2. Supply Chain Planning and Demand Fulfillment

This chapter summarizes the background of the DMC problem. Moving from rather general aspects to more specific topics, Section 2.1 will start with an overview of **supply chain planning** (SCP). This section will be used to introduce the concept of **hierarchical planning** and to characterize the key relations between the individual SCP tasks.

Since the discussion in this thesis is limited to MTS environments, all production processes are driven by forecasts and final items are sold from stock. In such MTS environments, the interrelations between three major SCP tasks need to be characterized to illustrate the background of the DMC problem:

- In Section 2.2, an overview of key aspects of **demand planning** will be presented; in particular, **hierarchical forecasting** will be addressed.

- Then, in Section 2.3, **master planning** will be discussed. Based on the inputs provided by demand planning, master planning determines a mid-term forecast-driven aggregate plan for procurement, production, distribution and sales.

- **Demand fulfillment**, i.e. the order-driven processes in an MTS supply chain will be covered in Section 2.4. In particular, an overview will be provided illustrating how current demand fulfillment systems handle customer heterogeneity in MTS environments, and a comprehensive review of the existing literature contributions will be given.

Finally, Section 2.5 offers a brief summary and concluding remarks. Overall, the analysis will confirm that the main issues associated with the DMC problem have not yet been addressed thoroughly in the literature, preparing the ground for the contributions in the subsequent chapters.

2.1. Supply Chain Planning

The following sections will provide an overview of planning concepts in a supply chain. As a starting point, the objectives and tasks of *supply chain management* will be introduced in Section 2.1.1, allowing for a further characterization of the DMC problem as an intra-organizational channel coordination problem. An appropriate planning concept, both for inter-company and intra-organizational supply chains, is *hierarchical planning*. It takes

S. Vogel, *Demand Fulfillment in Multi-Stage Customer Hierarchies,* Produktion und Logistik,
DOI 10.1007/978-3-658-02864-0_2, © Springer Fachmedien Wiesbaden 2014

care of the interrelations between the individual planning tasks at different levels. A brief introduction to hierarchical planning will be provided in Section 2.1.2.

In Section 2.1.3, the interrelations between the individual planning levels will be discussed. An intuitive framework for this is the *supply chain planning matrix* (SCPM). It arranges the SCP tasks along the dimensions *planning level* and *supply chain processes*. Unfortunately, an important characteristic of all supply chains, the customer order decoupling point (CODP) is not reflected in the SCPM. The CODP separates forecast-driven from order-based tasks, and it will be covered in Section 2.1.4.

Almost all planning tasks in today's supply chains can be supported with software modules which are part of APS. APS are powerful implementations of the hierarchical planning logic. Since some of the ideas developed in the subsequent chapters to solve the DMC problem may be integrated into such systems, Section 2.1.5 will provide a short overview of APS.

2.1.1. Supply Chain Management

In a very basic sense, a *supply chain* consists of "two or more parties linked by a flow of goods, information, and funds." (Tsay et al., 1999, p. 301). These parties are typically involved in four types of key activities: *Procuring* necessary raw materials, transforming them into semi-finished and finished products in a series of *production* steps and finally, *distributing* and *selling* these products to the end customers (Lee and Billington, 1993). These activities have to be aligned closely to ensure that individual customer needs can be fulfilled in the best possible manner.

This alignment is usually referred to as *supply chain management*. The **objective of SCM** is to coordinate these aforementioned activities and to manage the relationships between the involved entities. The ultimate goal is to deliver superior customer value at fewer costs to the whole supply chain (see Christopher, 1998, p. 18). The breadth of tasks involved suggests that SCM comprises both a design and an execution perspective.[1]

A key characteristic of SCM is its focus on the collaboration of multiple parties. SCM has risen to prominence as traditional production settings based on vertically integrated companies have been gradually replaced by a sequence—or chain—of multiple parties working together. It is their joint effort which is required in modern industrial production settings.

Many drivers for the establishment of supply chains and of SCM originate from the market environment, for example the globalization of many markets. This qualitative change from traditional production environments to supply chains has been accompanied by a trend towards better customer orientation, resulting in an explosion of product variants, shorter product life cycles and more complex products. Many companies have responded to this challenge by specializing and concentrating on their core competencies (see Prahalad and Hamel, 1990). This has resulted in the participation of more and of many separate economic and legal entities in the overall process of value creation.

[1] For more comprehensive discussions on the term SCM and particularly its relationship to logistics, see Cooper et al. (1997) or Mentzer et al. (2001).

Therefore, most SCM initiatives focus on facilitating cooperation and on coordinating decisions at the interfaces to other enterprises along the logistical chain (Zimmer, 2002, p. 1).[2]

At an aggregate level, the key **tasks of SCM** are to reduce costs, particularly with respect to inventory, to gain efficiency in operations and to improve customer service (Lee and Billington, 1995). Essentially, SCM is concerned with determining the trade-offs between these apparently conflicting goals. Improving customer service and operations while at the same time preventing inventory levels from soaring requires a sophisticated and coordinated effort. For this purpose, SCM incorporates a broad spectrum of managerial decisions. Long-term strategic deliberations must be addressed while at the same time important tactical planning activities must be taken care of. This breadth of planning tasks can be handled with the hierarchical planning framework which will be discussed shortly, in Section 2.1.2.

The DMC Problem as an Intra-Organizational Channel Coordination Problem

As stated above, at the heart of most SCM issues lies the problem of coordinating multiple entities fulfilling a variety of tasks. A typical example is the *channel coordination* problem. In a particular sales and distribution channel, a number of independent supply chain entities such as manufacturer, wholesaler and retailer are collectively involved in bringing a particular product to market. The entities are independent since there is often no central authority which can exert discretionary power. Hence coordination is required as the entities differ in terms of their objectives, information endowments or general market power.

The main planning problem consists of incentivizing the independent entities to cooperate for their mutual benefit. A typical phenomenon in uncoordinated sales and distribution channels is *double marginalization* (see Spengler, 1950). Here, independent price setting decisions are made both by the wholesaler and by the retailer. Since each party only focuses on individual profit maximization, this not only jeopardizes overall supply chain profits, but also leads to individually disadvantageous results (e.g. see Corbett and Tang, 1999). Due to the lack of a central coordinating authority, such problems can only be mitigated by proper mutual contracts which align incentives and which encourage information sharing (e.g. see Cachon, 2003). Channel coordination is achieved if all parties involved independently make decisions which maximize joint profits (Barnes-Schuster et al., 2002, p. 173).

A typical by-product of missing channel coordination is the build-up of large and costly inventory positions at each of the involved supply chain entities. These excess inventories result from distorted aggregate demand signals and disproportionate ordering. Such distortions can occur if the demand variability increases from the perspective of more

[2] The related term ‘demand chain management’, while more appropriate to describe the market-related activities (see Selen and Soliman, 2002), was never accepted in literature and practice.

upstream supply chain entities. Forrester (1958, 1961) was the first to observe this phenomenon which is commonly referred to as the *bullwhip effect*. In a seminal paper, Lee et al. (1997a) identified and analyzed four major sources of the bullwhip effect in supply chains:

- **Myopic processing of demand signals** as end customer demand is invisible to intermediate supply chain entities,
- **rationing games** due to proportional allocations and unrestricted return policies,
- **order batching** and
- frequent **price variations**.

The literature on the bullwhip effect and on its prevention has increased beyond measure in recent years. Many theoretical contributions highlight the importance of centralized managerial control to solve such coordination problems. Yet, this is only rarely feasible. In the absence of a central coordination authority, most authors agree that increased transparency, information sharing and an alignment of incentives constitute key measures to prevent the bullwhip effect in supply chains and to ensure channel coordination (e.g. Lee et al., 1997b).

However, this result is not limited to supply chains consisting of separate (legal) entities. Problematic situations which are conceptually similar to the channel coordination problem may also arise in *intra-organizational* settings. If individual entities with private information and selfish behavior make uncoordinated decisions, phenomena which are similar to the bullwhip effect may also occur within organizations. In particular, this is the case for the DMC problem which was introduced in the previous chapter. This problem is characterized by a two aspects which closely mirror the root causes of the bullwhip effect in supply chains:

- Due to the salesforce composite forecasting method, the **true demand signal** from the leaf nodes is typically *not observable* at higher levels in the customer hierarchy.
- **Decentral information** in the customer hierarchy may lead to shortage or **rationing gaming**.[3]

While some sales staff in a customer hierarchy may also possess pricing power, there often exist centrally enforced pricing policies.[4] Furthermore, order batching within a sales organization is typically less of a problem than between independent entities in an inter-organizational supply chain. Hence, the effects order batching and price variations are less likely to be encountered in the DMC problem. Nevertheless, the other two effects

[3] Houlihan (1985) was the first to discuss the rationing or shortage gaming phenomenon also in an intra-organizational context. Note that some authors use the term 'Houlihan effect' when referring to rationing and shortage gaming in general, e.g. Disney and Towill (2003).

[4] Recall that the discussion in this thesis primarily focuses on MTS environments where prices are often set uniformly for all markets.

may result in distorted aggregate demand signals. This can be seen as an analogy to the bullwhip effect in an inter-organizational setting.

Countermeasures against this ‘internal bullwhip effect’ are similar to the inter-company case: The activities of the individual agents in the customer hierarchy have to be coordinated, information transparency needs to be improved and incentives must be aligned. In contrast to the traditional channel coordination problem, however, the individual agents in a customer hierarchy already have an established, hierarchical relationship. To some extent, this may allow for a tighter control of the resulting superior-subordinate relationships. But important information asymmetries remain and need to be dealt with. Overall, the DMC problem can therefore be interpreted as an *intra-organizational channel coordination problem*.

As can be seen, SCM is not limited to an inter-company setting. Many similar coordination problems also occur within larger firms with distributed decision-making. While the analogy discussed above only addresses the sales and demand fulfillment tasks in customer hierarchies, also the other SCM tasks performed by the different entities in a multi-divisional firm need to be aligned. An adequate planning concept to solve such coordination problems in inter-company and intra-organizational supply chains is hierarchical planning. In the following section, this concept will be introduced briefly.

2.1.2. Hierarchical Planning

According to Ijiri et al. (1968), planning can be understood “as the process of developing a strategy for changing or responding to changes in one’s environment” by identifying and evaluating alternatives. In an SCM context, Fleischmann and Meyr (2003) defined supply chain planning (SCP) “as a generic term for the whole range of those decisions on the design of the supply chain, on the mid-term coordination and on the short-term scheduling of the processes in the supply chain.” This definition exhibits two key characteristics: First, the large problem of planning an entire supply chain actually consists of many individual, but closely related subproblems. These subproblems are referred to as *planning tasks*. Second, this definition illustrates that several of these planning tasks can be grouped at certain *planning levels*.

The term planning level requires a definition. Mesarovic et al. (1970, p. 52) observed that ‘level’ is a rather generic term, and they distinguished between three different notions of levels in a planning context:

- **Strata**: Levels in the sense of strata refer to different *degrees of abstraction*. Strata may be used to differentiate between the extents to which certain features are included in a planning model.

- **Layers**: Layers refer to different *degrees of decision complexity* which result from vertically decomposing a comprehensive decision problem into one or multiple usually simpler subproblems.

- **Echelons**: Echelons refer to different *organizational levels*, i.e. the mutual relationships between different decision units in larger organizations.

While these three aspects are inextricably linked in most practical problems, the process perspective inherent in planning suggests putting a strong focus on the notion of layers in defining a planning level. Hence, the following definition of a planning level will be adopted. It is based on Emery (1964, p. 20), who summarized earlier work:

Definition 2. *A planning level is a particular vertical partitioning of a larger problem. A certain plan, addressing the entire large problem or parts of it, lies at a lower planning level if it partitions the behavior described by plans at higher levels into finer details.*

In many cases, such a partitioning may simply reflect the decisions that need to be made at different points in time.[5] Often, lower planning levels consist of multiple plans which collectively address the entire problem at the higher level. Emery (1964, p. 20) pointed out that such an apportionment corresponds to a consistent "one-to-many transformation" between the high-level plan and its associated lower-level plans. Consistency implies that the different lower-level plans are indistinguishable in terms of the variables which have been used in defining the high-level plan.

An early differentiation between different types of planning levels was introduced by Anthony (1965). He suggested partitioning a larger planning problem by grouping individual planning tasks according to the time during which these decisions have an effect. In particular, he observed that some decisions are more concerned with the broader aspects of the overall system behavior than others. The related decision periods are longer. The result is the familiar differentiation between *long-term*, *mid-term* and *short-term* planning levels which is typically used in SCP. With each of these three major planning levels, a number of key SCP tasks are associated (see Miller (2002, Ch. 1.1) and Voß and Woodruff (2006, pp. 4–5)):

- **Long-term** or **strategic planning** is concerned with setting the long-term objectives of a company or of an entire supply chain and with defining a strategy which allows meeting these objectives. Such decisions have major implications over a long period of time and are thus associated with high risk and many uncertainties. Typical strategic supply chain decisions pertain to the potential markets to serve and to finding ways to differentiate from competitors. From a design point of view, strategic planning requires making choices regarding the structure of the supply chain network and its key links. Such decisions typically have an impact over several years and are made by senior management, usually based on aggregated internal and also external data (Miller, 2002, p. 2). Decisions at a strategic planning level are the least structured ones, are associated with high levels of uncertainty and are often difficult to formalize in quantitative terms (Steven, 1994, pp. 54–55).

- **Mid-term planning** focuses on the efficient allocation and utilization of the resources which were established by long-term planning. At a mid-term level, SCP

[5] This perspective will be referred as a *decision-time hierarchy*, see Section 3.2.

tasks can be split along the four main functional areas procurement, production, distribution and sales planning. The time frame of mid-term planning covers at least one full seasonal cycle, i.e. usually at the minimum one year. Production planning—often the most important mid-term supply chain planning task—is typically split into two sub-tasks, particularly in the case of multi-site production environments. While **master planning** focuses on aligning and optimizing production plans across multiple sites, **production planning and scheduling** has a more limited scope and addresses lot-sizing, machine assignment, scheduling and sequencing decisions at the level of a single plant (Fleischmann and Meyr, 2003, p. 481).

Mid-term decisions are usually made by middle managers and lower-level senior executives (Miller, 2002, p. 4). An important decision which already has to be made at a tactical planning level is the development of specific inventory allocation policies. In case of foreseeable shortages during the mid-term planning horizon, these policies are used to determine which customers will be served with priority (Miller, 2002, p. 183).

- **Short-term planning** ensures that individual tasks per functional area are performed efficiently and effectively (Miller, 2002, p. 5). In most supply chains, this includes routine sequencing and lot-sizing decisions, but also distribution and transportation planning to deliver goods or to pick up material. In contrast to mid-term and long-term planning, the horizontal interrelationships between individual planning tasks at the short-term level are less crucial and the use of integrated decision models is less common. Instead, there is typically a close vertical relationship between short-term planning and **execution**. Operational short-term plans have a short planning horizon in the range of days, up to several weeks.

The above assignment of individual planning tasks to planning levels represents an ideal planning situation. In practice, the actual assignment is rather fuzzy and strongly depends on the particular supply chain (type) considered (Fleischmann and Meyr, 2003, p. 471).

Given the many interdependencies between the individual planning tasks at all planning levels, all decision problems should be considered simultaneously to find a solution which is optimal from a global perspective. However, designing and solving a monolithic model covering all major supply chain planning tasks is typically not feasible. Such a *simultaneous planning* model requires significant amounts of data and thus will have enormous memory requirements. Moreover, it will exhibit a high computational complexity, rendering it impossible in most practical cases to actually determine the optimal solution.

Another major problem of simultaneous planning is the uncertainty which is associated with the required long-term and mid-term forecasts. For example, production decisions for all individual final items have to be made for several years in advance. Since the accuracy of forecasts typically improves with shorter lead times, such a monolithic model could theoretically be executed again at later points in time with updated data. However, this is highly problematic. Most updated decisions can no longer be implemented in the short run as they will be inconsistent with prior decisions. Furthermore, higher-level planning tasks have longer re-planning frequencies than short-term tasks. For example, supply network adjustments will be revised at most annually whereas lot-sizing decisions will usually be updated daily or weekly. A common schedule to revise all planning tasks at all planning levels will introduce undesirable nervousness in the planning system. Overall, simultaneous planning approaches are no feasible option in practice.

An alternative of the other extreme is *successive planning*. In a successive planning approach, the entire problem is clustered into several smaller subproblems with the objective of minimizing the interdependences between them. These subproblems will then be solved sequentially. Usually, this sequential planning approach will come at the cost of over-simplifying the interrelationship between the individual subproblems. In practice, only a one-dimensional (forward) flow of information is assumed between the subproblems while the impact of other subproblems is either estimated or ignored altogether (Steven, 1994, p. 12). This simplifies the planning situation considerably and usually permits determining feasible and often optimal solutions to each subproblem. However, the successive planning approach leads to a suboptimal overall solution.

A compromise between the simultaneous and successive planning approach is the so-called *hierarchical planning* concept (Fleischmann and Meyr, 2003, p. 457). In hierarchical planning, a larger planning problem is broken along the lines of hierarchically linked planning levels. At each planning level, only certain subproblems of the overall problem are solved. Moving down the planning hierarchy, one obtains a more detailed explanation of a complex planning problem. Contrariwise, moving up in the hierarchy leads to a deeper understanding of the overall problem and its significance (Mesarovic et al., 1970, p. 42). Lower planning levels are associated with a high degree of detail, a high re-planning frequency as well as a short planning horizon whereas the opposite applies to higher planning levels.

The key strength of a hierarchical planning concept lies in its ability to allow for *decision postponing*. While long-term and aggregate decisions with a long time horizon such as supply network planning have to be made early (i.e. at higher planning levels), decisions affecting more detailed issues may be moved to lower planning levels. These detailed decisions (e.g. lot-sizing or transportation planning) are thus postponed to later points in time when better decisions based on updated and more accurate information can be made.[6] However, it is important to account for interdependencies between these planning levels and to ensure that decisions made at a lower planning level are not in contradiction with prior decisions at higher planning levels (this is referred to as *consistency*). Decisions at higher planning levels should only restrict the decision space at the lower levels, but not pre-determine a particular decision for the short-term problems. The key challenge lies in ensuring that the decision spaces conceded to the lower planning levels always permit the generation of feasible detailed plans. While this splitting of the overall problem into multiple hierarchically aligned partial solutions usually does not necessarily lead to an

[6] In Section 3.2, this approach will be characterized as a *decision time hierarchy*.

optimum, it provides at least a feasible, consistent and in many cases quite good overall solution (Steven, 1994, p. 1).

The hierarchical planning concept was originally proposed by Hax and Meal (1975) as hierarchical production planning (HPP) for a tire manufacturer. This initial publication has spurred an enormous amount of follow-up work, was subsequently extended to various other industries and broadened in scope to include other supply chain processes. Nevertheless, all hierarchical planning systems are still built upon five major principles (Stadtler and Fleischmann, 2012): Decomposition, coordination, aggregation, model building and model solving.[7] Each principle will now be characterized in more detail.

Decomposition: As illustrated, monolithic models are usually difficult to solve in practice. Furthermore, neither is such a model readily accepted by managers in charge of specific SC tasks. Hence, hierarchical planning always entails a decomposition of the overall problem into a set of interrelated subproblems and corresponding smaller models.[8] In contrast to successive planning approaches, this decomposition leads to a hierarchical structure which typically exploits existing responsibilities and information channels (Steven, 1994, p. 1). Decomposition—or hierarchization—is thus closely linked to the existing organizational structure of a company or of an entire supply chain. Therefore, hierarchical planning facilitates the split of a larger planning problem into multiple decision areas along the lines of responsibility of individual departments or of separate legal entities.

Coordination: In contrast to successive planning approaches, the interrelations between individual subproblems in hierarchical planning are closely coordinated. Each subproblem belongs to a specific planning level and is linked to the next lower planning level in a series of top-down *instructions*. The subproblem at the higher planning level controls and restricts the decision space of the problem at the lower level by these instructions. This way, a high level of integration can be enforced, contributing to the consistency of the overall plan. Two types of instructions can be differentiated (see Stadtler, 1988):

- Primal instructions (e.g. target production quantities, available capacities or inventory levels) primarily limit the solution space and thus guarantee the solvability of the lower-level subproblem.

- Dual instructions (e.g. lot-sizing costs, inventory costs, more generally: transfer prices) directly affect the objective functions of the lower-level subproblems.

A number of other types of links may exist between individual subproblems besides simple unidirectional top-down instructions (see Steven, 1994, pp. 36–37). For example, a higher level of reciprocity can be ensured by an asymmetrically bidirectional link (e.g. one-way instructions with a feedback mechanism) or a truly symmetric, mutual link. The

[7] Steven (1994) and Mesarovic (1970) discuss similar principles.

[8] As will be discussed later in Section 3.2, such a decomposition is a prime example for a so-called *constructional hierarchy*.

latter is often the case between subproblems at the same planning level, e.g. between master planning and mid-term distribution planning.

Aggregation: Loosely speaking, aggregation refers to the grouping of similar objects into one (Steven, 1994, p. 43), usually with the objective of reducing complexity. The reverse operation to aggregation is referred to as disaggregation. A more thorough definition of these operators will be provided later in Section 3.1.2.

At higher planning levels, aggregation significantly reduces the complexity of the plan and the uncertainties of input data, e.g. by balancing lower-level demand forecast fluctuations. Typical dimensions are the aggregation of time, geographies, products and capacities:

- *Aggregation of time:* In mid-term planning, typically weekly or monthly figures of the expected demand are used rather than considering data at the level of days. As will be shown in Section 2.2.5, demand figures aggregated over time are less volatile and easier to forecast.

- *Aggregation of geographies:* Production planning can often be facilitated by combining the demands from several smaller regions. The actual geographical origin of the demands only needs to be considered at the later stages of distribution and transport planning.

- *Aggregation of products:* Several similar end items are combined to *product families* which in turn may be aggregated to *product types*. Often, there is no need for setup changes when producing items from the same product family. Items from the same product type can often be produced on the same production line. This facilitates aggregate production planning, and more detailed product data will only be considered when making scheduling decisions.

- *Aggregation of capacities:* Similarly, rather than considering the individual machine capacities per production line per minute, an aggregate figure is the amount of production capacity at a particular plant per month. The latter figure is often appropriate when deciding at which plant in a network production should occur.

Aggregation may be accomplished in a number of different ways:[9]

- *Perfect* (or consistent) *aggregation* is a fully commutative operation, i.e. it may be reversed without loss of information (Switalski, 1988, p. 384).

- Since perfect aggregation usually involves significant practical problems, Axsäter (1979) and Axsäter and Jönsson (1984) proposed an alternative approach termed *approximative aggregation*. Acknowledging that some loss of information is often unavoidable in practice, approximative aggregation merely requires that the results of both the aggregated and detailed model should coincide as much as possible.

[9] For a general framework on aggregation and disaggregation methodology and a literature overview, see Rogers et al. (1991).

http://www.springer.com/978-3-658-02863-3

Demand Fulfillment in Multi-Stage Customer Hierarchies
Vogel, S.
2014, XX, 377 p. 70 illus., Softcover
ISBN: 978-3-658-02863-3