections 12.2-12.4, the references cited therein, and the specific references cited in each chapter of this thesis.

The generalized hybrid flowshop structure is found in many different industries. We hereafter give a brief description of the production process of a Carpet facility located in Belgium.

4.2 A SOLUTION FROM A CARPET FACILITY

THE PROCESS

The company produces thousands of (carpets) items of different types along with different sizes, designs, colors and qualities. Production management of this carpet facility is driven by customer orders as well as demand forecast. Demand forecast is used to constitute production stocks in anticipation of future and unknown customer needs.

The production process of this carpet industry can be regarded as being made up of three main production stages: the tufting or weaving stage, the coloring stage, as well as the finishing and cutting stage.

- **Tufting or weaving stage.** The main operation of this stage, i.e. tufting or weaving, consists in the making of the carpet, either by tufting threads or yarns on a synthetic canvas or by weaving the carpet using threads only. Carpet items have different sizes, designs and qualities. Qualities differ, regarding the type of threads used and the density of the carpet, i.e. the number of threads used by line. The produced carpets are 3 to 5 meters wide and up to 25 meters long. The operation at this stage needs a major changeover, i.e. time needed to prepare the machine for processing the next item. However, a large amount of this changeover can be masked, i.e. made while the machine is processing the previous item.
- **Coloring stage.** The coloring operation is applicable only to carpets made by non-colored threads, which leads some items not to undergo this stage, or equivalently having zero processing time at this stage. The coloring operation produces the finished product. This operation differentiates the final product, in a way that a carpet type corresponds to many different finished products and thus can be produced in large quantities at the first stage. The specific color is put only when a specific customer need is known or when the inventory level of a specific item with the specific color is below its stock safety level. The setup time operation needed for changing the color, though drastically reduced, is still critical if the sequencing is not handled effectively, e.g. by sequencing two incompatible colors successively (i.e. black followed by white) which leads to a large setup time for cleaning the machine.
- **Finishing and cutting stage.** The last operation puts either foam rubber or glue to the back of the carpet to firmly maintain the tufted or woven threads. It is also directly followed by the cutting operation to fit customer needs or standard cuts. These two operations (finishing and cutting) are non-dissociated operations because they are successive operations processed by one inline-machine, applying the foam rubber or glue, drying and immediately cutting the carpet. At this stage, the sequencing is also critical since there are different types of foam rubber mixtures with different costs. And mixing the right quantity for the right lot of carpets is very important so as to reduce setup costs and times.

THE EXISTING PLANNING AND SCHEDULING SYSTEMS

The Carpet Company guarantees the delivery of a set of specific items upon customer orders arrival and, therefore, needs to constitute a safety stock for these items. Based upon previous sales, a forecast is carried out so as to determine future requirements for these items, i.e. make-to-stock. For the remaining items, delivery dates or due-dates are negotiated when a customer places an order. Their manufacturing is then considered in future production plans, i.e. make-to-order. Demands of finished products, i.e. forecasted demands and customer orders, along with their due dates, i.e. periods by the end of which they should be ready for delivery, constitutes what the company calls the Master Production Schedule (MPS). Stock-safety-levels as well as forecasted demand and customer orders trigger overall production.

The planning manager uses a manual and a simple lot sizing procedure so as to aggregate item demands into product families and to generate the production plan. He proceeds as follows:

- Based upon forecasts, known customer orders and stock levels, the planning manager builds a production plan over a four week planning horizon. The production plan mainly concerns the first stage, i.e. the tufting, because of the large number of machines at this stage and the external complicating factors at the upstream of this stage, e.g. the grouping of raw materials orders. The production plan consists in loading each specific machine or pool of machines composing the stage over the four weeks horizon. The sequencing is carried out at the beginning of each week. No analytical method, but only his experience, is used for solving the planning and scheduling problems. Statistics show a large stock amount of tufted and woven carpets constituted below of this stage.
- The loading of the second and third stages are driven by due dates which are set by forecast and customer orders, i.e. MPS. The planning manager makes the coordination of these two stages at an aggregate level. The loading, at these stages, is carried out for the first subsequent period only. Each stage receives the corresponding work orders that need processing within the next period. Sub-items (i.e. tufted or woven -but not colored- and colored carpets) required for processing at the coloring or finishing stage, need to be available at the beginning of the corresponding stage and week. Two different persons carry out the scheduling of the two last stages, respectively.

The planning and scheduling systems implemented in this Carpet Company produce large amounts of work in process inventories, which can explain the large average lead-time. The average lead-time is about three weeks, while the total processing time plus setup, handling and moving times do not cover more than two to three days, on average. Although an aggregate and non-optimal coordination does exist at the first stage, these two drawbacks are mainly due to the lack of a global aggregate coordination. They might also be due to the lack of coordination between stages at the scheduling level, which results in local suboptimization.

In the remainder of this section and as solution examples for integrated production planning and scheduling systems, we briefly recall an integrated planning and scheduling approach in a job shop environment proposed by Lasserre (1992). We also present a brief description of a loading and scheduling model in an HFS facility proposed by Elmaghraby and Karnoub (1997). Eventually, we summarize our theoretical approach whose detailed description is given in Section 5.

4.3 SOLUTIONS FROM A JOB SHOP ENVIRONMENT

Lasserre (1992) has proposed an integrated planning and scheduling approach in a job shop environment. He uses a two-level model. The upper level (a lot sizing model), including capacity constraints driven from a given sequence (y) of jobs on the machines, computes a feasible plan $P(y)$. Actually, in the proposed mathematical model, the capacity constraints are formulated as a job sequence, whose completion time should not exceed the due dates of the jobs composing the sequence. With a given and fixed sequence y , the lot sizing module is able to compute the required capacity and hence deduces a feasible production plan $P(y)$ according to the sequence (y). The scheduling level, given fixed lot sizes $P(y)$, determines possibly a new sequence (y'). The procedure alternates then between solving

- a lot sizing problem using a fixed sequence (y) which yields a plan $P(y)$,
- a scheduling problem using fixed lot sizes from the plan $P(y)$ which yields a new sequence (y').

The procedure starts with the second step, i.e. scheduling, and uses demands as initial lot sizes. The procedure can always stop with a feasible plan with a final sequence (y'). By iterating, it aims at finding the best feasible plan, i.e. one with the lowest possible cost.

Dauzère-Péres and Lasserre (1994) have carried out numerical tests when the lot-sizing model is reduced to the case where items have neither setup times nor setup costs. The lot-sizing model is then a pure linear programming model. The scheduling has been solved through a set of heuristics, i.e. priority rule-based dispatching heuristics and a simplified version of the shifting bottleneck procedure (see Adams et al., (1988) or later on in Chapter III).

Lasserre claims that the ideal procedure (where the problem at each level is solved optimally) converges in a finite number of iterations to a local optimal solution of the whole problem. In Lasserre (1992) and Dauzère-Péres and Lasserre (1994) details of this procedure are explained.

Lambrecht and Vanderveken (1979) have also proposed an integrated lot sizing and sequencing approach in a job shop environment. Their procedure is a greedy lot sizing heuristic, which sets the lot sizes period by period. No *time* backtracking is used. The heuristic moves forward in time to fix the lot sizes Q_{it} in period t for all products only if the lot sizes Q_{it-1} in period $t-1$ are proven feasible according to the *sequencing system*.

Each machine k for each period t is assigned an estimated capacity B_{kt} expressed in time units. It is also assigned a time capacity L_{kt} within which products need to be completed. B_{kt} is equal to L_{kt} at the start of the lot sizing heuristic. At each period t , as soon as all Q_{it} are known, the lots of the period are sequenced. Q_{it} is considered infeasible if the completion time of machine k , on which products in lots of size Q_{it} are sequenced, is larger than L_{kt} . If Q_{it} is infeasible, B_{kt} is adjusted to reflect a better estimate of the available capacity and the lot sizing heuristic re-computes new values for all Q_{it} in period t .

4.4 A SOLUTION FROM A TEXTILE FACILITY

Elmaghraby and Karnoub (1997) modeled the planning and control functions in a Textile plant as a planning and scheduling problem in a hybrid flowshop. The Textile Company has a three-stage hybrid flowshop. Each stage is made up of a set of different groups of parallel machines. The machines are identical within a group of parallel machines and different from group to group. At each stage, each job first needs assignment to a specific group of machines and afterwards to a specific machine within that group.

Elmaghraby and Karnoub (1995) solved the main problem by dividing the global problem into two parts: a 'capacity loading function' and a 'scheduling function', both undertaken over a finite horizon of six weeks. The loading was optimized via Linear Programming (LP). The scheduling was implemented via a number of heuristics.

The LP model (a pure LP) maximizes the contribution to profit and minimizes holding and backorders costs. It generates the production plan, i.e. fixes lot size of each item, over the six weeks horizon and handles the loading of stages according to their average estimated capacity. The global loading system balances the capacity utilization (or the production) at all stages equally at each period, i.e. the used capacity is the same at all stages. This balancing operation helps to reduce the work in process inventories.

The LP model handles the capacity of the shop floor by taking the aggregate capacity of each group of parallel machines into account. The aggregate capacity of each stage is then based upon the aggregate capacity of the groups of parallel machines composing the stage. These group aggregate capacities are explicitly specified in the LP formulation.

Furthermore, the LP fixes at each stage, the group of parallel machines on which each lot has to be produced. Therefore, the lower level scheduling system only needs to schedule the jobs (lots) on each group of parallel machines by assigning each job to a specific machine and sequencing jobs on each machine. The scheduling module uses simple heuristics. It schedules the shop one stage at a time and begins by the scheduling of the last stage and goes backward until the first stage has been reached. For a complete description of this approach the reader can refer to Elmaghraby and Karnoub (1997).

4.5 OUR APPROACH

We propose a different approach to deal with the production planning and scheduling in an HFS facility. The HFS considered is composed of several stages where each stage is made up of identical parallel machines.

This problem is inspired from the carpet industrial case mentioned above. The major drawbacks of the planning and scheduling system used in this facility are the lack of a global planning system as well as the lack of tight coordination between the planning and scheduling subsystems and between the scheduling of the different stages.

We do not pretend here to solve the *whole planning and scheduling problem* for this industrial case, which remains a huge problem to deal with. However, we tackle *two main aspects* of it.

(1) First, we present in the remainder of this chapter a global planning and scheduling approach. Its main issue is to tackle the whole problem in an integrated and a global way while allowing an effective use of the available capacity. One important advantage of this approach is a close interaction between the planning and scheduling subsystems. No numerical implementation will be carried out for this global approach, since we decided to tackle only the scheduling part of the problem.

(2) The second aspect is the study of efficient scheduling algorithms, which will be carried out through Chapters 2, 3, 4 and 5. We will study the case where at each stage the machines are identical and jobs have release dates and tails, but without setup times. The choice of this research orientation comes from the fact that a review of the literature (see later Chapter 2) has shown that even this case is not well studied. And, because our approach relies heavily on the scheduling subsystem and its efficiency (see Section 5), we decided to carry out an in depth study of this part of the global system.

Note however, that our global approach remains essential to solve the original planning and scheduling problem. In order to do so, what still needs to be accomplished after the contribution of this thesis is:

- (i) to solve the proposed planning problem (see later section 5).
- (ii) to adapt the scheduling methods studied in this thesis to the case where the machines are not identical and jobs have setup times, and

Hence, this thesis has been inspired by this industrial case, and its contribution has to be considered as a preliminary and theoretical step towards the design of an integrated planning and scheduling system for that facility. Next, we present our global approach.

We also advocate the planning and scheduling separation but with a closer or tighter interaction and integration between the two levels. The monolithic approach remains very hard to solve in general. Since one main critical issue is the accurate evaluation of capacities, as highlighted in Section 3, we therefore propose a planning model where the aggregate capacity is evaluated by scheduling. We present here the general principles of our approach. Its detailed description is given in Section 5.

The proposed procedure proceeds iteratively by first producing a higher-level production plan over the planning horizon. It uses the capacity of one stage (defined as the bottleneck stage) only to represent the overall shop capacity. A lot sizing procedure builds, over the planning horizon, a production plan using the aggregate capacity of the bottleneck stage, as well as aggregate production and holding costs. A special procedure carries out then the checking of the feasibility of the production plan, over the planning horizon, by using scheduling algorithms. If the plan is not feasible at the scheduling level, capacity is corrected and reconsidered at the planning level with respect to the observed available capacity. This procedure is repeated iteratively until the higher-level production plan is considered feasible. The scheduling of the first period is then effected and the corresponding plan is then implemented. Scheduling is carried out via a set of heuristics.

Our approach evaluates the capacity of the overall shop floor by using the capacity of one stage only. The idea behind this principal is that, in our approach, the global production load is balanced over all stages so that the production at each stage is consumed by the subsequent stage within one period. This balanced production load over all stages leads to the work in process inventories reduction, and to lead time reduction. The stage representing the capacity of the shop floor is the bottleneck stage. If it is known, its estimated capacity is used, otherwise we use the capacity of the final stage. At the beginning of the procedure, the capacity of the bottleneck stage is measured by summing up the availability time of the composing machines. Scheduling algorithms are used to check the load balancing or the feasibility of the planned load.

The idea of adjusting the estimate of the machine capacities, at the lot sizing level, using sequencing rules has been used in a job shop environment by Lambrecht and Vanderveken (1979), see Subsection 4.3.

The main difference between our approach and their approach is that our planning procedure first fixes the loads for all periods and then checks the feasibility for each period separately. If at some period t the corresponding load is proven infeasible, the corresponding estimated capacity is adjusted and the whole planning procedure is carried out for all periods from the very beginning. The approach used by Lambrecht and Vanderveken (1979) reconsiders only the infeasible lot sizes and once a lot size at some period is proven feasible it is never modified.

The main differences between our approach and the one proposed by Lasserre (1992) are as follows: the lot-sizing model in our approach does not consider sequencing at the planning level, whereas his approach explicitly imposes a fixed sequence of jobs so as to compute the capacity consumption. His approach makes the lot-sizing model more constrained with a sequence (y) , which may not be a good sequence for the currently being computed planned load (the currently fixed lots). Actually, the lot-sizing model might produce a better objective function value if it is less constrained. Note also, that by imposing a job sequence in the formulation the lot-sizing model is much harder to solve. In our approach at each iteration the capacity, represented by an estimated parameter, is fine-tuned according to the observed possibilities at the scheduling level.

The main differences between our method and the one proposed by Elmaghraby and Karnoub (1997) are as follows: our lot-sizing model loads the shop as if it was made up of one stage only. Whereas in their model the aggregate capacity of each group of machines is explicitly specified in the LP formulation. Also, our global procedure (planning and scheduling) is an iterative one, in the sense that the planning and the schedule procedures inter-react and are rerun several times until the load, fixed by the planning level, is considered feasible by the scheduling level.

While in classical methods available capacity, at the planning level, is considered as a fixed parameter, in our approach, its estimation is fine-tuned by scheduling and by the proposed iterative process.

Note also that our approach can still find the global feasible optimum of the global planning and scheduling problem if at some iteration the lot-sizing model is optimally solved with the best fine-tuned estimated capacity (representing the accurate available capacity) and the feasible schedule of the checked load is optimal. Indeed, an optimal feasible schedule of a checked load implies that capacity is wisely used, i.e. capacity waste is minimized. Unfortunately, polynomial time exact methods are unlikely to exist either for the lot-sizing problem or for the scheduling problem and only heuristics can be used.

Our solution is then presented as a decomposition-based heuristic using two main levels. In the next section we give a detailed description of this theoretical approach. It is referred to as theoretical since no global empirical implementation has been carried out so far. Actually, this is out of the scope of this project since we have concentrated our study on an in-depth analysis of the scheduling part of the system, only.

5. AN INTEGRATED PLANNING AND SCHEDULING MODEL

The objectives in our specific planning approach of the HFS are:

1. reduce the inventory levels between stages
2. reduce the inventory levels between periods of the planning horizon
3. reduce setup times (that causes loss in capacities)
4. reduce the production or setup costs
5. use the available capacity effectively

The first objective is met by considering the capacity of the bottleneck stage as representing the overall HFS capacity, and by finishing the production of any item within one period. The second, third and fourth objectives will be explicitly considered in the planning procedure. We attempt at achieving the last objective by our proposed interaction between the planning and scheduling as well as by the effectiveness of the used scheduling procedure.

Before introducing our decomposition based heuristic, we state the planning problem by giving its mathematical formulation. Since the capacity of the overall shop is estimated by the capacity of one stage, we are considering only the production of one stage in this formulation.

5.1 MATHEMATICAL FORMULATION OF THE PLANNING PROBLEM

Problem:

We need to satisfy a set of demands $\{d(i, t), i=1, \dots, n, t=1, \dots, T\}$, where n is the number of items and T is the planning horizon. $d(i, t)$ could be any forecasted demand and/or customer order of item i , that need to be produced and be ready for delivery by the end of period t . Period t is then considered as the due date of demand $d(i, t)$. At each period t , the considered stage, representing the HFS capacity, has an estimated aggregate capacity of $c(t)$. The latter is expressed as an available amount of time units, e.g. for one week and five parallel machines, $c(t)=24 \text{ hours} * 7 \text{ days} * 5 \text{ machines} = 840 \text{ hours}$, which is the maximum time available for processing the items on the 5 machines within the whole week.

Each item i has:

- processing time $p(i)$ expressed in time units, e.g. 8 hours.
- inventory holding cost $h(i, t)$ per unit of inventory at the end of period t
- setup time $st(i)$ and an aggregate setup cost $sc(i, t)$ if item i is produced in period t .

$$\text{Let } D = \sum_{i,t} d(i, t)$$

$Z(i,t)$ be the quantity of product i to produce in period t

$Y(i,t)$ be a binary variable set to one if product i is produced in period t and zero otherwise.

Objective:

Our task lies in building a production plan covering the T -period horizon, so that the overall setup and holding costs are as low as possible and so that the proposed load at each period t does not exceed the corresponding estimated capacity $c(t)$. In

other words, the load $Z(i, t)$, for all i , at each period t , could be produced within one period.

Objective function:

$$\text{Minimize } \sum_{t=1}^T \sum_{i=1}^n h(i, t) \left(\sum_{\tau \leq t} Z(i, \tau) - \sum_{\tau \leq t} d(i, \tau) \right) + \sum_{t=1}^T \sum_{i=1}^n (sc(i, t) * Y(i, t)) \quad (1.0)$$

Subject to:

$$\sum_{\tau \leq t} Z(i, \tau) - \sum_{\tau \leq t} d(i, \tau) \begin{cases} \geq 0 & \text{if } t < T \\ = 0 & \text{if } t = T \end{cases} ; \quad \text{for all } i, t \quad (1.1)$$

$$\sum_{i=1}^n (p(i) * Z(i, t) + st(i) * Y(i, t)) \leq c(t); \quad \text{for all } t \quad (1.2)$$

$$D * Y(i, t) - Z(i, t) \geq 0; \quad \text{for all } i, t \quad (1.3)$$

$$Y(i, t) \in \{0, 1\}, Z(i, t) \geq 0; \quad \text{for all } i, t \quad (1.4)$$

- The objective function (1.0) minimizes overall holding and setup costs
- Constraint set (1.1) ensures the satisfaction of the demand for all periods, by imposing that cumulative production exceeds cumulative demand in all periods. Backlogging is not permitted, but can easily be integrated in this formulation.
- Constraint set (1.2) limits the production time (including processing and setup times) in each period t to the available capacity $c(t)$.
- Constraint set (1.3) states that a product cannot consume capacity at some period unless it is produced during this period.

This formulation assumes that the production of an item (processed at all HFS stages) has to be finished within one period. Doing so reduces the work in process inventories, which is an important issue in the design of production plans for HFS production facilities.

The planning problem, as formulated above is a classical lot-sizing problem. In fact, it is a capacitated, multi-item lot-sizing problem with setup costs and setup times which is an NP-hard problem. Lot-sizing problems have been largely studied in the production management and operations research literature, from the simplest problems that are solved with the Wilson formula to the most complicated ones like the multi-items, multi-level, lot-sizing problems which is an NP-hard problem, Florian et al., (1980), Bitran and Matsuo (1986). Moreover, Maes et al., (1991) have shown that even the feasibility problem, when lot-sizing problems consider setup times explicitly is NP-complete. For a survey and the study of some lot-sizing problems, the reader can refer to Graves et al., (1993) and the references cited therein.

In order to solve the production planning and scheduling of the HFS, we propose a Global Procedure (GP) made up of two global heuristics. The first one, a Global Planning Heuristic (GPH), carries out the planning part of the system over a T -period horizon, by providing a heuristic solution to formulation (1.0)-(1.4). The second one, a Global Scheduling Heuristic (GSH), is used for scheduling the whole HFS production facility. Moreover, GPH uses a *scheduling module* for checking the feasibility of the planned load over the T -period, see later the “*feasibility checking procedure*”. This module potentially feeds back the planning system with accurate information on the available aggregate or average capacity. Once the plan has been fixed, the GPH is used to build a scheduling

plan for the first period. The global procedure is now presented in detail in the following subsection.

5.2 GLOBAL PLANNING HEURISTIC

The Global Planning Heuristic (GPH) proceeds as follows: it carries out the planning step by using the capacity of the bottleneck stage if an obvious one exists or the last stage otherwise. The capacity of the bottleneck stage is used at the planning level because it better represents the aggregate capacity of the overall shop floor. This stage is called the First Planned Stage (FPS). To plan FPS one can use a simple “Single Stage Heuristic” (SSH) which is a heuristic for solving a capacitated, single-level, multi-item, lot-sizing problem when the shop floor is made up of one stage only.

As a heuristic for this problem one can adapt, for example, the “ABC” heuristic proposed by Maes and Van Wassenhove (1986) for solving the multi-item single-level capacitated lot-sizing problem, the heuristics presented and tested in Kirca and Kokten (1994) or in Shaw and Wagelmans (1998).

Aggregate information. SSH needs aggregate information about costs and capacities. Aggregate costs can be computed by summing up costs over all stages. Indeed, if item i is needed by time t and produced by period $(t-3)$, the setup cost $sc(i, t-3)$ of that item is the sum of its setup costs over all stages at period $t-3$. Since we want to finish the overall production of this item within one period, the setup costs occur at all stages within one period, i.e. period $(t-3)$. In the same way, its aggregate holding cost (the holding cost of the end-item) occurs for three periods since its production is completed by the end of period $(t-3)$.

Feasibility checking procedure. As soon as lot sizes of the items (i.e. $Z(i, t)$, for all i, t) are known, a ‘Heuristic for Checking the Planned Load’ (HCPL) is carried out to verify whether these lots can be produced with respect to the available capacity at the scheduling level. The checking procedure is carried out separately for each period t of the planning horizon. For each period, and with the corresponding planned load (set of lots or jobs), this procedure schedules all stages composing the HFS, as if we were at the scheduling level. Since we have a set of lots or jobs to be produced within each period, scheduling this set of jobs over all stages, within one period, allows us to check whether the HFS available capacity is large enough to produce what has been planned for this period.

If at some period t , the schedule length at some stage, i.e. difference between the completion time of the last job and the starting time of the first job at some stage, is larger than one period we conclude that the production plan at period t is infeasible. Figure 1.3(a) gives an overview of an ideal production plan over 4-week horizon in a 3-stage HFS. In Figure 1.3(a) we can observe that at each period the load is balanced over the 3 stages and that the length of the schedule of each stage does not exceed the *length* of one period. Figure 1.3(b) gives the case where the length of the schedule of stage 1 in period 2 exceeds one period, and the consequences that might occur in subsequent stages and periods. The plan in such a case is considered infeasible in period 2.

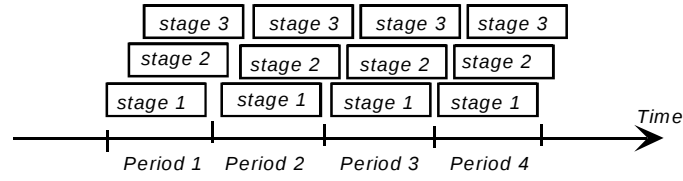


Figure 1.3: (a) A production plan over 4-weeks horizon at the stage level in a 3-stage HFS
Ideal production plan

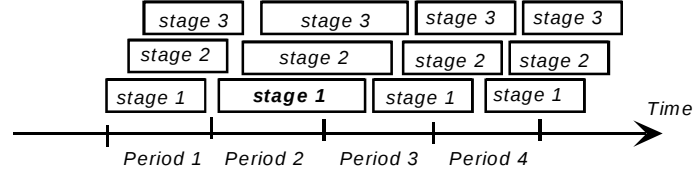


Figure 1.3: (b) A production plan over 4-weeks horizon at the stage level in a 3-stage HFS
Infeasible production plan

The checking process then continues until the plan has proved feasible by scheduling or infeasibility is encountered, i.e. demands of a set of items can not be met unless the capacity is violated.

Recovering feasibility. During the checking process and at some periods, the capacity of the shop floor may not be large enough to complete the planned items within one period. In this case two main solutions can be used: recover feasibility or reconsider the production plan, i.e. reconsider the $Z(i, t)$.

- A first solution would consist of recovering feasibility by postponing in time the production of items with low setup costs, *should capacity in subsequent periods permit*. If at some period t the corresponding load ($Z(i, t)$ for all i) could not be totally produced within one period, and if at some period $r > t$, the capacity is not totally used, one can then postpone in time the production of some lot of item j ($Z(j, t) > 0$) from period t to period r . The selection of item j is an important and non trivial decision. One major condition has to be satisfied. Due dates of the items composing lot of item j should be larger or equal to r . Note that postponing in time the production of a set of items, from period t to period $r > t$, reduces the overall cost if the induced setup cost at period r is smaller than their holding cost from period t to period r . Note also that production capacities in period t and r are influenced by the choice of the postponed item j . See Maes and Van Wassenhove (1986) for an example of a recovering feasibility procedure.
- If, after carrying out the recovering feasibility routine, the planned load is still infeasible at some periods, capacity of the FPS at these periods is then decreased. The capacity correction is made according to the difference between the planned capacity $c(t)$ and capacity utilization in the detailed schedule, i.e. $c(t)_{k+1} = c(t)_k - \epsilon_k$, where k is the iteration at which the plan is infeasible and ϵ_k is the capacity correction based on $c(t)_k$ and on the observable capacity violation at the scheduling level in period t . The whole planning procedure is then carried out from the very beginning using $c(t)_{k+1}$. One should pay careful attention to the computation of ϵ_k . The smaller are the ϵ_k 's the more accurate is the estimation of the available capacity, but the more time consuming is the overall procedure. The larger are the ϵ_k 's the larger is the probability of falling into the underestimation of the available capacity case, see Subsection 3.3.

The planning and checking procedures are then repeated until the planned load of FPS, at each period, has come to fit the available capacity, i.e. permits us to find a feasible schedule at all stages in each period. At this step we are sure that the production plan proposed by the planning model over the T-period horizon is feasible, and that overall setup costs as well as inventory costs are as low as possible. Indeed, depending on the efficiency of the *planning heuristic* used, the setup and holding costs are as low as possible and the *feasibility checking procedure* ensures us that the fixed loads are feasible at the scheduling level. Figure 1.4 summarizes the Global Planning Heuristic.

The planning system operates in a rolling horizon manner and progresses period by period. At each period, such information as inventory levels and capacity modification are gathered from the shop floor and used as input to the planning system along with new demands, i.e. new demand forecasts and customers orders.

1. Run the SSH for FPS (either the bottleneck or the last stage)
2. Run HCPL for checking planning feasibility
 - Check feasibility in all periods
 - If the plan is infeasible try to recover feasibility
 - Find infeasible periods
 - Update the aggregate capacities of infeasible periods
 - Go to 1
 - Else stop

Figure 1.4: Global Planning Heuristic

5.3 GLOBAL SCHEDULING HEURISTIC

The Global Scheduling Heuristic (GSH) only tackles the first period of the planning horizon. It produces a schedule for the lots or jobs determined by the upper level planning system, i.e. at each stage assigns jobs to machines and sequences the assigned jobs on each machine. At this level the procedure can be handled by an effective and time consuming method (a heuristic or an exact method), because it is only carried for one period and might be run once and for all, or a few times if rescheduling is necessary.

In order not to deteriorate the objectives reached at the planning level, we need to produce (schedule) the planned load for the first period, i.e. the lots $Z(i, 1)$ for all items, within one period, i.e. completing the set of jobs at each stage within one period. Minimizing the makespan is pursued as an objective since it attempts at producing the whole planned load as soon as possible, as well as at reducing work in-process inventories.

The critical and most time-consuming part of our Global Procedure is its scheduling part. The latter might be used several times at GPH level by the ‘Heuristic for Checking the Planned Load’ (HCPL) and one or few times at GSH. As a matter of fact, we need several types of heuristics. Good and fast heuristics for the HCPL since this procedure might be carried out many times. Also, the effectiveness of these heuristics will permit carrying out an accurate checking of the available capacity. We also need efficient heuristics for GSH, first, in order to respect the load constraints at period 1 and, second, possibly for saving some capacity so as to be able to face non-predictable events, like

urgent customers orders and machines breakdowns. The remaining chapters in this thesis are dedicated to the study of scheduling solutions for HFS production facilities. Figure 1.5 presents an overview of our decomposition-based heuristic used to deal with production planning and scheduling of an HFS facility.

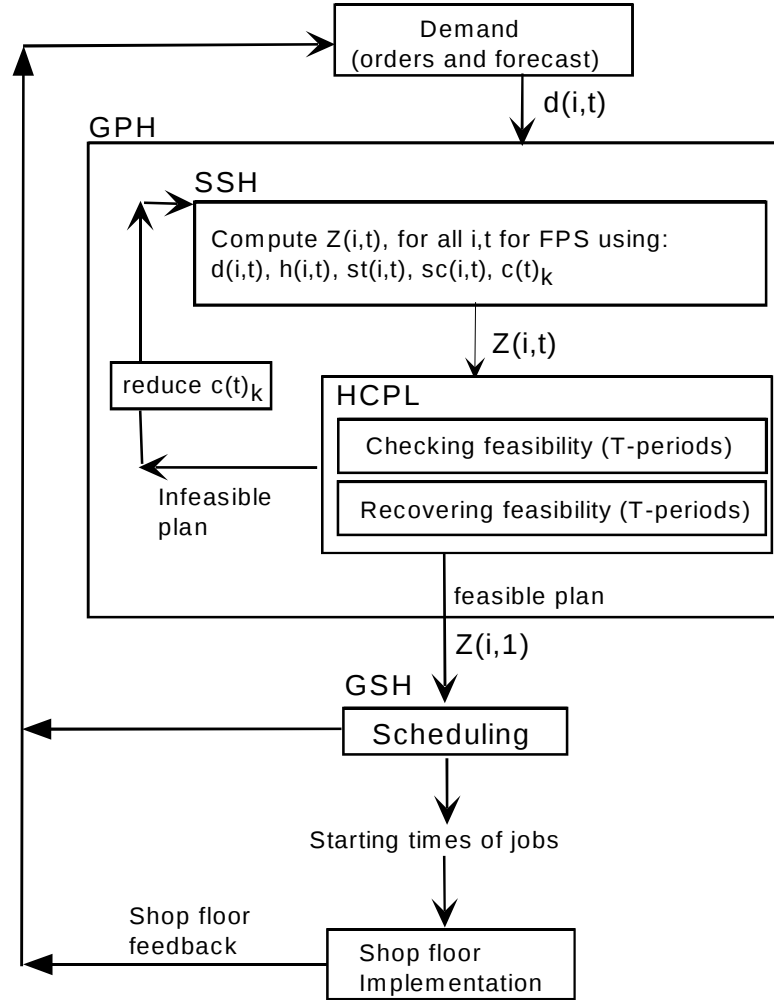


Figure 1.5: An overview of the decomposition-based heuristic for the HFS planning and scheduling

6. CONCLUSION

We have presented in this chapter a general view of production planning and scheduling problems in a hybrid flowshop facility, and proposed a decomposition-based heuristic. The main issue in our approach is the effective use of the available capacity. Our approach is an iterative one, composed of two main procedures. The first one carries out the planning part of the system by using the estimated capacity of the bottleneck stage as representing the overall shop capacity. The second one schedules the planned load at each stage so as to check the feasibility of the plan produced by the first procedure. The latter potentially feeds back the planning system with accurate information on the aggregate capacity utilization. Scheduling is achieved by global scheduling heuristics, which can be any effective solution for scheduling the hybrid flowshop.

At the beginning of our research a review of the literature (see later in Chapter II) has shown that the HFS scheduling has not been well-studied and effective algorithms for finding good solutions for large problems did not exist. Also, because scheduling is so

critical to the approach proposed in this chapter, our research has then been oriented towards the scheduling of the HFS. In the remaining chapters we study branch and bound techniques for solving these scheduling problems by designing new branching schemes, upper bounds as well as lower bounds. These three topics will be dealt with in separate chapters.

The next chapter introduces the studied HFS scheduling subproblem as well as some scheduling solutions, which will be used for designing global HFS scheduling algorithms.

REFERENCES

1. Adams J., E. Balas, D. Zawack, (1988). The Shifting Bottleneck Procedure, *Management Science*, vol. 34, 391-401, n°3, March 1988.
2. Anthony R.N., (1965). *Planning and control systems: A framework for analysis*, Harvard University, Graduate School of Business Administration, Cambridge, MA.
3. Baker Kenneth, (1993). Requirements Planning, 571-627. Chapter 11, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin. *Logistics of production and Inventory*, Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
4. Bitran G.R., and A.C. Hax, (1977). On the design of hierarchical production planning systems, *Decision Sci.*, 8, 28-55.
5. Bitran G.R., E.A. Haas and A.C. Hax, (1981). Hierarchical production planning: A single stage system, *Oper. Res.*, 29, 717-743.
6. Bitran G.R., E.A. Haas and A.C. Hax, (1982). Hierarchical production planning: A two stage system, *Oper. Res.*, 30, 232-251.
7. Bitran G.R., and A.C. Hax, (1982). Disaggregation and resource allocation using convex knapsack problems, *Management Sci.*, 27, 431-441.
8. Bitran. G.R., and H. Matsuo, (1986). The multi-item capacitated lot size problem: Error bounds on Mannes's Formulation', *Management Science*, 32/3, 350-359, 1996.
9. Bitran G.R., and Devanath Tirupati, (1993). Hierarchical Production Planning, 523-568. Chapter 10, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin, *Logistics of production and Inventory*, Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
10. Dautère-Péres S., J.B. Lasserre, (1994). Integration of lot sizing and scheduling decisions in a job shop, *European Journal of Operational Research*, 75, 413-426, 1994.
11. Elmaghraby S.E., (1978). The Economic Lot Scheduling Problem: Review and extensions, *Management Science*, 24, 587-598, 1978.
12. Elmaghraby S.E., and R.E. Karnoub, (1997). Chapter 6: Production control in hybrid flowshops: an example from textile manufacturing. The planning and scheduling of production systems Methodologies and applications, edited by A. Artiba and S.E. Elmaghraby, Chapman & Hall, 163-198, 1997.
13. Florian M., J.K. Lenstra, and A.H.G. Rinnooy Kan, (1980). Deterministic production planning: Algorithms and complexity, *Management Science*, 26/7, 122-20, 1980.
14. Fontan G. and S. Imbert, (1985). Integration between the two levels of decision making in a job shop, presented at EURO VI, Bologna, 1985.
15. Gelders, L.F. and L.N. Van Wassenhove, (1981). Production planning: A review, *European J. Oper. Res.*, 101-110.

16. Graves, S.C., A.H.G. Rinnooy Kan and P.H. Zipkin, (1993). Logistics of production and Inventory, Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
17. Groenevelt Harry, (1993). The Just-in-Time System, 629-670. Chapter 12, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin, Logistics of production and Inventory. Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
18. Hax A.C., and H.C. Meal, (1975). Hierarchical integration of production planning and scheduling, in: Murray Geisler (ed.), TIMS Stud. In Management Sci., vol. 1, Logistics, pp 53-69.
19. Hax A.C., (1978). Aggregate production planning, in: J. Moder and S. Elmaghraby (eds) Handbook of Operations Research, Van Nostrand Reinhold, New York.
20. Hax A.C., and D. Candea, (1984). Production and Inventory Management, Prentice-Hall, Englewood Cliffs, NJ.
21. Karmarkar Uday S., (1993). Manufacturing Lead time, Order Release and Capacity Loading, 287-329. Chapter 6, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin, Logistics of production and Inventory, Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
22. Kirca O., M. Kokten, (1994). A new heuristic approach for the multi-item dynamic lot-sizing problem, European Journal of Operational Research, 75(2), 332-341, 1994.
23. Lambrecht M.R. and H. Vanderveken (1979). Production Scheduling and Sequencing for Multi-Stage Production Systems, OR Spektrum 1, 103-114, by Springer-Verlag.
24. Lasserre J.B., (1992). An integrated model for job-shop planning and scheduling, Management Science, 38, 1201-1211, 1992.
25. Leisten R., (1998). An LP-aggregation view on aggregation in multi-level production planning, Annals of Operations Research, 82, 413-434, 1998.
26. Maes J., L.N. Van Wassenhove, (1986). A simple heuristic for the multi-item single-level capacitated lot-sizing problem, Operations research letters, Volume 4, n°6, 265-273, 1986.
27. Maes J., J.O. McClain, L.N. Van Wassenhove, (1991). Multilevel capacitated lot-sizing complexity and LP-based heuristics. European Journal of Operation Research, 53, 131-148, 1991.
28. Ouenniche J., F. Boctor, (1998). Sequencing, lot-sizing and scheduling of several products in job shops: the common approach, International Journal of Production Research, 36 (4), 1125-1140, 1998.
29. Pinedo, M., (1995). Scheduling: Theory, Algorithms, and Systems, Prentice-Hall, Englewood Cliffs, New Jersey 07632, 1995.
30. Potts C.N., L.N. Van Wassenhove, (1992). Integrating scheduling batching and lot sizing: a review of algorithms and complexity", Journal of the Operational Research Society, 43, 395-406, 1992.
31. Shaw D.X., A.P.M. Wagelmans, (1998). An algorithm for single-item capacitated economic lot sizing with piecewise linear production costs and general holding costs, Management Science, 44(6), 831-838, 1998.
32. Thomas L. Joseph and John O. McClain, (1993). An over view of Production Planning, 333-369. Chapter 7, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin, Logistics of production and Inventory, Handbooks in OR & MS, vol. 4, Elsevier Science Publishers B.V., 1993.
33. Zipkin P.H., (1991). Computing optimal lot sizes in the Economic Lot Scheduling Problem, Operational Research 39, 56-63 1991.