Production Planning Models and Systems

Motivation

In the two industrial production planning systems, the Kellogg Company and GW, mentioned or analyzed in Chapter 1, we observed some important requirements for the new production planning model and tool.

- It was supposed to remedy important weaknesses of the current planning system (the inability to model and plan capacity utilization accurately, because of neglected machine preparation times in the case of the Kellogg Company, and because of neglected machine cleaning times in the MPS model for GW).

- It needed some coordination with the global planning system in charge of supporting all planning decisions from the strategic and long-term horizon level to the very detailed and short-term level. We heard about the tactical plan and the operational plan for Kellogg, and the MRP, ERP, and MPS for the GW Company case, and we observed that these planning levels are not independent (decisions at one level act as constraints at another level).

- It required a high level of integration in the decision processes in place, to avoid manual replanning and make sure that the decisions suggested by the model truly support and have an impact on the real planning decisions.

This need for improvement, coordination, and integration can be observed in almost all industrial projects. In order to develop an effective planning model, the modeler must be aware of the planning process and system used, of the limitations of the current system, of the architecture and structure of the existing system, and of the decision processes used by the planning teams.

This is our motivation for the inclusion of this chapter on production planning models and systems.
Objective

To do a useful job, the modeler must have sufficient knowledge about existing planning models, systems, and processes to be able to evaluate the current system, and in order to design improved, coordinated, and integrated solutions.

The general objective of this chapter is to provide this necessary knowledge. More specifically, the objective is to

- describe or survey the structure of the planning systems used by many – or most – companies,
- learn the general principles of the planning procedures, and
- study some generic classes of production planning models encountered in such systems.

We also provide some analysis and criticism of the planning models and methods used in these systems to help readers to develop some evaluation criteria to measure their performance, and to identify situations where the optimization approach may help to improve the productivity and flexibility of manufacturing systems.

Contents

In this chapter:

- In Section 2.1 we first give mathematical formulations of some of the classical production planning models considered in ERP (enterprise resource planning) or MRP systems;
- Then we analyze in detail in Section 2.2 the well-known generic MRP planning procedure used to solve these models by
  - describing its inputs and its structured data model,
  - presenting the single-item decomposition planning heuristic that forms the basis of most MRP planning systems, and
  - analyzing the limitations of the MRP decomposition approach;
- Next in Section 2.3 we take a broader view and define the planning tasks of APS (Advanced Planning Systems), which subsume the well-known manufacturing, planning, and control systems; material requirements planning (MRP-I); manufacturing resource planning (MRP-II); and hierarchical production planning (HPP); and
- Finally, to illustrate the planning tasks and the planning process along the supply chain, we describe in Section 2.4, without mathematical models or formulations, the generic strategic network design and supply chain master planning problems as further examples of procurement, production, and distribution planning problems.
2.1 Some Production Planning Models

The purpose of this section is to provide further examples and mixed integer programming formulations of production planning models. The formulations described here correspond to classical models in ERP or MRP systems. The next section describes the global structure of such systems.

Modeling Elements

There are a number of modeling elements present in many or most production planning problems. Production planning deals mainly with the determination of production lots or batches, specifically the size of batches and the time of production, in order to meet some demand over a given finite horizon, called the planning horizon. Demand is usually generated from forecasts in a make-to-stock environment, or by customer orders in a make-to-order environment, or often by a combination of the two.

In order to define feasible and economical production plans, several other characteristics of the manufacturing system are usually taken into account: the availability of resources (machine hours, workforce, subcontracting, etc.), the production and inventory costs, and other performance measures such as customer-service level.

The simplest such production planning model is presented next. It is known as the single-item uncapacitated lot-sizing model (LS-U). It corresponds to the planning of a single item to meet some dynamic demand over a discretized planning horizon. It contains all the modeling elements cited above, apart from the fact that there are no resource capacity restrictions. Our tiny economical example in Chapter 1 is one instance of this LS-U model.

There are also modeling elements that are present in some, but not all, models. Such elements usually make the models more complex and more difficult to solve.

- For instance, the products may compete for the allocation of capacity from some shared resources. This has been illustrated with the mixer or the packaging lines in our industrial example in Section 1.2, and is typical of the Master Production Schedule (MPS) model presented hereafter. This MPS approach is often used to plan the production of finished products.
- In some other cases, the products interact through multi-level product structures. In other words, a product can be an output of some production stage and also an input of some other production stage, or it may be delivered from an external supplier. This creates some precedence constraints between the supply and the consumption of that product. These restrictions are usually modeled through inventory balance constraints. Examples of such models are the Material Requirements Planning (MRP) model, or the MPS/MRP integrated model described later in this Section. This MRP model is used to integrate the production and procurement plans of all products and components.
Finally, there are other elements needed to refine the model, or to model capacity utilization in a more precise way. For instance, the demand satisfaction process may allow demand for finished products to be backlogged. In this case, it is possible — but penalized because it has a negative impact on customer satisfaction — to deliver to a customer later than required. This occurs, for example, when a factory does not have enough capacity to deliver to all customers on time.

In some other cases, it is necessary to model capacity utilization more precisely in order to guarantee to obtain feasible production plans. For instance, the capacity consumed when a machine starts or finishes a production batch, or when a machine switches from one product to another, may need to be considered. In these cases, we obtain models with set-up times, start-up times, changeover times, or models with sequencing restrictions. This was the case for the mixer in our industrial example in Chapter 1. On the other hand, such models may be too complex to be solved with set-up or start-up time restrictions, and then simpler models involving only set-up or start-up costs may be worth considering.

Uncapacitated Lot-Sizing Model

The first model is the single-item, single-level, uncapacitated lot-sizing model. This model is the core subproblem in production planning because it is the problem solved repeatedly for each item (from end products to raw materials) in the material requirements sequential planning system (see Section 2.2).

We use the index $t$, with $1 \leq t \leq n$, to represent the discrete time periods, and $n$ is the final period at the end of the planning horizon. The purpose is to plan the production over the planning horizon (i.e., fix the lot size in each period) in order to satisfy demand, and to minimize the sum of production and inventory costs.

Classically, as in our tiny economical example in Chapter 1, the production costs exhibit some economies of scale that are modeled through a fixed charge cost function. That is, the production cost of a lot is decomposed into a fixed cost independent of the lot size, and a constant unit or marginal cost incurred for each unit produced in the lot. The inventory costs are modeled by charging an inventory cost per unit held in inventory at the end of each period. Any demand in a period can be satisfied by production or inventory, and backlogging is not allowed. The production capacity in each period is not considered in the model, and is therefore assumed to be infinite.

For each period $t$, with $1 \leq t \leq n$, the data $p_t$, $q_t$, $h_t$, and $d_t$ model the unit production cost, the fixed production cost, the unit inventory cost, and the demand to be satisfied, respectively. For simplicity we suppose that $d_t \geq 0$ for all periods $t$. The decision variables are $x_t$, $y_t$, and $s_t$. They represent the production lot size in period $t$, the binary variable indicating whether there is a positive production in period $t$ ($y_t = 1$ if $x_t > 0$), and the inventory at the end of period $t$, respectively.
2.1 Some Production Planning Models

The natural formulation of this uncapacitated lot-sizing problem can be written as follows, using the demand satisfaction and set-up enforcement (variable upper bound) generic constraints described in Chapter 1.

\[
\min \sum_{t=1}^{n} (p_t x_t + q_t y_t + h_t s_t)
\]

subject to

\[
s_{t-1} + x_t = d_t + s_t \quad \text{for all } t \tag{2.2}
\]

\[
s_0 = s_n = 0 \tag{2.3}
\]

\[
x_t \leq M_t y_t \quad \text{for all } t \tag{2.4}
\]

\[
x \in \mathbb{R}^n_+, \quad s \in \mathbb{R}^{n+1}_+, \quad y \in \{0,1\}^n \tag{2.5}
\]

where \(M_t\) is a large positive number, expressing an upper bound on the maximum lot size in period \(t\). Constraint (2.2) expresses the demand satisfaction in each period, and is also called the flow balance or flow conservation constraint. This is because every feasible solution of LS-U corresponds to a flow in the network shown in Figure 2.1, where \(d_{14} = \sum_{i=1}^{4} d_i\) is the total demand. Constraint (2.3) says there is no initial and no final inventory. Constraint (2.4) forces the set-up variable in period \(t\) to be 1 when there is positive production (i.e., \(x_t > 0\)) in period \(t\). Constraint (2.5) imposes the nonnegativity and binary restrictions on the variables. The objective function defined by (2.1) is simply the sum of unit production, fixed production, and unit inventory costs.

![Figure 2.1. Uncapacitated lot-sizing network \((n = 4)\).](image)

**Master Production Scheduling Model**

The next model is known as the multi-item (single level) capacitated lot-sizing model. It corresponds to the simplest Master Production Scheduling problem solved to plan the production of finished products in a Manufacturing
Planning and Control System (MPCS) (see Section 2.3). Our GW example in Section 1.2 is another example of such a MPS model.

The purpose is to plan the production of a set of items, usually finished products, over a short-term horizon corresponding at least to the total production cycle of these items. For each item, the model is the same as the \textit{LS-U} model in terms of costs and demand satisfaction. In addition, the production plans of the different items are linked through capacity restrictions coming from the common resources used.

We define the indices \( i \) with \( 1 \leq i \leq m \) to represent the set of items to be produced, \( k \) with \( 1 \leq k \leq K \) to represent the set of shared resources with limited capacity, and \( t \) with \( 1 \leq t \leq n \) to represent the time periods. The variables \( x, y, s \) and the data \( p, q, h, d \) have the same meaning for each item \( i \) as in the model \textit{LS-U}. A superscript \( i \) has been added to represent the item \( i \) for which they are each defined.

The data \( L^i_t \) represent the available capacity of resource \( k \) during period \( t \). The data \( \alpha^{ik} \) and \( \beta^{ik} \) represent the amount of capacity of resource \( k \) consumed per unit of item \( i \) produced, and for a set-up of item \( i \), respectively. The coefficient \( \beta^{ik} \) is often called the set-up time of item \( i \) on resource \( k \), and represents the time spent to prepare the resource \( k \) just before the production of a lot of item \( i \). Together with \( \alpha^{ik} \), it may also be used to represent some economies of scale in the productivity factor of item \( i \) on resource \( k \).

The natural formulation of this multi-item capacitated lot-sizing model, or basic MPS model, can be written as follows,

\[
\min \sum_{i} \sum_{t} (p^i_t x^i_t + q^i_t y^i_t + h^i_t s^i_t) \tag{2.6}
\]

subject to

\[
s^i_{t-1} + x^i_t = d^i_t + s^i_t \quad \text{for all } i, t \tag{2.7}
\]

\[
x^i_t \leq M^i_t y^i_t \quad \text{for all } i, t \tag{2.8}
\]

\[
\sum_{i} \alpha^{ik} x^i_t + \sum_{i} \beta^{ik} y^i_t \leq L^k_t \quad \text{for all } t, k \tag{2.9}
\]

\[
x \in \mathbb{R}^{mn}_+, \quad s \in \mathbb{R}^{m(n+1)}_+ \quad \text{for all } i, t \quad y \in \{0,1\}^{mn} , \tag{2.10}
\]

where constraints (2.6)-(2.8) and (2.10) are the same as for the \textit{LS-U} model, and the generic constraint (2.9) expresses the capacity restriction on each resource \( k \) in each period \( t \).

\section*{Material Requirements Planning Model}

As a last example model, we describe the \textit{multi-item multi-level capacitated lot-sizing model}, that can be seen as the integration of the previous MPS model for finished products, and the \textit{LS-U} models for all intermediate products and raw materials, into a single monolithic model. It is often referred to as the Material Requirements Planning model, or the integrated MPS/MRP model.
The purpose of this model is to optimize simultaneously the production and purchase of all items, from raw materials to finished products, in order to satisfy for each item the external or independent demand coming from customers and the internal or dependent demand coming from the production of other items, over a short-term horizon.

The dependency between items is modeled through the definition of the product structure, also called the bill of materials (BOM). The product structures are usually classified into Series, Assembly or General structures; see Figure 2.2.

![Figure 2.2. Types of product structures in multi-level models.](image)

The indices, variables, and data are the same as before, except that, for simplicity, we also use the index $j$ with $1 \leq j \leq m$ to identify items. For item $i$, we use the additional notation $D(i)$ to represent the set of direct successors of $i$, that is, the items consuming directly some amount of item $i$ when they are produced. Note that for series and assembly structures, these sets $D(i)$ are singletons for all items $i$, and for a finished product $i$, we always have $D(i) = \emptyset$. For $j \in D(i)$, we denote by $r^{ij}$ the amount of item $i$ required to make one unit of item $j$. These $r^{ij}$ values are indicated along the edges $(i, j)$ in Figure 2.2. This parameter $r$ is used to identify the dependent demand, whereas $d^i_t$ corresponds to the independent demand. For each item $i$, we denote by $\gamma^i$ the lead-time to produce or deliver any lot of $i$. More precisely, $x^i_t$ represents the size of a production or purchase order of item $i$ launched in period $t$, and delivered in period $t + \gamma^i$.

The natural formulation for the general product structure capacitated multi-level lot-sizing model, or the monolithic MRP model, is

$$\min \sum_i \sum_t (p_i x^i_t + q_i y^i_t + h_i s^i_t)$$

subject to
where the only difference with respect to the previous MPS model resides in the form of the generic demand satisfaction or flow conservation constraint (2.12). For each item $i$ in each period $t$, the amount delivered from production or vendors is $x^i_{t-\gamma^i}$ ordered in period $t - \gamma^i$, and the demand to be satisfied is the sum the independent demand $d^i_t$ and the dependent demand $\sum_{j \in D(i)} r^{ij} x^j_t$ implied by the production of direct successors $j \in D(i)$.

Because of the multi-level structure, the presence of single item $LS-U$ models as submodels is less obvious, but we show in Part IV how to reformulate this model in the form of single-item $LS-U$ models linked by capacity and product structure restrictions. This reformulation is known as the echelon stock reformulation, and plays a very important role because it allows one to use all the results on the reformulation of single-level problems when treating multi-level problems.

### 2.2 The MRP Planning Model

Many industrial production planning models are variants or extensions of the the generic MRP model (2.11)–(2.15), described in Section 2.1, which is typical of discrete parts manufacturing systems. Provided that the BOM structure allows one to describe the product structure, which is usually the case for discrete parts manufacturing, this model potentially plans the procurement or production of all components needed to satisfy external customer demand over a medium-term horizon.

The numerous extensions or adaptations to this basic model correspond usually to better or refined models to include overtime, product or component substitutes in BOMs, alternate routings or machine selection to perform production operations, shipping and transportation to and from other sites, buying or subcontracting of some components, productivity and capacity utilization, and so on.

Nevertheless, the basic MRP model (2.11)–(2.15) is the kernel of many or most multi-item single-facility production planning models, and is solved in most integrated planning systems (see Section 2.3 for a general introduction to such systems). Moreover, most MRP and ERP planning systems use the same basic or trivial decomposition approach based on $LS-U$ in order to solve this model or, at least, to provide feasible solutions.
In this section, we describe this simple but generic MRP model and its inputs, using the standard operations management terminology for production planning models. We also describe the traditional and heuristic MRP decomposition approach, and discuss its weaknesses.

In such boxes, we establish the link between the generic MRP planning model and its inputs described here, and the mathematical programming formulation (2.11)–(2.15).

A major difference between the traditional MRP approach and the modeling/optimization approach is that the latter forces the user/modeler to make a clear distinction, and avoid some confusion, between the data required as input to the model and the model formulation itself (decisions, constraints, and objective), and also between the model formulation and the algorithm used to build a feasible or optimal production plan.

2.2.1 The Planning Model and Its Inputs

The data required to define and implement the MRP model are now described.

Independent Demand over the Planning Horizon

The main objective of production planning is to meet the so-called independent demand, which is defined for each facility as the demand coming from external sources. This comprises demand from customers for the main finished products, but also spare parts demand and demands from the distribution system or from other facilities.

The independent or external demand for item \( i \) in period \( t \) is represented by \( d_i^t \) in Equation (2.12).

In a make-to-stock (MTS) production policy, this independent demand must be already in stock when the customer demand arrives at the facility. Therefore, all the procurement and production activities must be carried out in anticipation of this demand, and be based on demand forecasts. This policy is typically used for standard products, with little product variety or diversity, such as fast-moving consumer goods and many standard items of household equipment.

In a make-to-order (MTO) or assemble-to-order (ATO) production policy, some activities can still be performed after the external ordering of the products. The delivery lead-time is the time promised to customers for delivery. Therefore, at the time of ordering, the facility must hold enough raw materials or semi-finished products in inventory in such a way that the remaining production lead-time required to terminate the finished products ordered is
less than (or equal to) the commercial lead-time. This implies that planning is decomposed in two phases or two separate problems. The upstream phase, also called anticipation or "push" phase, plans the procurement and production from raw materials up to some semi-finished products, and is based on demand forecasts for these semi-finished products. This is similar to MTS planning. The downstream phase, called the final assembly, on-order phase, or "pull" phase, schedules the production from the semi-finished products held in inventory up to the finished products, and is based on effective customer orders. This decomposition is illustrated in Figure 2.3. This approach is typical of production systems where there exists a large variety of finished product variants, based on a limited variety of raw materials or semi-finished products. This makes it more economical to hold these semi-products in inventory, but imposes a positive commercial lead-time to complete production. This is, for instance, the policy used by Dell to assemble its PCs.

![Figure 2.3. MTS and ATO production policies.](image)

Formulation (2.11)–(2.15) is used to represent either a MTS policy, or the push phase of an ATO policy.

For all production policies, the planning horizon must be long enough to cover at least the total or cumulative lead-time, including procurement, production, and satisfaction of demand. This is necessary if one is to reach a high customer-service level, defined as the fraction of customer demands delivered on time, because we need to order the right materials now from
our suppliers (i.e., the right quantity of each material) to be included into the finished products that will be delivered one lead-time from now. In other words, the total lead-time represents the required anticipation time in the planning process or, equivalently, the minimal planning horizon length. Then, the planning model will be solved and used in a *rolling horizon* manner. That is, the solution proposed for the early time periods will be implemented, the model data and parameters will be updated for the subsequent time periods, the model will be solved again, and so forth.

In formulation (2.11)–(2.15) the number of time periods \( n \) is at least as large as the total cumulative lead-time from the ordering of raw materials to the completion of finished products, expressed in number of periods.

**Bill of Materials (BOM) to Compute Dependent Demand**

The *bill of materials* defines the product structure by specifying for each component (finished or semi-finished product) all of its direct predecessor components (raw materials or semi-finished products), as well as the number of each required per unit of the successor component. This BOM information allows one to transform the finished product or external time-phased demand—forecasts or orders—into detailed time-phased requirements for all components in the production system.

![Bill of Materials Diagram](image)

**Figure 2.4.** The bill of materials for finished product \( FP \).

A BOM example is given in Figure 2.4, where

- each unit of finished product \( FP \) is obtained by assembling one unit of \( A \) with one unit of \( C \),
- each unit of \( A \) is itself directly obtained from two units of \( B \) and three units of \( C \), and
items $B$ and $C$ are raw materials.

We use this simple example to illustrate the MRP planning process. In this example, a total of four units of raw material $C$ are required to produce each unit of $FP$, three units of $C$ per unit of item $A$ are consumed when a production order of item $A$ is performed, and one unit of $C$ per unit of $FP$ is used when an order of $FP$ is released.

The BOM structure is modeled in Equation (2.12) by $r^{ij}$ for all items $i$ and all $j \in D(i)$, that is, all direct successors $j$ of $i$, where $r^{ij}$ is the number of units of $i$ required per unit of $j$.

The demand for intermediate products (such as $A$ or $C$ in the example) coming from the production orders of their successors is called dependent demand, as opposed to independent for finished products, because it depends entirely on the production plans of successor items. Such plans are controlled by the planner, whereas independent or external demand is not. For instance, item $C$ will be consumed only when a production order of $A$ or $FP$ is started (a decision under the control of the planner), but not when a finished product $FP$ is ordered or delivered. This distinction between dependent and independent demand is crucial and constitutes the basis of the MRP planning process.

The dependent demand for item $i$ in period $t$ is represented by

$$\sum_{j \in D(i)} r^{ij} x^j_t$$

in Equation (2.12).

**Procurement and Production Lead-Times**

Procurement and production activities cannot be performed instantaneously. In order to build realistic production plans, procurement or production lead-times, lead-times for short, are taken into account for all components in the BOM structure. They represent the total time needed to complete a procurement or production order, including preparation, administration, waiting, production, quality control and tests, and delivery, and are measured as an integer number of time periods. This information is written next to each component in Figure 2.4.

In the MRP planning model, such lead-times are constant over time, are independent of the order sizes, and are an input of the planning process.

The constant procurement or production lead-time for item $i$ is represented by $\gamma^i$ in Equation (2.12).

We can now rephrase our minimum length condition on the planning horizon. The number of time periods $n$ is at least as large as the sum of $\gamma^i$ values along any path in the BOM graph.
However, there is an important difference between the value of $\gamma^i$ used in the planning optimization model (Equation (2.12)) and the value of the production lead-time used in the MRP planning process.

In the optimization model, $\gamma^i$ is the minimum lead-time required for a batch of item $i$ to be produced, minimum in the sense that no queue time (waiting time for the availability of machines or resources) is included. This holds because the explicit capacity constraints of the optimization model guarantee that there are enough capacity and resources to produce each lot without any delay, and therefore no safety queue time is required.

In contrast, the MRP planning process does not take capacity restrictions directly into account, and the constant production lead-time principle forces the planner to take a worst-case approach. The minimum production lead-time $\gamma^i$ must be augmented by some safety lead-time to guarantee the feasibility of the production plans. For instance, in Figure 2.4, the lead-time for $A$ has been fixed to two periods because item $A$ is sometimes produced in large lots or on machines that are heavily loaded. Therefore a lead-time of two periods is reserved for all production orders, even though most of these orders are of small size, and will be released when there is enough capacity to complete them effectively after one time period. As a side effect, this also increases the level of work-in-progress inventory.

We come back later to the dramatic effects of the necessary inflation of production lead-times in the MRP planning process.

**Routing of Components**

In addition to the product structure defined by the BOM and to the production lead-times, the routing of products through different work centers, as well as the time and capacity consumed at each work center by a production order, are described in order to model and control capacity utilization.

![Figure 2.5. The routing of semi-finished product A.](image-url)
The simplest routing model consists of the decomposition of the production order of each BOM component into a sequence of production operations. This is illustrated in Figure 2.5 where the production of component $A$ is shown to require three successive operations (cutting, assembling, and painting).

The corresponding routing data for component $A$ is given in Table 2.1. The sequence of operations is defined by numbering the operations. For each operation in the sequence, and for important or critical resources (manpower, machines, departments as a whole, etc.), the unit production time ($\alpha_{ik}$ minutes per unit) and the resource preparation or set-up time ($\beta_{ik}$ minutes per order or batch) are defined. This set-up time is independent of the batch size. Usually, transportation and transfer times between operations are also modeled.

<table>
<thead>
<tr>
<th>Operation Number</th>
<th>Operation Description</th>
<th>Resource</th>
<th>Unit Time ($\alpha_{ik}$, [min])</th>
<th>Set-up Time ($\beta_{ik}$, [min])</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Cutting</td>
<td>Mach S100</td>
<td>1.5</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>Transfer</td>
<td>Forklift</td>
<td>–</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>Assembling</td>
<td>Mach ASS</td>
<td>0.5</td>
<td>10</td>
</tr>
<tr>
<td>30</td>
<td>Painting</td>
<td>Mach PPP</td>
<td>2.5</td>
<td>30</td>
</tr>
</tbody>
</table>

For instance, according to Table 2.1, a production order of 10 (resp., 20) units of component $A$ requires 130 (resp., 175) minutes in total, assuming that 20 (resp., 40) units of component $B$ and 30 (resp., 60) units of component $C$ are available, and assuming that the resources are also available when needed. Even if the lot size of component $A$ is almost always below 20 and requires thus less than 200 minutes, the production lead-time for component $A$ has been fixed to two periods – two days or almost 1000 minutes – simply because the machines used are not always available when they are required to produce $A$.

These routing data are first used to model capacity utilization.

The unit production time of item $i$ on resource $k$ is denoted $\alpha_{ik}$ in Equation (2.14). Similarly, the set-up or preparation time of resource $k$ to produce one batch of item $i$ is denoted $\beta_{ik}$ in Equation (2.14).

The routing information allows one to compute the minimal lead-time required for each production order, as well as their load profiles (i.e., the evolution of the load over time) induced by the production plans in each work center and on each critical resource.
Capacity of Resources

To perform finite capacity planning, one needs additional information on the actual or usable capacity of each resource in each time period. The actual capacity is defined as the number of effective production hours that can be performed on the resource during the time period. This capacity will be compared with the load profiles computed from the production plans and routing data.

Usually, the available capacity is obtained as the product of the gross capacity (i.e., the office or worked hours), and the productivity factor (i.e., the fraction of worked hours that are effectively used for production). This productivity factor accounts for unavoidable breaks, interruptions, disturbances, or inefficiencies during the utilization of the resource.

Table 2.2. The Usable Capacity

<table>
<thead>
<tr>
<th>Resource Description</th>
<th>Gross Capacity [hours/day]</th>
<th>Productivity Factor</th>
<th>Usable Capacity [hours/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mach S100</td>
<td>8</td>
<td>0.95</td>
<td>7.6</td>
</tr>
<tr>
<td>Forklift</td>
<td>8</td>
<td>0.85</td>
<td>6.8</td>
</tr>
<tr>
<td>Mach ASS</td>
<td>16</td>
<td>0.85</td>
<td>13.6</td>
</tr>
<tr>
<td>Mach PPP</td>
<td>8</td>
<td>0.95</td>
<td>7.6</td>
</tr>
</tbody>
</table>

The only capacity information needed in production planning models is the net or usable capacity. These data are illustrated in Table 2.2 for the resources used in the routing of component A, where the productivity factors are higher for the automated cutting machine S100 and painting cell PPP than for the resources and operations requiring some manual intervention. There are 16 gross hours per day for assembly because two identical machines are available during one shift.

The net capacity on resource $k$ in time period $t$ is represented by $L_t^k$ in Equation (2.14).

Inventory Records

For all components, the independent and dependent time-phased demand define together the so-called gross requirements, corresponding to the total consumption, by external customers or internally by the production orders, of the components over time. This consumption requirement can be satisfied either from current inventory or from additional production or purchase orders. In order to compute the amounts that still need to be produced or purchased, the inventory status of each component must be known. This includes
the on-hand inventory, which is the physical inventory in the warehouses;
bullet the allocated or reserved inventory, which is the part of the on-hand inventory that is reserved for production orders that have already been released, and is therefore not available any more to satisfy the gross requirements;
bullet the back-orders, which correspond to overdue or late component orders, and will be satisfied or delivered at the next reception; and
bullet the on-order inventory, which is the quantity of components already ordered (purchase or production) but not yet received, and for each such released order the scheduled receipt time period is known.

The available inventory is the inventory status used in production planning models, and is defined as the on-hand inventory minus the allocated inventory. It is often called inventory. The inventory position is defined as the available inventory augmented by the on-order inventory minus the back-orders. It is the most useful inventory status for inventory control, but it is rarely directly used in production planning models.

The planned available inventory of item $i$ at the end of period $t$ is represented by the variable $s^i_t$ in formulation (2.11)–(2.15). The on-order inventory of item $i$, scheduled to be received in period $t$, corresponds to the fixed quantity $x^i_{t−\gamma^i}$ released in the past (typically with $t−\gamma^i \leq 0$).

The planned back-orders of item $i$ at the end of period $t$ will be represented by adding a new backlogging variable $r^i_t$ in the formulation of the flow balance equation (2.12).

The net requirements of a component are the time-phased requirements obtained by subtracting the available inventory, and the on-order inventory when its reception is scheduled, from the gross requirements. They represent the amount still to be purchased or produced in order to satisfy the total or gross requirements.

The inventory status of each component is central and crucial information for the reliability of MRP systems. They are updated very regularly to incorporate the most recent events or transactions (order release, order reception, physical removal from stock, etc.) in order to reflect accurately the real situation on the shop floor and in the warehouses.

**Planning Rules**

Finally, the product database has to contain some more information relative to the definition and parameters of the planning rules used. Typically, it contains

bullet the rules and parameters for safety stocks, where the safety stock of a component is defined as the minimum stock to be held at the end of each planning period in order to be able to cover small variations of demand or consumption during the realization of the plan;
• the rules and parameters for safety times, where the safety time of a component is the time added to the component lead-time to cover unpredictable lead-time variations during the realization of the plan;
• the single-item lot-sizing rules and parameters for each component; such rules are used to transform the computed net requirements into economical procurement and production plans satisfying the requirements; we describe below the role of such single-item plans in the global MRP planning process; and
• component data required to use the lot-sizing rules: the procurement or production cost, the inventory holding cost, and so on.

| The unit production cost, fixed set-up cost, and per unit and per period inventory holding cost are represented, respectively, by $p_i^t$, $q_i^t$, $h_i^t$ in the objective function (2.11).
| The safety times are part of the lead-time parameter $\gamma_i^t$ in Equation (2.11).
| There is no safety stock in formulation (2.11)–(2.15). Such safety stocks can be represented as simple lower bounds on the inventory variables $s_i^t$.

2.2.2 The Planning Process: Single Item Decomposition

So far we have studied the MRP model as defined by its inputs – products, BOM, routing, resources, capacity, inventory – and its mathematical representation. Now, the challenge is to design a solution approach for the mathematical programming problem (2.11)–(2.15).

Unfortunately, this model is usually too large to be solved directly, for the following reasons.

• Short time intervals/buckets are required to model demand satisfaction and capacity utilization accurately.
• Long planning horizons, and thus a large number of time periods, are required to cover the global procurement and production cycle.
• Capacity utilization needs to be tracked for all the critical resources.
• All the intermediate items need to be modeled in order to guarantee the feasibility of the planned flow of materials.

Therefore, decomposition approaches have been proposed to solve the planning model, leading to suboptimal production plans. The typical approach used in ERP/MRP planning systems is illustrated in Figure 2.6 for a MTS production policy and consists of the following steps.

(i) Master Production Scheduling (MPS)

The process starts with the computation of the Master Production Schedule, which, in a make-to-stock setting, is the production plan (lot or batch sizes
per period) for finished products. This means that the MPS is only concerned with the plan of the last production operation yielding the finished product.

The MPS is built to satisfy the combination of firm customer orders – some are usually available for the very short term – and forecasts of customer orders throughout the planning horizon, as well as the required inventory levels at the end of the planning horizon. This last requirement is in anticipation of some future peak demand period, or simply to cover demand up to the next production batch for low-demand items. The MPS must take into account the existing inventory, the scheduled receipts of already released orders, as well as some safety stock requirements to cover forecasting errors.

The MPS mechanics are illustrated in Table 2.3 for the finished product FP from the BOM Figure 2.4, where the planning horizon has been fixed to six time periods; we are currently at the end of period 0, and all inputs to the MPS process are indicated in italics.

In this example:

- The gross requirements are defined, by convention, as the maximum of firm orders and forecasts in each time period, and correspond to updated forecasts.
- The required ending inventory plays the same role as an additional demand forecast for period 6.
- The net requirements are the minimal additional production quantities needed to satisfy the gross requirements, or equivalently the minimal quantities needed for the projected inventory to reach the safety stock level.
- The MPS is chosen to correspond to the net requirements and, therefore, the projected inventory corresponds to the safety stock after the consumption of the initial stock (and where the projected inventory in each period
Table 2.3. MPS Planning Process for Product FP

<table>
<thead>
<tr>
<th>Planning Parameters</th>
<th>Time Periods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ending inventory = 10</td>
<td>Firm customer orders</td>
<td>17</td>
<td>9</td>
<td>2</td>
<td>15</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Demand forecasts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gross requirements</td>
<td>17</td>
<td>25</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Scheduled receipts</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Net requirements</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Safety stock = 5</td>
<td>Projected inventory</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Current inventory = 7</td>
<td>MPS planned orders (end)</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Lead-time = 1</td>
<td>MPS planned orders (start)</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Available to promise</td>
<td>10</td>
<td>11</td>
<td>38</td>
<td>40</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

is equal to initial inventory plus scheduled receipt plus finished MPS orders minus gross requirements).

- The planned MPS orders have to start one period (the lead-time) before their completion.
- The *available to promise (ATP)* row gives the basic information needed to accept new customer orders; it indicates how many units of FP become available to satisfy new customer orders in each period.

In this example, we have just tried to minimize the finished product inventory by producing as little as possible. More economical plans can be built by minimizing production and inventory costs, but this would remain a single-item single-stage production plan.

(ii) Rough Cut Capacity Planning (RCCP)

The above approach can be used to determine, or even optimize, the MPS for each finished product individually. However, such finished products usually share some critical scarce resources, and some consolidation of the MPS plans is needed. This is carried out in parallel to the MPS process and is known as Rough Cut Capacity Planning. Its role is to check globally or "roughly" the feasibility of the MPS with respect to capacity utilization.

In the simplest case, the MPS is established without considering the capacity restrictions and RCCP consists of the computation of approximate load profiles implied by the MPS for some critical resources or for some aggregate view of the capacity (e.g., by department) using historical capacity utilization factors or simplified BOM structures. If the load exceeds the capacity, the planner has to adapt the MPS or increase the capacity manually.

In more sophisticated systems, the consolidation and modification of the MPS or the increase of capacity are suggested by the system.

In all cases, this approach remains approximate or rough because this capacity planning process does not take into account production stages other than the final one, and in particular does not consider
the current inventory and in-progress orders at various stages, as if net requirements were equal to gross requirements at all stages but the final one, and

- the size and timing of production orders required at various production stages to produce the components consumed by the MPS.

Therefore, a detailed finite capacity verification step can only take place once the detailed production plans for all components are known.

(iii) Final Assembly Scheduling (FAS)

In the case of an assemble-to-order (ATO) production policy, a similar approach is used but the MPS is established for the decoupling items. The decoupling items are the semi-finished products at the interface between the push and pull planning phases; they are thus the last items produced to stock.

![Diagram](image)

Figure 2.7. Planning models for an ATO policy.

For the MPS, the only modification with respect to the MTS policy is the need to compute customer demand forecasts at the level of the intermediate decoupling items, rather than at the finished product level. In other words, to behave as if customers were ordering directly the semi-finished products to be assembled.

Then, assuming that these decoupling items are available in stock when needed, the Final Assembly Schedule (FAS) determines when to realize the operations required to transform the intermediate items into the finished prod-
ucts, in order to meet firm customer orders on time. This approach is illustrated in Figure 2.7.

(iv) Material Requirements Planning (MRP)

The MPS and RCCP fix the production plan for all finished products, or decoupling items. Using a similar approach, that is, planning the production to meet uncertain forecasts, does not make sense for the other items in the BOM structure. One can do much better.

Once the production plan for finished products is fixed, one knows exactly when and in what quantity the components entering in the final production stage are required. This information has been called the dependent demand. So, we can replace uncertain forecasts by certain dependent demands, computed using the BOM structure. This eliminates the major source of uncertainty from the planning process, and hence the major reason to hold huge safety stocks. Then, we can plan the production of these components to meet their dependent demand. These production plans determine in turn the dependent demand of their immediate predecessors.

This process can be repeated, level by level in the BOM structure, all the way through, from the finished products back to the raw materials. It is known as the Material Requirements Planning process. Its sequential aspect is illustrated in Figure 2.8 on the BOM structure from Figure 2.4, assuming a MTS policy. Observe for instance that the total dependent demand and the production plan of item $C$ can only be computed after the production plans of both $FP$ and $A$ have been fixed.

![Figure 2.8. The MRP planning process.](image-url)
For each item in the BOM structure subject to dependent demand, this sequential MRP planning process involves the following steps.

**MRP Process: Step 1.** Computation of the gross requirements.
These are time-phased requirements equal to the sum of dependent and independent demand. For some items such as spare parts, there can be a mix of a dependent and independent demand. In this case, forecasts must be computed for the independent part of the demand. The dependent demand is derived directly from the production plans of the direct successors in the BOM.

**MRP Process: Step 2.** Netting or computation of net requirements.
The net requirements are time-phased requirements. They correspond to the minimal additional (i.e., in addition to available stock and scheduled receipts) production quantities needed to satisfy the gross requirements.

**MRP Process: Step 3.** Planning or uncapacitated lot-sizing.
This last step consists in solving the single-item planning subproblem (LS-U) to determine the production plan meeting the net requirements, and satisfying some criterion. A production batch of an item in a period is called a *suggested production order*, or a *suggested procurement order*, or simply a *suggested order*.

Production plans or suggested orders are computed in MRP systems by using so-called *lot-sizing planning rules*. For instance the *lot for lot* (LFL) planning rule consists in taking the suggested orders equal to the net requirements, in every time period. This means that one produces exactly the demand, and therefore one minimizes the inventory level or cost. Other heuristic planning rules try to balance the set-up and inventory costs by grouping net requirements over several time periods in a static way (economic order quantity (EOQ), period order quantity (POQ)) or a dynamic way (part period balancing (PPB), least unit cost (LUC), or least period cost (LPC)).

Finally, this single-item lot-sizing problem with the objective of minimizing the sum of unit production costs, set-up costs and inventory costs (i.e., LS-U) can be solved to optimality by dynamic programming and mixed-integer programming approaches. This single-item subproblem plays a central role in our optimization approaches, and is studied extensively in the sequel.

In all these solution methods, the single-item lot-sizing problem is solved as an uncapacitated problem, simply because the problem is solved separately for each item. This makes it impossible to take joint capacity restrictions into account, and this is also why the lead-time is fixed and independent of the production order sizes.

The MRP mechanism is illustrated in Table 2.4 for the raw material C from the BOM Figure 2.4, using the usual *MRP record* presentation. According to the sequential MRP process, we assume that production plans are available
for \( FP \) and \( A \), and all data available prior to the computation of the MRP record are given in italics.

<table>
<thead>
<tr>
<th>Planning Parameters</th>
<th>Time Periods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orders for ( FP ) (start)</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Orders for ( A ) (start)</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Safety stock = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current inventory = 150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan. rule: EOQ=130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lead-time = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross requirements</td>
<td>80</td>
<td>100</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Scheduled receipts</td>
<td></td>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net requirements</td>
<td></td>
<td></td>
<td></td>
<td>60</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Projected inventory</td>
<td>70</td>
<td>90</td>
<td>20</td>
<td>70</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Suggested orders (end)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suggested orders (start)</td>
<td>130</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the MRP record in Table 2.4:

- There is no independent demand for item \( C \).
- The dependent demand for item \( C \) is three times the suggested orders of \( A \) plus the planned orders of \( FP \) (see the BOM structure in Figure 2.4).
- The initial inventory is large enough to cover the gross requirements up to period 3, and there are only net requirements in periods 4 and 5.
- The planning rule used is the fixed-order size rule (EOQ), and an order of size 130 (this order size is computed using the EOQ formula as the best compromise between inventory and set-up costs for the average net requirement observed) is suggested each time the projected inventory becomes negative.
- The suggested MRP orders have to start two periods (the lead-time duration) before their completion.

This MRP planning process automatically computes suggested orders for all components in the product structure. The MRP records are updated regularly to take into account all transactions that have occurred and have modified the status of the production system, such as new customer orders, new order releases, order reception from suppliers, and so on.

In this dynamic context, the role of the planner (i.e., the user of the MRP system) is first to check the availability of the components and of the resources to perform the orders suggested in the coming or next few periods, and then to release the corresponding orders to the shop floor or to the supplier. In some cases, the MRP system makes infeasible or inadequate suggestions, mainly because it does not take capacity into account during the MRP process, and the planner has to adapt or improve the suggested plan manually. In such cases, the modified orders are transformed into firm suggested orders or blocked orders to prevent the MRP system changing them on the next run or automatic update.
(v) Capacity Requirements Planning (CRP)

The above approach determines the production plan for each component individually. As for MPS and RCCP, some consolidation of the MRP plans is needed. This is done after the MRP computations and it is known as *Capacity Requirements Planning*. Its role is to check the feasibility of the orders suggested by the MRP with respect to capacity utilization.

As for RCCP, there are several versions of CRP. In the simplest case, the CRP consists in the computation of detailed load profiles implied by the MRP orders. This is done by starting each MRP suggested order at its earliest start date, or at its latest finish date, and loading each work center or each resource according to the detailed description of the sequence of operations in the routing data. Once this is done for all suggested and in-progress orders, if the load exceeds the capacity in a work center, the planner has to adapt the suggested orders—start earlier or later to smooth the load—and to create firm suggested orders, or has to increase the capacity, manually. Hence, CRP identifies capacity problems, but does not resolve them.

In more sophisticated systems, the modification of the proposed orders or an increase of capacity are automatically suggested by the system.

In all cases, this approach remains very heuristic and suboptimal. Uncapacitated production plans are first generated, and then locally adapted to become feasible, by moving orders backward and forward in time or by increasing the capacity (overtime, alternate routing, etc.).

### 2.2.3 Limitations of MRP and the Optimization Answer

Although MRP systems are very powerful integrated production management and information systems, their planning modules implementing the myopic decomposition approach described above suffer from very severe limitations.

![Production Planning Example](image)

**Figure 2.9.** A two-level serial production planning example: (a) the minimum cost flow model; (b) its MRP solution with cost = 675; (c) an optimal solution with cost = 575.
This decomposition approach is called myopic because it does not exploit any knowledge about the model in the decomposition.

In other words, the decomposition is carried out in the same naive way for all models. The planning problem is decomposed into uncapacitated single-item subproblems, that are solved independently and sequentially, without backtracking, from finished products to raw materials. Capacity restrictions are taken into account only after the calculation of the production plan, and mainly to compute capacity requirements or to adapt the production plans locally (i.e., using minor modifications) with the hope of making them feasible.

In particular, in terms of productivity optimization, the major drawbacks of this myopic approach are the following.

**Drawback 1:**

**Single-Level Decomposition ⇒ Suboptimal Productivity**

**(Inventory and Production Costs)**

The level-by-level decomposition of the product structure leads to suboptimal solutions with respect to the global minimum cost objective function. This is illustrated in the following simple example, the simplest one can imagine, involving only two items.

Suppose that we have an instance of the MRP model (2.11)–(2.15) with three periods and a serial BOM structure with 2 levels, one item at each level, where one unit of the raw material \( i = 2 \) is required to produce one unit of the finished product \( i = 1 \). For simplicity, we assume that the lead-time \( \gamma^i \) is zero for each item.

The external or independent demand for the finished product is \( d^1 = (10, 15, 20) \). There is no external demand for the raw material. There is a fixed ordering cost of \( q^1 = 200 \) for the raw material, and fixed production cost of \( q^1_t = 100 \) for the finished product for all \( t \). The unit production cost is constant over time, and thus constant in all solutions, and therefore not considered. The inventory cost is \( h^i_t = 5 \) for all \( i, t \). There are no capacity restrictions. This planning problem can be viewed as the fixed charge minimum cost network flow problem represented in Figure 2.9a.

In the MRP approach, we first determine the MPS for the finished product \( i = 1 \) in order to satisfy its external demand. The optimal solution, minimizing the set-up and inventory costs, is to produce 25 units in period 1, stock 15 from period 1 to period 2, and produce 20 units in period 3 with a total cost of 275. This production plan defines the internal or dependent demand for the raw material: 25 units of raw material have to be available in period 1, and 20 units have to be available in period 3. We then solve the MRP subproblem for the raw material, and find that it is optimal to order 45 units in period 1, and stock 20 units from period 1 to period 3 at a cost of 400. Note that an alternate optimal solution for the raw material is to order twice (25 units in period 1 and 20 units in period 3), and avoid the inventory costs. So the
MRP process has produced a global production plan with a cost of 675. This MRP solution is represented in Figure 2.9b.

The optimal solution with a cost of 575 is represented in Figure 2.9c.

Furthermore, when determining the MPS (which corresponds to solving the single-level minimum cost flow subproblem for item \( i = 1 \)), the worst possible solution for item 1 (which is to produce the 45 units in period 1, and satisfy the demands in periods 2 and 3 from stock) forms part of the globally optimal solution shown in Figure 2.9c. This holds because it avoids the very costly procurement of the raw material. Such interactions between the items are simply ignored in the MRP decomposition process.

This example illustrates the difficulty of optimizing the production plans by solving independent single-level subproblems sequentially.

**Drawback 2:**

**Single-Item Decomposition \( \Rightarrow \) Infinite Capacity Planning \( \Rightarrow \) Suboptimal Productivity (Capacity Utilization Plans)**

The main characteristic of the MRP process is the decomposition into independent single-item planning subproblems. Because the resources are usually shared by several or many items, this decomposition scheme does not allow one to take capacity restrictions directly into consideration, that is, into consideration when the production plan is drawn up. In other words, the capacity available for item \( i \) depends on the production plans of some other items, and is therefore not known when planning item \( i \).

Therefore finite capacity planning in MRP is carried out as follows. First, *infinite capacity* production plans (i.e., production plans defined as if capacity were infinite) are determined for all components (MPS and MRP). Next, these plans are translated into capacity requirements (RCCP at the MPS level, and CRP at the MRP level). Finally the plans are heuristically, and often manually, adjusted when some resources are overloaded. This clearly defines suboptimal capacity utilization plans. There is no reason to believe that the best or even good plans can be obtained in this way. The bottleneck (i.e., the most heavily loaded resource) capacity should be accounted for initially in the planning procedure, and exploited optimally, in order to optimize the global productivity.

**Drawback 3:**

**Infinite Capacity Planning \( \Rightarrow \) Constant Lead-Times \( \Rightarrow \) Increased Inventory, Decreased Flexibility**

Another consequence of infinite capacity planning is the impossibility of determining the production cycle and production lead-times as part of the output of the planning process.

As already explained in Section 2.2.1, in a finite capacity planning process it is possible to build realistic or feasible production plans without adding
safety waiting times to the minimum production lead-times ($\gamma^i$ in the MRP optimization model). This is done by taking work center capacity and routing data explicitly into account, and by only releasing orders for which enough capacity is available.

Unfortunately, in an infinite capacity planning approach, the load of the resources cannot be estimated or anticipated. Therefore, the effective production lead-time for each operation is the sum of the technical or minimum production lead-time $\gamma^i$ and the waiting time for the availability of the resources. This waiting or queue time clearly depends on the resource load, and consequently varies over time for each resource. Because these waiting times cannot be anticipated, a worst-case approach has to be taken, and the constant lead-time used in MRP is inflated by a large enough safety time to guarantee that the lead-time can be met in all cases. This safety time is useful in the rare cases when the resources are heavily loaded, and useless in all the other cases.

A first consequence is that production orders are most often completed well in advance of the due-date or requirement date. Thus the safety times translate into increased work-in-progress inventory. A second and indirect consequence is that the total production cycle is augmented by the safety times at all production stages, the MPS time horizon is augmented accordingly, and the whole MRP planning process is based on longer-term forecasts. As long-term forecasts are usually much worse, larger end-product safety stocks are needed to protect the system against larger forecast errors. Finally, this longer MPS horizon requires more anticipation, and reduces the flexibility of the production system.

Summary

In summary, the myopic MRP decomposition scheme leads to important productivity and flexibility losses, two of the key levers in all manufacturing strategies, which is exactly the opposite of what is expected from a good planning system, and the opposite of what was initially expected from MRP systems. Indeed, the starting idea of MRP was to distinguish the dependent demand, which is computable, from the uncertain independent demand, for which forecasts are needed, with the objective of knowing when and how much is needed of each component, and thereby opening the way to a reduction of the global inventory levels.

The Optimization Approach

The observed limitations all relate to the MRP decomposition approach and planning process, and not to the MRP model itself. The MRP model formulated and discussed above adequately represents the planning problem faced by many companies, but a global solution and optimization approach is needed
in order to reach the desired goal of improving the productivity and flexibility simultaneously.

This global optimization approach depends on the two main modeling ingredients to which we hope to contribute: the expertise needed to build correct and adequate mathematical models, and the expertise required to improve the initial problem formulations and to design optimization software allowing one to solve larger instances globally, without resorting to myopic decomposition.

2.3 Advanced Planning Systems

The purpose of this section is to describe the general context of production planning and supply chain planning models and systems.

2.3.1 Supply Chain Planning

A supply chain (SC) consists of a set of organizations, often legally separated, linked by materials, information, and financial flows, that produce value in the form of products and services for the ultimate customer. It can also consist of the geographically dispersed sites of a single and large company. Along this supply chain, raw materials have to be purchased, intermediate and finished products have to be produced or transformed, and finished products have to be sold and distributed.

Therefore a SC is usually modeled as a network composed of vendor nodes; plant nodes where products are produced or transformed; distribution center nodes where products are received, stored, and dispatched but not transformed; market nodes where products are sold or consumed; and transportation arcs connecting the nodes and supporting both the physical and information flow.

Supply Chain Planning (SCP) is defined as an integrated planning approach used to organize the SC activities.

- This multi-dimensional integration is concerned with the functional integration of the primary activities – purchasing, manufacturing, warehousing, transportation – and support activities that constitute the value chain of the SC.
- It is also concerned with the inter-temporal integration – often called hierarchical planning – of these activities over strategic, tactical, and operational planning horizons. Strategic problems deal with the management of change in the production process and the acquisition of the resources over long-term horizons based on aggregated data. Tactical problems analyze the resource allocation and utilization problems over a medium-term planning horizon using aggregate information. This consists in making decisions about, for instance, materials flow, inventory, capacity utilization,
and maintenance planning. Operational problems aim at planning and controlling the execution of the production tasks. For instance, production sequencing and input/output analysis models fit into this category. This integration is critical to success because the design of the SC must take into account the operations performed under this design, and because a company cannot maintain competitive operations and position with poor strategic decisions regarding its technology or the location of its plants and facilities.

- Finally it is concerned with the spatial integration of these activities.

Integrated planning is made possible because of the recent advances in information technology (IT). Focusing only on the procurement and manufacturing or production functions of the supply chain, Manufacturing Planning and Control (MPC) systems are developed to cope with these complex planning environments, and integrate these planning problems into a single integrated management system.

**Figure 2.10.** An MRP-II system.

For instance, Figure 2.10 describes how the tactical and operational planning problems are integrated in *Manufacturing Resources Planning (MRP-II)*
systems, an example of an MPC system. In these systems, medium-term \textit{aggregate or master planning} consists in deciding about capacity utilization, and aggregate inventory levels to meet the forecast demand over a medium-term horizon of about one year. A medium-term horizon is usually needed to be able to take into account some seasonal pattern in demand. MPS consists of planning the detailed short-term production of end-products in order to meet forecast demand and firm customer orders, taking into account the capacity utilization and aggregate inventory levels decided at the master planning stage. Here the time horizon is usually expressed in weeks and corresponds to the duration of the production cycle. MRP-I establishes the short-term production plans for all components (intermediate products and raw materials) from the production plan of end-products decided at the MPS stage, and from the product structure database (bills of materials). Then, \textit{shop-floor control systems} (for manufactured components) and \textit{vendor follow-up systems} (for purchased components) control the very short-term execution of the plans decided at the MRP-I stage. The time horizon at this stage is usually of a few days.

Other well-known integrated production planning concepts and systems fit into this general manufacturing, planning, and control framework. For instance, the MRP-II system represented in Figure 2.10 subsumes the original MRP-I system, and follows the \textit{Hierarchical Production Planning (HPP)} principles.

Such MPC systems are based on transactional databases. However, the existence and storage of transactional data, as well as faster and cheaper data communication, do not automatically lead to improved decisions. The effective application of IT in SC management requires the building of effective decision-support systems. These are called \textit{analytical IT systems}, as opposed to \textit{transactional IT systems}.

Optimization planning models are an essential component of these analytical systems because they are able to evaluate and identify provably good plans and optimize the trade-off between financial and customer satisfaction objectives. In supply chain planning, as well as in operations management in general, the financial objectives are usually represented by transportation costs for purchasing and delivering products, production costs for machines, materials, manpower, start-ups and overheads, inventory holding costs, opportunity costs of the capital tied up in the stocks, insurance, and so on. Customer-service objectives are represented by the ability to deliver the right product, in the right quantity, at the right date and place.

2.3.2 Advanced Planning Systems and the Supply Chain Planning Matrix

The analytical IT or "computerized" planning systems, based on the transactional data gathered from an Enterprise Resource Planning transactional System, are called \textit{Advanced Planning Systems (APS)}. The structure of the
planning tasks of such APS is described in Figure 2.11, and is known as the Supply Chain Planning Matrix (SCPM).

Figure 2.11. The Supply Chain Planning Matrix ([70]).

The main characteristics of an APS are the following:

- **Integral or global planning**: coordination of the planning of the entire supply chain;
- **Optimization focus**: the definition of alternatives, objectives, and constraints for all the planning tasks; and
- **Hierarchical approach**: the decomposition into planning modules, and their vertical and horizontal coordination by information flows.

These characteristics are reflected in Figure 2.11. In most of the applications, traditional MRP or ERP systems do not share these characteristics. They are restricted to the production function, and they do not optimize. Moreover, when they consider various planning horizons, they use essentially a sequential or independent approach for the different planning tasks. In other words, with respect to the APS as in Figure 2.11, there is no real bottom-up coordination with respect to planning horizons, and no left–right coordination with respect to planning functions.
This SCP matrix is also used by APS software providers to offer a set of software modules covering the matrix as much as possible. The typical module architecture of such systems is depicted in Figure 2.12.

![Figure 2.12. Architecture of Advanced Planning Systems ([70]).](image)

Of course, the SCPM defines the general structure of an APS, but a single APS consisting of a fixed combination of software modules cannot respond to the management requirements of all supply chains. Typologies of supply chains are defined in the literature in order to identify supply chains having or sharing the same major characteristics, and therefore sharing the same planning requirements and tasks. One such typology is based on supply chain functional attributes – related to the functions of procurement, production, distribution, and sales – and structural attributes – related to the topography, integration, and coordination of the SC. For instance, this typology has been used to design APSs for specific industries, such as the computer assembly and consumer goods industries.

In the context of APS and its planning matrix, the objective of this book is to give a state-of-the-art description of the modeling and reformulation theory needed to design efficient optimization- or mathematical programming-based algorithms to support the supply chain planning tasks. In other words,
we focus on one of the major characteristics of the APS approach, namely optimization.

2.4 Some Supply Chain Planning Problems

We describe here briefly, without mathematical formulations, two generic classes of supply chain planning problems. They extend the scope of production planning models presented so far by considering the entire supply chain rather than a single production facility or plant.

2.4.1 Strategic Network Design Problems

The purpose of supply chain strategic network design problems is to configure the supply chain network as a whole, from suppliers through production, warehousing, and distribution facilities, down to end customers, which can be downstream subsidiaries.

Decisions

The main decisions to be taken at this stage are the status of the nodes and arcs in the supply chain network. For each node, the decision is usually whether to install a facility at a specific location or site, and also the amount of product processed (bought, produced, transformed) at the node. For the arcs, the decision is whether to use a specific route for a given product to link some nodes in the network using a given transportation mode, and also the amount of product flowing through that route. These major decisions are either considered as static, requiring single-period decisions and a single-period model, or dynamic, involving a multiple-period model. In the case of a multiple-period model, similar decisions can be taken in each time period, and inventory arcs are added in the network at some specific nodes, typically modeling production and storage facilities.

Restrictions

These decisions have to be taken in order to satisfy the forecast demand of the customers. Therefore, the main constraint in this problem is the flow conservation constraint for each product at each location or node in each time period. This means that during each time period, the amount of product received from the local suppliers plus the amount available from initial inventory, the input flows, cover exactly the amount shipped to local customers or to other facilities and the amount put into final inventory, output flows, at each node.

Moreover, there are usually capacity restrictions attached to the various activities. These can be supply capacity, production, processing or transportation capacity, or storage capacity restrictions. Capacity installation (the
amount to install) and expansion (the increase of capacity) become decisions to be taken in these problems.

**Objective**

The objective of the problem is to design a network able to satisfy customer demand and to maximize the after-tax discounted yearly profit of the corporations involved in the supply chain. This includes all costs and revenues in the supply chain, namely revenues from sales and supply, manufacturing, warehousing, inventory, and transportation costs. When the supply chain is composed of different legal entities or covers several countries, this also requires decisions to be taken regarding the product transfer prices between these entities, and the modeling of the legal restrictions on the pricing mechanisms.

**Model Type**

The manufacturing costs and investment costs often exhibit economies of scale with respect to the amount produced or the capacity installed. This is usually approximated or modeled using fixed charge cost functions, that is, cost functions with a fixed component to be paid if there is production/investment and a linear component directly proportional to the amount produced or capacity installed. In such cases, the resulting model is a mixed integer programming model, very often linear.

The general structure of these problems is of the multi-period, multi-product, multi-echelon or level, capacitated fixed charge network flow type. This comprehensive modeling and optimization approach was used to design the supply chain of Digital Equipment Corporation.

**Challenges**

Solving such complex and often large-scale models to optimality is still challenging. This is particularly true when transfer prices have to be incorporated in these problems because this feature often makes the model nonlinear.

**2.4.2 Supply Chain Master Planning Problems**

The purpose of supply chain master planning problems is to optimize and synchronize the materials flow along the complete supply chain over a medium-term horizon. The main purpose is to adapt supply and production levels to demand for aggregated products, taking the capacity of bottleneck resources into account, with a centralized view considering all relevant costs and constraints. This global supply chain perspective for the mid-term decisions allows one to reduce inventory levels by improved coordination and by removing redundant buffers between supply chain entities.
This problem takes the design of the supply chain network as fixed by a higher-level, longer horizon, planning module. The results of master planning impose restrictions on lower level detailed planning modules, which are very often functionally decomposed into short-term procurement, production, distribution, and transportation modules. Feedback mechanisms have to be implemented in order to coordinate these three planning levels.

**Decisions**

The main decisions are the aggregate production and distribution plan for all supply chain entities. In particular, production quantities are decided for each product group, each time period and location (plant or warehouse). Similarly, transportation quantities are decided for each link in the supply chain network, each product group and each time period.

These problems are always dynamic, multiple-period problems because their major objective is to optimize the trade-off between variations in processing levels and variations in inventory levels over time, in order to minimize the cost of satisfying the global supply chain demand for all products. Inventory levels over time are a consequence of the production and transportation decisions.

The main difference with respect to the strategic network design problem is the level of detail for the decisions modeled. Usually, the length of the planning periods is shorter, more detailed product groups are modeled by incorporating intermediate and storable products, and production and storage facilities are modeled in more detail. For example, set-up times and changeover times, the time or capacity consumed when a machine starts a production batch or when a machine switches from one product to another, are incorporated in master planning when they have a significant impact on capacity utilization. This level of detail is required in order to exploit the flexibility of the supply, production, and distribution processes in satisfying demand. To facilitate the coordination with short-term planning, it is often the case that the short-term horizon (the first few days or weeks) within the medium-term horizon is modeled using smaller time period intervals.

**Restrictions**

As for the strategic problem, the decisions are taken in order to satisfy the forecast demand. Therefore, the main constraint is again the flow conservation constraint for each product at each location in each time period.

The other main constraints are the capacity restrictions on supply, production, transportation, and inventory levels.

**Objective**

The objective of the problem is to optimize the trade-off among inventory costs, production, and transportation costs.
Model Type

Again, the manufacturing and transportation costs exhibit economies of scale with respect to the amount produced, and are modeled using fixed charge cost functions.

The general structure of these problems is of the multi-period, multi-product, multi-echelon or level, capacitated fixed charge network flow type. For instance, impressive returns with this master planning optimization approach have been obtained at the Kellogg Company.

Notes

Sections 2.1 and 2.2 In addition to the detailed planning case studies provided in this book, and to the generic planning models described here, we refer the reader to Voss and Woodruff [186] for an introduction to the modeling and solution of MRP optimization problems. For another general survey on production planning, we refer to Graves et al. [78].

Section 2.3 Our general definitions of supply chains and supply chain management are adapted from Christopher [38], Shapiro [149], and Stadtler and Kilger [155].

The presentation of the structure of Manufacturing Planning and Control Systems is derived from Vollmann et al. [185], integrating original characterizations of MRP-I systems by Orlicky [127] and Hierarchical Production Planning approaches by Hax and Meal [87]. The reader should refer to Vollman et al. [185] and Browne et al. [32] for a general description of MRP systems, to Hopp and Spearman [91] for a critical analysis of MRP systems, to standard operations management texts such as Silver et al. [151] and Johnson and Montgomery [93] for a more extensive treatment of the heuristic lot-sizing rules, and to Chopra and Meindl [37] for a modern textbook on Supply Chain Management.

The important distinction between analytical and transactional IT systems is emphasized in Shapiro [149] and Fleischmann et al. [70].

The description of the general architecture of Advanced Planning Systems comes from Fleischmann et al. [70], where a complete description of the Supply Chain Planning Matrix and its planning modules can be found. A similar structure focusing on the difference between transactional and analytical IT systems can be found in Shapiro [149].

The typology of supply chains we refer to is defined by Meyr at al. [120]. They illustrate how to use this typology to design an APS for the computer assembly and the consumer goods industries (see also Fleischmann and Meyr [69]).

Section 2.4 Our description of the generic supply network design problem is inspired by the more complete review on the subject by Goetschalckx [76].
Its application to Digital Equipment Corporation can be found in Arntzen et al. [12].

A more complete introduction to the required coordination between the master planning SC module and the other SC modules, through disaggregation and feedback mechanisms, can be found in Rohde and Wagner [146].

The application of master planning to the Kellogg Company, and its impressive returns, can be found in Brown et al. [31].