Tilburg University

Planning supply chain operations: definition and comparison of planning concepts

Kok de, A.G.; Fransoo, J.C.

Published in:
Design and Analysis of Supply Chains

DOI:
10.1016/S0927-0507(03)11012-2

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Kok de, A. G., & Fransoo, J. C. (2003). Planning supply chain operations: definition and comparison of planning concepts. In A. G. Kok de, & S. C. Graves (Eds.), *Design and Analysis of Supply Chains* (pp. 597-675). (Handbooks in Operations Research and Management Science). North-Holland Publishing Company. https://doi.org/10.1016/S0927-0507(03)11012-2

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. Sep. 2024

A.G. de Kok and S.C. Graves, Eds., *Handbooks in OR & MS, Vol. 11*
© 2003 Elsevier B.V. All rights reserved.

Chapter 12

Planning Supply Chain Operations: Definition and Comparison of Planning Concepts

Ton G. de Kok and Jan C. Fransoo

Department of Technology Management, Technische Universiteit Eindhoven, 5600 MB Eindhoven, The Netherlands

1 Introduction

In this chapter we discuss the Supply Chain Operations Planning problem. Positioning Supply Chain Operations Planning (SCOP) in the context of Supply Chain Management (SCM), the objective of SCOP is to **coordinate the release of materials and resources** in the **supply network** under consideration such that **customer service constraints are met at minimal cost.**

We boldfaced the coordination of releases of materials and resources, since this distinguishes SCOP from other SCM decision processes discussed in this volume. First, by deciding on the releases of both materials and resources, we take into account material and resource constraints simultaneously. Second, we explicitly consider coordination of all release decisions in a multi-item, multi-period setting, i.e., operational coordination over time. Especially this operational coordination over time distinguishes SCOP from other Supply Chain Planning (SCP) activities, such as setting seasonal stock levels based on aggregate supply-demand balancing or planning availability of resources in general. The SCOP problem incorporates the outcomes of earlier planning decisions and generates material and resource release decision that are executed by the shopfloor control function. In the context of the SCP matrix presented in Stadtler and Kilger (2000), SCOP overlaps both mid-term and short-term planning, i.e., it translates mid-term planning decisions into short-term execution decisions.

Another consequence of the focus on operational coordination is the incorporation of demand uncertainty and throughput time uncertainty in the models discussed in this Chapter to the extent possible to date. In that sense our approach differs from the more common approach in the context of SCP to formulate a deterministic instance of a particular SCP problem and develop heuristics or optimization methods to solve the instance. We will discuss the difference between these two approaches in more detail in Sections 4 and 6.

We boldfaced 'supply network' since we discuss in this chapter the SCOP problem for general supply network structures. This implies that we assume that items considered in the SCOP problem may be assembled from other multiple items and may be assembled themselves into other multiple items. In that respect we extend the analysis presented in Chapters 10 and 11 of this volume, albeit that the generality of the problem posed prohibits an analysis of similar elegance.

Notice that nowadays the supply network involves multiple organizations, e.g., an OEM and 1st and 2nd tier suppliers. Furthermore, the SCOP problem involves multiple units within an organization, e.g., sales, marketing, production and purchasing, or different manufacturing or warehousing locations. This observation has important consequences for the formulation of quantitative models that formalize the SCOP problem. In our formulations we assume that there exists a centralized objective function, and that information is shared across the supply chain. Furthermore, we assume that each of the echelons in the supply chain assumes responsibility for maintaining a certain lead time, as this is not considered part of the SCOP task. We will discuss this decomposition further in the next section, where we argue that this hierarchical decomposition and ordering of planning decisions is the only possible way both from a theoretical and from a practical perspective. Note that we will thus neither consider multi-agent situations in SCP, nor gaming situations (which have only been studied until now in a supply chain contracting setting, see Chapters 6 and 7 of this volume). We would however like to point out that some form of central coordination is possible even under limited information exchange using the models discussed in this chapter (see Fransoo, Wouters, and de Kok (2001)).

We boldfaced the objective of meeting customer service level constraints at minimal cost, because our objective is to compare, where possible and by selection of appropriate case situations, the various SCOP concepts proposed in the literature to date. To our knowledge such a comparison has not yet been undertaken and we have found that it generates deeper insights into the nature of the SCOP problem. Surprisingly, the comparative study also generated further insight into the design of supply networks, in particular the positioning of inventory capital in a given supply chain structure.

The outline of this introductory section is as follows. In Section 1.1 we introduce the SCOP problem from a practical perspective in more detail. Thereafter we define the variables and notions that enable to formulate the SCOP problem as a quantitative optimization problem. These variables and notions are used throughout this chapter. In Section 1.2 we discuss the material aspect of the SCOP problem, whereas in Section 1.3 we discuss resource aspects. In Section 1.4 we briefly discuss the concept of so-called planned lead times. In Section 1.5 we present two optimization problems that are used as a basis for comparison between different SCP concepts. In Section 1.6 we deal with the Customer Order Decoupling Point (CODP) concept and its relevance for the SCOP problem.

1.1 The SCOP problem

The SCOP function is responsible for the coordination of activities along the supply chain, by making decisions on the quantities and timing of material and resource releases. In this chapter, we explicitly model the supply chain as a network, i.e., activities that transform inputs into outputs using available resources are preceded by multiple transformation activities and succeeded by multiple transformation activities. Note that a transformation activity is a general designation of any type of relationship between two items in a supply chain, and can be both referring to physical transformation activities such as manufacturing or assembly activities and to non-physical transformation activities such as transportation from one location to another. Typical activities to be considered are

1. Manufacturing activities, i.e., activities that physically transform physical inputs into physical outputs.
2. Transportation activities, i.e., activities that move physical outputs from one location to another.
3. Planning activities, i.e., all administrative activities that are required for enabling a manufacturing or transportation activity to take place, such as process planning, transportation contracting, creation of purchase orders, etc.

From the SCP point of view it is essential to identify all relevant activities and their mutual relationships. In particular planning activities can be executed parallel to manufacturing and transportation activities. On the other hand it is well possible that planning activities determine a major part of the overall supply chain throughput time, e.g., when negotiation of price is essential for economic viability or acquisition of information about future demand is of paramount interest, or when letters of credit need to be obtained before actual manufacturing can start.

In each SCOP situation, the definition of these three types of activities, including their characteristics, is the starting point for defining the SCOP problem. In this subsection we concentrate on the representation of the manufacturing activities and transportation activities in relation to the SCOP problem. Reason for this starting point is two-fold:

1. Manufacturing activities and transportation activities are usually well-defined processes of which the main characteristics like processing times, resource requirements and process yields can be easily determined relatively.
2. Supply chain planning in itself is a planning activity at a specific hierarchical level. This implies that the choice for a particular SCP concept impacts planning activities at both lower and higher hierarchical levels. We postpone a detailed discussion to Section 2.

When considering physical transformation activities and their mutual relationships, we find that two generic aspects determine these relationships:

1. One transformation activity's output is another transformation activity's input.
2. One transformation activity shares one or more resources with another transformation activity

In Section 1.2 we focus on the first aspect, defining a Bill Of Material (BOM) structure and all related variables. In Section 1.3 we discuss the second aspect, defining a Bill Of Process (BOP) structure and all related variables.

1.2 Bill of material structure

The physical supply network structure is defined by parent-child relationships between items. 'Item' is the generic term for any input into and any output from transformation activities. In the case of a manufacturing operation a set of items is transformed into one or more items. In this chapter, we restrict ourselves to the situation that a transformation process outputs only a single item, which in turn can be used in multiple other transformation processes. This assumption is valid in many situations, e.g., in discrete part manufacturing. There are situations, especially in process industries and reverse manufacturing, where the manufacturing of one item implies that another item is manufactured at the same time. Such items are called by-products. For a discussion of this phenomenon and the resulting complexities, we refer to Spengler, Püchert, Penkuhn and Rentz (1997). It is important to note that a transportation activity transforms one item into another by changzing the location of the material involved. Generally speaking an item is equivalent to a material/location combination. We omit here the time-aspect of an item; we assume the item does not change over time, e.g., due to engineering changes.

Let us consider a supply network consisting of N items. For each item i, $i=1,2,\ldots,N$ we define a_{ij} as the number of items i required to produce one item j ($i=1,2,\ldots,N$, $j=1,2,\ldots,N$).

The matrix (a_{ij}) is called the Bill Of Material (BOM). In the context of MRP-literature the BOM is usually associated with a single end-item or so-called MPS-item (MPS is Master Production Schedule) [cf. Orlicky (1975)]. Our definition of BOM comprises that definition in the sense that if our supply network would produce a single end-item for a market then both definitions of BOM would be identical.

An end-item is an item that is not used in any other item. Such an end-item is delivered to customers of the supply network. We define E as the set of end-items, i.e.,

$$E \quad \{i|a_{ij}=0,\ i=1,2,\ldots,N,\ j=1,2,\ldots,N\}$$

We define the set I of intermediate items as

$$I \quad \{i | \exists\, 1 \leq j \leq N \text{ with } a_{ij} > 0,\ i = 1, 2, \ldots, N\}$$

For notational purposes it is convenient to introduce the following sets associated with each item,

$$V_i \quad \{j |\, a_{ij} > 0, \quad j = 1, 2, \ldots, N\}$$
$$W_i \quad \{j | a_{ji}, > 0, \quad j = 1, 2, \ldots, N\}$$

Hence V_i is the set of successors of item i and W_i is the set of predecessors of i, $i = 1, 2, \ldots, N$. With the above definitions we have characterized completely the material structure of the supply chain.

One set of decisions that results from solving the SCOP problem is the set of *material release decisions*. In order to rigorously describe these decisions we must introduce some notation. The variables defined relate to the solution of the SCOP problem at a particular point in time. We assume that the SCOP problem is solved periodically at equidistant moments in time. Typical periods used in practice are days, weeks and months.

Let us define period t as the time interval $(t-1, t]$. At time t, $t = 0, 1, 2, \ldots,$ release decisions are taken. We define for $i = 1, 2, \ldots, N$:

$D_i(t)$ independent demand for item i in period t, i.e., demand in period t for item i, that is not derived from demand for items in $I \cup E$

$G_i(t)$ dependent demand for item i in period t, i.e., demand in period t for item i, that is derived from demand for items in $I \cup E$

$p_i(t)$ quantity of item i that becomes available at the start of period t from the transformation activity generating item i

$r_i(t)$ quantity of item i released at the start of period t immediately after receipt of $p_i(t)$

$I_i(t)$ physical inventory of item i at the start of period t, immediately before receipt of $p_i(t)$

$B_i(t)$ backlog of item i at the start of period t, immediately before receipt of $p_i(t)$

$J_i(t)$ net inventory, i.e., physical inventory minus backorders, of item i a t the start of period t, immediately before receipt of $p_i(t)$

Notice that independent demand is demand generated by customers of the supply network. Such demand is usually not known beforehand and must be forecast. We define the item set P as,

$$P \quad \text{Set of items } i \text{ with } D_i(t) > 0 \quad \text{for some } t > 0.$$

Furthermore notice that $\{r_i(t)\}$ are the set of decision variables that constitute one part of the core outcome of the SCOP problem, viz. the material release decisions.

1.3 Bill Of process structure

Manufacturing and transportation activities are executed by resources. Execution of an activity at a resource requires some processing time. In general such a processing time may vary over time due to many different causes, most of which cannot be controlled. Hence in general we represent the processing time by a random variable.

We assume that the processing time at a resource depends on the item that is the output of the transformation activity. It can be easily seen that this assumption is without any loss of generality. We furthermore assume that a resource can only execute one transformation process at a time. This assumption may be a restriction, e.g., in process industries it is quite common that one transformation activity generates multiple items (e.g., by-products), which implies that multiple transformation processes are run in parallel on the same resource. In our definition, this would have to be solved by pre-allocating a specific portion of the available resource capacity to a specific item.

As stated above we assume that the SCP process is executed periodically, e.g., daily or weekly. This implies that we define resource availability as available capacity in units of time during a period. Thus we define

C_{kt} Amount of capacity available in units of time of resource k in period t, $k = 1, \ldots, K$, $t \geq 1$,

where K is the number of available resources. Let us define the following variables associated with resource usage,

U_k Set of items that can be processed on resource k
c_i Time required to process one unit of item i on its resource

For the sake of simplicity we assume that an item can be processed on one resource only. In many cases the analysis can be extended to the situation where item i can be processed on multiple resources, which implies the definition of additional variables and constraints.

The decision variables related to the release of resources at the start of an arbitrary period are given by the set $\{q_i(t)\}$, where $q_i(t)$ is defined as

$q_i(t)$ Amount of item i processed in period t, $t \geq 1$.

We note here that we assume that an amount of item i processed in period t becomes available at the start of period $t+1$. This implies that

$$p_i(t) = q_i(t-1), \quad t \geq U.$$

This implies that we do not consider the possibility of random yield. For an extensive discussion of random yield we refer to Yano and Lee (1995). Based on our experience with SCOP problems in highly volatile environments with random yields we argue that current state-of-the-art literature on this subject is not applicable to multi-item multi-echelon networks. As a quick-and-dirty solution we propose to incorporate random yield through periodic updates of the state of the supply network, taking into account actual yields, and some safety stock provisions.

1.4 The planned lead time concept

As stated above SCP coordinates release of material and resources. The coordination of material would be more or less trivial if transformation activities would require negligible time. However, transformation activities require processing times on resources. Due to various sources of uncertainty, such as demand uncertainty and random processing times, the interactions between materials and resources result into *lead times* of item orders. The phenomenon of lead times in manufacturing has been extensively studied. For an overview of this literature, which strongly relies on queuing network theory, we refer to Suri, Sanders and Kamath (1993) and Buzacott and Shantikumar (1993). From this literature we learn that a lead time consists of a processing time and a waiting time. The waiting time is typically the major part of the lead time, caused by interaction between multiple item orders, which are using the same resources for the execution of the transformation process. Waiting of an item can occur both before and after the actual transformation activity.

Thus, in order to properly coordinate the release of materials and resources, the SCOP level has to take into account lead times. For each item we thus define its lead time L_i,

L_i throughput time between time of release of an order for item i and time at which the ordered items are available for usage in other items and/or delivery to customers

Given the periodic nature of the SCP process we assume that L_i is an integer number of time units. We assume that items i released at the start of period t are available for usage at the start of period $t+L_i$, i.e., in $(t+L_i-1, t+L_i]$.

In the context of SCP we are faced with the following core issue:

Is L_i endogenous or exogenous to the SCP concept?

This issue is extensively discussed in Section 2. It is concluded there that the lead time L_i is exogenous to the SCOP problem. Given the fact that, as stated above, lead times are related to resource utilization, the actual choice of L_i should be consistent with the resource availability and resource requirements that can be derived from the BOP and the exogenous demand characteristics. Such consistency should be derived from either empirical data or by applications of the above-mentioned results from queuing (network) models.

1.5 Performance measurement as a basis for comparison

The main objective of this chapter is to provide insights into the applicability of the various SCP concepts proposed in the literature. Typically, each scientific contribution in this area selects its own case to generate managerial insights. Contrary to common practice in combinatorial optimization there is no commonly accepted set of test problems. Furthermore, it may happen that analytical results are derived based on assumptions that need not hold. A typical example of the latter is the assumption that upstream availability is guaranteed, so that a decomposition of the supply chain model results. We have seen examples where the analysis that builds on this assumption eventually yields 'optimal' solutions that strongly violate the assumption required for the analysis. Such examples are discussed in Section 5.

In order to make a proper comparison of different SCP concepts we define a cost structure and a performance criterion. We define $C(t)$ as the cost incurred at the end of period t, $t \geq 0$,

$$C(t) = \sum_{i=1}^{N} h, I_i(t),$$

where h_i var value of item i $\forall i$

Notice that $C(t)$ is not really a cost function but represents the total supply chain inventory capital investment at the start of period t. We are interested in the long-run average value of $C(t)$,

$$\overline{C} = \lim_{t \to \infty} \frac{1}{t} \sum_{s=1}^{l} C(s).$$

We assume existence of $\overline{C}$, which holds true in the case of stationary stochastic demand. The comparisons of SCOP concepts in Sections 5 and 6 are restricted to that situation.

The long-run average supply chain inventory holding cost can be derived from multiplying $\overline{C}$ by the interest rate. Note that by taking $\overline{C}$ as a basis for comparison we circumvent discussions about proper interest rates, although a discussion about value added in the supply network cannot be circumvented.

Comparing capital investments suffices only when lot-sizing restrictions are irrelevant. More precisely, when we assume that each period for all items a positive quantity is released, we can refrain from considering fixed set-up or ordering costs. In many practical applications lot sizing is not an issue at SCOP level. Either because of sufficiently high manufacturing flexibility [cf. Bertrand (2003), Chapter 4 of this volume for an extensive discussion of flexibility in supply chain management context], or because of the time aggregation into weekly or monthly buckets, implying that lot sizing decisions are taken at the level below the SCOP level. In most cases the lot sizing restrictions have been determined at some level above the SCOP level, since lot sizing impacts the need for resources. Still, there are situations where lot sizing decisions must be taken at SCOP level. We will not deal with that situation due to the fact that to date virtually no results are available in literature on the analysis of general assembly networks under demand uncertainty and lot sizing restrictions. For an in-depth discussion of lot sizing and its position in the planning hierarchy we refer to Chapter 9, Fleischmann and Meyr (2003).

The capital investment is needed to ensure sufficient customer service. Note that we must define customer service for all items with independent demand, i.e., for all items in P. As performance criteria we choose α_i and β_i, $\forall i \in P$, defined as

α_i $\quad \lim_{t\to\infty} P\{I_i(t) > 0\}, \forall i \in P$, non-stockout probability

β_i $\quad \lim_{t\to\infty} 1 - \dfrac{[E(I_i(t) + p_i(t) - D_i(t))^+] - E[(-I_i(t) - p_i(t)^+]}{E[D_i(t)]}, \forall i \in P$, fill rate

Likewise the case of $\overline{C}$ we assume existence of α_i and β_i. Notice that α_i is identical to the P_1-measure defined in Silver, Pyke, and Peterson (1998) and β_i is identical to the P_2-measure defined there. For each SCOP concept $\mathcal{P}$ we want to solve the following problems:

Problem (P_α)

$$\min \overline{C}(\mathcal{P})$$

$$\text{s.t.} \quad \alpha_i(\mathcal{P}) \geq \alpha_i^*, i \in P$$

Problem (P_β)

$$\min \overline{C}(\mathcal{P})$$

$$\text{s.t.} \quad \beta_i(\mathcal{P}) \geq \beta_i^*, i \in P$$

It is implicitly assumed that the SCP concept $\mathcal{P}$ satisfies the set of material and resource constraints derived in Section 3. In both problems we express the dependence of both $\overline{C}$ and α_i and β_i on $\mathcal{P}$. Hence we want to minimize capital investments subject to the service level constraints α_i^* or β_i^*, $i \in P$. An alternative to the use of service level constraints is the introduction of penalty costs. Apart from the fact that penalty costs are hard to determine in practice, one is also compelled to replace capital investments by inventory holding costs in order to make a proper trade-off. As indicated above we want to circumvent a discussion of holding costs. For results on the equivalence of penalty costs and service level constraints we refer to Silver et al. (1998), Diks, de Kok and Lagodimos (1996), Diks (1997), Janssen (1998) and Van Houtum and Zijm (2000).

Apart from a comparison of SCP concepts, we will also discuss the differences in analysis of particular SCP concepts as reported in the literature. Especially for SCP concepts that assume stochastic demand, which are discussed in Section 5, we have found different simplifying assumptions needed for tractability. We will show that some of these assumptions may yield suboptimal or erroneous results when validating the model results with simulation or an exact analysis.

We emphasize here that the comparative approach presented here is, to our knowledge, the first of its kind. Furthermore, the supply chain operations planning problem with stochastic demand is complex and to date no insight exists into the structure of optimal policies for general supply chain structures. For single period problems with two components and two end-items Rosenblatt and Eynan (1996) are able to derive the optimal policy structure. A recent paper by Hillier (2000) discusses an Assemble-To-Order situation, where the supply chain structure consists of multiple components and multiple end-items. Stocks are only held at the component level. Using Stochastic Dynamic Programming Hillier derives the optimal policy for the multi-period problem and the infinite-horizon problem. For more complex supply chain structures no results are currently available in the literature.

All this implies that this state-of-the-art comparative study identifies more questions than answers. These questions are the basis for further research, which is, as stated above, discussed in Section 7.

1.6 The customer order decoupling point concept

In this introductory Section, it is relevant to discuss a concept from the Supply Chain Management literature that is relevant for the development of an appropriate SLOP model. This concept is known as the Customer Order Decoupling Point [CODP, cf. Bertrand, Wortmann and Wijngaard (1990), Silver et al. (1998), or Hoekstra and Romme (1991), from which the Decoupling Point concept originates]. The CODP is the point that indicates how deeply the customer order penetrates into the supply chain. It is the distinction between the order-driven and forecast-driven parts of the supply chain for a particular product market combination. Items that are kept in stock at the CODP are those items for which demand must be forecast due to the fact that future demand between the moment of release of items and the moment those items are received is (partially) unknown; the lead time of supplying the item is longer than the lead time requested by the customer. Downstream of the CODP, items are not kept in stock, since future demand for these items between release moments and receipt moments is known; the lead time of supplying the item from the CODP to the customer is shorter than the lead time requested by the customer. In our models, we only consider demand upstream from the

CODP. Demand *at the CODP* then is independent demand, either originating directly from the demand forecast of end-items (Make-to-Stack), demand forecasts of modules (Assemble-to-Order), or of components (Make-to-Order). All releases *downstream of the CODP* are based on actual customer orders and thus are not planned under uncertainty of demand (cf. Final Assembly Schedule, Orlicky (1975)). All releases *upstream of the CODP* are planned based on dependent demand.

It is interesting to note here that originally the CODP concept is applied to a single organization. When taking a supply chain view, i.e., a view comprising multiple organizations, many CODPs vanish due to the fact that demand originally considered to be independent and thereby requiring forecasting, becomes dependent demand. As we will see in the sequel this observation has a major impact on the performance of the supply network. Typically, items that are stocked at the CODP are assumed to be available with high probability to satisfy independent demand. When converting these items that face independent demand into items facing dependent demand only, such a requirement is no longer relevant. Trade-offs will reveal that a supply chain view implies low availability of these items to prevent unnecessary inventory capital investments.

In this chapter, we will only consider the releases of items that are kept in stock at the CODP and the items that are part of the BOM of those CODP-items. Therefore, without loss of generality, *we denote the CODP-items as end-items* in this text. In case these end-items are not sellable products, but modules or (sets of) components, the independent demand for these end-items can be derived from forecasts of (sets of) sellable products. Alternatively, one can use historical data and market intelligence to derive forecasts of end-items directly.

1.7 Structure of the chapter

This concludes our discussion of the basic notions that enable to define the SCOP problem in the form of quantitative models. The remainder of this Chapter is structured as follows. In Section 2 we discuss the position of the SCOP problem in the context of a hierarchical planning framework comprising aggregate planning, SCOP and detailed scheduling. An important aspect of the discussion in Section 2 is the motivation for incorporating so-called planned lead times into the formulation of the SCOP problem, since in the sequel of this Chapter we restrict to supply network models with deterministic item lead times. In Section 3 we derive generic material and resource release constraints, which we denote further as 'generic SCOP constraints', which are used in later sections to test the validity of various SCOP concepts proposed in the literature. In Section 4 we use the generic SCOP constraints to derive an LP formulation for the SCOP problem without lot sizing restrictions on the material releases. The LP formulation is a benchmark in the discussion of other SCOP concepts based on deterministic exogenous demand in a rolling schedule context. The exogenous demand is derived from a forecasting or sales planning process [cf. Chapter 9, Fleischmann and Meyr (2003) of this volume]. In Section 5 we discuss SCOP concepts based on quantitative models that explicitly incorporate stochastic demand. Where possible we compare the various concepts with respect to assumptions made regarding the supply network structure and regarding the item availability assumptions required for the quantitative analysis. In Section 6 we compare the LP-based SCOP concept with a so-called synchronized base stock (SBS) policy developed by De Kok and Visschers (1999) under infinite resource availability. The reason for selecting these two concepts is that they represent two really distinct modeling concepts for the SCOP problem for general supply networks and for both concepts we can derive solutions that ensure that predefined customer service level constraints are met with a high accuracy. The LP-based concept represents the class of SCOP concepts based on deterministic models in a rolling schedule context, widely available in standard software for SCOP as discussed by Fleischmann and Meyr (2003) in Chapter 9. The SBS concept seems to be to date the only concept representing SCOP concepts explicitly incorporating stochastic demand into the SCOP model, that is able to cope with the general multi-item multi-echelon models that result from the SCOP problem for general supply networks. The comparison of these two concepts yields fundamental insights into the SCOP problem that are extensively discussed. Finally, in Section 7 we summarize our findings and discuss further SCOP research challenges.

2 The hierarchical nature of SCP

In this chapter, we position the supply chain operations planning problem in a hierarchical framework. Hierarchical planning frameworks enable us to accurately model the consecutive planning and scheduling decisions made in manufacturing organizations. The SCOP problem is only one in a series of planning and scheduling problems to be solved by (groups of) manufacturing organizations to realize their objectives in terms of customer service, turnover, profit, ROI, etc. We start with a discussion of various research perspectives that underlie the development of hierarchical planning concepts developed in the past.

2.1 Hierarchies in planning

Decisions with regard to the different components of planning of supply chain operations have traditionally been analyzed independently from one another by researchers. Research addressing the scheduling problem, the (multi-echelon) inventory problem, and the aggregate capacity planning problem have hardly been interconnected while maintaining their own characteristics. On the contrary, in the late 1960s and early 1970s attempts have been made from each of these domains to expand the scope of research and apply their available specific methods to other components of the SCP problem. In these approaches, the specific nature of each of the components has however been disregarded, and the problems have developed into conceptually monolithic models. An illustration is the work on combining lot sizing and scheduling [see, e.g., Dauzère-Pérès and Lasserre (1994)], in which two models remain to exist, but a final solution is obtained by iterating between the two models.

Managers were however still faced with this multitude of different problems in the SCP domain. They solved these issues by organizing these decisions in a hierarchical manner. Meal (1984) analyzes and describes these hierarchies and links them to the hierarchical planning hierarchies introduced by Hax and Meal (1975) and Bitran and Hax (1977).

The idea for hierarchical production planning was captured formally by Anthony (1965). He introduced three levels of hierarchical control: Strategic Planning, Management Control, and Operational Control. The principal ideas for developing this planning hierarchy into a set of formal models supporting coordinated decision making at these levels were developed by Bitran, Hax, and Meal (BHM) in the early 1970s [Winter (1989)]. In their publications, generally the following terms are used for the models supporting the three decision levels: aggregate planning, family disaggregation, and item disaggregation. The BHM hierarchy is based on capacity coordination only. Material coordination is not considered and bills-of-material are not included, which precludes the use of their methodology in SCP. Originally, the work was motivated by and the models were based on discrete parts batch manufacturing, with later applications in continuous manufacturing and job shops [see Bitran and Tirupati (1993) for a review and McKay, Safayeni and Buzacott (1995) for a historical perspective]. Note that all these environments are primarily capacity oriented [Bertrand et al., 1990]. Essentially there are two types of constraints at each of the levels of the BHM formulation:

(1) Primary process constraints: these are 'hard' constraints that are derived from physical constraints in the process, such as resources.
(2) Decision process constraints: these are 'soft' constraints that are imposed upon a level by its immediate higher level in the decision hierarchy.

At the highest level, the aggregate resource constraints form the basis of the resource hierarchy, with the decision how much time to allocate in regular time and in overtime, in line with the original HMMS model [Holt et al. (1960)], although the costs in the BHM model are linear and not quadratic as in HMMS. A distinction between the primary process constraints and the decision process constraints is not made in the model formulation, leading to the fact that, e.g., a decision to produce a certain quantity of a product family is fixed, despite possibly 'better' feasible solutions once the more detailed planning starts at a lower level.

When multiple stages are introduced into the BHM formulation, perfect aggregation becomes very difficult [see the review of the work of Axsäter by Bitran and Tirupati (1993)] and the only way in which to devise a hierarchical planning procedure is by a very loose coupling. Graves, Meal, Dasu and Qui (1986) propose a two-level model, which at the aggregate level plans capacity and at the detailed level uses a base stock approach to coordinate the various stages of production. Bitran and Tirupati (1993) and Meal, Wachter and Whybark (1987) discuss the relationship between the BHM formulation of hierarchical planning and MRP. They conclude that the hierarchical planning and MRP systems are complimentary, in that hierarchical planning focuses primarily on determining capacity levels and capacity smoothing, while MRP determines the amount of material required at various points in the manufacturing process. Due to the interactions between material release and resource release decisions as formulated in the SCOP models to be discussed in the following sections, we can argue that this argument cannot be extended to supply chain operations planning, and in fact can only be upheld if the time periods considered in BHM's hierarchical planning model are an order of magnitude longer than the time periods in the MRP model and if resources are flexible.

Decisions with regard to the planning of supply chain operations have traditionally been taken at the operational level. Meal (1984) argues that this was necessarily decentralized due to the lack of good information processing technology. In this approach, which he names the 'conventional approach', operations planning decisions were an integral part of the decision making power of the line managers in all parts and at all levels in the organization. Decisions were only coordinated marginally, and certainly not in a systematic manner. Due to the emergence of large-scale information processing technology in the 1970s, initiatives were taken to create large-scale comprehensive models of planning operations. Meal (1984) calls this the 'centralized approach', which is based on a tendency to create central decision functions which are given the power to control in detail the planning decisions of the operational process in all parts of the organization.

There are a number of difficulties associated with these centralized monolithic decision models [see also Chapter 9 of this volume, Fleischmann and Meyr (2003)]. The models tend to be very big and complex. This makes the analysis of the models and finding an optimal solution very difficult and requires a decomposition of the model in order to be able to solve this. Model decomposition is a widely used strategy in solving optimization problems. Apart from the complexity in the mathematical sense, there are also a number of organizational and people-related difficulties associated with the centralized approach. The most important difficulty is that there appears to be no owner of the monolithic model. Responsibilities within organizations tend to be dispersed over a number of people. The monolithic model assumes it is a single organizational unit deciding about a large number of details across the entire organization. If we assume that the higher-level management would actually own the model and make these decisions, a number of people and model related difficulties come about:

(1) detailed figures do not mean much to higher-level managers
(2) detailed figures give a false sense of security because they may be highly unreliable, not only if they refer to some future state of the system (e.g., forecast of exogenous data), but also if they refer to the current state of the system (data quality problems)
(3) centralized planning takes away authority from local managers further down the hierarchy and reduces their responsibility, which is not in line with the dominant management philosophy of self-contained and autonomous groups. Apart from that, it is also contradictory to a principle from control theory, which states that responsibility and decision authority should be matched with the opportunity to control. This last issue is extensively discussed by McPherson and White (1994), who state that 'Planning at superior levels must be consistent with control capabilities at subordinate levels, while planning at subordinate levels must be consistent with achieving the superior goals of the hierarchy.'
(4) A model never captures the complete richness of a situation. As a consequence, a local planner down the hierarchy will always have more information and a better representation of the actual processes than a (higher-level) model.

All this leads to the fact that a decomposition of the problem is required in order to be able to find a solution to the planning problem that can also be implemented within an organization. If a decision problem is decomposed and a hierarchy is constructed, higher levels of the hierarchy will need to aggregate the lower level models. This aggregation is necessary to overcome the difficulties just listed. Furthermore, this decomposition will lead to more or less independent units along the supply chain, that are self-contained with regard to their control within the unit, but receive objectives and constraints to be taken into account from an aggregate and centralized control function. This is in line with the idea of separating goods flow control and production unit control, as developed by Bertrand and Wortmann (1981) and further elaborated on by Bertrand et al. (1990). A consequence of this approach is that lead times of the various production units are fixed and are input to the system rather than output. These lead times are then essentially modeled in exactly the same way as in MRP [Orlicky (1975)]. We will discuss this issue further in Section 2.4. Note that the fixed lead time we are discussing here is the internal lead time of the controlled part of the supply chain that needs to be distinguished from the external lead time promised to any customers of this supply chain. The external lead time must vary to reflect the work load changes over time. As a consequence of this approach, workload control is executed over the supply chain. In summary, we can state that hierarchical decomposition of the SCP problem has two essential characteristics, namely:

– aggregation, which is necessary to construct higher level models
– fixed lead times, which are needed as a control mechanism.

In the next two subsections we will discuss the concepts of effectuation lead times and information asymmetry which underlie the notion of obtaining supply chain control by working with planned lead times. This notion of control will be further elaborated on in Section 2.4.

2.2 Effectuation lead times

Asymmetry in the decision making hierarchy and the necessity to anticipate is primarily caused by the fact that it takes time to implement a decision. We will denote this time in the remainder of this chapter as *effectuation lead time*. The effectuation lead time is the time that passes between the moment a decision is made and the moment that the consequences of this decision can be observed in the operation of the supply chain. In SCP decisions, the length of the effectuation lead times can be determined based on the product and process structure: the bill of material and the bill of resources.

An example of an effectuation lead time is the procurement time of components. If the procurement lead time for a component i is L_i, then $r_i(t)$, the quantity procured at the start of period t is supposed to be available for further assembly or sales at the start of period $t+L_i$. The immediate decisions $\{r_i, (t)\}$ are dependent on the exogenous demand forecasts $\left\{\hat{D}_i(t, t+s)\right\}$, $s \geq 0$, defined as

$\hat{D}_t(t, t+s)$ forecast of exogenous demand for item i in period $t+s$ as decided on at the start of period t, $t \geq 0$, $s \geq 0$, $\forall i$

Assuming supply is reliable and L_i is realized, we may expect that

$$\hat{p}_i(t, t+L_i) = r_i(t),$$

where we define $\hat{p}_j(t, t+s)$ as

$\hat{p}_j(t, t+s)$ forecast of quantity of item i that becomes available at the start of period $t+s$ as determined at the start of period t, $t \geq 0$, $s \geq 0$, $\forall i$

reflecting the decision of the supplier to ship as late as possible. Note that $\hat{p}_i(t, t+L_i)$ is only a planned decision from the perspective of the organization ordering the item. For the supplier this may be either a firm decision, in case the supplier has to start immediately with processing and transporting the order for item i, or a planned decision, in case the effectuation lead time incorporates some slack time.

Suppose that at time t the planned decision is taken according to the above equation. At the start of period $t+1$ we generate new forecasts $\{\hat{D}_i(t+1, t+s)\}, s \geq 1$. It is now well possible that, e.g., due to a decrease in demand for item i it is decided to change the decision made earlier, i.e.,

$$\hat{p}_i(t+1, t+L_i) \neq \hat{p}_i(t, t+L_i)$$

Following the discussion of planned versus firm decisions, we can see that dependent on the incorporation of slack time into the procurement time, it is possible or not to change the earlier decision. Flexibility can be further modelled by the choice of the (time-dependent) resource constraints.

Information asymmetry itself can be described using the following example. Consider again item i with planned lead time, i.e., the effectuation lead time of the material order release decision, L_i. Often suppliers receive forecasts about future orders in some period t multiple times in order to take consecutive decisions on e.g., buying production equipment, hiring and training people and procurement of materials. As stated above the forecasts $\{\hat{D}_i(t-s, t)\}$ for period t made at the start of an earlier period t-s differ for different s. Thus the procurement orders derived from these forecasts change over time, so that the supplier's decision to buy production equipment is based on different information then the supplier's decision to procure materials. This asymmetry in information needs to be taken into account when designing the decision hierarchy or supply chain control structure. The effectuation lead time thus leads to differences between the moments in time that certain decisions must be taken. Also, it means that decisions are often taken a substantial time before the actual action in the physical process takes place. As a consequence the decision maker in fact feeds forward in terms of control theory rather than feeds back as is often suggested in hierarchical production planning frameworks. In order to feed forward, the decision maker essentially anticipates the events over the period of time until his decision is effectuated.

It should be realized that the effectuation lead time is not only related to the bills of materials and bills of resources, but is also a characteristic of the SCP and control system. In many cases, the time buckets at higher levels of decision making are larger than at lower levels (Meal (1984)). Further, the frequency at which decisions are made, revised or processed is less at higher levels of decision making (the hierarchical structures of Gershwin (1994) are based upon this premise). This means that changes in the actual (physical) process, e.g., changes in demand, may not be observed directly. Further, if they are observed, there may be a delay in processing the consequences of this observation. This processing time due to the decreased frequency of decision making at higher levels should be included in the effectuation lead time.

2.3 Asymmetry in SCP

After constructing a hierarchical planning structure, oftentimes the resulting planning situation is characterized by asymmetry of information. Essentially having different levels of control being owned by different organizational units leads to different information statuses.

A useful framework that describes this anticipatory decision cycle, is presented by Schneeweiss (1999) and discussed by Fleischmann and Meyr (2003). In Schneeweiss's model, a decision structure within an organization can be represented as a series of decision tandems, i.e., two decision levels interacting with each other by the first of the two levels (the top level) giving an instruction to the second of the two levels (the base level), and the base level responding by giving a reaction to the top level. Before giving its instruction, the top level anticipates the base level's reaction by either implicitly or explicitly modeling the behavior of the base level in the top level's model. This is called the anticipated base model. Schalla, Fransoo and de Kok (2001) have further analyzed the various types of anticipation that may exist. In general, the anticipated base model can be constructed based on aggregating information and/or on aggregating the base level model itself.

We will now first discuss the aggregation of the base model itself. Aggregation referring to the model part explicitly deals with complexity reduction. At the top level the decision making process is represented by an aggregate and simple model in order to reduce complexity and to distribute detailed decisions to lower planning levels. We can thus distinguish the following anticipation types with regard to the model:

- *Explicit Model*: The base level model as seen at the top level is exactly the same as the original base level model.
- *Implicit Model*: The base level model as seen at the top level is different than the original base level model.

Consequently, the terms explicit and implicit with regard to anticipation refer to the fact whether the top-level base model (including the objective functions) is exactly the same as the base-level base model. If this is the case, we call this explicit anticipation; if this is not the case, we call this implicit anticipation. Explicit anticipation thus uses a detailed model of the base level, whereas implicit anticipation uses an aggregate model of the base level.

The second type of aggregation to construct an anticipation model is aggregation of information. This type of aggregation is related to uncertainty and effectuation time. In the context of the questions related to the anticipation function, special attention regarding information is paid to the concept of information asymmetry. Information asymmetry basically entails the fact that when making a decision at a higher level, the amount and quality of information may be different from when the lower level decision is made (later), and again different from when the actual execution of the decision is taking place. The fact that information asymmetry exists, leads to the necessity to *anticipate* at a higher level decision what *may* happen at the lower levels decisions. We can thus distinguish the following anticipation types with regard to information:

- *Exact Information*: The base level information as seen at the top level is exactly the same as the original base level information.
- *Approximate Information*: The base level information as seen in the top level is different than the original base level information.

Consequently, the terms exact and approximate refer to the fact whether the top level model has exact information of the base level status. Note that in most cases some time elapses between the moment at which the top level makes its decision (instruction) and the moment at which the base level makes its (final) decision. This difference in *effectuation lead times of top level decisions and base level decisions* usually entails a difference in the information status between the top level and the base level, resulting in information asymmetry and – automatically – in approximate anticipation.

Taking all combinations between anticipation types referring to the modeling part and anticipation types referring to the information part, we can distinguish three types of anticipation, based on the various types of aggregation used to construct the anticipation function as discussed above. The three types are:

- Explicit Model l Exact Information (EE)
- Explicit Model 1 Approximate Information (EA)
- Implicit Model l Approximate Information (IA)

Note that the combination of exact information and implicit model does not make a lot of sense, since there does not seem to be a clear reason for constructing an implicit model (i.e., more aggregate than the detailed model) if exact information is available. Further note that information asymmetry more often than not will lead to the fact that approximate information is the only information that can be used in the anticipatory model at the higher level. Given the fact that only approximate information can be used, it is not a priori clear whether the use of an explicit and detailed model is better than the use of an implicit model.

It is interesting to note that the BHM hierarchical models hardly contain any anticipation of the lower levels by the higher level, as has been noted by Schneeweiss (1999). Neither do the BHM models contain the concept of effectuation lead time, although this concept is noted as a rationale for hierarchical planning in a book largely built on the BHM models (Miller (2001), p. 8, named as 'gestation period').

2.4 The need for control

Anticipation models need to capture the base level behavior in a sufficiently accurate manner. In this sense, accurate refers to the predictive quality of the anticipation model. When designing a decision structure, two different approaches can be taken when constructing the anticipation functions. The first approach is to try and capture the base level behavior as completely as possible by enriching the anticipation function by as many details as known about the base level. The second approach is to design the decision function at the base level in such a way that the actual anticipation becomes straightforward. In this case, the objective of the base level is to realize a set of targets set by the top level (see also McPherson and White (1994), for a discussion on this matter). We will refer to this situation as a controlled situation. An example of such a design is the reliance on planned lead times maintained by workload control methods (Bertrand and Wortmann (1981), Bertrand et al. (1990), Wiendahl (1987, 1995) and Van Ooijen (1991)).

It is neither obvious nor conclusive whether working with planned lead times is a correct approach, since current research is not conclusive and there are researchers who advocate the use of variable lead times. Kanet and Sridharan (1998) demonstrate results by which the use of detailed scheduling information in material procurement reduces the inventory of components that are controlled by MRP. Tardiff (1995) and Hopp and Spearman (2000) in their concept called Capacitated MRP (or MRP-C) calculate the expected Work-in-Process and then adjust the lead times of the products accordingly, by 'building ahead' those items that would be late due to longer lead times caused by higher WIP levels. Based on work by Buzacott (1989), a research line has been developed which integrates the capacity and material planning perspectives completely using so-called 'generalized kanban systems' [Frein, Di Mascola and Dallery (1995), see also Section 5.8]. The assumptions are however very strict, such as Poisson arrivals of items orders and FIFO dispatching at resource level. These assumptions are mostly not satisfied due to the periodic review nature of the SCP process, which leads to coordinated release decisions.

Apart from the mathematical complexity of applying detailed scheduling on a supply chain wide scale, approaches that use detailed scheduling information to update the supply chain plan abstain from two basic principles that we have outlined earlier, namely the organizational hierarchical concerns and the asymmetry in information. Since the scheduling decision is generally the domain of some lower-level organizational function than the SCP decision, taking this scheduling decision at a higher level may infringe upon this organizational design [Meal (1984)]. With regard to information asymmetry, note that the actual schedule will be constructed at a later stage when more information will be available. As a consequence, the actual schedule may be very different from the projected detailed schedule constructed to make the supply chain plan. In fact the more detailed scheduling of materials will then lead to additional constraints on the operational schedule. This issue is not addressed in the paper by Kanet and Sridharan (1998) discussed above. Given only slight variations in for example the operating times, the impact on the schedule in various environments may be substantial [see e.g., Lawrence (1997), for an example of this in a job shop]. Unfortunately, this interaction between the SCP level and the detailed scheduling level has not been researched under asymmetric information conditions, so we do not know the impact of this effect on the operational performance.

With regard to dynamically adjusting the information based on aggregate status (workload) information of the shopfloor, the impact is even less clear. Much will depend upon the actual stability of the workload prediction between the moment that the SCP decision is taken and the moment the actual execution takes place, i.e., the quality of the *expected* workload as an anticipator of the *actual* lead time. If this quality is good, then the method should work fairly well in the manufacturing environments for which it was designed. In situation with multiple items and multiple resources it is however very difficult to give accurate predictions of the lead time and to adjust the supply chain plan accordingly. This will lead to very complex models and will lead to performance and accuracy problems [Hopp and Spearman (2000)]. The situation is however very different when multiple actors conduct collaborative planning actions across independent companies in the supply chain. In that case, the planned lead times act as a coordination aid between the various actors in the supply chain and planned lead times allow for independent planning of parts of the supply chain.

2.5 Positioning SCOP in the hierarchy

From the exposition above, it can be concluded that SCOP needs to be positioned hierarchically above the unit control functions that are responsible for controlling lead time in a particular unit of the supply chain [cf. Bertrand et al. (1990)]. Supply Chain Operations Planning in most industries deals with a horizon up to several months typically with weekly time buckets. In some industries, e.g., bulk chemicals, this function may have a horizon as short as a couple of weeks with daily buckets, whereas in other industries, e.g., pharmaceuticals, the horizon may be as long as a couple of years with monthly time buckets. Everything is determined by the typical effectuation lead times of the industry and the lead times that customers are willing to accept. Next to the SCOP function, an order acceptance function (often called Available-To-Promise engine in current planning software) needs to be introduced in the control loop in order to control the total amount of work accepted by the supply chain, and to externalise the portion of the customer-perceived leadtime that is due to varying demand that cannot be processed within the fixed and controlled leadtime. Finally, a parameter setting function

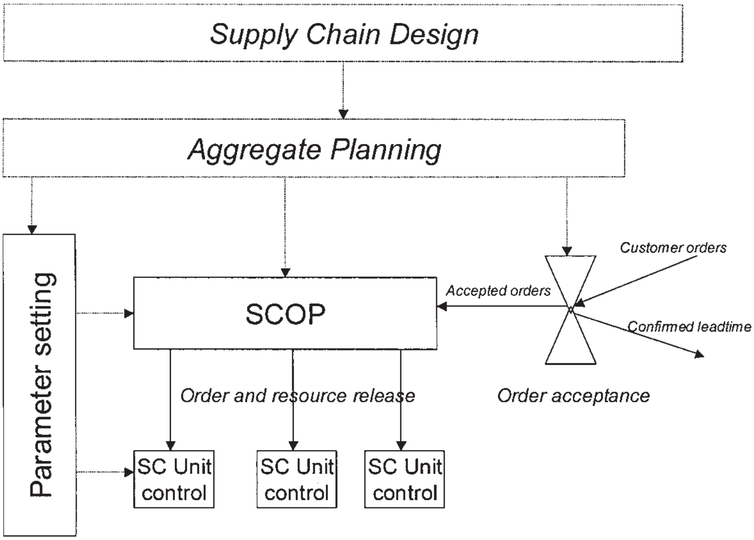

Fig. 1. Position of Supply Chain Operations Planning in the Planning Hierarchy.

needs to coordinate the safety stock, leadtime, and workload parameters of the Supply Chain. This system is depicted in Fig. 1.

Note that the functions discussed here only relate to the planning of operations, i.e., the release of materials and resource triggered by actual demand downstream. In this discussion, we abstain from other functions, such as supply chain design, the planning of seasonal or other controlled inventories, lotsizing, transportation planning, etc. For a full description of this hierarchy, we refer to Fleischmann and Meyr (2003). Along a timeline, it clearly shows that the various SCOP decisions need to be timed along the lead time characteristics of the supply chain, both the physical lead time and the information processing lead time. This is depicted in Fig. 2 and was discussed in detail in Section 2.2.

In the next section, we will further model the SCOP problem in detail, and formulate the constraints that determine the SCOP problem.

3 Constraints for SCOP

In this section we propose a modeling framework that comprises currently existing supply chain operations planning concepts. In Section 3.1 we derive the set of constraints that follow from the material structure (BOM). In Section 3.2 we derive the set of constraints that follow from the resource structure (BOP). We emphasize here that the constraints derived are induced by the underlying BOM and BOP and hold for any choice with respect to the SCOP concept. It may be that one chooses to ignore particular constraints as

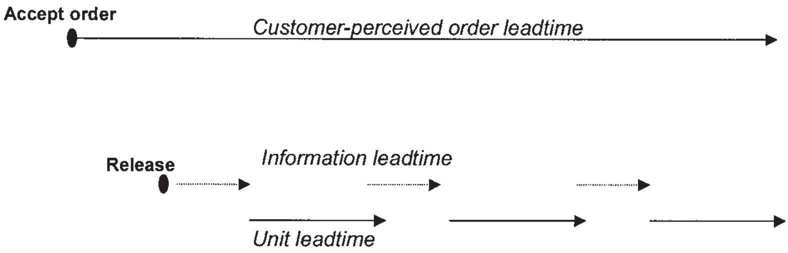

Fig. 2. Decision moments driven by leadtime structure.

irrelevant, implying that the result of the SCP process may be infeasible. The decision to ignore particular constraints should imply that either the constraints are never binding (which one might be able to prove) or the impact of the resulting infeasibility is negligible (which is likely to be much harder to prove).

3.1 Material constraints and their representation

The definitions in Section 1.1 characterize completely the material structure of the supply chain. We now exploit this structure to identify a set of constraints that any SCOP concept should satisfy. These constraints can informally be described as follows: any SCOP concept can only release items for usage in a transformation activity if it is physically available at the moment of release. This might seem an obvious statement, but the pitfall lies in the definition of 'release'. In this chapter a release decision at the start of period t is a decision that authorizes at the start of period t the usage of materials and resources for a transformation process. The release decision is the core decision in Materials. Requirements Planning (MRP-I), Manufacturing Resource Planning (MRP-II), Statistical Inventory Control (SIC) and Just-In-Time (JIT) and any other planning and control concept. It is assumed that such release decisions are coordinated, since they are mutually dependent. In the concepts mentioned above the release decision is taken without a check on availability of resources and materials. For example, in SIC [Statistical Inventory Control, cf. Silver et al. (1998)] a release decision is only based on the output item status (in this case its inventory position) associated with a transformation process and not with the input item(s) status. This implies that a release decision cannot be executed at the time it is authorized to do so. Hence execution delays occur that should be taken into account when coordinating decisions. Similarly, the top–down explosion process of MRP-I does not incorporate item availability checks. Only exception messages are generated, but the explosion process implicitly assumes that infeasibilities at lower levels in the BOM are resolved. It can be argued that such a fallacy has a major impact in situations where lots of items are assembled into multiple other items, which explains the current practice of human expediting in high volume consumer electronics supply chains.

Since SCOP coordinates release decisions between different parts of an organization or even between different organizations, typically such release decisions are taken periodically. In between the periodic release decisions preparatory process planning activities take place, such as demand forecasting, checking availability of resources, etc. In the sequel we therefore restrict to periodic SCOP concepts. Where appropriate, we discuss extensions to continuous SCP concepts.

Let us now formally derive the material release constraints. Given the definition of the physical inventory $I_i(t)$ and the backlog $B_i(t)$ it is clear that

$$I_i(t), B_i(t) \geq 0, \qquad t \geq 1, \forall i$$

It is obvious that a backlog exists iff. the physical stock is zero, i.e.,

$$I_i(t) B_i(t) = 0, \qquad t \geq 1, \forall i$$

The net inventory is defined as the physical inventory minus the backlog, i.e.,

$$J_i(t) = I_i(t) - B_i(t), \qquad t \geq 1, \forall i$$

Independent demand is demand generated by customers of the supply network, either directly at the CODP or indirectly by derivation from final product forecasts and an offset by the lead time for final assembly. Such demand is usually not known beforehand and must be forecast. Note that $E \subset P$, since end-items have independent demand, only. Yet there may be items in I that have independent demand, too. For example, a company producing hard disks may sell to OEM as well as to individual customers. The product sold to individual customers contains the product sold to the OEM as a subassembly. The subassembly is an item in *I*, while the OEM demand for the subassembly is independent.

For items with independent demand it is impossible to preclude backorders, unless some upper bound on the demand per period is known. In this chapter we assume that such an upper bound does not exist or the upper bound is so high compared to the average demand per period that it would be economically infeasible to guarantee no backordering of exogenous demand for an item. However, for items with dependent demand by definition the demand is known, since it is determined within the planning process itself. We argue that backordering of dependent demand does not make sense. Suppose at some period we take the decision to create a backorder for an item by deciding to release more material then available. In that case we only release all available material physically. The earliest moment in time that we may resolve the backordering situation is at the start of the next period. However, at the start of the next period we have exact information about demand during the current period and possibly better information about future demand. Thus, it is easy to see that the decision taken to create a (logical) backorder by releasing more material than available cannot be better than the decision to release exactly all available material.

Our argument only holds for SCOP concepts where all items in the supply chain 'know' information about future exogenous demand, either implicitly or explicitly. This applies to all SCOP concepts that subsume some centralized database with all current state information and forecast information. Top–down SCOP concepts like SIC and MRP-I use the explosion process to transfer exogenous demand information at the expense of incorrect order release decisions, that cannot be executed and require human intervention to resolve the resulting issues.

The above implies that we impose the following constraint on the evolution of the backorders over time, which holds for all items.

$$B_i(t+1) - B_i(t) \leq D_i(t), \qquad \forall i, t \geq 1. \tag{3.1}$$

The above equation states that the increase of the backlog cannot exceed the exogenous demand. It is easy to see that for an intermediate item i with $D_i(t) = 0$ for all $t \geq 1$,i.e. intermediate item i has no independent demand, we have that

$$B_i(1) = 0 \Rightarrow B_i(t) = 0, \qquad t \geq 1.$$

Dependent demand $G_i(t)$ for item i is generated by items in V_i. The dependent demand for item i at the start of period t consists of the sum of all released quantities of items in V_i at the start of period t. This implies that

$$G_i(t) = \sum_{j \in V_i} a_{ij} r_j(t), \qquad \forall i \in I$$

Clearly there must be sufficient inventory of item i to ensure immediate start of the execution of the transformation activities involved in the release decisions. The physical starting inventory at the beginning of period t equals $p_i(t) + \max(0, I_i(t) - B_i(t))$. Thus it follows that $G_i(t)$ must satisfy the following equation,

$$G_i(t) \leq p_i(t) + \max(0, I_i(t) - B_i(t)), \qquad \forall i, t = 1, \ldots, T$$

The set of Equations (3.1) states that the backlog from a period to the next period must not grow faster than the exogenous demand in that period, while the above set of equations states that a planning concept must not release more than physically available. The lemma below shows that both sets of equations are equivalent.

Lemma 1.

$$G_i(t) \leq p_i(t) + \max(0, I_i(t) - B_i(t)) \Leftrightarrow B_i(t+1) - B_i(t) \leq D_i(t)$$

The proof of the above Lemma 1 is straightforward and is derived from the inventory balance equations formulated below and the definition of the variables involved.

Furthermore we assume that all released quantities are non-negative, i.e., returns are not possible,

$$r_i(t) \geq 0, \qquad \forall i, t = 1, \ldots, T \tag{3.2}$$

This implies that the released quantities together constitute a feasible plan. Note that Equations (3.1) only state that one is not allowed to release more than physically available. One might decide to reserve availability for exogenous demand in case $i \in P$.This aspect is dependent on the SCOP concept and will be discussed in Section 4 and following.

It can easily be shown that the MRP/DRP-concept [cf. Silver et al. (1998)] as a planning concept does not satisfy constraints (3.1) (see above discussion of top–down planning logic). In Section 4 we discuss mathematical programming models that can be seen as an extension of the MRP/DRP-concept in that such models incorporate the feasibility constraints at the expense of (much) more computational effort.

Given the release decisions taken we can write the so-called *inventory balance* equations,

$$J_i(t+1) = J_i(t) - G_i(t) - D_i(t) + p_i(t), \qquad \forall i, t = 1, \ldots, T. \tag{3.3}$$

Using the definition of the net inventory we equivalently can write

$$I_i(t+1) - B_i(t+1) = I_i(t) - B_i(t) - G_i(t) - D_i(t) + p_i(t), \quad \forall i, t = 1, \ldots, T \tag{3.3'}$$

The dynamics of $I_i(t)$ and $B_i(t)$ determine the performance of the supply network. We want to emphasize here that the impact of the release decisions taken in the past as well as the impact of the planned lead time of item i is 'accumulated' in $p_i(t)$. This will be discussed in detail in Sections 3.2 and 3.3.

Summarizing, we have defined the material structure of a supply network by defining *gozinto* relations between items. From those relations we derived a set of (release) constraints that should be satisfied by any supply chain operations planning concept. When discussing the various supply chain operations planning concepts we pay attention to their adherence to these constraints.

3.2 Resource constraints and their representation

In this subsection we derive a set of necessary conditions with respect to capacity usage that any SCOP concept should satisfy, given the above information on resource availability and resource usage. It turns out that the derivation of such constraints is not as straightforward as the material constraints derived in the previous section. This can be explained as follows.

Given the released quantities $r_i(t)$ and the lead time L_i for orders of item i, we can formulate the following set of constraints,

$$\sum_{i \in U_k} c_i r_i(t) \leq C_{kt} + L_{i-1}, \qquad k = 1, \ldots, K, t \geq 1,$$

implying that the total capacity requirements for resource k associated with the item orders released at the beginning of period t should not exceed the available capacity of resource k during period $t+L_i-1$. These conditions are sufficient, yet not necessary, for ensuring that orders for item i released at the start of period t are made available for usage at the start of period $t+L_i$. The constraints above are necessary, only, if we require that items released at the start of period t are processed in period $t+L_i-1$. This is equivalent to the decision rule to produce as late as possible with respect to the lead time L_i. In general it follows from the planned lead time L_i of an order for item i, that the item order released at the start of period t, i.e., the time interval $(t-1, t]$, must be processed on its associated resources in the time interval $(t-1, t+L_i-1]$ in order to guarantee availability for usage at the start of period $t+L_i$. Since material required for processing orders for item i must be released earlier it follows that

$$\sum_{s=1}^{t} r_i(s) \geq \sum_{s=1}^{t} q_i(s). \tag{3.4a}$$

The right hand side of (3.4a) denotes the cumulative amount of item i processed up to and including period t. The left hand side of (3.4a) denotes the cumulative amount of item i released up to and including period t. Here we assume without loss of generality that at the start of period 1 the system is empty, i.e., no orders are released before time 0, no stocks are available.

To ensure that the order released at the start of period t is available for usage in period $t+L_i$ we must process the materials associated with $r_i(t)$ in the periods $t,\ldots,t+L_i-1$. From this it follows that

$$\sum_{s=1}^{t} r_i(s) \leq \sum_{s=1}^{i+L_i-1} q_i(s). \tag{3.4a}$$

From the definition of $q_i(t)$ we find that

$$\sum_{i\in U_k} c_i q_i(t) \leq C_{kt}. \tag{3.4b}$$

Combining the above we find

$$\sum_{s=1}^{t}\sum_{i\in U_k} c_i r_i(s) \leq \sum_{s=1}^{t+L_i-1} C_{ks}, \qquad k=1,\ldots,K, t\geq 1. \tag{3.5}$$

The necessity of condition (3.5′) is obvious. The sufficiency follows from the fact that we may assume a FIFO allocation of capacity from which we can construct a feasible allocation by allocating orders released as soon as possible to the resources. In a rolling schedule context we can rewrite the necessary and sufficient conditions so that capacity consumed before period t is subtracted from the left-hand side of (3.5′), and capacity available before period t is subtracted from the right-hand side of (3.5′). In the sequel we will use equations (3.4a), (3.4b) and (3.5) since they explicitly relate material release quantities $\{r_i(t)\}$ and material processing quantities $\{q_i\ (t)\}$. The latter quantities provide useful information about capacity usage.

Extension to the situation where item i can be processed by multiple resources implies the definition of variables that indicate which amount of the order released at the start of period t is processed by a specific resource. For our further comparison of the different supply chain planning concepts we can restrict to the situation where each item is processed at a single resource.

3.3 Planned lead times and the relationship between $\{r_i(t)\}$, $\{q_i(t)\}$ and $\{p_i(t)\}$

In the inventory balance Equations (3.3′) we use the variables $\{p_i(t)\}$ that denote the amounts of material that become available for usage at the start of period t. Clearly, these variables are related to the material release quantities $\{r_i(t)\}$ and the material process quantities $\{q_i(t)\}$, since these are decisions that have to be taken before amounts can be made available for usage. It turns out that we have a considerable degree of freedom here. As stated in Section 1.3 we have that

$$p_i(t) = q_i(t-1), \qquad t \geq 0,$$

which implies that we assume that an amount of item i processed in period $t-1$ becomes available at the start of period t. This provides maximum flexibility within the decision space bounded by constraints (3.4a), (3.4b) and (3.5). However, this may imply that materials are available *earlier* than planned for, according to the moments of order release and the planned lead times. This may be seen as favorable, yet this may imply that materials are available *too early*. We must be aware of the fact that the SCOP model is a representation of part of reality only, implying that materials not modeled at SCOP level are only available at the due dates derived from the planned lead times. And even if it would represent all materials, than still we are faced with uncertainty of processing and future demand.

The concept of planned lead times is a means to create certainty about future material availability. In our view we should formulate the SCOP constraints such that they reflect the conceptual ideas behind the planned lead times concept. This is ensured by defining $\{p_i(t)\}$ as

$$p_i(t) = r_i(t - L), \qquad t \geq 0$$

Assuming that planned lead times are realistic and thereby due dates are met with high probability by the shopfloor level, this definition is in line with the constraints (3.4a) and (3.4b). Notice that this is the typical assumption when we consider uncapacitated systems as in classical inventory management theory [cf. Silver et al. (1998)].

By the above definition of $\{p_i\ (t)\}$ we reformulate the inventory balance equations as follows,

$$I_i(t+1) - B_i(t+1) = I_i(t) - B_i(t) - G_i(t) - D_i(t) + r_i(t - L_i), \quad \forall i, t = 1, \ldots, T \tag{3.3}$$

3.4 Summary

In this section we defined the supply chain operations planning problem in detail and derived necessary and sufficient material and resource constraints:

Necessary and sufficient material constraints

$$B_i(t+1) - B_i(t) \leq D_i(t), \qquad \forall i, t \geq 1 \tag{3.1}$$

$$r_i(t) \geq 0, \qquad \forall i, t = 1, \ldots, T \tag{3.2}$$

$$I_i(t+1) - B_i(t+1) = I_i(t) - B_i(t) - G_i(t) - D_i(t) + r_i(t - L_i), \quad \forall i, t = 1, \ldots, T \tag{3.3}$$

Necessary and sufficient resource constraints

$$\sum_{s=1}^{t} r_i(s) \geq \sum_{s=1}^{t} q_i(s), \qquad \forall i, t = 1, \ldots, T \tag{3.4a}$$

$$\sum_{s=1}^{t} r_i(s) \leq \sum_{s=1}^{t+L_i-1} q_i(s), \qquad t = 1, \ldots, T \tag{3.4b}$$

$$\sum_{i \in U_k} c_i q_i(t) \leq C_{kt}, \forall k, \qquad t = 1, \ldots, T \tag{3.5}$$

$$q_i(t) \geq 0, \qquad \forall i, t = 1, \ldots, T \tag{3.6}$$

Conditions (3.4a), (3.4b) and (3.5) have been derived for the special case where each item can only be processed on a single resource. Still, the above-defined constraints provide a basis for comparison of different supply chain concepts. In particular they enable to identify what assumptions are made on material and resource availability and usage.

In the next section we use the generic SCOP constraints to develop SCOP concepts based on the optimization of a deterministic SCOP model in the rolling schedule framework.

4 Mathematical programming models for supply chain planning

In this section we derive the basic mathematical programming formulation of the supply chain operations planning model in a rolling schedule context. We use the product structure and the resource constraints presented in Section 3 as a representation of the primary process to be planned. Special attention is paid to the fact that exogenous demand must be forecast in order to derive a sensible problem formulation. We show how the generic supply chain planning constraints derived in Section 3 are incorporated into the mathematical programming formulation. Since the SCOP problem is a stochastic problem by nature we next address the issue of safety stocks, i.e., buffer stocks required to cope with end-item demand uncertainty. In particular we present a theorem that enables us to apply mathematical programming models in a rolling schedule context in such a way that the long-run average costs to maintain customer service level constraints can be determined. After that we give an overview of the main contributions to the literature on mathematical programming models for supply chain planning over the last 10 years.

4.1 Rolling schedule context

In reality a planning concept does not only generate immediate release decisions, but also provides information on future release decisions. These future release decisions are provisional because they will be affected by future unknown events. In particular we do not know future demand. To cope with this we must forecast future demand. Incorporating the fact that we must forecast future demand we reformulate the generic supply chain planning constraints. Towards this end we define the following variables,

$\hat{D}_i(t, t+s)$	exogenous demand for item i in period $t+s$ as determined at the start of period $t, t \geq 1, s \geq -t, \forall i$
$\hat{G}_i(t, t+s)$	endogenous demand for item i in period $t+s$ as determined at the start of period $t, t \geq 1, s \geq -t, \forall i$
$\hat{B}_i(t, t+s)$	backlog of item i at the start of period $t+s$ as determined at the start of period $t, t \geq 1, s \geq -t, \forall i$
$\hat{r}_i(t, t+s)$	quantity of item i released at the start of period $t+s$ as determined at the start of period $t, t \geq 1, s \geq -t, \forall i$
$\hat{q}_i(t, t+s)$	quantity of item i processed in period $t+s$ as determined at the start of period $t, t \geq 1, s \geq -t, \forall i$

Note that for $-t < s < 0$, these variables represent actuals, i.e., realized quantities taken from historical data and exogenous to the problem. For $s \geq 0$, these variables represent forecasts, i.e., estimates of future quantities, made at the start of period t. Note that $\hat{r}_i(t, t+s)$, $\hat{q}_i(t, t+s)$, for $s \geq 0$, are the decision variables of the supply chain operations planning problem at the start of period t. In the sequel we assume that there is a time origin 0 at which the SCOP problem is solved first and the initial state of the system at time 0 is known. This is important because some equations formulated below are formulated in terms of decisions taken from time 0 onwards, i.e., from period 1 onwards.

4.2 LP formulation of the supply chain planning problem

In Section 3 we have derived a set of material and resource constraints that each supply chain operations planning concept should satisfy in order to generate feasible solutions with regard to the current demand forecast. This suggests that we want the decision variables $\hat{r}_i(t, t+s)$, $\hat{q}_i(t, t+s)$, for $s \geq 0$, to satisfy these generic supply chain operations planning constraints. Then the planning problem to be solved at each time t must satisfy the following equations.

LP constraints

$$
\begin{aligned}
\hat{I}_i(t, t+s+1) - \hat{B}_i(t, t+s+1) &= \hat{I}_i(t, t+s) - \hat{B}_i(t, t+s) \\
&\quad - \sum_{j=1}^{N} a_{ij}\hat{r}_j(t, t+s) - \hat{D}_i(t, t+s) + \hat{r}_i(t, t+s-L_i), \\
&\quad \forall i, s = 0, \ldots, T-1
\end{aligned}
$$

$$\hat{B}_i(t, t+s+1) - \hat{B}_i(t, t+s) \leq \hat{D}_i(t, t+s), \qquad \forall i, s = 0, \ldots, T-1$$

$$\sum_{w=1-t}^{s} \hat{r}_i(t, t+w) \geq \sum_{w=1-t}^{s} \hat{q}_i(t, t+w), \qquad \forall i, s = 0, \ldots, T-1$$

$$\sum_{w=1-t}^{s} \hat{r}_i(t,t+w) \leq \sum_{w=1-t}^{s+L_i-1} \hat{q}_i(t,t+w), \qquad \forall i, s = 0, \ldots, T-1$$

$$\sum_{i \in U_k} c_i \hat{q}_i(t, t+s) \leq C_{k,L+s}, \quad k = 1, \ldots, K, \quad s = 0, \ldots, T-1.$$

$$\hat{r}_i(t, t+s) \geq 0, \qquad \forall i, s = 0, \ldots, T-1,$$

$$\hat{q}_i(t, t+s) \geq 0, \qquad \forall i, s = 0, \ldots, T-1,$$

$$\hat{I}_i(t, t+s) \geq 0, \qquad \forall i, s = 0, \ldots, T-1,$$

$$\hat{B}_i(t, t+s) \geq 0, \qquad \forall i, s = 0, \ldots, T-1.$$

As remarked above the model formulation includes decision variables with decisions taken before period t. Obviously we have that

$$\hat{r}_i(t, t+s) = r_i(t+s), \qquad s < 0, \forall i, t \geq 1,$$
$$\hat{q}_i(t, t+s) = q_i(t+s), \qquad s < 0, \forall i, t \geq 1.$$

The decisions implemented at the start of period t are given by $\{r_i\ (t)\}$ and $\{q_i\ (t)\}$, which are derived from the equations below.

$$r_i(t) = \hat{r}(t, t), \qquad \forall i, t \geq 1,$$
$$q_i(t) = \hat{q}(t, t), \qquad \forall i, t \geq 1.$$

In the sequel we assume that the planning decisions ($\{r_i(t)\}$, $\{q_i(t)\}$) are executed according to plan.

The above set of linear equations constitutes the basis for the formulation of a mathematical programming model. In fact we can formulate an LP model that can be solved by standard algorithms, such as the simplex method. The LP-formulation requires the definition of a linear objective function. Let us discuss the derivation of such a linear objective function in the context of supply chain planning under stochastic exogenous demand. In the sequel we assume that $P = E$, i.e., only end-items have exogenous demand. The results below can be extended straightforwardly to the situation, where $P \neq E$.

First notice that the concept of service level constraints does not make sense in the context of a deterministic model instance embedded in a rolling schedule concept. In order to still ensure that within the deterministic setting of the problem priority is given to satisfaction of exogenous demand, we introduce linear backorder costs. Assuming that expensive items are more important than cheap items we assume that the cost per item i backlogged at the end of a period is proportional to h_i, the cost per item held on stock at the end of a period. This yields the following objective function,

$$\sum_{i=1}^{N}\left(\sum_{s=1}^{T} h_i \hat{I}_i(t, t+s) + \sum_{s \in E} \theta h_i \hat{B}_i(t, t+s)\right) \tag{O$_1$}$$

We assume that $\theta \gg 1$.

The above formulated objective function does not completely solve our problem. We want to determine feasible plans that satisfy end-item service level constraints at low cost, since it is easily seen that the above problem yields the same solution for any value of θ larger than some value θ_0. This implies that with this objective function we obtain customer service levels that may not satisfy our objective.

Apparently the customer service level restrictions require additional decision variables. Following the inventory management literature (cf. Silver et al. (1998)) we introduce the concept of safety stocks in order to cope with short-term demand uncertainty

v_i safety stock parameter of item i, $i = 1, 2, \ldots, N$

We note here that in general the safety stocks depend on t and $t+s$, since the demand forecast may show seasonality and trends. Since our purpose is to explicitly compare different SCOP concepts, we must restrict to the stationary demand situation, implying a constant safety stock.

In order to control the customer service levels we modify the objective function as follows,

$$\sum_{i=1}^{N}\left(\sum_{s=1}^{T} h_i (\hat{I}_i(t, t+s) - v_i)^+ + \sum_{s \in E} \theta h_i (v_i - \hat{I}_i(t, t+s))^+\right) \tag{O$_2$}$$

At first glance the objective function (O$_2$) does not represent the real inventory holding costs and backorder costs. However, we should keep in mind that the MP problem formulation is only an attempt to model the supply chain operations planning problem under stochastic exogenous demand. In that sense any such formulation results into a heuristic with respect to the original optimization problem. Still objective function (O$_2$) reflects the trade-off between inventory holding and backorder costs. On top of that the safety stock parameters control the service levels. This becomes even more evident from the following lemma.

Sample path lemma

Suppose a sample path $\{D_i(t)\}$of the demand process and a sample path $\{\hat{D}_i(t, t+s)\}$ of the forecasting process are given. Furthermore assume that for all end items

$$I_i(0) = v_i, \qquad i \in E$$

Then the solution to the problem expressed in terms of the material order releases $\{\hat{r}_i(t, t+s)\}$ and processed quantities $\{\hat{q}_i(t, t+s)\}$ with objective function (O$_2$) subject to the LP constraints is the same for each value of v_i, $i \in E$, for all $t \geq 1$ and for all $s \geq 0$.

The proof of the sample path lemma is based on induction. Given the initial inventory levels it is clear that the objective function (O$_2$) implies an optimal solution $(\{\hat{r}_i(1, 1+s)\}, \{\hat{q}_i(1, 1+s)\})$ that is the same for any value of v_i. This implies that $(\{r_i(1)\},\{q_i(1)\})$ are the same for any value of v_i. But then $I_i\,(1){-}B_i(1){-}v_i$ is the same for any value of v_i. This argument can be repeated for any value of t. For a formal proof we refer to Køhler-Gudum and De Kok (2002).

Corollary to sample path lemma

The problems P_α and P_β for the SCOP concept defined by the LP constraints and objective function (O$_2$) have a unique solution $\{v_i\}_{i \in E}$, where each v_i, $i \in E$, can be determined independent of all other v_j, $j \in E, j \neq i$.

Noticing that the objective function (O$_1$) is identical to (O$_2$) with $v_i = 0$, $i \in E$, the corollary to the sample path lemma justifies the following procedure.

i. Run a discrete event simulation of the system with $v_i = 0$, where at the start of each period $t = 1, 2, \ldots,$ we solve the LP that follows from the forecasts $\{\hat{D}_i(t, t+s)\}$, the set of linear constraints given above and the linear objective function (O$_2$).
ii. From the discrete event simulation compute the empirical distribution function of $J_i(t){-}v_i$.
iii. Given this empirical distribution function compute v_i*, such that the required end-item service level is achieved.
iv. Run another simulation with v_i* in order to compute $\overline{C}(\mathcal{P})$.

We note here that step (iii) can be executed for most well-known performance measures, such as non-stockout probability at the end of a period, fill rate and average backlog [cf. Køhler-Gudum and De Kok (2002)]. In the comparison of SCOP concepts in Section 6 we apply this procedure with service criterion α the probability of a non-negative stock at the end of an arbitrary period.

Though the service level constraints determine the safety stocks v_i, $i \in E$, finding the optimal safety stocks v_i, for all items $i \in I$ constitutes an extremely complex nonlinear optimization problem. In Section 6 we argue that choosing

$v_i = 0$ for all items $i \in E$ yields a useful heuristic solution, yet clearly more research is required to validate this heuristic.

We notice here that the above approach can be applied to many other MP problems, so that alternative rolling schedule approaches for specific stochastic planning and scheduling problems can be compared. One of the (current) main issues with this approach is the CPU time required to accurately compute the stationary distribution of $J_i(t)$, assuming it exists. Especially when capacity constraints are tight one is confronted with similar issues as typical for the simulation of high load queueing (network) systems. Typically, several millions of periods (customers) are required to obtain sufficient accuracy for comparison purposes. Combined with the solution of an LP-problem in each period, this may result in extremely long computation times (several hours to several days per problem!). In the comparison study reported in Section 6 we circumvent this issue by restricting the analysis to non-capacitated systems.

4.3 Alternative MP formulations for the SCOP problem

The SCOP problem can be seen as a multi-item multi-level capacitated lot sizing problem (MLCLSP). Literature reviews of mathematical programming approaches for production planning and supply chain planning can be found in Shapiro (1993) and Baker (1993), and Erenguc, Simpson and Vakharia (1999). In this section, we will review some papers, each of which can be seen as representative of a class of approaches for the MLCLSP.

The MLCLSP formulation is a more general formulation than the SCOP problem. In fact, in the MLCLSP literature, no explicit reference is made to the planning problem as such. This means that the formulation can be used for a variety of planning problem in the SCP hierarchy. If the MLCSLP formulation is used for the SCOP problem, it is assumed that the quantities planned are also the quantities released, i.e., the MCLSP formulation does not distinguish between $\hat{r}_i(t, t+s)$ and $\hat{q}_i(t, t+s)$. In most cases, therefore, it makes more sense to use the MLCLSP formulation at a higher level of planning than SCOP. In this higher-level plan, the determined quantities are then in fact aggregate quantities to be detailed out at a later stage.

If lot sizing restrictions are not considered at the SCP level under consideration, the MLCLSP formulation typically reduces to the LP formulation given above. However, in the paper by Billington, McClain and Thomas (1983) one finds an LP formulation of the SCP problem that differs from our formulation presented earlier in this section. Billington et al. (1983) formulate a periodic planning problem, where the periods are typically short periods of time, e.g., hours or shifts. Furthermore they do not consider planned lead times. Instead they introduce a so-called *minimum lead time*, which should be interpreted as the minimum time involved in the transformation process to make an item available for usage. This time represents a delay and during this time no resources are used. A valid interpretation is that first an item is produced by some resource after which it is transferred to a stock point. The transfer time equals the minimum lead time. Billington et al. (1983) consider the item lead times, consisting of both waiting and processing times and the minimum lead time, as endogenous to the model. At first sight such an approach seems superior to ours, yet our discussion in Section 2 formulated this as an issue for further research.

Erenguc et al. (1999) give an excellent overview of MP formulations of the SCOP problem and discuss various issues by distinguishing between supplier stage problems, plant stage problems and distribution stage problems. Based on their literature survey they formulate a number of MP models for the SCP problem that include the notion of planned lead times, yet no distinction is made between material order release variables and material processing variables. In our view this is an important distinction that models manufacturing flexibility in accordance with the planned lead time concept. As follows from the discussion in Section 3 capacity checks on $\{r_i(t)\}$ instead of $\{q_i(t)\}$ yield inefficient usage of available resources.

Özdamar and Barbarosoglu (2000) represent a class of heuristics that is based on Lagrangean relaxation of either the capacity constraints (3.5) or the inventory balance equations (3.3) or both. Relaxing the capacity constraints and introducing Lagrange multipliers associated with those constraints reduce the MLCLSP to a number of independent uncapacitated problems. If also the inventory balance equations are relaxed one obtains a number of single item lot sizing problems that can be solved by e.g., the Wagner-Whitin algorithm [cf. Silver et al. (1998)]. The solutions to these undependent problems are tied together by an iterative update procedure for the Lagrange multipliers based on the calculation of subgradients. Özdamar and Barbarosoglu (2000) propose simulated annealing for solving the relaxed uncapacitated lot sizing problems. They present a number of combinations of different Lagrangean relaxations and simulated annealing. A computational study identifies the best combination and shows the practical applicability of this method. The model presented in Özdamar and Barbarosoglu (2000) is similar to the model presented in Billington et al. (1983), i.e., planned lead times are not considered, but lead times of material orders are time dependent outputs of the algorithm. In principle it is possible to modify the analysis in Özdamar and Barbarosoglu (2000) to take into account planned lead times. The paper by Barbarosoğlu and Özgür (1999) is similar to Özdamar and Barbarosoglu (2000) in that problem decomposition is proposed based on the Lagrange multiplier technique. It is interesting to mention their observation that the decomposition relates to an organizational decomposition where specific parts of the organization are responsible for particular sets of items. The decentralized decision making that results from this decomposition is supported by a central agent that ensures the exchange of relevant information between the different parts of the organization. This observation is related to our discussion in Section 2 about the distinction between problem decomposition, which is strongly based on organizational considerations, and model decomposition, which is strongly based on algorithmic efficiency considerations. Apparently both points of view may be aligned.

Belvaux and Wolsey (2001) propose MP formulations of the MLCLSP that lend themselves to relatively efficient solution with commercial mixed integer programming software such as CPLEX and XPRESS. The focus of their paper is the derivation of problem-specific (yet generic to the MLCLSP) sets of necessary inequalities, i.e., cutting planes, that considerably improve the performance of MIP solvers. They also emphasize the usefulness of the echelon stock concept [cf. Section 5] when dealing with multi-level problems. The echelon concept enables to reformulate the multi-level multi-item problem as a set of relaxed single-item problems that provide additional constraints that can be used for further efficiency of the MIP solution procedures. Belvaux and Wolsey (2001) remark that a distinction should be made between SCP problems where planning occurs infrequently and SCP problems where planning occurs frequently: big bucket problems versus small bucket problems, respectively. In the former case typically many setups for many different items occur during the single planning period under consideration, while in the latter case a small number of setups for a limited number of items occurs and multiple planning periods must be considered. The models to be applied in the two different cases are different. The SCOP problems considered in this chapter should be considered as small bucket problems. The SCP model formulation proposed by Belvaux and Wolsey (2001) does not include planned lead times, nor does it distinguish between material order release variables and material processing variables. Still, the equations derived in their paper as well as the techniques proposed can be easily modified to include these two phenomena.

This concludes our brief survey of MP formulations for the SCOP problem. In the next section we discuss the recent literature on stochastic models for the SCOP problem for general supply networks. In Section 6 we compare the LP-formulation of the SCOP problem as a representative of the class of MP models with a specific stochastic model on the basis of the required supply network inventory capital to achieve target customer service levels.

5 Stochastic demand models for supply chain planning

In this section we discuss various SCOP concepts for stochastic demand models as proposed in the literature. We assess these concepts on the basis of the SCOP constraints derived in Section 3. The discussion is restricted to incapacitated supply chains, since results for capacitated systems are only available for single item, single stage systems [see e.g., De Kok (1989)], serial systems [Tayur (1993)], or for divergent systems where only the most upstream stage is capacitated [see De Kok (2000)]. Due to the fact that the analysis of general supply networks with stochastic demand in the current literature is based on simplifying assumptions, we discuss the validity of these assumptions. In the context of Supply Chain Operations Planning under stochastic demand we focus on the determination of safety stocks, since these parameters determine an important part of the supply network inventory capital.

For ease of presentation we assume that a_{ij} is 0 or 1. For most of the policies we can easily extend the analysis to general a_{ij} values. Also we assume without loss of generality that $P = E$.

5.1 Echelon concept

Stochastic multi-echelon models can be distinguished by the state variables used to derive the item orders. In Axsater (2003), Chapter 10 of this volume, the distinction between installation stock policies and echelon stock policies has been discussed in detail. In this chapter we restrict ourselves to echelon stock policies. The main reason for this is that in a SCOP context installation stock policies typically violate SCOP constraints (3.1), since dependent demand is backordered. For general supply networks we define the echelon stock, echelon inventory position and some associated concepts in order to formally define item order release policies.

Firstly, we define

$O_i\,(t)$ Cumulative amount of orders outstanding at the start of period t.

Then we can define the echelon inventory stock $X_i(t)$ and the echelon inventory position $Y_i(t)$ of item i recursively as follows,

$$X_i(t) = J_i(t), \qquad \forall i \in E$$
$$Y_i(t) = X_i(t) + O_i(t), \qquad \forall i \in E$$
$$X_i(t) = J_i(t) + \sum_{j \in V_i} Y_j(t), \qquad \forall i \in I$$
$$Y_i(t) = X_i(t) + O_i(t), \qquad \forall i \in I$$

The echelon inventory position has the following important interpretation. The echelon inventory position of item i represents the coverage of future demand for item i up to and including the periods in time that the last order for item i becomes physically available for sales to customers. Typically, an item becomes physically available for sales to customers as part of a sellable item, i.e., an item in E. Also, we emphasize that in that sense an item becomes available for sales in different future periods in time, since that period depends on the sellable item under consideration and the assembly steps required to convert the item into this sellable item. The interpretation of the echelon inventory position as a coverage of future demand for sellable items is important for the synchronization of order releases for items that are assembled into the same sellable items. As we will see we can use this interpretation for the development of order release policies for general supply networks. For the special case of pure assembly systems we find that the concept of coverage of future demand is in line with optimal policies found in the literature. Let us first discuss the class of pure base stock policies as proposed by Magee (1958).

5.2 Pure base stock policies

In Chapter 11 of this volume, Song and Zipkin (2003) discuss stochastic models for assembly systems. In particular they focus on pure base stock policies. For sake of self-containedness of this chapter we define pure base stock policies below. Let

S_i Base stock level of item i

A pure base stock policy operates as follows:

$$r_i(t) = S_i - Y_i(t). \tag{5.1}$$

Song and Zipkin (2003), Chapter 11 of this volume, restrict their analysis to so-called Assemble-To-Order (ATO) systems, i.e., the general assembly system consists of two levels only: a finite product level and a component level. Only components can be kept in inventory. Through this restriction pure base stock policies are feasible. If we extend the general system to more than two levels or to the case where finite products can be kept in inventory as well, pure base stock policies are no longer feasible, even in the infinite capacity case.

It is easy to see that pure base stock policies in general violate constraints (3.1), which state that the increase of the backlog cannot exceed the *exogenous* demand. This is due to the fact that equation (5.1) is not constrained by upstream availability considerations. It implies that dependent demand can be backordered. In that sense it suffers from the same problem as the MRP I-logic. Incorporation of upstream availability constraints is non-trivial as can be seen from the analysis in Agrawal and Cohen (2001), Hausman, Lee and Zhang (1998), amongst others, where even the allocation of components to final products in an ATO-setting makes analysis of remnant component stocks due to lack of other components intractable. In Sections 5.3 and 5.5 we discuss synchronized base stock policies that can circumvent this problem at the expense of inadequate exploitation of component commonality.

A key result for ATO systems under pure base stock policies derived by all authors dealing with this is the following [cf. Agrawal and Cohen (2001), Hausman et al. (1998)].

ATO key theorem

Let α_i be defined as the probability that the demand for item $j \in E$ in period t can be satisfied immediately. Furthermore assume that if an item i is out-of-stock at the end of period t then all items $j \in V_i$ are allocated part of the shortage of item i. Then

$$\alpha_i = P\left\{ \sum_{s=t-L_i+1}^{t} D_i(s) \leq S_i, \quad \forall i \in W_j \right\}.$$

The results follows from the fact that the demand for item $j \in E$ in period t can be satisfied immediately if and only if all component stocks of items $j \in W_j$ are positive at the end of period t. Computation of α_j is complicated due to the correlation between the random variables $\sum_{s=t-L_i+1}^{t} D_i(s), i \in W_j$. Hausman et al. (1998) and Agrawal and Cohen (2001) assume the demand for final products to be normally distributed, so that they can apply results for multivariate normally distributed random variables. Song (1998) explores the combinatorial nature of the probabilistic expression for α_j in the case of continuous review (compound) Poisson demand and derives computationally efficient upper and lower bounds [cf. Song and Zipkin (2003), Chapter 11 of this volume].

From the ATO key theorem we can derive the following property of pure base stock policies.

Pure base stock policy property

Let j_1, and j_2 be two different end-items. Then

$$W_{j_1} \subset W_{j_2} \Rightarrow \alpha_{j_1} \geq \alpha_{j_2}.$$

Although this property is obvious and can be extended to multiple level systems, it shows an important drawback of pure base stock policies, even in situation where they yield feasible solutions. The property states that if an end-item contains a subset of the components of the other end-items then the non-stockout probability of the former is at least as high as the non-stockout probability of the latter. Considering typical ATO settings one finds that high-end end-items have additional features compared with low-end end-items. This implies that high-end end-item service levels must be lower than low-end end-item service levels. This is typically not desired because of economic reasons. As a consequence we find:

Under pure base stock policies arbitrarily chosen customer service levels for different end-items cannot be satisfied with equality.

This is different from what we have been able to show for the mathematical programming formulation of the supply chain operations planning problem in Section 4 and this is likely to cause higher supply chain capital investments than deemed necessary. In the next subsection we will discuss the generic approach proposed by De Kok and Visschers (1999) that circumvents this problem. We will compare the two generic approaches on the basis of some small-scale ATO examples. We notice here that the approach by De Kok and Visschers (1999) can be applied to any network structure and to combinations of Assemble-To-Order and Make-To-Stock.

5.3 Modified base stock policies for pure assembly systems

As stated above under pure base stock policies it may be possible that the quantity released cannot be met due to lack of material. As shown by Rosling (1989) and Langenhof and Zijm (1990) the echelon-order-up-to-policies for pure assembly systems, i.e., each item has at most one parent, can easily be modified by taking into account the availability of the child items.

Let us derive this modified base stock policy for pure assembly systems. First of all note that for pure assembly systems we have exactly one end-item and each item i has exactly one successor $suc(i)$. Thus we can uniquely define the cumulative lead time L_i^c of item i,

$$L_i^c = L_i, \qquad i \in E,$$
$$L_i^c = L_i + L_{suc(i)}, \qquad i \in I.$$

Given the definition of L_i^c we can state that $Y_i(t)$ represents the coverage by item i of the end-item demand from the start of period t until the start of period $t + L_i^c$ just before releasing the item ordered at the start of period t. Now notice that for all items with a longer cumulative lead time that at the start of period t we know exactly their coverage of end-item demand from the start of period t until the start of period $t + L_i^c$. Define

$Z_{ij}(t)$ coverage of end-item demand by item j from the start of period t until the start of period $t + L_i^c$, $L_j^c > L_i^c$.

Given the state information $(Y_i(t), \{Z_{ij}(t)\})$ we can define the modified base stock policy as follows,

$$r_i(t) = \max\left(0, \min\left\{S_i, \min_{\{j|L_j^c \geq L_i^c|\}} \{Z_{ij}(t)\}\right\} - Y_i(t)\right).$$

It is shown in Rosling (1989) Langenhof and Zijm (1990) that the policy described through the above equation is cost-optimal. In De Kok and Seidel (1990) and Van Houtum and Zijm (1991) simple computational schemes are given to determine the optimal echelon order-up-to-levels.

Thus for pure assembly systems we can find easy-to-implement optimal order release policies. This seems relevant for the SCOP problem in the manufacturing of complex and expensive products, such as manufacturing equipment and aircraft. Recognizing that such products have a modular structure where typically the forecast-driven activities relate to items common to all variants and the order-driven activities relate to variant-specific items, the supply network that faces stochastic demand may be seen as a pure assembly system. For such a system Dellaert, de Kok and Wang (2000) study a nonoptimal base stock policy for pure assembly systems. This non-optimal policy is inspired by a Make-To-Order environment (e.g. assembly of Public Telephone Exchanges), where the final assembly lead time is shorter than component purchasing lead times, so that the latter are held on stock. The base stock policy releases a production order for the main assembly, after which this main assembly is pushed through a number of assembly stages. Given the planned throughput time at each assembly stage for each component, purchase orders are released based on the projected demand according to the production order for the main assembly. This is clearly non-optimal from an inventory management point of view, since the orders for the components should be based on the latest customer demand information, as is the case for the optimal policy given above. However, in practice other considerations, like workforce planning, play a role. The push policy provides early information about resource requirements, whereas the optimal policy decides at each assembly stage how much to release immediately after receiving the latest demand information. In Dellaert et al. (2000) insight is given into the circumstances under which the push policy yields near-optimal inventory costs.

5.4 *Optimal and near-optimal base-stock policies for divergent systems*

Another special supply chain structure is the divergent structure, where each item has exactly one child, but may have multiple parents. The most upstream item, the root item, has a single supplier with infinite material availability. In Diks and de Kok (1998) the structure of the average-cost optimal policies for divergent systems is derived under the balance assumption. The costs considered are linear holding and penalty costs incurred at the end of a period. The penalty costs are incurred for each end-item short. The balance assumption states that in case the cumulative orders from parent items exceed the available stock of the item, then the optimal allocation policy guarantees that each parent item is allocated a non-negative quantity of the available stock. It can be easily verified by discrete event simulation that even in a two-echelon divergent system with identical end-items the balance assumption is violated by the optimal allocation policy, which is an equal fractile policy [cf. Eppen and Schrage (1981) Axsater (2003)]. Below we discuss this issue of imbalance in more detail.

The optimal policies under the balance assumption are base-stock policies and satisfy so-called generalized Newsboy equations. In order to formulate the necessary and sufficient conditions for base stock policies under the balance assumption we introduce the following notation:

For each item we can define its associated end-items,

p_i penalty cost incurred for each unit short of item i and the end of a period, $i \in E$
U_i all items on the path from the root item (inclusive) to item i (exclusive), $i = 1, 2, \ldots, N$
E_i set of end-items downstream of item i
α_k^i non-stockout probability of $k \in E_i$ under the optimal policy under the balance assumption for the subtree of the divergent system with item i as root item, $i = 1, 2, \ldots, N$

It follows from the definition of E_i, that

$$E_i = \{i\}, \qquad i \in E$$
$$E_i = \bigcup_{j \in V_i} E_j, \qquad i \in I$$

Diks and de Kok (1998) proof the following theorem:

Generalized newsboy equations theorem
Under the balance assumption the optimal base stock levels S_j and optimal allocation policies satisfy

$$\alpha_k^i = \frac{\sum_{m \in U_i} h_m + p_k}{h_k + \sum_{m \in U_k} h_m + p_k} \qquad \text{for every } k \in E_i,\ i = 1, 2, \ldots, N.$$

From the Generalized Newsboy Equations Theorem in theory one can recursively compute the optimal base-stock levels S_i and optimal allocation policies. However, this turns out to be computationally infeasible for realistic problem instances. The main issue here is the non-linearity of the optimal allocation functions. Diks and de Kok (1999) propose to assume *linear* allocation functions. In order to define these linear allocation functions we introduce the following variables:

q_j fraction of shortage allocated to item j
$X_{t,i}$ echelon stock of item i at time t immediately before allocation
$I_{t,j}$ echelon inventory position of item i at time t immediately after allocation

A linear allocation rule associated with item i and its parent items $j \in V_i$ is defined by

$$I_{t,j} = S_j - q_j \left(\sum_{m \in V_i} S_m - X_{t,i} \right)^+.$$

Note that $\left(\sum_{m\in V_i} S_m - X_{t,i}\right)^+$ is the shortage at time t, since it indicates the difference between the cumulative base stock levels of all parent items of i and the echelon stock of i.

Diks and de Kok (1999) derive a generalized Newsboy equation theorem for optimal linear allocation policies, implicitly assuming that in each stage of the recursive procedure linear allocation policies can be found that solve these generalized Newsboy equations. Though this is not true in general, the generalized Newsboy equation theorem for linear allocation policies yields a recursive heuristic to efficiently compute (S_i, q_i) for all $i=1, 2, \ldots, N$. Diks and de Kok (1999) show that the heuristic yields policies that "almost" solve the recursive set of generalized Newsboy equations, thereby suggesting that the policies found are close-to-optimal.

Likewise with the optimal allocation policies it cannot be guaranteed that linear allocation policies satisfy the balance assumption. This motivated Van der Heijden (1997) to determine the linear allocation policies that minimize the probability of imbalance. Based on proxies for the imbalance probability he derived a remarkably simple expression for the allocation fractions q_i, $\mathrm{i}=1, 2, \ldots, N$. Defining

$$
\begin{aligned}
D_k \quad &:= \quad \text{demand per period of item } k \in E \\
\mu_j \quad &:= \quad \sum_{k\in E_j} E[D_k] \\
\sigma_j \quad &:= \quad \sigma\Big(\sum_{k\in E_j} D_k\Big),
\end{aligned}
$$

we find the following expression:

$$
q_j = \frac{\mu_j^2}{2\sum_{m\in V_i}\mu_m^2} + \frac{\sigma_j^2}{2\sum_{m\in V_i}\sigma_m^2}, \qquad j \in V_i\,,\; i = 1,2,\ldots,N
$$

In the formula above we correct an error in the analysis in Van der Heijden (1997): from his formula (22) he minimizes the variance of each expected negative allocation to a successor separately instead of minimizing the sum of the probabilities of a negative allocation to a successor. The formula above coincides with the one derived in Van der Heijden (1997) when assuming that $\mu_j = \mu$ for all $j \in V_i$. We assume here that the demands for different end-items $k \in E$ are independent. Extensive discrete event simulation experiments show that indeed the linear allocation policies based on the above formula for q_j (Van der Heijden (1997) coins the term Balanced Stock rationing) yield such low probabilities of imbalance that imbalance can be ignored. This is extremely important because this implies that the analytical results and the policies obtained in Diks and de Kok (1999) are applicable, i.e., average costs and customer service levels are accurately computed. Further numerical study is required to find out whether the combination of base stock control policies and Balanced Stock rationing yields cost-effective policies as compared with alternative policies proposed in the literature. We refer to Axsater (2003) for a further discussion of the issue of imbalance. We conclude here with the statement that for the combination of base-stock policies and linear allocation policies it is possible to efficiently compute base-stock levels and allocation fractions. The policies computed seem to be close-to-optimal and performance characteristics, such as costs and customer service levels, are accurately computed. The analysis of divergent systems is the basis of a class of policies that can be applied to general assembly networks.

5.5 Synchronized base stock policies for general supply networks

The optimal policy derived for pure assembly systems cannot be applied to assembly systems where items have multiple successors (parents). We can identify two related root causes for this statement:

(1) In case of a shortage the above policy does not define the procedure for allocation of this shortage to successors.
(2) The state variables $Z_{ij}(t)$ cannot be defined since we cannot uniquely define L_i^c in case of multiple successors.

In De Kok and Visschers (1999) a class of policies is proposed that introduces uniquely defined state variables similar to $Z_{ij}(t)$ and allocation mechanisms derived from the analysis of divergent systems [cf. Van der Heijden, Diks and de Kok (1997)], so that it is possible to generate feasible item order releases in a straightforward way. Furthermore within this class of policies it is possible to characterize the optimal policy under i.i.d. exogenous demand, and near-optimal policies can be found numerically. For pure assembly systems this class of policies coincides with the modified base-stock policies described above. To understand the idea behind the class of policies proposed for general assembly systems let us consider in more detail our material co-ordination problem.

The lead time structure identifies at which moments in time item orders must be released in order to have them available for (sub)assembly activities required for production of the end-items at the start of period t. A natural order of decisions made over time thus arises. We can identify the item(s) that must be released first, which ones thereafter and so on. It is important to understand that as soon as an item is ordered, this ordering decision restricts the future demand that can be satisfied. In fact, the ordering decision leads to a particular echelon inventory position and, as stated above, this echelon inventory position covers future end-item demand. The problem however is that it is not clear how this coverage is actually used over time by the various end-items. This is not only due to uncertainty in future demand, but also due to the interactions between items caused by lack of availability. If some item is missing for an assembly operation, then another item is no longer needed as well. In a sense people in practice find themselves coping with a material co-ordination problem where solving the problem of a particular item creates a problem for another item, and so on.

The main reason for this complexity is that in a general supply network there is no clear hierarchy in decision making about order releases as we found above for the case of a pure assembly systems. Such a clear hierarchy also exists if the supply network would be a pure divergent system, i.e., each item is transformed into multiple items without a need for other items. The approach proposed in De Kok and Visschers (1999) is based on an *artificial hierarchy* derived from the structure of the general supply network. This structure is determined by the BOM and the planned lead times. Below we restrict ourselves to the main ideas behind this hierarchical planning concept. The artificial hierarchy enables us to define state variables that unambiguously define the item order releases.

We define the cumulative lead time of an item as follows:

$$L_i^c = L_i, \qquad i \in E,$$
$$L_i^c = L_i + \max_{j \in V_i} L_j, \qquad i \in I.$$

Now we define the root node s as

$$s = \arg\left(\max_i L_i^c\right),$$

i.e.

$$L_s^c \geq L_i^c, \qquad i \in I \cup E.$$

Without loss of generality we assume that s is unique and that all cumulative lead times are different. Now we develop a hierarchical procedure that decides on all item order releases related to the end-items in E_s. The hierarchy is derived from the cumulative lead times L_i^c. Define the set of items $\hat{C}_i$ as follows,

$$\hat{C}_i = \left\{ j | L_j^c > L_i^c, E_j \cap E_i \neq \phi \right\}.$$

We assume without loss of generality that

$$E_j \cap E_i = E_i, \qquad \forall j \in \hat{C}_i,$$

i.e., all items that are used in the same end-items as item i, but are ordered earlier than item i, are common to end-items in E_i. In case this restriction does not hold, we can find a partition of E_i, and a one-to-one related collection of subsets of $\hat{C}_i$ for which the above holds for each one-to-one related pair of subsets and apply the principles below to each of the subsets. Finally we define

$$E(\hat{C}_i) = \bigcap_{j \in \hat{C}_i} E_j.$$

The first decision in the hierarchy is to order item s at the start of period t according to a pure basic stock policy, i.e.,

$$r_s(t) = S_s - Y_s(t).$$

Let us consider item i to be ordered at the start of period t. In principle we would like to order according to a base stock policy, i.e., bring the echelon inventory position to S_i to cover future end-item demand. Notice that decisions have already been taken in the past for items $j \in \hat{C}_i$ that affect this decision. In fact we assume that our decision hierarchy in the past determined

$Z_{\hat{C}_i}(t)$ coverage of future end-item demand during periods $t, t+l, \ldots, t+L_i^c$, for all items in $E(\hat{C}_i)$ at the start of period t.

Given our assumptions stated above we have that E_i is a subset of $E(\hat{C}_i)$. Therefore we distinguish between two situations:

(i) $E_i = E(\hat{C}_i)$
(ii) $E_i \neq E(\hat{C}_i)$

In situation (i) we have that $Z_{\hat{C}_i}(t)$ is fully dedicated to future demand of end-items in E_i. Our target coverage equals S_i, but it does not make sense to increase the coverage above $Z_{\hat{C}_i}(t)$. Thus we release an order for item i as follows,

$$r_i(t) = \max\Big(0, \min\Big(S_i, Z_{\hat{C}_i}(t)\Big) - Y_i(t)\Big).$$

In situation (ii) $Z_{\hat{C}_i}(t)$ is intended to cover future demand for other end-items than those in E_i, alone. The problem is that we must decide how much to order for item i, thereby allocating quantities of the components in $\hat{C}_i$ to item i, while it may well be that we need not order yet any other items related to $E(\hat{C}_i) \setminus E_i$. In this case we maintain our hierarchy in decision making by introducing an *artificial base stock level* $S_{E(\hat{C}_i)\setminus E_i}$ that relates to end-items in $E(\hat{C}_i) \setminus E_i$. This implies that the target coverage of future demand for all end-items in $E(\hat{C}_i)$ equals $S_i + S_{E(\hat{C}_i)\setminus E_i}$. In case $Z_{\hat{C}_i}(t)$ is below this target level, then we must decide about the rationing of the deficit. This yields the following order release policy for item i,

$$r_i(t) = \max\Big(0, S_i - q_i\Big(S_i + S_{E(\hat{C}_i)\setminus Ei} - Z_{\hat{C}_i}(t)\Big)^+ - Y_i(t)\Big).$$

Here we use a linear rationing policy, where q_i is the fraction of the deficit allocated to end-items in E_i. Notice that situation (i) is a special case of situation (ii) with $q_i = 1$ and $S_{E(\hat{C}_i)\setminus E_i} = 0$.

In situation (ii) we create coverage by a set of items for future demand related to end-items in E_i and $E(\hat{C}_i) \setminus E_i$. The set of items associated with E_i is $\hat{C}_i \cup \{i\}$. The set of items associated with $E(\hat{C}_i) \setminus Ei$ is $\hat{C}_i$, which implies that $\hat{C}_i$ may again play the role of the set of items that will restrict the order release decision for an item to be ordered later. In fact, the creation of an artificial order-up-to level $S_{E(\hat{C}_i)\setminus E_i}$ always implies such a situation. Each order-up-to-level thus relates to a *decision node*, which is *uniquely* determined by the *combination* of a set of items and a set of end-items. A set of items can be associated with multiple decisions nodes, c.q. (artificial) order-up-to-levels, and a set of end-items as well can be associated with multiple decision nodes.

In case $Z_{\hat{C}i}(t) > S_i + S_{E(\hat{C}_i)\setminus E_i}$ the excess coverage $Z_{\hat{C}i}(t) - (S_i + S_{E(\hat{C}_i)\setminus E_i})$ with respect to future end-item demand in periods $t, t+1, \ldots, t+L_i^c$ is not used to cover this future demand and hence will be available to cover end-item demand after period $t + L_i^c$. As a consequence this excess coverage results in future excess stocks of all items in $\hat{C}_i$ at the end of period $t + L_i^c$. Hence a decision node can be seen as a stockpoint, where physical inventory relates one-to-one to excess future stocks of the items associated with this decision node.

We have stated that the procedure above holds for any supply network. Informally speaking, the above approach creates a number of divergent systems of decision nodes. Our restrictions on the sets E_i given above are restrictions on the possible combinations of item sets and end-item sets. Once understanding the principles it is rather straightforward to remove these restrictions. In Sections 5.6 and 6 we present examples for which we derive the divergent system(s) of decision nodes.

As stated above the policy described above extends the optimal policy for pure assembly systems described in Rosling (1989) to a (nonoptimal) policy for general assembly systems, i.e., multi-item multi-echelon systems. The main idea behind the approach is the artificial hierarchy that enables synchronization of order release decisions over time. Hence we define these policies as *synchronized base stock policies*.

An important distinction between the pure base stock policies and the synchronized base stock policies given here is that in the former case each item has a uniquely defined base stock level, while in the latter case multiple base stock levels, each associated with a decision node, may be associated with a single item. In addition the allocation fractions defined above constitute another set of decision variables. This implies that the synchronized base stock policies provide many more degrees of freedom. This explains why pure base stock policies do not allow for any combination of target customer service levels, while synchronized base stock policies can be found that satisfy any set of customer service level constraints with equality.

We note here that multiple divergent systems emerge when an order must be released for an item that is not contained in any of the end-items for which we know the limits on future coverage due to earlier decisions. In that case this item becomes the root node for another divergent tree of decision nodes and we can apply a pure base stock policy for this item.

It is important to note that synchronized base stock policies are *not cost-optimal*. In fact in situation (ii) we allocate part of the future coverage $Z_{\hat{C}_j}(t)$ specifically to item i without a real need for taking that decision. This allocation could have been postponed until the moment items are physically assembled into a (sub) assembly or end-item. In De Kok and Visschers (1999) it is shown for product families satisfying a particular constraint on the product and lead time structure, that this simultaneous ordering and allocation decision seems to hardly affect the performance of the system in terms of cost and customer service level. The comparison in Section 5.6 seems to underline the effectiveness of the synchronized base stock policies. The comparison in Section 6 provides even stronger support to the SCOP concept described above. Yet clearly more research is required to make any conclusive statement.

Due to the fact that the general assembly network is translated into a set of divergent systems we can apply the algorithms proposed by Diks and de Kok (1999) in order to determine order-up-to-levels and rationing fractions that satisfy the service level constraints [cf. Section 5.4]. This implies that *given the concept of synchronized base stock policies* as defined above we can derive close-to-optimal policies for this concept even for large-scale systems. In De Kok (2002a) efficient algorithms are proposed for determining these close-to-optimal policies based on the relationship between generalized Newsboy equations and finite-horizon ruin probabilities.

5.6 *Comparison of pure and synchronized base stock policies*

To provide insight into the differences between the pure base stock policy models discussed in the literature and the synchronized base stock control policy model, we discuss an example. We consider an ATO system consisting of 3 end-items and 3 modules. Fig. 3 shows the product structure. End-item 1 is a base-version, consisting of module 6, only. End-item 2 and 3 have additional features, represented by modules 4 and 5 respectively. The costs of items 4, 5 and 6 are 1, 1 and 8, respectively.

The synchronized base stock policies are related to a logical mapping of the above structure into (a set of) divergent structures. In the case of this simple example we find a single divergent structure as given in Fig. 4.

Here E_k denotes the set of end-items associated with decision node k, and C_k denotes the set of items associated with decision node k. The triangle associated with E_k and C_k expresses the fact that the release decisions taken with respect to successors of decision node k may result into an excess coverage of future demand for end-items by the items associated with decision node k [see Section 5.5]. This excess is *planned* at the moment the release

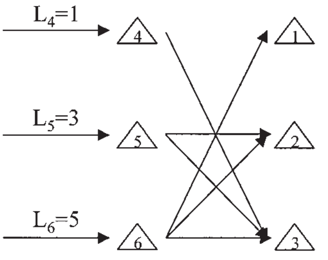

Fig. 3. Example product structure

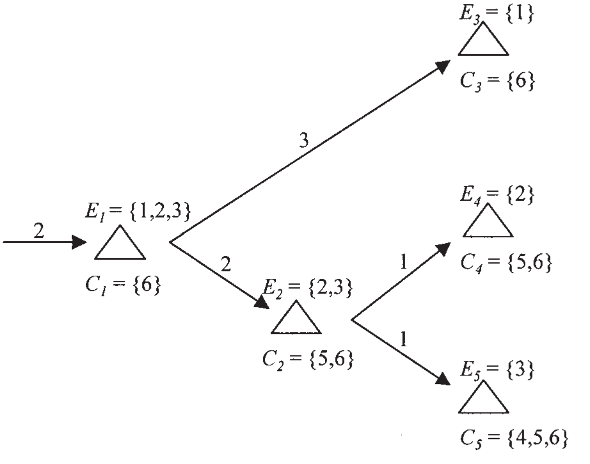

Fig. 4. Divergent structure underlying synchronized base stock policies for example product structure.

decisions associated with decision node k are taken, but eventually the excess results in physical stock of the items associated with decision node k. This implies that the long-run average physical stock of each item i can be computed from this associated divergent multi-echelon inventory system by adding the long-run average physical stocks of all decision nodes k with $i \in C_k$.

From this divergent structure we obtain the following decision hierarchy:

- First we order, say at time 0, component 6 to be used in all end-items 1, 2 and 3.
- After two time units, i.e., at time 2, we order item 5 after deciding first on allocation of the echelon stocks associated with item 6, thereby determining how much of future availability of item 6 is allocated to end-item 1 and how much to end-items 2 and 3 together.
- After another two time units, i.e., at time 4, we order item 4 after deciding first on allocation of the echelon stocks associated with items 5 and 6

dedicated to items 2 and 3, thereby determining how much of future availability of item 5 and 6 is allocated to end-item 2 and how much to end-item 3. At this point in time all items have been allocated to satisfy demand during period 6, which starts at time 5.

In Table 1 below we give the results for the situation where item 3, the high-end-item consisting of all three modules, has a target α of 99%, item 2 has a target a of 95% and item 1 has a target α of 90%. The demand per period for the end-items is i.i.d. with mean $E[D]$ and $\sigma(D)$.

The base stock levels are denoted by S, the average stock investment of a module by X. The index *PBS* denotes the pure base stock policies, the index *SBS* denotes the synchronized base stock policies. The *SBS* policies have been computed from the generalized Newsboy equations derived in Diks and de Kok (1999), while the PBS policies have been derived from a brute force cost minimization based on the results in de Kok (2002b).

The results are at first sight negative for the pure base stock policies, since the amount of stock investment required is 4% higher than the stock investment requirements in case of the synchronized base stock policy. A closer look reveals clearly why this is the case. In order to guarantee a 99% service level for item 3, items 1 and 2 both have a service level of 99%. These high service levels are enabled by high stocks of all items. Yet more insight can be gained when we realize the following. The main reason for the extreme difference in stock requirements is caused by the definition of the α-measure in the case of pure base stock policies and the assumption underlying the explicit expression given in the ATO theorem: *In case of a shortage of a particular item it is assumed that all end-items share part of the shortage*.

Although this seems obvious at first sight, it implies the ordering of the α-values for end-items in case one item's component (or module) set is a subset of another item's component set and it implies a worst-case scenario. In reality typically one would like to share shortages, yet not to the extreme that any shortage of a component causes a backorder for all end-items using this

Table 1
Performance base stock policies in case of high service of high-end-item

Item	$E[D]$	$\sigma(D)$	α^*	α_{PBS}	α_{PBS}
1	100	25	0.9	0.99	0.9
2	100	25	0.95	0.99	0.95
3	100	25	0.99	0.99	0.99

Component charactristics

Item	S_{PBS}	S_{SBS}	X_{PBS}	X_{SBS}
4	200	192	100	83
5	810	800	210	166
6	1740	1733	1923	1891
System			2233	2140

Table 2
Performance base stock policies in case o low service of high-end-item

Item	$E[D]$	$\sigma(D)$	α^*	α_{PBS}	α_{SBS}
1	100	25	0.99	0.99	0.99
2	100	25	0.95	0.95	0.95
3	100	25	0.9	0.9	0.9

Component characteristics

Item	S_{PBS}	S_{SBS}	X_{PBS}	X_{SBS}
4	142	162	43	47
5	708	770	109	126
6	1740	1741	1923	1952
System			2075	2125

component. What seems to be a quite natural assumption leading to an elegant expression for the non-stockout probabilities, in fact may yield high supply chain capital investments. Apparently, it is important to explicitly define a component shortage allocation mechanism that ensures lower supply chain capital requirements. Since the main reason for high supply chain investments is caused by the fact that the high-end-items have the highest service, we expect a more favorable situation for the case that we want high-end-items to have a lower service level than the low-end end-item. The results in Table 2 confirm this. In this case the pure base stock policies outperform the synchronized base stock policies by 2.4%. This is likely to be caused by the fact that synchronized base stock policies allocate common components too early, as follows from the logical decision hierarchy that is depicted in Fig. 4.

It should be clear that in practice we set α-service level constraints on the basis of an (implicit) trade-off between supply chain stock capital investments and penalty costs for disservice. It is likely that high-end service levels are higher than low-end service levels. The discussion above implies that pure base stock control policies do not allow for sufficient degrees of freedom (or means of control) to set the α-service levels accordingly. The potential supply chain cost savings are an interesting subject for further research. This concludes the discussion of pure base stock policies as a basis for SCOP.

5.7 *Heuristic analysis of general network structures*

In Section 5.2 we argued that pure base stock policies cannot be used for supply chain operations planning, since these policies do not satisfy the material availability constraints (3.1). We must be aware of the fact that even in the case of infinite resource capacity, the constraints (3.1) are of such complexity that it is not at all clear how to incorporate them into a supply chain operations planning concept that takes into account demand (and supply) uncertainty. We emphasize again here that this is the most important reason to resort to mathematical programming based supply chain planning concepts in a rolling schedule context. However, such concepts do not answer important questions such as

- Where and how much safety stocks should be held?
- What is the impact of demand uncertainty on supply chain capital requirements to achieve the required customer service levels?

Such questions are more of a strategic and tactical nature, yet the answers have an immediate consequence for the operational performance of the supply chain.

In the literature we find several approaches that discuss the issue of safety stock positioning. For an extensive discussion of these approaches we refer to Axsater (2003) and Song and Zipkin (2003) (Chapters 10 and 11, respectively, of this volume). It seems appropriate to briefly discuss these approaches in the context of general assembly structures. For such structures it seems that there are two generic assumptions underlying the heuristic analysis of such systems:

Decomposition assumption. *Safety stock parameters are set at such high levels that every material order released can be satisfied from stock on hand without checking upstream availability*.

The decomposition assumption allows an analysis of general assembly systems, where costs are derived by adding costs derived from an item-by-item single-item single-echelon model and customer service levels are derived from single-item single-echelon models for items in E.

Assembly assumption. *If an assembly order has to wait for material then this is caused by the shortage of exactly one item required for assembly*.

The assembly assumption simplifies the analysis considerably due to the fact that it enables the translation of an assembly system into a weighted sum of serial systems. The weights relate to the probabilities that particular items are short.

The decomposition assumption seems stronger than the assembly assumption. It can be verified by discrete event simulation that when item service levels for items in I are above 95%, then the (heuristic) analysis of the performance of the supply chain in terms of costs and service yields an acceptable accuracy. Realizing that 95% service levels are also a prerequisite for the assembly assumption to hold, we might argue that if the assembly assumption holds, then the decomposition assumption holds as well, in the sense that both yield more or less the same results. Further research is however required to verify this line of thought.

Analysis under the decomposition assumption

The typical focus of the heuristic analysis under the decomposition assumption is on the analysis of a supply network cost function under some service level restriction and the derivation of properties of optimal policies [e.g., see Lee and Tang (1996)]. We notice however that the optimal policies found, by implication of the decomposition assumption, keep high stocks at all upstream stages. It follows from Whybark and Yang (1996) (and the results of Section 6) that truly optimal policies tend to concentrate inventory capital at downstream stages, unless a high value is added at the most downstream stages. We believe that this observation may have considerable impact with respect to the benefit of postponement of item diversity. Item diversity is postponed if the item is made more common to downstream items. Postponement strategies therefore allow for reduction in upstream item stock investments while maintaining end-customer service levels. The decomposition assumption may yield exaggeration of the benefits of postponement since these benefits are derived from reduction of upstream safety stocks that guarantee high intermediate service levels. For an in-depth discussion of the benefits of postponement in Supply Chain Management we refer to Lee and Swaminathan (2003), Chapter 5 of this volume.

We notice here that postponement strategies should not be confused with strategies that allow for an upstream shift of the CODP [cf. Section 1]. In that case changes in the transformation and transportation processes allows holding e.g. modules instead of end-items. An upstream shift of the CODP typically has a large impact on stock investments, since (specific) end-item stockpoints with high service requirements are eliminated completely.

In Graves and Willems (2003), Chapter 3 of this volume, the decomposition approach proposed by Inderfurth (1994) and Graves and Willems (2000) based on Simpson (1958) is extensively discussed. Minner (2000) discusses the distinction between full-delay models and no-delay models. The full-delay models assume that if an item is not available, then one has to wait with an order release until the item becomes available. This is in line with our discussion in Section 3 and in fact is equivalent to the set of constraints (3.1). In no-delay models it is assumed that if the dependent demand from downstream orders exceeds the available inventory, then there is some outside source (not considered in the model in terms of costs) that provides the material. In fact, this is the assumption that Simpson (1958) proposed and enables an elegant analysis of complex supply networks. Graves and Willems (2000) argue that this assumption is reasonable if safety stocks are set such that they cover the maximum demand during a so-called coverage time. The coverage time is the sum of the lead times of items that are consecutively upstream of the item under consideration.

The analysis of no-delay models reduces to the determination of the optimal cover times. In fact the concept of cover times enables a generalization of the decomposition assumption in that if a cover time of an item covers multiple upstream stages, then at these stages no safety stocks are kept and the safety stock is only held for the item under consideration. We note here that the no-delay assumption converts the analysis of a stochastic demand model into the analysis of a deterministic model with cover times as decision variables.

From the analysis in the above-mentioned papers one finds items where no stocks are held at all and items for which the stock equals the safety stock associated with the cover time derived and the demand uncertainty associated with the item. Using dynamic programming Minner (2000) is able to analyze general supply chain structures. He shows by an example of a serial supply chain that the optimal policies for no-delay models and full-delay models may differ considerably. He discusses heuristics that may bridge the gap between full-delay models and no-delay models. However, more research is required to better understand the applicability of no-delay models. Hereby we implicitly assume that full delay models better represent the reality modeled in a supply chain operations planning context. We emphasize here that the main issue here is the external validity of either approach. Since in reality people intervene in situations where demand exceeds availability the only way to test applicability of models like the ones discussed in this Chapter is empirical. Only then we can identify situations where either full-delay or no-delay models perform best.

An interesting contribution to the analysis of the SCOP problem based on the decomposition assumption is given in Graves, Kletter and Hetzel (1998). They consider a general supply network under dynamic demand. The main contribution of the paper is the introduction of so-called forecast revisions $\Delta F(t, t+s)$,

$$\Delta F(t, t+s) = \hat{D}(t, t+s) - \hat{D}(t-1, t+s), \qquad s \geq 0.$$

As before $D(t, t+s)$ denotes the forecast of the demand in period $t+s$ made at the start of period t. It is assumed that the random variables $\Delta F(t, t+s)$ are i.i.d. with respect to t. The forecast revisions for an arbitrary t may be correlated. These assumptions seem quite reasonable. Graves et al. (1998) give a detailed analysis of the single stage model. They assume that item order release revisions $r_i\ (t, t+s){-}r_i\ (t{-}1, t+s)$ can be expressed as a linear function of the forecast releases. This assumption enables the formulation of an optimization problem focused on minimizing the variance of the item order releases subject to a constraint on the variance of the net inventory, which relates to the level of customer service. The decision variables are the weights that determine the linear relationship between item order release revisions and forecast revisions. Applying the decomposition assumption the results for the single stage system are combined into an analysis of a general assembly system. The approach was successfully applied to a real life case study.

The approach described in Graves et al. (1998) has a strong resemblance with the seminal work by Holt, Modigliani, Muth and Simon (1960) who study linear decision models in the context of aggregate planning. In both approaches the linear relationship between production schedule revisions and forecast revisions is used as a means to implicitly model finite capacity.

Even though Graves et al. (1998) focus on the design of the supply network, their innovative modeling of the demand process deserves further study in the context of the SCOP problem (see Section 7).

Analysis under the assembly assumption

The assembly assumption has been extensively used in literature. It enabled the exact analysis of spare part networks, where end-items consist of modules that may fail due to components that may fail, etc. In this context the assembly assumption states that if a product fails this is due to exactly one module. In turn, this module failed due to exactly one component, etc. For an extensive survey of relevant literature we refer to Sherbrooke (1992) and Axsater (2000).

In the context of safety stock positioning in general supply networks recently Ettl, Feigin, Lin and Yao. (2000) developed a framework of analysis building on the assembly assumption. They assume continuous review installation base stock policies for all items. In line with the discussion in Sections 1 and 2 they assume so-called nominal lead times that are equivalent to the planned lead times assumed in this chapter. Demand is modeled as a compound Poisson process. The authors emphasize that the compound Poisson process is a means to approximate the real life demand process and provides the degrees of freedom to obtain a good fit.

The objective of Ettl et al. (2000) is to find optimal base stock levels that minimize costs subject to customer service level constraints. The analysis in the paper is roughly as follows: First of all, it is easy to see that the external demand process and the installation base stock policies determine the demand process for each item, irrespective of the base stock levels of its parent items. This enables to compute on an item-by-item basis the waiting time distribution of an arbitrary order for the item, given its base stock policy. This waiting time is derived from the analysis of the number of outstanding orders, which in turn is derived from the analysis of an $M^{[X]}/G/\infty$ queue. The analysis yields for each item, the first two moments of the waiting time distribution of an arbitrary order for this item. Next the assembly assumption converts the multi-item multi-stage problem into a set of interrelated single stage problems that can be analyzed subsequently. Thus expressions can be derived for holding costs and service levels. This yields an overall objective function that is optimized using a conjugate gradient search technique with the base stock levels as decision variables.

A careful study of the numerical results reported in Ettl et al. (2000) and, more extensively, in Feigin (1998) reveals that the optimal solution yields extremely low upstream fill rates (between 0.1 and 0.7). This puts forward a fundamental issue. The approach is based on the assembly assumption, yet the optimization procedure suggests optimal solutions that strongly violate the major assumption underlying the approximate analysis. Although the actual performance of the heuristic may be very good or even close-to-optimal, from a standpoint of mathematical rigor it implies that the solution obtained is infeasible and the optimization problem should be reformulated including constraints that ensure adherence to the assembly assumption. This problem is addressed implicitly in Ettl et al. (2000) in that they apply discrete event simulation to verify the solutions obtained. It means that the quality of the solutions cannot be supported by the analytical characteristics of the approach. In our view this fundamental issue deserves further research.

This concludes our discussion of base stock policies in the context of the SCOP problem. Before summarizing our conclusions we would like to discuss briefly another interesting class of control policies.

5.8 *Combined Kanban and base stock policies for Supply Chain Planning*

This class of policies for the control of manufacturing systems has been proposed by Buzacott and Shantikumar (1993) and Frein et al. (1995) and others: a combination of base stock control policies and Kanban control policies. Though there are differences between the concepts proposed by the various authors the idea behind the policies are essentially the same. Below we give an informal discussion of this type of policies and focus on their relevance for SCOP. We will denote these concepts as Combined Kanban Base Stock Control (CKBCS), without having the ambition to add another acronym to this part of the literature, but to avoid that we are not precise enough to describe any of the proposed mechanisms.

First of all let us consider the standard Kanban policies. A standard Kanban policy can be defined through two parameters: the number of cards circulating between two stages and the quantity per card. To clarify this further let us consider an arbitrary item i. Assume that item i is processed on resource k_i after which it is used by items $j \in V_i$ that are processed on resources k_j. Then we can define

- K_{ij} Number of Kanban cards associated with items i and j circulating between resource k_i and resource k_j
- Q_{ij} The quantity associated with each Kanban card to be released of item i on behalf of item j
- S_i Base stock level of item i

As seen before the base stock level should be seen as a target inventory level associated with item i. Yet the state variable that defines the inventory level can have different definitions. We have seen examples such as the echelon inventory position and the local inventory position, but in principle alternatives exist. Typically, one could vary the definition of the echelon of item i, comprising more or less of the supply chain downstream of item i.

In principle CKBSC policies operate as follows: If the inventory level of item j is below its target level S_j and at least one Kanban card associated with items i and j is available, then a Kanban card is sent from resource k_j to resource k_i, releasing an amount Q_{ij} of item i. From this we see that the base stock level S_i should ensure availability of item i, while the parameters K_{ij} and Q_{ij} ensure that the maximum amount of work in process of item i, i.e., the total amount of outstanding orders of item i, cannot exceed $\sum_{j\in V_i} K_{ij}Q_{ij}$

It can be shown that CKBSC policies can emulate all currently known inventory control policies by appropriate choices of both the definition of the inventory state variable and the values of the parameters S_i, K_{ij} and Q_{ij}. Although the discussion in Buzacott and Shantikumar (1993) and Frein et al. (1995) restricts to continuous review systems, in our view the principles of CKBSC policies carry over to the periodic review setting of supply chain planning.

When testing the CKBSC policies against our generic supply chain planning constraints we find that the feasibility constraints (3.1) may be violated. The reason for this is that a Kanban card is sent from a resource k_j without checking availability of the material required. One solution to this problem is to ensure that the number of outstanding orders never exceeds the base stock level, so that always inventory is available. This implies that $\sum_{j\in V_i} K_{ij}Q_{ij} \leq S_i$. In the CKBSC concepts proposed in the literature the opposite is assumed, i.e., the maximum amount on order is at least equal to the base stock level. If this is the case we have to develop procedures that impose the feasibility constraints (3.1). It is likely that such a modification will greatly complicate the analysis of the performance of CKBSC systems. Currently, this analysis is strongly related to the performance analysis of queuing network systems. For further details we refer to Frein et al. (1995) and Buzacott and Shantikumar (1993). It should be noted that such an analysis strongly relies on the assumption of continuous review and a FCFS discipline for dealing with priorities in case of material shortage. Given the fact that the supply chain operations planning problem is by nature periodic and priorities should be based on costs structures, it is clear that further research is required to develop the framework of CKBSC policies for a supply chain planning. Given the richness of the framework as such we consider this direction for further research as quite promising. However, the state-of-the-art of analysis of CKBSC policies does not allow for a comparison with other supply chain planning concepts, as we intend to undertake in Section 6.

5.9 Concluding remarks

In this section we discussed quantitative models for the incapacitated SCOP problem that explicitly include demand uncertainty. We have shown that pure base stock policies violate the feasibility constraints (3.1). We extensively discussed the synchronized base stock policies introduced by De Kok and Visschers (1999) that satisfy the feasibility constraints and allow for an exact analysis and the numerical computation of near-optimal policies within this class. We identified policies that combine Kanban control and base stock control as interesting candidates for further SCOP research. We discussed various heuristic methods to compute control policies minimizing costs subject to service level constraints. The heuristics based on the decomposition assumptions yield solutions that contradict the findings in Whybark and Yang (1996) about optimal positioning of safety stocks, i.e., safety stocks should be concentrated downstream (see also Section 6). The assembly-assumption-based heuristic of Ettl et al. (2000) yields solutions that, although they seem to violate the assembly assumption, are close to optimal.

In the next section we compare the two classes of control concepts for which we have been able to show that all SCOP constraints are taken into account, viz. the LP-based control concept discussed in Section 4 and the synchronized base stock policies discussed in this section. We restrict our comparison to the uncapacitated SCOP problem, since there are no results for the capacitated SCOP problem under stochastic demand.

6 Comparison of supply chain planning concepts for general supply chains

In this section we compare supply chain operations planning concepts from the two main classes of supply chain planning concepts, i.e., MP-based concepts in a rolling schedule context (discussed in Section 4) and concepts based on stochastic models that are applicable to general supply chain structures (discussed in Section 5). Firstly in Section 6.1 we briefly discuss in the impact of a (dynamic) forecasting process on the release decisions generated by the two classes of concepts. Thereafter, we present a numerical study that provides insight into various aspects of the SCOP problem. We restrict this numerical comparison to the situation with infinite resource availability. The main reason for this restriction is that stochastic model-based concepts do not incorporate finite capacity resources. Furthermore we assume stationary stochastic demand. We expect that the results obtained are applicable to the situation with stationary forecast errors, yet this requires further research.

Because our focus is on managerial insights into the SCOP problem and the characteristics of the two different SCOP concepts, we restrict our comparison to a relatively simple case situation. One should be aware of the fact that the structural complexity of the SCOP problem discussed in this chapter, i.e., multi-item, multi-echelon, general BOM relationships, is enormous. By carefully selecting case situations we obtain useful insights. These insights were confirmed by results from a comparative study based on a real-world case [cf. De Kok (2001)].

In Section 6.2 we present the case situation. In Section 6.3 we present the results of our comparative study. In Section 6.4 we recapitulate our conclusions into a number of managerial insights and issues.

6.1 Supply Chain Planning concepts and forecasting

As has been made clear in Section 4 the key dynamic input to an SCOP concept is the forecast of exogenous demand for items in E (like before we assume $P=E$). A practically relevant question that seems to have an obvious affirmative answer is the following: Are the immediate release decisions affected by the forecasts of exogenous end-item demand? Given the set of equations in Section 4 this may be obviously true, but we concluded that the answer depends on the planning concept used. It can be easily proven that a planning concept based on base-stock policies *with fixed base-stock levels* generates immediate order release decisions that do not depend on the forecasts generated. Similarly, it can be proven that if the base-stock levels are dependent on the forecast, for example, when the base stock levels represent a fixed number of weeks coverage of future demand, then the immediate order release decisions depend on the forecast. It can also be verified that the supply chain operations planning concept based on LP discussed in Section 4 generates immediate order release decisions that depend on the forecasts of exogenous demand. It seems natural to assume that the SCOP concept should generate immediate order release decisions that depend on the forecast. In this context it is interesting to notice that recently a growing number of retail and manufacturing companies seems to resort to classical end-item inventory management policies with fixed reorder and/or order-up-to-levels. There seems to be empirical evidence that current forecasting (or sales planning) processes generate a forecasting accuracy that justifies these decisions. For an interesting discussion of this particular issue we refer to Aviv (2001).

6.2 Comparison of the LP-based SCOP concept and the SBS concept

In this section we compare the LP(-based SCOP) concept presented in Section 4 with the SBS concept presented in Section 5. The LP concept represents the class of deterministic optimization models in a rolling schedule setting, while the SBS concept represents the class of stochastic models for SCOP. We first describe the case example in detail.

Case description

In this section we describe the case we used for a comparative study. We subsequently present the BOM structure, the demand process, the cost structure and the performance measures used in our comparison.

The example product structure.

In order to compare the two SCOP concepts we create a test bed. We only consider the specific product structure consisting of 11 items given in Fig. 5. As stated above we found that the results obtained for this test bed are typical.

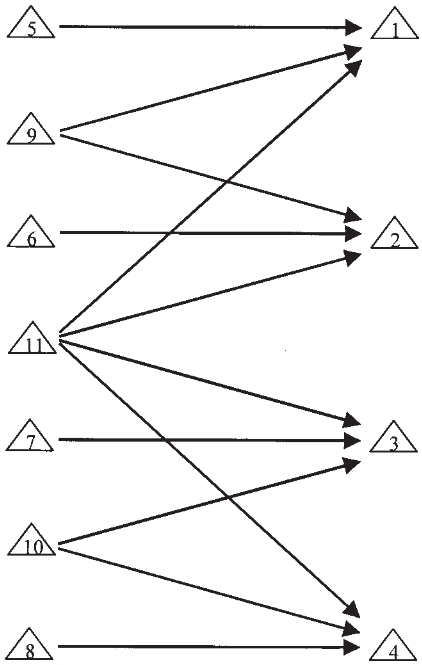

Fig. 5. The example product structure

We consider an 11-item product structure consisting of four end-items 1 to 4. All products contain common component 11. Products 1 and 2 share component 9, while products 3 and 4 share component 10. Each product contains a specific component.

As stated above we do not consider finite resources. Implicitly, such finite resources can be incorporated into the BOM through the planned lead times. High resource utilizations result into long planned lead times, low resource utilizations result into short planned lead times. In the context of the discussion in Section 2 on planning hierarchies one might state that the choice of the planned lead times is such that the order release decisions generated at SCOP level can be realized at shopfloor level with high probability. That is the due date of an order derived from the moment of its release and the planned lead time can be met with high probability.

Therefore, it is of interest to vary the lead times of the different items. Note that the lead time of the end-items 1 to 4 relates to an assembly process. We may consider the lead times for items 5 to 11 as procurement lead times. A long procurement lead time thus relates to a supplier with tight capacity in a remote location. The planned lead time structure is given by the following variables:

L_f planned lead time end-items
L_s planned lead time specific components

L_{sc} planned lead time semi-common components
L_c planned lead time common component

This implies the following equations,

$L_i = L_f$, $i = 1, 2, 3, 4$
$L_i = L_s$, $i = 5, 6, 7, 8$
$L_i = L_{sc}$, $i = 9, 10$
$L_i = L_c$, $i = 11$

In order to get insight into the impact of the supply chain structure we vary the planned lead times (L_s, L_{sc}, L_c)as follows,

(1, 2, 4) common component long lead time, specific component short lead time
(4, 2, 1) common component short lead time, specific component long lead time
(1, 4, 2) semi-common component long lead time, specific component short lead time

We note here that the planned lead time structure impacts the divergent structures that emerge when applying the SBS concept. Below we present the divergent structures associated with each of the three planned lead time structures.

Demand process and cost structure

As stated above we assume that the demand for the end-items is stationary. More precisely, demand for end-item i in consecutive periods is i.i.d. We also assume that the demand processes for different end-items are uncorrelated. We define

$E[D_i]$ average demand per period for item i, $i = 1, 2, 3, 4$
cv_i^2 squared coefficient of variation of demand per period item i, $i = 1, 2, 3, 4$

We set $E[D_i]$ equal to 100 for all end-items. To get insight into the impact of demand variability on the choice of an SCOP concept we vary cv_i^2 as 0.25, 0.5, 1 and 2. Unless stated otherwise, we assume identical demand parameters for all end-items.

Cost structure

As explained in the introduction to this chapter we want to compare different SCOP concepts on the basis of the supply chain inventory capital required to achieve a prespecified customer service level. Based on the cost structure in high volume electronics supply chains we developed a base case cost structure as follows. Analogously to the definition of planned lead times we define

h_f added value end-items
h_s added value specific components
h_{sc} added value semi-common components
h_c added value common components,

implying that

$h_i = h_f,\quad i = 1,2,3,4$
$h_i = h_s,\quad i = 5,6,7,8$
$h_i = h_{sc},\quad i = 9,10$
$h_i = h_c,\quad i = 11$

In the base case we assume that $(h_f, h_s, h_{sc}, h_c,) = (\$10, \$10, \$30, \$50)$. Hence the common component is expensive, while the added value of assembly is only 10% of the total value of the end-item. An example of such a situation is the manufacturing of TVs. Typically the Cathode Ray Tube is 50% of the total cost, a Printed Circuit Board may account for 30% of the cost, while additional materials such as a housing account for another 10% of total cost.

Customer service levels

In the introduction we defined the two most commonly used customer service levels in practice, the non-stockout probability α and the fill rate β. In our comparative study we will discuss results for the α-service measure, only. In our base case comparisons the customer service objective is to achieve a non-stockout probability of 95%, i.e.,

$$\alpha^* = 0.95$$

Evaluation of the SCOP concepts

The above case description has been the basis for a numerical study where discrete event simulation was used to compute the performance of the two SCOP concepts. Let us describe the steps along which we derived the numerical results in more detail.

For the SBS concept described in Section 5 we determined the base-stock levels and allocation policies analytically. Towards this end we followed the procedure in De Kok and Visschers (1999) (see Section 5.5) to derive the divergent structures presented in Fig. 6a–c. Given these divergent structures we computed near-optimal linear allocation policies and base stock levels based on the algorithms given in Diks and de Kok (1999). Thereafter we checked these analytical results with discrete event simulations. The

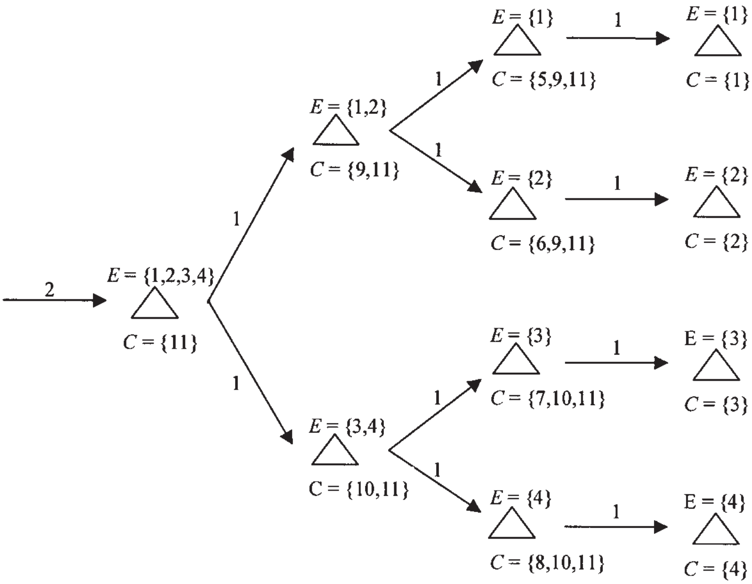

Fig. 6. (a) Decision node network for $(L_f, L_s, L_{sC}, L_c) = (1, 1, 2, 4)$.

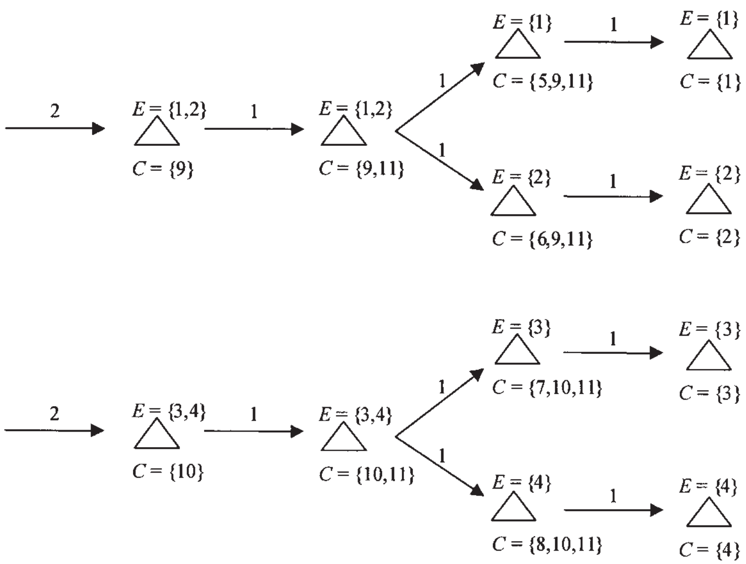

Fig. 6. (b) Decision node network for $(L_f, L_s, L_{sC}, L_c) = (1, 1, 4, 2)$.

simulation run length in all experiments was 100,000 time periods. The point estimates obtained did not change with longer simulations.

For the analysis of the LP concept we had to fully rely on discrete event simulation. In each period simulated we solved the LP problem described in

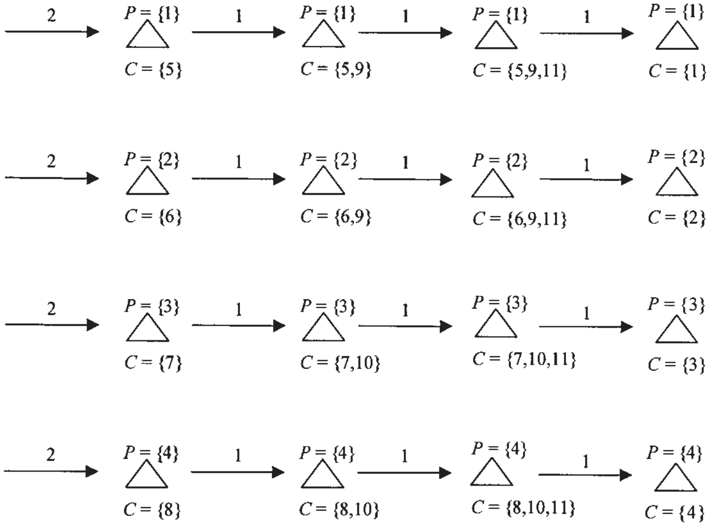

Fig. 6. (c) Decision node network for $(L_f, L_s, L_{sC}, L_c) = (1, 4, 2, 1)$.

Section 4 using CPLEX [cf. http://www.ilog.com]. As described in Section 4 we used an initial run with end-item safety stocks equal to zero to compute for each end-item the empirical distribution of the difference between net stock and safety stock. From this distribution we determined the safety stocks that would ensure the required service level. A second simulation run with the required safety stocks yielded the performance of the LP concept. For both simulation runs we used a run length of 25,000 time periods. This run length proved to be sufficiently long for all experiments to obtain the required accuracy with respect to determination of the end-item safety stocks guaranteeing 0.95 service levels.

6.3 *Quantitative comparison of LP and SBS*

For our base case we generated twelve different SCOP scenarios by combining three planned lead time structures and four squared coefficients of variation. The results are given in Table 3.

From Table 3 we conclude that the analytical results for the supply chain inventory capital obtained with the SBS concept, found in the column with heading SBS_{ana}, coincide with the results obtained with discrete event simulation, found in the column with heading SBS_{sim}. As expected the procedure described above for the LP concept yields the target α-levels (α_{LP}). Likewise the SBS policies computed yield the required customer service levels (α_{SBS}).

Furthermore we find that the SBS concept considerably outperforms the LP-based concept. This in itself is an important and striking result. We should realize ourselves that most commercial software for Supply Chain Planning

Table 3
Supply chain stock capital comparison, identical end-item demand

cv_i^2	(L_f, L_s, L_{sc}, L_c)	Supply chain inventory capital				Customer service	
		SBS_{ana}	SBS_{sim}	LP	$\Delta\%$	α_{SBS} (%)	α_{LP} (%)
0.25	(1,1,2,4)	72,188	71,682	83,225	16	95	95
0.25	(1,4,2,1)	76,154	76,476	83,762	10	95	95
0.25	(1,1,4,2)	74,162	73,550	84,424	15	95	95
0.5	(1,1,2,4)	105,114	104,448	119,645	15	95	95
0.5	(1,4,2,1)	112,226	112,316	121,586	8	95	95
0.5	(1,1,4,2)	108,079	107,616	122,916	14	95	95
1	(1,1,2,4)	152,583	152,203	173,056	14	95	95
1	(1,4,2,1)	165,264	165,328	180,211	9	95	95
1	(1,1,4,2)	157,294	157,034	177,166	13	95	95
2	(1,1,2,4)	217,664	218,551	258,651	18	94	95
2	(1,4,2,1)	246,637	245,998	265,952	8	95	95
2	(1,1,4,2)	228,967	228,789	261,499	14	94	95

currently available is at best employing the LP-based concept of Section 4 and most likely LP is embedded in a set of heuristics to generate a feasible solution to the deterministic SCOP model described in Section 4. This statement is in line with the findings of Stadtler et al (2000) and Fleischmann and Meyr (2003), Chapter 9 of this volume.

Interestingly, the difference between the two concepts is not very sensitive to the demand variability. One might expect that LP performs better if demand variability is low. The difference between the two concepts is mainly determined by the lead time structure. The difference is largest for the case with a long lead time for the common component and smallest for the case with a long lead time for the specific component. Apparently the SBS concept better exploits the commonality in the former situation.

To gain intuition for the surprising observation from Section 4 we present for both concepts the allocation of supply chain inventory capital among stocks of common component, semi-common components, specific component and end-item.

The results in Table 4 indicate that the LP-based concept does not seem to find the right balance in allocating stock among the different items. In case commonality can be exploited, i.e., when (semi-)common components have a longer lead time than the specific component the LP-based concept tends to withhold too much stock in (long-lead time) components. A possible explanation for this is that the LP-based concept aims at satisfying the end-item demand forecast. After satisfying this forecast remaining stocks of components are not used for assembly of end-items, because stocks of components are cheaper than stocks of assembled end-items. Especially if exogenous demand is low in a number of consecutive periods, then the LP concept will tend to build up stock upstream. The base stock levels computed

Table 4
Allocation of inventory capitol along the supply chain

cv_i^2	(L_f, L_s, L_{sc}, L)	End-item		Specific		Semi-common		Common	
		SBS (%)	*LP* (%)	*SBS* (%)	*LP* (%)	*SBS* (%)	*LP* (%)	*SBS* (%)	*LP* (%)
0.25	(1,1,2,4)	95	86	0	1	1	4	3	10
0.50	(1,1,2,4)	97	86	0	1	1	4	2	10
1.00	(1,1,2,4)	99	87	0	1	0	3	1	9
2.00	(1,1,2,4)	100	88	0	1	0	3	0	8
0.25	(1,1,4,2)	93	87	0	1	5	7	2	6
0.50	(1,1,4,2)	94	88	0	1	5	6	1	5
1.00	(1,1,4,2)	96	89	0	1	4	6	0	5
2.00	(1,1,4,2)	97	89	0	1	3	6	0	5
0.25	(1,4,2,1)	88	91	6	3	5	3	1	2
0.50	(1,4,2,1)	90	93	6	2	4	3	0	2
1.00	(1,4,2,1)	90	93	7	2	4	3	0	2
2.00	(1,4,2,1)	91	94	6	2	3	2	0	2

under the SBS concept are typically such that even during low demand periods inventory capital is pushed towards the locations where customer demand must be met. Informally speaking, the base-stock policies have a *just-in-case* character, whereas the LP-based concept has a *just-too-late* character.

Extensive numerical studies indicate that under the SBS concept the sums of base stock levels at each echelon in the divergent structures of decision nodes underlying the SBS concept tends to increase slightly upstream. If these sums had been equal at all echelons then no stocks would be held at all upstream stages. The results in Table 4 seem to indicate that we are close to that situation. This seems to imply that when commonality can be exploited, then only little common upstream stocks are required to reap the benefits.

In case the specific component has the longest lead time, commonality cannot be exploited and in that case the LP-concept seems to stock too much end-item. In that case it may be that the SBS concept identifies indeed that commonality cannot be exploited and favors to hold more (less expensive) components.

The rationale behind the extremely low component stock levels with both concepts, and in particular for the base-stock policies, is that holding back stock at component level does not contribute to immediate customer service. Apparently this outweighs the so-called portfolio effect for common component stocks, i.e., demand for the common component is relatively more stable than demand for individual products.

Another interesting observation from our computational study is that under the LP concept the safety stocks for the identically distributed end-items strongly differ. Our explanation is that for our base case with identical added values for all end-items the LP-problem solved each period is strongly degenerate. Thereby it depends on the particular implementation of the algorithm (e.g. choice of tie-breaking rules), which end-item is favored over the other with respect to the allocation of items. Apparently the CPLEX-solver used is not allocating these mismatches evenly over time among the end-items. Of course, the base-stock policies are identical for all end-items. Given this observation with respect to the allocation of stocks, it is interesting to compare the average stock levels in case the values of end-items are different. In Table 5 we present some results that support our conclusion that the LP concept does not handle component shortages properly. In Table 5 we define Δ_1 as the standard deviation of the three different safety stocks for end-items 1, 2, and 3 normalized by their mean, while Δ_2 is defined as the relative difference between the average safety stock for products 1, 2, and 3 and the safety stock of the lower cost end-item 4.

It follows from Table 5 that the modified base-stock concept yields identical safety stocks for identical products ($\Delta_1 = 0$), while the LP-concept yields considerably different safety stocks ($\Delta_1 > 0$). For different end-items we find that the Δ_2, the difference between the low-cost end-item safety stock and the high-cost end-items safety stock, is much bigger for the LP concept than for the SBS concept.

The results in Tables 3–5 show that the LP-based concept rations shortages among items inappropriately. Identical products are not rationed similarly due to tie-breaking rules needed to deal with the degeneracy of the associated SCOP problem. In case of different products, LP rations shortages according to a priority list based on holding and penalty costs. The priority list implies that the item first on the list is satisfied first (if possible), after that the item second on the list is satisfied, etc. until no inventory is left. Lagodimos (1992) has shown that such a priority rationing mechanism is suboptimal. The linear rationing rules used in the SBS concept ensure that shortages are shared among all products. Apparently this is superior. Informally speaking, LP is a *greedy* approach that is inferior to the *balanced* SBS approach. The importance of careful rationing of shortages of material has, to our

Table 5
Safety stock differences for non-identical end-items

$(cv_1^2, cv_2^2, cv_3^2, cv_4^2)$	(h_1, h_2, h_3, h_4)	(L_f, L_s, L_{sc}, L_c)	LP		SBS	
			Δ_1 (%)	Δ_2 (%)	Δ_1 (%)	Δ_2 (%)
(0.25,0.25,0.25,0.25)	(20,20,20,**10**)	(1,1,4,2)	23	48	0	14
(0.25,0.25,0.25,0.25)	(20,20,20,**10**)	(1,1,2,4)	11	73	0	14
(0.50,0.50,0.50,0.50)	(20,20,20,**10**)	(1,1,2,4)	9	58	0	14
(0.50,0.50,0.50,0.50)	(20,20,20,**10**)	(1,1,4,2)	17	39	0	14
(1,1,1,1)	(20,20,20,**10**)	(1,1,2,4)	11	39	0	13
(1,1,1,1)	(20,20,20,**10**)	(1,1,4,2)	16	29	0	14

knowledge, not been identified in the Mathematical Programming literature on the SCOP problem. An explanation for this may be, that this problem only reveals itself when applying MP-based concepts in a rolling schedule context under stochastic demand. We conjecture here that similar rationing problems should be addressed when resources are constraining the release of orders.

We also considered more complicated product structures and the superiority of the synchronized base-stock concept seems even stronger, i.e., stock capital investment differences become greater as the structure gets more complicated. For a more detailed discussion of a real-world case consisting of 57 items we refer to De Kok (2001).

6.4 Managerial insights

From our comparison of the LP concept and the SBS concept we draw a number of conclusions leading to managerial insights. First we summarize some generic insights that hold for both SCOP concepts discussed. Thereafter we summarize the major distinctions between the two concepts.

Product flow towards the CODP

From the definition of the Customer Order Decoupling Point (CODP) given in Section 1 we derive that the CODP in the networks considered in our comparison is at the end-item level. From the results presented above we conclude that inventory capital is concentrated at end-item level. Since little inventory capital is held at controlled intermediate stock points, we may conclude that optimal solutions can be characterized by items flowing in the supply network towards the CODP's. Intuitively this is in line with notions such as Just-In-Time (JIT), where the objective is to create perfect flow. We should note however that JIT, more precisely Kanban, is a usage-driven control concept, which requires a high degree of usage stability, otherwise such a pull concept cannot guarantee high customer service with low inventories. The optimal SCOP concept enables item flows even under situations with highly volatile demand, because parameters can be set such that low upstream inventories are guaranteed. For example, for the synchronized base stock policies by setting an item base stock level close to the sum of the base stock levels of its successors, one creates almost permanent "shortages" from the successors' point of view, whereby all available item inventory is allocated among the successor items.

An intuitive explanation for this characteristic of optimal policies (within the concepts considered) is as follows. When allocating inventory capital among all items in the supply network a trade-off must be made between customer service and inventory capital cost. By definition customer service is realized only by availability at the CODPs. When balancing inventory capital between the CODPs and upstream item level one expects that, because of the portfolio effect, an increase of inventory capital at (common) upstream item levels enables a decrease of inventory capital at the CODP level. However such a decrease must not lead to a decrease of customer service below the target levels. The results presented above *apparently* show that the pull of inventory towards the CODPs to ensure high service is stronger than the pull of inventory towards upstream items to reap the benefits of the portfolio effect.

Stochastic control concept outperforms MP-based control concept

The results clearly show the superiority of the SBS concept over the LP-based rolling schedule concept. This is an observation with far reaching consequences. Currently, all commercially available SCP software is based on objective functions and constraints that are within the MP realm. Apparently a control policy based on optimization of a deterministic model, within which stochastic demand is only represented by average values, yields release decisions over time that are suboptimal. Again this is counterintuitive to most people not familiar with stochastic models.

The good news is that the SBS policies are remarkably simple from a computational point of view. Supposing the SBS concept can be extended to real-life situations, this may enable interactive SCOP. Such is currently infeasible due to the running times of MP-based software [cf. Schalla (2001)].

Another feature of the SBS policy is that it is based on a (to some extent artificial) decision hierarchy that provides insight into the SCOP problem that cannot be derived from an MP modeling approach. The hierarchy, expressed through the divergent systems of decision nodes associated with sets of components and end-items, gives guidelines for the different aggregation levels for which forecasts of future demand are needed. By choosing the appropriate aggregations of future demand over time it may be possible to exploit the (likely higher) accuracy of aggregate forecasts.

The consequence of dealing with uncertainty outside the MP models is that safety stocks and safety lead times are human input. The above discussion of (safety) stock positioning in the supply network shows that setting safety stocks without decision support of a stochastic model to capture the complex interactions between items is beyond a human's capability.

Clearly, the results obtained have been derived for stationary demand, only. In our view the stationary demand process is not the main reason for the results obtained. Both classes of policies can easily be modified to take into account non-stationary demand. In that case we must assume stationary forecast errors, which is no real restriction. The forecast revision process proposed in Graves et al. (1998) seems to be an interesting candidate for modeling non-stationary demand. Further research is needed to support our claim.

Hybrid approaches

The comparison has been restricted to uncapacitated problems. The reason for this is that no stochastic control concepts are available in the literature that apply to general capacitated networks. The LP-based rolling schedule can be applied to capacitated problems. The insights discussed above seem to indicate the need for the development of a hybrid approach. If the rationing mechanism in the SBS concept is one of the main reasons for its superiority, it seems logical to implement the linear rationing rules into the LP-based concept in the form of additional constraints. In addition the linear rationing mechanism could be applied in case of binding resource constraints, as well. These are topics for further research.

This concludes the analysis of SCOP concepts. In the next section we discuss topics for further research that we identified from this research.

7 Summary and issues for further research

In this chapter we have discussed various Supply Chain Operations Planning concepts defined in the literature. We provided a generic setting for the SCOP problem by defining decision variables and state variables. We motivated a generic formulation of the SCOP problem driven by a hierarchical planning approach. We argued that the concept of planned lead times is the building block for making a clear distinction between different planning levels that does justice to the modeling of information asymmetry. Information asymmetry is inevitable because related planning decisions are made at different moments in time and are made in different parts of an organization with a different view of the state of the system.

We have formulated sets of constraints for release of materials and resources that have been used as the basis for an assessment of these SCOP concepts. We identified two major classes of SCOP concepts that have been developed using two different modeling perspectives:

- Mathematical Programming models embedded in a rolling schedule approach
- Stochastic models that incorporate random demand

Based on our perspective of the SCOP problem we formulated a capacitated multi-item multi-stage LP model that lends itself for straightforward optimization using commercially available software. Extensions to the problem with lot sizing restrictions have been briefly discussed.

An extensive discussion of stochastic models identified that no literature is available on uncapacitated general supply networks, and then presented the various approaches for uncapacitated models. It turned out that in the literature mostly base stock control concepts have been proposed, either based on the installation stock concept or based on the echelon stock concept. We discussed the infeasibility of pure base stock policies for the SCOP problem for general supply networks. The reason for this is the violation of material availability constraints, i.e., pure base stock policies generate material orders without checking availability of upstream inventories. A synchronized base stock (SBS) control concept has been discussed that solves this problem at the expense of suboptimality. The suboptimality is caused by the integration of material allocation and material ordering decisions in the sense that availability of upstream item inventories is checked after taking allocation decisions of these upstream items among their successors in case of shortages. It has been shown that the SBS concept allows for an exact analysis of general supply networks and further permits the alignment of supply chain design and supply chain planning under the assumption of stationary demand.

The high level of complexity of the SCOP problem for general supply networks under random demand explains why currently available literature mostly proposes a heuristic analysis. We distinguished between heuristics that assume 100% upstream availability (decomposition assumption) and heuristics that assume that in case of lack of availability, there is exactly one item causing this (assembly assumption). We found that the decomposition assumption implies relatively high inventory levels at all echelons of the supply network, which is in contrast with the characteristics of near-optimal solutions for divergent networks that indicate that the inventory levels at upstream echelons should be low. We note here that the above statement does not apply to the analysis of Inderfurth and Minner (1998) and Graves and Willems (2000). This is because their analysis allows covering potential shortages at some stockpoint by downstream and upstream stockpoints. The heuristic developed by Ettl et al. (2000) that was based on the assembly assumption yields optimal solutions that violate the assembly assumption, because it finds solutions with low upstream fill rates. Interestingly, Ettl et al. (2000) report that the solutions found could not be improved by discrete event simulation based methods.

Based on a set of cases we compared the SBS concept with the LP-based concept. Surprisingly, the SBS concept outperformed the LP-based concept considerably. This indicates the importance of further study into stochastic models for the SCOP problem. Explanation for the superiority of the SBS concept was found in the way LP tends to prioritize items in case of shortages of upstream availability instead of rationing among the items that need this upstream availability. Also LP tends to keep too much inventory capital upstream, because a deterministic objective function identifies upstream stages as attractive due to lower cumulative added value.

Based on our discussion we have identified several areas for further research. We will subsequently discuss empirical validation of SCOP models, capacitated stochastic demand models for SCOP in general networks, incorporation of non-stationary demand into the SBS concept, comparison of SCOP concepts and integration of SCOP and shopfloor scheduling.

7.1 Empirical validation of SCOP models

Our discussion of the various SCOP models is strongly based on a mathematical perspective. This implies that we discussed the SCOP model definition and assessed the validity of certain assumptions that enabled us to simplify the analysis of the model. Wherever possible, we used a mathematically rigorous approach. If this was not possible we used discrete event simulation to assess validity. Although this is a scientifically sound approach, it fails to answer a fundamental question, which is: *Are the results of the model analysis in line with empirical data from the real-life SCOP problem?* In theory it may be that a mathematically sound analysis of an SCOP problem yields worse results than a heuristic analysis of the same problem. If such is the case this raises new and scientifically relevant research questions about modeling. In general we find that most of the research about SCOP in general supply networks lacks evidence that managerial insights obtained are supported by empirical validation. Even if successful cases are reported it is questionable whether there is a causal relationship between the use of specific models for SCOP and this success. We advocate the careful design of scientifically sound empirical studies, where such a causal relationship can be tested. The existence of ERP systems allows for the analysis of transactional data over time, from which we can obtain insight into the behavior of demand, inventories and lead time. Such research will be time consuming and difficult, due to the fact that the experimental setting cannot be fully controlled. Such research will be rewarding in terms of deeper insights into real-world SCOP problems and the contribution of quantitative modeling to its solution.

7.2 Incorporation of non-stationary demand

From experiences with real-life SCOP problems we conclude that in many situations we are faced with non-stationary demand. Examples are seasonality of demand, new product introductions and old product phase-outs. Graves et al. (1998) provide an interesting model for such non-stationarity. It assumes the capability of humans (possibly supported by software) to update forecasting information in such a way that the revisions of the forecast for a particular period into the future from one period to the next are i.i.d. Such models of non-stationary demand may still allow for a mathematically rigorous analysis of the SCOP problem.

If demands are correlated over time the question arises even for serial supply chains whether base stock policies are optimal and if so, whether the optimal policy can be found by solving the generalized Newsboy equations as derived in Diks and de Kok (1998). The issue of demand correlation over time is quite relevant, since most standard forecasting methods, including simple exponential smoothing, induce forecast errors that are correlated over time.

7.3 Capacitated models with stochastic demand

The above issue of empirical validation is quite relevant for the development of models for capacitated supply networks. The generalized Kanban control policies discussed in Section 5.8 have been proposed for the control of manufacturing systems. Typically resources are modeled as queues.

Hence the analysis of such policies integrates material control and resource usage. However, SCOP *plans* resource usage in order to smooth capacity requirements, while maintaining the due dates set by the planned lead times. This implies that a queuing network analysis based on continuous review and FCFS at resources does not represent properly the planning characteristics and periodic nature of SCOP.

Still, queuing network analysis may be a starting point for both the determination of planned lead times and a heuristic analysis of the capacitated SCOP problem under stochastic demand. The main idea behind this is that in most real-life situations capacity is hard to define. Processes can be speeded up if necessary; resources can be reallocated to provide more capacity to specific capacity requirements. This observation is one of the reasons to decompose the SCOP problem and the short-term scheduling problem. Thus capacitated SCOP models may be based on a similar decomposition of the control of material order releases and resource releases. This would result into a queuing network analysis of the resources that provides realistic planned lead times. The justification of this idea requires an experimental and empirical validation as discussed above.

We discussed non-stationarity of demand. Similarly, resource availability may be non-stationary due to preventive maintenance and holidays. The non-stationarity of resources does not fit a typical queuing (network) analysis. It seems to us that this problem is a white spot in the literature on stochastic models.

7.4 Comparison of SCOP concepts

The fact that there is a surge in the implementation of commercial software (such as Advanced Planning and Scheduling (APS) Systems) for SCOP motivates a scientific assessment of the various SCOP concepts implemented. Our experience shows that many people involved with the implementations of APS Systems do not really understand the mismatch between the optimality notions from deterministic optimization and the optimality notions from stochastic models. The discussion in Section 6 provides solid ground for a critical assessment of software based on LP, MIP, and rule-based optimization, since the common denominator of these approaches is the deterministic world view within a rolling schedule concept. In our opinion the results of our comparison surfaced some planning principles that allows for implementation in commercial software, yet further research is necessary to test the hybrid SCOP concepts suggested in Section 6.

We argued in this chapter that lead times need to be exogenous to the SCOP concept, implying that the system needs to take care of controlling lead times such that they are more or less fixed. In a supply chain context, it has not yet been researched whether this actually does provide better results than working with variable lead times.

Finally, we notice that none of the research conducted has been studying the SCOP concept performance under dynamic conditions. Dynamic conditions do not only refer to the non-stationarity of demand (or forecast error) and resource availability, but also to the dynamics in the planning process. In our exposition in Section 2, we have discussed the importance of anticipating future events. We know that estimates will never be completely correct. Further, we know from the research in the field of Systems Dynamics [e.g., Sterman (2000)] that even small differences between anticipated values of variables, perceived values of variables and actual values of variables may yield very unstable systems and uncontrolled planning situations. This dynamic behavior of the concepts discussed and proposed in this chapter, will need to be investigated.

Acknowledgements

This chapter would not have been possible without the support of many of our colleagues (and friends). First, we thank the Operations Planning and Control group at Eindhoven: Henk Akkermans, Will Bertrand, Rob Broekmeulen, Nico Dellaert, Karel van Donselaar, Simme Douwe Flapper, Geert-Jan van Houturn, Gudrun Kiesmüller, Henny van Ooijen, Graham Sharman and Vincent Wiers. Second, we are indebted to the PhD students who contributed all their intellectual capabilities during the 1990s to SCM research: Jos Verrijdt, Erik Diks, Jannet van Zante-De Fokkert, Fred Janssen, Wenny Raaymakers, Ebbe Negenman, Mark Euwe, Anastasia Schalla and Sanne Smits. Third, we would like to thank our German colleagues with whom we have debated while working on this chapter and who helped us to set our research agenda for the next five years: Bernhard Fleischmann, Herbert Meyr, Hartmut Stadtler and Horst Tempelmeier. And last, but not least, we would like to express our thanks to the imaginative minds we met in the industry for their invaluable support, that enabled us to implement our ideas so that we could gather empirical evidence of their validity: Wim Vrijland, Steve Martin, Fred Janssen (again!), Jan van Doremalen, Mynt Zijlstra, Han Langereis, Erik van Wachem, Jeroen Bordewijk, Dave Grandt, Alessandro de Luca and the members of the European Supply Chain Forum.

References

Agrawal, N, M. Cohen (2001). Optimal material control in an assembly system with component commonality. *Naval Research Logistics* 48, 408–429.

Anthony, R.N. (1965). Planning and control systems; *A framework for analysis,* Graduate School of Business Administration, Harvard University, Boston.

Aviv, Y. (2001). The effect of collaborative forecasting on supply chain performance. *Management Science* 47, 1326–1343.

Axsater, S. (2000). *Inventory Control*, Boston, Kluwer.

Axsater, S. (2003). Supply chain operations: Serial and distribution inventory systems, in: A. G. de Kok, S. C. Graves, (eds). *Handbooks in Oper. Res. And Management Sci*, Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Chapter 10, North-Holland Publishing Company, Amsterdam, The Netherlands.

Baker, K.R. (1993). Requirements planning, in: S.C. Graves, A.H.G. Rinnooy Kan, P.H. Zipkin (eds.), *Logistics of Production and Inventory*, Amsterdam, North-Holland Publishing Company Amsterdam, The Netherlands, pp. 571–628.

Barbarosoğlu, G., D. Özgür (1999). Hierarchical design of an integrated production and 2-echelon distribution system. *European Journal of Operational Research* 118, 464–484.

Belvaux, G., L.A. Wolsey (2001). Modeling practical lot-sizing problems as mixed integer programs. *Management Science* 47, 993–1007.

Bertrand, J.W.M. (2003). Supply Chain Design: Flexibility Considerations, in: A. G. de Kok S. C. Graves (eds.), *Handbooks in Oper. Res. And Management Sci*, Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Chapter 4, North-Holland Publishing Company, Amsterdam, The Netherlands.

Bertrand, J.W.M., J. C. Wortmann (1981). *Production control and information systems for component manufacturing shops*, Amsterdam, Elsevier.

Bertrand, J.W.M., J.C. Wortmann, J. Wijngaard (1990). *Production control: a structural and design oriented approach*, Amsterdam, Elsevier.

Billington, P.J., J.O. McClain, L.J. Thomas (1983). Mathematical programming approaches to capacity constrained MRP systems: review, formulation and problem reduction. *Management Science* 29, 1126–1141.

Bitran, G.R., D. Tirupati (1993). Hierarchical production planning, in: S.C. Graves, A.H.G. Rinnooy Kan, P.H. Zipkin (eds.), *Logistics of Production and Inventory*, Amsterdam, North-Holland, pp. 523–568.

Bitran, G.R., A.C. Hax (1977). On the design of hierarchical production planning systems. *Decision Sciences* 8(1), 28–55.

Buzacott, J.A. (1989). Queuing models of Kanban and MRP controlled production systems. *Engineering Costs and Production Economics* 17, 3–20.

Buzacott, J.A. J.G. Shantikumar, (1993). *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewoods Cliffs, N.J.

Dauzere-Peres, S., J.B. Lasserre (1994). Integration of lotsizing and scheduling decisions in a job-shop. *European Journal of Operational Research* 75, 413–426.

De Kok, A.G. (1989). A moment-iteration method for approximating the waiting-time characteristics of the GI/G/1 queue. *Prob. Eng. and Inf. Sc*. 3, 273–287.

De Kok, A.G. (2000). Capacity allocation and outsourcing in a process industry. *International Journal Of Production Economics* 68, 229–239.

De Kok, A.G. (2001). Comparison of Supply Chain Planning concepts, Working Paper TUE/TM/LBS/101-03. Eindhoven: Technische Universiteit Eindhoven.

De Kok, A.G. (2002a). Ruin probabilities with compounding assets for discrete time finite horizon problems, independent period claim sizes and general premium structure, BETA Working Paper 82. Eindhoven: Technische Universiteit Eindhoven accepted for publication in: Insurance, Mathametics and Economics.

De Kok, A.G. (2002b). Evaluation And Optimization Of Strongly Ideal Assemble-To-Order Systems, in: Shanthikumar, J.G., Yao, D.D. and Zijm, W.H.M. (eds), Stochastic Modeling and Optimization of Manufacturing Systems and Supply chains, Kluwer, International series in Operations Research and Management Science, 66, 2003.

De Kok, A.G. and Seidel, H.P. (1990). Analysis of Stock Allocation in a 2-echelon Distribution System, Technical Report 098, Eindhoven: CQM.

De Kok, A.G., J.W.C.H. Visschers (1999). Analysis of assembly systems with service level constraints. *International Journal of Production Economics* 59, 313–326.

Dellaert, N.P., A.G. de Kok, W. Wang (2000). Push and pull strategies in multi-stage assembly systems. *Statistica Neerlandica* 54, 175–189.

Diks, E.B. (1997). Controlling Divergent Multi-echelon Systems, Ph.D. thesis, Eindhoven University of Technology, The Netherlands.

Diks, E.B., A.G. de Kok (1998). Optimal control of a divergent N-echelon inventory system. *European Journal of Operational Research* 111, 75–97.

Diks, E.B., A.G. de Kok (1999). Computational results for the control of a divergent *N*-echelon inventory system. *International Journal of Production Economics* 59, 327–336.

Diks, E.B., A.G. de Kok, A.G. Lagodimos (1996). Multi-echelon systems: A service measure perspective. *European Journal Operational Research* 95, 241–263.

Erenguc, S.S., N.C. Simpson, A.J. Vakharia (1999). Integrated production/distribution planning in supply chains: an invited review. *European Journal of Operational Research* 115, 219–236.

Eppen, G., L. Schrage (1981) Centralized ordering policies in a multi-warehouse system with lead times and random demand, in: L.B. Schwarz, (eds.), Multi-level Production-Inventory Control Systems: theory and Practice, North-Holland, Amsterdam.

Ettl, M., G.E. Feigin, G.Y. Lin, D.D. Yao (2000). A supply network model with base-stock control and service requirements. *Operations Research* 48, 216–232.

Feigin, G.E. (1998). Inventory planning in large assembly supply chains, in: S. Tayur, R. Ganeshan, M. Magazine (eds.), *Quantitative Methods for Supply Chain Management*, Boston, Kluwer Academic Publishers, pp. 760–788.

Fleischmann, B., H. Meyr (2003) Planning hierarchy, modeling, and advanced planning systems, in: A. G. de Kok, S. C. Graves, (eds.), *Handbooks in Oper. Res. And Management Sci*,. Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Chapter 9, North-Holland Publishing Company, Amsterdam, The Netherlands.

Fransoo, J.C., M.J.F. Wouters, A.G. de Kok (2001). Mufti-echelon multi-company inventory planning with limited information exchange. *Journal of the Operational Research Society* 52, 830–838.

Frein, Y., M. Di Mascola, Y. Dallery (1995). On the design of generalized kanban control systems. *International Journal of Operations and Production Management* 15(9), 158–184.

Gershwin, S.B. (1994). *Manufacturing Systems Engineering*, Prentice Hall, Englewood Cliffs.

Graves, S.C., S.P. Willems (2000). Optimizing strategic safety stock placement in supply chains. *Manufacturing and Service Operations Management* 2, 68–83.

Graves, S.C., Willems, S.P. (2003). Supply Chain Design – safety stock placement, inventory hedges for buffering against demand and supply uncertainty, in: A. G. de Kok, S. C. Graves, (eds.), *Handbooks in Oper Res. And Management Sci*., Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Chapter 3, North-Holland Publishing Company, Amsterdam, The Netherlands.

Graves, S.C., D.B. Kletter, W.B. Hetzel (1998). A dynamic model for requirements planning with application to supply chain optimization. *Operations Research* 46, S35–S49.

Graves, S.C., H.C. Meal, S. Dasu, Y. Qui (1986). Two-stage production planning in a dynamic environment, in: S. Axsater, Ch. Schneeweiss, E. Silver (eds.), *Multi-stage production planning and inventory control*, Berlin, Springer, pp. 9–43.

Hausman, W.H., H.L. Lee, A.X. Zhang (1998). Order response time reliability in multi-item inventory systems. *European Journal of Operational Research* 109, 646–659.

Hax, A.C., H.C. Meal (1975). Hierarchical integration of production planning and scheduling, in: M.A. Geisler (ed.), *Logistics*, Amsterdam, North Holland Publishing Company, Amsterdam, The Netherlands pp. 53–69.

Hillier, M.S. (2000). Component commonality in multiple-period assemble-to-order systems. *IIE Transactions* 32, 755–766.

(1991S. Hoekstra, J.H.J.M. Romme (eds.), *Integral logistic structures: developing customer-oriented goods flow*, London, McGraw-Hill.

Holt, C.C., F. Modigliani, J.F. Muth, H.A. Simon (1960). *Planning, Production, Inventories and Workforce*, Prentice Hall, Englewood Cliffs.

Hopp, W., M. Spearman (2000). *Factory Physics*, 2nd ed., Irwin McGraw-Hil Bostonl.

Inderfurth, K. (1994). Safety stocks in multistage divergent inventory systems: a survey. *International Journal of Production Economics* 35, 321–329.

Inderfurth, K., S. Minner (1998). Safety stocks in mufti-stage inventory systems under different service measures. *European Journal of Operational Research* 106, 57–73.

Janssen, F.B.L.S.P. (1998). Inventory Management Systems, unpublished PhD. Thesis, Tilburg University, Tilburg.

Kanet, J.J., S.V. Sridharan (1998). The value of using scheduling information in planning material requirements. *Decision Sciences* 29(2), 479–497.

Køhler-Gudum, C.K., and A.G. De Kok (2002). A safety stock adjustment procedure to enable target service levels in simulation of generic inventory systems, BETA Working Paper 71. Eindhoven: Technische Universiteit Eindhoven.

Lagodimos, A.G. (1992). Multi-echelon service models for inventory systems under different rationing policies. *International Journal of Production Research* 30, 939–958.

Lagodimos, A.G. (1993). Models for evaluating the performance of serial and assembly MRP systems. *European Journal of Operational Research* 68, 49–68.

Langenhof, L.J.G., W.H.M. Zijm (1990). An analytical theory of multi-echelon production/ distribution systems. *Statistica Neerlandica* 44, 149–174.

Lawrence, S.R. (1997). Heuristic, optimal, static and dynamic schedules when processing times are uncertain. *Journal of Operations Management* 15(1), 71–82.

Lee, H.L., C. S. Tang (1996). Modelling the costs and benefits of delayed product differentiation. *Management Science* 43, 40–53.

Lee, H.L., J.M. Swaminathan (2003). Design for postponement, A.G. de Kok, S. C. Graves (eds.),. *Handbooks in Oper. Res And Management Sci.*, Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Chapter 5, North-Holland Publishing Company, Amsterdam, The Netherlands.

Magee, J.F. (1958). *Production Planning and Inventory Control*, New York, McGraw-Hill.

McKay, K.N., F.R. Safayeni, J.A. Buzacott (1995). A review of hierarchical production planning and its applicability for modern manufacturing. *Production Planning & Control* 6(5), 384–394.

McPherson, R.F., White, K.P., Jr. (1994). Management control and the manufacturing hierarchy: Managing integrated manufacturing organizations. *International Journal of Human Factors in Manufacturing* 4(2), 121–144.

Meal, H.C. (1984). Putting production decisions where they belong. *Harvard Business Review* 62(2), 102–111.

Meal, H.C., M.H. Wachter, D.C. Whybark (1987). Material requirements planning in hierarchical production planning systems. *International Journal of Production Research* 25(7), 947–956.

Miller, T. (2001). *Hierarchical Operations and Supply Chain Planning*, London, Springer.

Minner, S. (2000). *Strategic Safety Stocks in Supply Chains*, Berlin, Springer.

Orlicky, J.A. (1975). *Material Requirements Planning*, New York, McGraw-Hill.

Özdamar, L., G. Barbarosoglu (2000). An integrated Lagrangean relaxation-simulated annealing approach to the multi-level multi-item capacitated lot sizing problem. *International Journal of Production Economics* 68, 319–331.

Rosenblatt, M.J., A. Eynan (1996). Component commonality effects on inventory costs. *IIE Transactions* 28, 93–104.

Rosling, K. (1989). Optimal inventory policies for assembly systems under random demands. *Operations Research* 37, 565–579.

Schalla, A.J., J.C. Fransoo, and A.G. de Kok (2001). Hierarchical anticipation in Advanced Planning and Scheduling Systems. Working Paper TUE/TM/LBS/01-02. Eindhoven: Technische Universiteit Eindhoven.

Schneeweiss, C. (1999). *Hierarchies in Distributed Decision Making*, Berlin, Springer.

Shapiro. J.F. (1993). Mathematical Programming Models and Methods for Production Planning and Scheduling, in: Graves, S.C., A.H.G. Rinnooy Kan, and P.H. Zipkin (eds.), *Logistics of Production and Inventory*. Amsterdam: North-Holland, 371–444.

Sherbrooke, C. C. (1992). *Optimal Inventory Modeling of Systems, New Dimensions in Engineering*, New York, Wiley.

Silver, E.A., D.F. Pyke, R. Peterson (1998). *Inventory Management and Production Planning and Scheduling*, New York, Wiley.

Simpson, K.F. (1958). In-process inventories. *Operations Research* 6, 863–873.

Song, J.S. (1998). On the order fill rate in a multi-item, base-stock inventory system. *Operations Research* 46, 831–845.

Song, J.S. and P.H. Zipkin (2003). Supply Chain Operations: Assemble-to-Order Systems. A. G. de Kok, S. C. Graves, (eds.) *Handbooks in Oper. Res. And Management Sci.* Vol. 11, *Supply Chain Management: Design, Coordination and Operation*, Ch. 11. North-Holland Publishing Company, Amsterdam, The Netherlands.

Spengler, Th., H. Puchert, T. Penkuhn, O. Rentz (1997). Environmental integrated production and recycling management. *European Journal of Operational Research* 97, 308–326.

(2000H. Stadtler, C. Kilger (eds.), *Supply Chain Management and Advanced Planning: Concepts, Models, Software and Case Studies*, Berlin, Springer.

Sterman, J.D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Boston, Irwin McGraw-Hill.

Suri, R., J.L. Sanders and M. Kamath (1993). Performance evaluation of production networks, in: Graves, S.C., A.H.G. Rinnooy Kan, and P.H. Zipkin (eds.), *Logistics of Production and Inventory*. Amsterdam: North-Holland, 199–286.

Tardiff, V. (1995). Detecting Scheduling Infeasibilities in Multi-Stage Finite Capacity Production Environments. Unpublished PhD Dissertation, Evanston: Northwestern University.

Tayur, S.R. (1993). Computing the optimal policy for capacitated inventory models. *Communications in Statistics-Stochastic Models* 9, 585–598.

Van der Heijden, M.C. (1997). Supply rationing in multi-echelon divergent systems. *European Journal of Operational Research* 101, 532–549.

Van der Heijden, M.C., E.B. Diks, A.G. de Kok (1997). Stock allocation in general multi-echelon distribution systems with (RS) order-up-to-policies. *International Journal of Production Economics* 49, 157–174.

Van Houtum, G.J., W.H.M. Zijm (1991). Computational procedures for stochastic multi-echelon production systems. *International Journal of Production Economics* 23, 223–237.

Van Houtum, G.J., W.H.M. Zijm (2000). On the relation between cost and service models for general inventory systems. *Statistica Neerlandica* 54, 127–147.

Van Ooijen, H.P.G. (1991). Controlling different flow rates in job-shop like production departments. *International Journal of Production Economics* 23, 239–249.

Whybark, D.C., S. Yang (1996). Positioning inventory in distribution systems. *International Journal of Production Economics* 45, 271–278.

Wiendahl H.-P.(1987). Belastungsorientierte Fertigigungssteuerung Grundlagen, Verfahrensaufbau, Realisierung. Muenchen: Hanser (in German).

Wiendahl, H.-P. (1995). *Load-Oriented Manufacturing Control*, Berlin, Springer.

Winter, R. (1989). Der Ansatz des Massachusetts Institute of Technology zur Mehrstufigen Produktionsplanung. Arbeitsbericht 89-01. Frankfurt: Institut für Wirtschaftsinformatik, Johann Wolfgang Goethe Universitat (in German).

Yano, C., H. Lee (1995). Lot sizing with random yields: a review. *Operations Research* 43, 311–334.