

IBM Research Report

Supply and Demand Synchronization in Assemble-To-Order Supply Chains

**Markus Ettl, Pu Huang, Karthik Sourirajan¹,
Thomas R. Ervolina, Grace Y. Lin²**
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

¹School of Industrial Engineering
Purdue University
West Lafayette, IN 47906

²IBM Business Consulting Services
Chappaqua, NY 10514



Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Supply and Demand Synchronization in Assemble-to-Order Supply Chains

Markus Ettl

*IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
msettl@us.ibm.com*

Pu Huang

*IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
puhuang@us.ibm.com*

Karthik Sourirajan

*School of Industrial Engineering, Purdue University, West Lafayette, IN 47906
souriraj@ecn.purdue.edu*

Thomas R. Ervolina

*IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
ervolina@us.ibm.com*

Grace Y. Lin

*IBM Business Consulting Services, Chappaqua, NY 10514
gracelin@us.ibm.com*

1. Introduction

In 2000 AMR Research identified the benefits of the 21st century supply chain and introduced the concept of Demand-Driven Supply Networks (DDSN). A Demand-Driven Supply Network is a system of technologies and business processes that senses and responds to real time demand

across a network of customers, suppliers, and employees. The DDSN principles require that companies shift from a traditional push-based supply chain to a pull-based, customer-centric approach. Leading companies that have adopted the DDSN business strategy have become more *demand sensing*, have more efforts on *demand shaping* and focus on a profitable *demand response* (e.g., O'Marah and Souza 2004; Cecere et al. 2005).

Lee (2004) describes how leading companies approach a DDSN strategy to attain sustainable competitive advantage. He observes that top-performing supply chains possess three different qualities: *agility* (the ability to respond quickly to short-term change in demand and supply and manage external disruptions more effectively); *adaptability* (ability to adjust the design of the supply chain to meet structural shifts in markets and modify supply network strategies, products, and technologies); and *alignment* (ability to create shared incentives that align the interests of businesses across the supply chain). Similar principles are exercised in the Sense and Respond Value Net model described in Lin *et al.* (2002; 2004), an event-driven model with proactive sensing and intelligent responding for collaboratively optimizing the performance of end-to-end value networks. The model combines timely decision support with risk and resource management, extended supply chain optimization, business process automation and partner alliances into an integrated management system. It enables value network partners to adapt to changing business environments and respond quickly with the best available policies for achieving financial or operational business objectives. Although the articles by Lee and Lin *et al.* do not directly address DDSN, the recommended tasks are directly in line with the process of building DDSN capabilities.

In this chapter we describe an event-driven demand and supply planning process that incorporates the DDSN principles of demand shaping

and profitable demand response to drive better operational efficiency. The proposed business process, called *availability management process*, extends its focus beyond aligning demand, supply and financial metrics and directly applies demand and supply data to better respond to changes in the marketplace. It aims at creating marketable product alternatives to mitigate misalignments of supply and demand and to enable companies to take full advantage of a “sell-what-you-have” strategy. In contrast to traditional enterprise planning applications that generally require weeks to adapt to changes in demand, event-driven availability management can quickly highlight unexpected events such as component shortfalls, excess inventories, delayed shipments, or spikes in customer demand. As a result, production managers detect problems and opportunities earlier and can develop a focused response that avoids the time and expense of completely regenerating a production plan. A well executed availability management process benefits the customers through improved delivery times, and it benefits the enterprise through higher inventory turns, fewer supply overages and shortfalls, and reduced inventory liability exposure.

We further describe an analytical optimization model that supports availability management and determines profitable product offerings that minimize inventory liabilities and lost sales risks over the entire supply chain. The model provides dynamic, real-time sales recommendations based on current availability, price, performance and customer demand information. This enables on demand up-selling, alternative-selling and down-selling to better integrate the supply chain horizontally, connecting the interaction of customers, business partners and sales teams to the procurement and manufacturing capabilities of the company. The optimization is most effective in an assemble-to-order (ATO) environment where end products are configured from pluggable sales building blocks.

The remainder of this chapter is organized as follows. In section 2 we present the underpinning principles of availability management and discuss the advantages and disadvantages of different business processes to implement availability management. In section 3 we survey the relevant literature. In section 4 we formulate the problem of finding marketable product alternatives in a given product portfolio that best utilize the available component supply as an optimization problem and develop an efficient computational procedure for solving the problem. Numerical findings and discussions of results are presented in section 5. This produces several insights into how advanced availability management can help to proactively coordinate supply and sales, and it quantifies several business benefits in the context of assemble-to-order manufacturing. Section 6 concludes this chapter.

2. Business Processes for Availability Management

Availability management is the overarching task of balancing the planning of supply and demand and the execution of supply and demand. To achieve this task, companies have developed business processes that coordinate the flow of subtasks along with the use of information technology (IT) and decision support. Both the IT infrastructure (connectivity to Enterprise Resource Planning applications, Manufacturing Execution Systems, or legacy systems) and the decision support capabilities are essential to the successful implementation of an availability management system. Although the individual design and implementation of a business process for availability management varies from firm to firm, there are several key factors which govern the design of any such process. Before we present an optimization-based decision support model for availability management,

we discuss three different business process designs and discuss the determining factors that let firms choose one process over another.

2.1 Available-to-Promise (ATP)

The planning side of availability management is administered by a Sales and Operations Planning (S&OP) process. The goal of S&OP is to generate a single plan of product availability that incorporates a company's revenue targets, unbiased demand forecasts, and the capacity of its supply chain. This plan is called the Available-to-Promise (ATP) schedule. ATP states the common marching orders to which the sales and marketing teams will gear up to sell, the supply chain will plan to procure supply, and the finance teams will target revenue and earnings. The typical cycle time of the S&OP process is between two and four weeks.

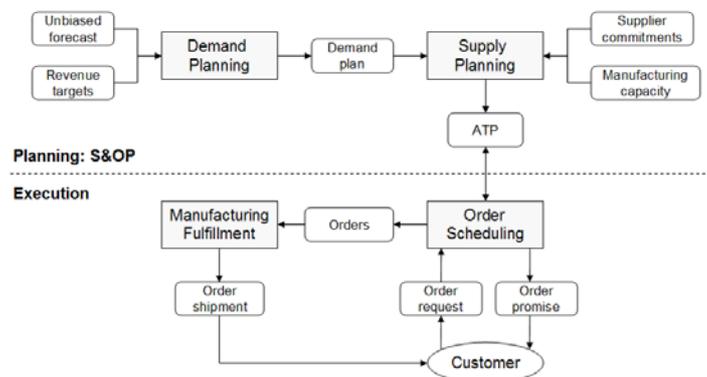


Figure 1. Availability Management with Available-to-Promise (ATP)

The execution side of availability management deals with a real-time stream of customer orders where each order must be scheduled or promised. As a customer request arrives, the order scheduling process must promise an availability date to the customer. This task involves checking

the contents of the order against the ATP schedule, determining an availability promise date to the customer, and decrementing the ATP to accurately reflect the supply committed to new customer orders. The ATP schedule must be fully supply-supported because order shipment dates are promised based on the ATP supply. The ATP schedule is the main linkage between planning and execution of the availability management process as shown in Figure 1.

The ATP process utilizes a technique called implosion (e.g., Dietrich *et al.* 2005; Ball *et al.* 2004) that takes into account supplier commitments and manufacturing capacities to generate an optimized ATP schedule. Due to the squared set nature of component usage in production bills-of-materials, shortages of one component can cause overages of another component. For example, if the bill-of-material of an end product calls out two components and one component is in shortfall, the ATP schedule will be reduced to the tightest constraint which in turn causes excess of the other component. The implosion technology is geared towards optimizing around supply chain shortages and constraints and it does not accommodate excess supply. Because ATP has no means of addressing excess supply, a separate (non-integrated) business process is often created to manage excess and overages, e.g. by exercising buy-back agreements with component suppliers.

2.2 Sequential Model for Available-To-Sell

The second business process design for availability management is a two-stage sequential process that utilizes ATP in conjunction with *Available-to-Sell (ATS)*, a new concept for managing supply shortages and overages that is increasingly gaining traction with manufacturers in the electronics industry. ATS is designed to identify alternative end products that con-

some excess supply while minimizing additional procurement investments to square up supply. In addition, ATS seeks to ensure that the additional product availability is sellable and not in conflict with sales and marketing goals.

ATS starts after the supply planning process is completed and an ATP schedule is generated. The ATP schedule is exploded through the bills-of-materials to generate a consumption outlook at the component level. This newly exploded demand is compared to existing inventories and supplier commitments to determine excess component supply. Next, an ATS optimization run is performed to generate a list of alternative product offerings with their corresponding availability quantities. The alternative offerings are subsequently analyzed by the sales and marketing teams, and a final set of ATS product offerings is selected. This selection is combined with the original ATP to create a “conditioned” ATP schedule. Figure 2 illustrates the integration of ATS into the overall availability management process.

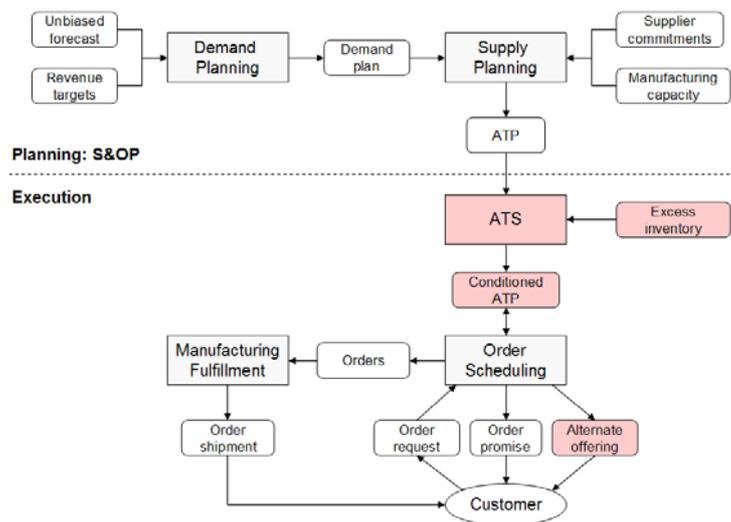


Figure 2. Sequential Model for Available-to-Sell

The ATS process gathers excess inventory and product related information to establish the conditioned ATP schedule. Generally there are three types of excess inventory considered: excess materials, build outs, and current product overage. Excess materials are saleable components or features that are compiled from long-term inventories, possibly including short-term overages. Build outs refer to transitioning products approaching withdrawal that have build-out targets established that require demand monitoring and possible action. While excess material denotes current long-term inventory, build-outs are considered as a preventive way of reducing future long-term overages. The last type is current product overage which is driven by current product offerings that are not tracking to their build and commit levels in the current planning horizon. Current product overage entails an action preventing future short-term overage.

The recommended ATS product offerings may take advantage of up-sell, alternate-sell or down-sell opportunities. An up-sell opportunity is where a customer or business partner is sold a more richly configured solution above the customer's initially selected price range. Incentives may be used to entice the customer to agree to an up-sell. An alternative-sell relates to a sale of a similar product that falls within the selected price range. An alternative-sell is performed when an up-sell is not available or the customer opts for a similarly priced product. A down-sell opportunity refers to a sale of a product that falls below the price range selected by the customer.

The sequential business process described above is most effective in a *complex-configured* supply chain environment that is characterized by high product variety, long sales cycles, long component lead-times, multiple levels of bills-of-materials, and expensive and custom-built components. The ATP process ensures that the available supply is allocated to the

tical drives etc. Because there is less differentiation between products, the ATP schedule does not have to protect high profit margin products. In such an environment component supply is generally more flexible and the simplified product structure makes it more probable that recommended product offerings will drive customer demand.

3. Related Literature

The models developed in this paper span several streams of literature. We first review the literature on available-to-promise (ATP) systems for order promising and order fulfillment. Ball *et al.* (2005) present a general optimization framework for ATP models that have been described in the literature, as well as several examples of actual ATP business practices in the electronics industry. Chen *et al.* (2002) present a mixed integer programming model that provides an ATP order promising and fulfillment solution for a batch of orders that arrive within a predefined time interval. A variety of constraints, such as raw material availability, production capacity, material compatibility, and customer preferences, are considered. Ervolina and Dietrich (2001) describe an application of the implosion technology for ATP order promising in an assemble-to-order (ATO) or configure-to-order (CTO) environment. The goal is to create a feasible production plan that can be used to schedule (or promise) orders against. Chen-Ritzo (2006) studies the availability management process for CTO system with order configuration uncertainty. For the ATP generation problem, she formulates a two stage stochastic linear program with recourse and solves it using a sample average approximation method. She also studies the issue of how to manage the order promising and scheduling process and introduces a component rationing problem which determines threshold

levels in the ATP to reserve component supply for future higher-profitable orders. Dietrich *et al.* (2005) describe an ATS implosion model that finds marketable products that consume the excess while minimizing additional purchase. In their paper the set of new product configurations is predetermined. No explicit demand statement is provided and the configurations are assumed to have infinite demand. Xu and Akcay (2004) formulate a two-stage stochastic integer program with recourse for allocating constrained components and selecting base-stock levels to maximize the fraction of orders assembled within a quoted maximum delay. They develop a heuristic method based on a simple, order-based component allocation rule. Ettl *et al.* (2006) formulate the problem of finding marketable product alternatives in a given product portfolio as a non-linear program and develop an efficient column generation procedure for solving the problem. They present a single-period model that creates alternative product offerings that use surplus inventory or substitutions, and a multi-period model that assists planners in managing major product and technology transitions. Kapoor *et al.* (2005) and Buckley *et al.* (2005) describe the implementation of a Sense and Respond Value Net that empowers an event-driven availability management process in an assemble-to-order supply chain at IBM.

We next review the literature on product substitution and reverse logistics. Balakrishnan and Geunes (2000) describe a manufacturing planning method with flexible bills-of-materials and component substitution. A dynamic programming solution method is developed to find production and substitution quantities that satisfy demands at minimum total cost, comprising setup, production, substitution, and inventory holding cost. Because supply is unconstrained, the model does not address matching of demand and supply. Another particular type of substitution is the so-called

downward substitution where high-end products can substitute for low-end products when the latter are out of stock. Hale *et al.* (2001) study a downward substitution problem in an ATO system with two end-products where each product is composed of two components. Substitutions are carried out at the component level. Chen and Plambeck (2005) study a single-item production-inventory system with periodic replenishments where customers may accept substitutes or choose not to buy when a product is out of stock. They show that learning about the demand distribution and customer substitution behavior influences the optimal inventory levels.

Several authors have studied disassembly decisions in reverse logistics supply chains. In reverse logistics, products which by themselves are no longer marketable may be disassembled to recover components or subassemblies. The recovered components may serve as spare parts or as components in new configurations that are resold in secondary markets (e.g., Dekker 2003; Fleischmann *et al.* 2005). Gupta and Taleb (1994) and Taleb *et al.* (1997) develop effective disassembly configurations with common components among products and limited inventory of products available for disassembly. Meacham *et al.* (1999) determine optimal disassembly configurations to meet a specified demand for recovered components and subassemblies from an available supply of recovered products. Products are represented through their bills-of-materials. The authors propose a fast column generation algorithm to determine maximum revenue disassembly configurations for individual products.

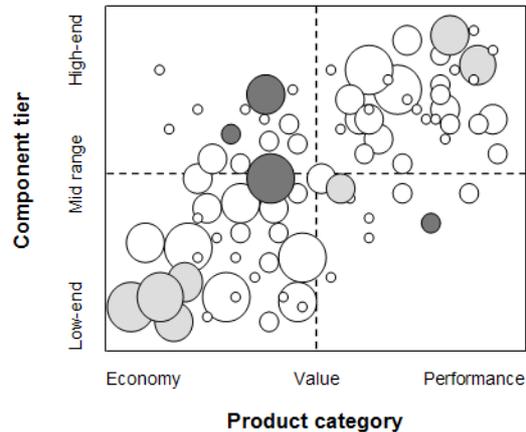
4. Mathematical Optimization Model for Available-To-Sell

In this section, we formulate a mathematical programming model that enables optimization-based decision support for availability management

and present an efficient solution algorithm that uses decomposition and column generation. The model helps demand and supply planners gain detailed insights into supply consumption and demand satisfaction to reduce excess inventory, inventory liabilities, and pending orders for improved revenue attainment. The analysis is based on optimizing financial objectives that include liability costs for excess inventory and penalty costs for pending (unfilled) customer orders. Orders that are left pending across financial quarters are particularly undesirable because they directly affect a company's top-line revenue.

Input to the model is a portfolio of existing product offerings and their demand forecasts. The product portfolio is grouped into product categories such as economy, value and performance products. The goal is to build enough volume to satisfy the demand forecast for each existing product offering. If that can not be achieved, the model creates a conditioned ATP schedule with alternate product offerings for each product category. An alternate offering created for a product category can be used to fulfill unsatisfied demand for an existing offering in the same category with additional substitution cost incurred. The components used to configure the end products are grouped into commodity groups with each component belonging to exactly one commodity group. A bill-of-materials that describes the component consumption is given for each existing product offering. Product configuration rules for each product category restrict the components that can be used in every commodity group to create new product offerings for this category. Product configuration rules represent technical and manufacturing restrictions that dictate how various parts can be assembled into a product. For example, the power consumption requirements of a specific component could limit the product category in which this

component can be used and would thus be expressed as a product configuration rule.



Legend:

- Original product offering
- ◐ Original product offering w/ revised allocation
- Alternative product offering

Figure 4. Illustration of a Conditioned ATP Schedule

The optimization model compares and analyzes product alternatives to create a set of recommended new product configurations. Figure 4 illustrates a conditioned ATP schedule generated by the optimization model. Each circle depicts a single product offering in the product portfolio. The size of the circle represents the volume that is allocated to a product; larger circles imply higher volume. The position of a circle on the grid indicates whether the corresponding product offering is assembled from low-end, mid-range or high-end components. For example, product offerings in the “Economy” category are predominantly assembled from low-end commodities and are thus shown in the lower left quadrant; product offerings in the “Performance” category are predominantly assembled from high-end commodities and are shown in the upper right quadrant. The light col-

ored circles represent existing product offerings with optimized allocations that deviate from their initial demand forecast. The dark circles depict alternative product offerings that were selected by the optimization model to supplement the original portfolio.

4.1 Notation

Before we discuss the formulation, let us define the notation required.

Products, commodities, and components

- I : Number of components (indexed by i , $i = 1, \dots, I$)
 K : Number of commodity groups (indexed by k , $k = 1, \dots, K$)
 P : Number of product categories (indexed by p , $p = 1, \dots, P$)
 M_p : Number of existing product offerings in category p (indexed by m , $m = 1, \dots, M_p$)

Demand and Supply

- D_m^p : Demand forecast for the existing product offering m in category p
 S_i^{\min} : Downside supply flexibility for component i (minimum supply quantity supported by the component supplier)
 S_i^{\max} : Upside supply flexibility (maximum supply quantity supported by component supplier)

Bills-of-Materials

- B_{ik} : 1 if part i belongs to commodity group k , 0 o/w
 u_{im}^p : 1 if part i is used to assemble existing product offering m in category p , 0 o/w (BOM)
 w_i^p : 1 if part i can be used to assemble products in category p , 0 o/w

Costs

h_i : Unit holding cost of component i (liability cost)

b_m^p : Unit lost-sales (or backorder) cost of existing product offering m in category p

c^p : Unit substitution cost for product category p . Building one unit of a new offering in product category p to fulfill the demand forecast of an existing offering in the same category will incur the cost c^p

Decision Variables

X_m^p : Build volume of existing product offering m in category p

N_p : Number of models to build for product category p (indexed by n , $n = 1, \dots, N_p$)

v_{in}^p : 1 if component i is used to assemble new product offering n in product category p ; 0 o/w

Y_{mn}^p : Build volume of new product offering n in product category p that is used to substitute existing offering m in the same product category

S_i : Quantity of component i ordered from supplier based on the build volumes

4.2 Problem Formulation

We can now formulate the optimization problem. The objective given by eqn. (1) is to minimize the total supply chain cost which consists of three components:

1. *Lost-sales costs (or backorder costs)*; if the build volume allocated to an existing product offering falls short of the demand forecast, a lost-sales cost is incurred.
2. *Product substitution costs*; costs incurred for using a new product offering to (partially) fulfill demand for an existing product offering.
3. *Inventory holding costs (or component liability costs)*; costs incurred for holding excess component inventories.

$$\begin{aligned} \text{Min } Z(X_m^p, v_{in}^p, Y_{mn}^p) = & \sum_{p=1}^P \sum_{m=1}^{M_p} b_m^p \left(D_m^p - X_m^p - \sum_{n=1}^{N_p} Y_{mn}^p \right) + \sum_{p=1}^P c^p \sum_{m=1}^{M_p} \sum_{n=1}^{N_p} Y_{mn}^p \\ & + \sum_{i=1}^I h_i \left(S_i - \sum_{p=1}^P \sum_{m=1}^{M_p} u_{im}^p X_m^p - \sum_{p=1}^P \sum_{n=1}^{N_p} v_{in}^p \sum_{m=1}^{M_p} Y_{mn}^p \right) \end{aligned} \quad (1)$$

Let us now formulate the constraints. The consumption of component i , S_i , is bounded by a maximum and minimum order quantity:

$$S_i \leq S_i^{\max}, \forall i \quad (2)$$

$$S_i \geq S_i^{\min}, \forall i. \quad (3)$$

Given the demand forecasts D_m^p for existing offering m in category p , the total build volume for this offering (including the volumes substituted by new product offerings) cannot exceed the demand forecast D_m^p :

$$D_m^p - X_m^p - \sum_{n=1}^{N_p} Y_{mn}^p \geq 0 \quad (4)$$

The number of components consumed must be less than or equal the number of components committed by the supplier:

$$S_i - \sum_{p=1}^P \sum_{m=1}^{M_p} u_{im}^p X_m^p - \sum_{p=1}^P \sum_{n=1}^{N_p} v_{in}^p \sum_{m=1}^{M_p} Y_{mn}^p \geq 0 \quad (5)$$

The components used in a new product offering must not violate the selection menu specified for the product category of each new product offering:

$$v_{in}^p \leq w_i^p, \forall i, p, n \quad (6)$$

Each new product offering must be a squared set configuration, i.e., it must use one and only one component from each commodity group:

$$\sum_{i=1}^I B_{ik} v_{in}^p = 1, \forall k, p, n \quad (7)$$

Finally, we also have non-negativity and integrality constraints on the decision variables.

$$X_m^p \geq 0, \forall p, m \quad (8a)$$

$$S_i \geq 0, \forall i \quad (8b)$$

$$Y_{mn}^p \geq 0, \forall p, m, n \quad (8c)$$

$$N_p \geq 0 \ \& \ Integer, \forall p \quad (8d)$$

$$v_{in}^p \in \{0,1\}, \forall i, n, p \quad (8e)$$

It is important to note that, even if the number of new product configurations in product category p , N_p , were fixed, there is a nonlinear term ($v_{in}^p Y_{mn}^p$) in the objective function (1). It is possible to linearize this term using standard techniques and convert the problem into a mixed integer program (e.g., Barnhart *et al.* 1998). However, given that industry-size problems involve hundreds of components and dozens of product categories (which results in thousands of binary variables), finding the optimal solution of the mixed integer program is likely to take a prohibitively long time. Because speed of execution is essential for a timely resolution of supply–demand imbalances, we have developed an efficient iterative procedure based on column generation that is described next.

4.3 Computational Algorithm

The basic idea is that new product offerings are columns that are introduced into the problem one at a time in an iterative algorithm. The problem is decomposed into two sub-problems: a master problem (*MP*) that aims at finding the optimal build volumes for a given set of new and existing product offerings, and a set of slave problems (SP_m^p , $p = 1, \dots, P$ and $m=1, \dots, M_p$) that generates new product offerings and supplies them to *MP*. The two sub-problems are solved iteratively until the optimal solution is reached. The master problem is defined in eqn. (9):

Master Problem (MP):

$$\begin{aligned} \text{Min } Z(X_m^p, Y_{mn}^p) = & \sum_{p=1}^P \sum_{m=1}^{M_p} b_m^p \left(D_m^p - X_m^p - \sum_{n=1}^{N_p} Y_{mn}^p \right) + \sum_{p=1}^P c^p \sum_{m=1}^{M_p} \sum_{n=1}^{N_p} Y_{mn}^p \\ & + \sum_{i=1}^I h_i \left(S_i - \sum_{p=1}^P \sum_{m=1}^{M_p} u_{im}^p X_m^p - \sum_{p=1}^P \sum_{n=1}^{N_p} v_{in}^p \sum_{m=1}^{M_p} Y_{mn}^p \right) \end{aligned} \quad (9)$$

subject to constraints (2), (3), (4), (5), (8a), (8b) and (8c)

Note that the bills-of-materials of the new product offerings, v_{in}^p , are not decision variables in *MP* and are fixed along with N_p . Thus, the master problem (9) is a linear program that can be solved very efficiently even for large problem sizes.

Let λ_i , $i = 1, \dots, I$, denote the shadow prices pertaining to the liability constraint (5), and β_m^p , $m = 1, \dots, M_p$ and $p = 1, \dots, P$, denote the shadow prices pertaining to the build volume constraint (4), in the optimal solution to *MP*. We can now formulate the slave problems SP_m^p for $p = 1, \dots, P$ and $m=1, \dots, M_p$ as follows:

Slave Problem (SP_m^p):

$$\text{Min } Z_2^p(\bar{v}_i^p) = \sum_{i=1}^I (-h_i + \lambda_i) \bar{v}_i^p - b_m^p + \beta_m^p + c^p \quad (10)$$

subject to constraints (6), (7), (8d) and (8e)

We use a standard column generation procedure to solve the problem (e.g., Barnhart 1998). In the initial step of the algorithm, MP is solved without any new product offerings and its solution is fed into the slave problems. The slave problems subsequently try to improve the solution by introducing new product offerings for every product category. Any new product offerings are then added back into MP after updating the corresponding values of N_p for $p = 1, \dots, P$, and the master problem MP is resolved. This process iterates until no further improvement is possible.

Although the slave problem SP_m^p is an integer program, it has a special structure that can be exploited to solve it easily: for each commodity group, first consider the parts that belong to this group and are allowed in the selection menu w_i^p . Compare their $(-h_i + \lambda_i)$ values and choose the one with the minimum $(-h_i + \lambda_i)$ value. It is easy to see that repeating this procedure for all commodity groups produces a squared set product offering that minimizes $Z_2^p(\bar{v}_i^p)$. This procedure is efficient in terms of the computational requirement per iteration. To further improve the runtime performance of the algorithm, we modified the above procedure to introduce multiple columns (new product offerings) per iteration instead of just one. This is achieved by simultaneously introducing all columns that yield a negative $Z_2^p(\cdot)$ value.

5. Numerical Study

We implemented the solution procedure described in the previous section and applied them to an ATO system for mid-range server computers. In this section we present and discuss our numerical findings. The numerical study focuses on (1) evaluating the benefits of the sequential and integrated availability management approach (in terms of expected backorder and liability costs) relative to a conventional ATP system; and (2) demonstrating the additional benefit of supply quantity flexibility in conjunction with advanced availability management.

5.1 Comparisons between ATP and ATS

The example scenario for the numerical study is depicted in Figure 5. The data set resembles a real-world problem compiled from actual server supply chain data. The product portfolio consists of three product families that represent low-end, mid-range and high-end server computers.

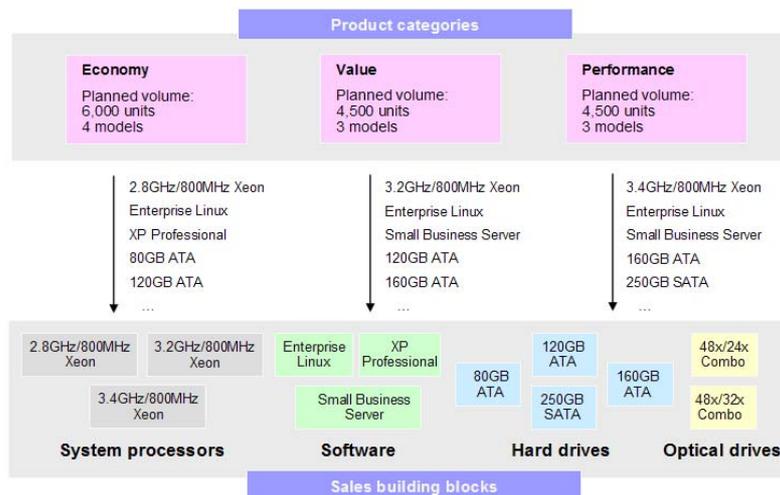


Figure 5. Assemble-to-Order Product Structure for Example Scenario

Each product family comprises a number of predefined product configurations with bills-of-materials as depicted in Table 1. Product configurations P_1 to P_4 are low-end systems, P_5 to P_7 are mid-range systems, and P_8 to P_{10} are high-end computers. Each configuration is assembled from components of the four commodity groups: system processors, hard drives, software, and optical drives. For example, product P_1 is assembled from a 2.8GHz system processor, an 80GB hard drive, XP Professional, and a 48x/32x optical drive. Notice that all products within the same product family use the same system processor, i.e., 2.8GHz for low-end systems, 3.2GHz for mid-range systems, and 3.4GHz for high-end configurations.

Components	Low-End (Economy)				Mid-Range (Value)			High-End (Performance)			Supply Requirements	Supply Commitment	
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}			
SYSTEM	2.8GHz/800MHz Xeon	1	1	1	1	-	-	-	-	-	-	6,000	2,500
PROCESSORS	3.2GHz/800MHz Xeon	-	-	-	-	1	1	1	-	-	-	4,500	10,000
	3.4GHz/800MHz Xeon	-	-	-	-	-	-	-	1	1	1	4,500	2,500
HARD DRIVES	80GB 7200rpm ATA	1	1	1	-	-	-	-	-	-	-	4,500	4,000
	120GB 7200rpm ATA	-	-	-	1	-	-	1	-	-	-	3,000	5,000
	160GB 7200rpm ATA	-	-	-	-	1	1	-	1	-	-	4,500	4,000
	250GB 7200rpm SATA	-	-	-	-	-	-	-	-	1	1	3,000	2,000
SOFTWARE	XP Professional	1	1	-	1	-	-	-	-	-	-	4,500	4,000
	Enterprise Linux	-	-	1	-	1	1	-	-	1	-	6,000	7,000
	Small Business Server	-	-	-	-	-	-	1	1	-	1	4,500	4,000
OPTICAL DRIVES	48x/24x Combo	-	1	-	-	1	-	1	1	1	-	7,500	6,000
	48x/32x Combo	1	-	1	1	-	1	-	-	-	1	7,500	9,000
Demand forecast		1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500		
ATP schedule		1,000	1,500	-	-	1,500	1,500	1,500	1,000	-	1,500		

Table 1. Bill-of-Materials, Supply Requirements, and Supply Commitments

We assume that the top-level demand forecast for configuration m in product family p is $D_m^p = 1,500$ units. To determine the supply requirements for each component, the demand forecast is exploded through the bills-of-materials in a standard ATP calculation. The two rightmost columns of Table 1 show the supply requirement and a sample supply commitment pertaining to the top-level demand forecast. The supply commitment indicates a suppliers' capability to deliver to the manufacturer's supply requirements. For now, assume that the minimum supply quantity

S_i^{\min} is equal to the maximum supply quantity S_i^{\max} , i.e., there is no supply flexibility.

Comparing the supply requirements with the supply commitment indicates supply constraints on 2.8GHz and 3.4GHz processors. To mitigate the constrained supply, the micro supplier committed a higher than requested supply volume of 10,000 units for the 3.2GHz processor. Notice that the supply commitment matches the supply requirements at the commodity group level, although the individual component mix deviates from the requirements. Given the supply commitment, the conventional ATP implosion matches the available supply to the demand forecast and provides an optimized ATP schedule as shown in the bottom row of Table 1. As a result of the supply constraints on the 2.8GHz and 3.4GHz processors, the ATP schedule incurs 3,500 backorders of low-end systems and 2,000 backorders of high-end systems.

Next, we apply the sequential and integrated availability management model to the example scenario. Table 2 shows the selection menu (w_i^p) for the three product families.

Selection menu		Low-End (Economy)	Mid-Range (Value)	High-End (Performance)
SYSTEM PROCESSORS	2.8GHz/800MHz Xeon	1	-	-
	3.2GHz/800MHz Xeon	1	1	1
	3.4GHz/800MHz Xeon	-	-	1
HARD DRIVES	80GB 7200rpm ATA	1	-	-
	120GB 7200rpm ATA	-	1	-
	160GB 7200rpm ATA	-	1	-
	250GB 7200rpm SATA	-	-	1
SOFTWARE	XP Professional	1	-	-
	Enterprise Linux	1	1	1
	Small Business Server	-	1	1
OPTICAL DRIVES	48x/24x Combo	1	1	1
	48x/32x Combo	1	1	1

Table 2. Selection Menu for Different Product Categories

To analyze the performance of the sequential model, we first explode the ATP schedule generated above through the bills-of-materials to obtain

the parts consumption for each technology component. The exploded component demand is compared to the supplier commitments to determine unallocated supply. Next we apply the optimization model described in section 4 to create alternative product offerings. The demand statement is given by the backorders pertaining to the ATP schedule. The master problem begins with the initial set of product configurations, the new demand forecasts and the new component supply commitments. Once an initial production plan is determined, the algorithm executes a slave problem to generate new product configurations until the inventory costs can not be reduced any further.

The sequential algorithm creates two alternative product configurations, R_{11} and R_{31} . The bills-of-materials of the new offerings are shown in the shaded columns of Table 3. The conditioned ATP schedule is displayed in the bottom row. As a result of optimization, the order backlog for low-end systems is reduced from 3,500 to 2,000, and the order backlog for high-end systems is reduced from 2,000 to 1,500.

Components		Low-End (Economy)					Mid-Range (Value)			High-End (Performance)			
		P ₁	P ₂	P ₃	P ₄	R ₁₁	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	R ₃₁
SYSTEM PROCESSORS	2.8GHz/800MHz Xeon	1	1	1	1	-	-	-	-	-	-	-	-
	3.2GHz/800MHz Xeon	-	-	-	-	1	1	1	1	-	-	-	1
	3.4GHz/800MHz Xeon	-	-	-	-	-	-	-	-	1	1	1	-
HARD DRIVES	80GB 7200rpm ATA	1	1	1	-	1	-	-	-	-	-	-	-
	120GB 7200rpm ATA	-	-	-	1	-	-	-	1	-	-	-	-
	160GB 7200rpm ATA	-	-	-	-	-	1	1	-	1	-	-	-
	250GB 7200rpm SATA	-	-	-	-	-	-	-	-	-	1	1	1
SOFTWARE	XP Professional	1	1	-	1	1	-	-	-	-	-	-	-
	Enterprise Linux	-	-	1	-	-	1	1	-	-	1	-	1
	Small Business Server	-	-	-	-	-	-	-	1	-	-	1	-
OPTICAL DRIVES	48x/24x Combo	-	1	-	-	-	1	-	1	1	1	-	-
	48x/32x Combo	1	-	1	1	1	-	1	-	-	-	1	1
Base forecast		1,500	1,500	1,500	1,500	-	1,500	1,500	1,500	1,500	1,500	1,500	-
Conditioned ATP (sequential model)		1,000	1,500	-	-	1,500	1,500	1,500	1,500	1,000	-	1,500	500

Table 3. Conditioned ATP Schedule Generated by the Sequential Model

We next analyze the performance of the integrated model. Recall that in the integrated process a conditioned ATP schedule is generated in a single optimization run based on the initial demand forecast, supplier com-

mitments and product configurations. The results are displayed in Table 4. The integrated algorithm creates four alternative product configurations with bills-of-materials shown in the shaded columns. We can see that the order backlog for all three families is significantly reduced: 500 backorders for low-end systems (down from 2,000 in the sequential model), zero backorders for mid-range systems (down from 500) and 1,000 backorders for high-end systems (down from 1,500).

Components	Low-End (Economy)						Mid-Range (Value)				High-End (Performance)			
	P ₁	P ₂	P ₃	P ₄	R ₁₁	R ₁₂	P ₅	P ₆	P ₇	R ₂₁	P ₈	P ₉	P ₁₀	R ₃₁
SYSTEM PROCESSORS	2.8GHz/800MHz Xeon	1	1	1	1	-	-	-	-	-	-	-	-	-
	3.2GHz/800MHz Xeon	-	-	-	-	1	1	1	1	1	-	-	-	1
	3.4GHz/800MHz Xeon	-	-	-	-	-	-	-	-	-	-	1	1	1
HARD DRIVES	80GB 7200rpm ATA	1	1	1	-	1	1	-	-	-	-	-	-	-
	120GB 7200rpm ATA	-	-	-	1	-	-	-	1	1	-	-	-	-
	160GB 7200rpm ATA	-	-	-	-	-	-	1	1	-	-	1	-	-
	250GB 7200rpm SATA	-	-	-	-	-	-	-	-	-	-	-	1	1
SOFTWARE	XP Professional	1	1	-	1	1	-	-	-	-	-	-	-	-
	Enterprise Linux	-	-	1	-	-	1	1	-	1	-	1	-	1
	Small Business Server	-	-	-	-	-	-	-	1	-	1	-	1	-
OPTICAL DRIVES	48x24x Combo	-	1	-	-	-	-	1	-	1	1	1	-	1
	48x32x Combo	1	-	1	1	1	1	-	1	-	-	-	1	-
Base forecast	1,500	1,500	1,500	1,500	-	-	1,500	1,500	1,500	-	1,500	1,500	1,500	-
Conditioned ATP (Integrated model)	-	-	1,000	1,500	2,500	500	1,000	1,500	1,500	500	1,500	-	1,000	1,000

Table 4. Conditioned ATP Schedule Generated by the Integrated Model

Table 5 compares the excess inventory (unallocated supply) derived from the traditional ATP schedule and the conditioned ATP schedules. We observe that the traditional ATP schedule leaves a substantial portion of the component supply unallocated, in particular 5,500 units of 3.2GHz system processors and 5,000 units of 80GB and 120GB hard drives. The reason for the large overage is that none of the existing products in Table 1 is configured with a 3.2GHz processor and a low-end hard drive. To mitigate the overage, both the sequential and the integrated model create alternative product configurations that utilize a 3.2GHz processor and an 80GB or 120GB hard drive in their bill-of-materials. As a result the excess supply is reduced from 22,000 to 14,000 units in the sequential model and 6,000 units in the integrated model.

Components	Supply Commitment	ATP	ATS (sequential model)	ATS (integrated model)
		Excess inventory	Excess inventory	Excess inventory
SYSTEM PROCESSORS	2.8GHz/800MHz Xeon	2,500	-	-
	3.2GHz/800MHz Xeon	10,000	5,500	3,500
	3.4GHz/800MHz Xeon	2,500	-	-
HARD DRIVES	80GB 7200rpm ATA	4,000	1,500	-
	120GB 7200rpm ATA	5,000	3,500	3,500
	160GB 7200rpm ATA	4,000	-	-
	250GB 7200rpm SATA	2,000	500	-
SOFTWARE	XP Professional	4,000	1,500	-
	Enterprise Linux	7,000	4,000	3,500
	Small Business Server	4,000	-	-
OPTICAL DRIVES	48x/24x Combo	6,000	500	500
	48x/32x Combo	9,000	5,000	3,000

Table 5. Excess Inventory (in units) for ATP and ATS Models

If we assume that the inventory liability cost for component i is $h_i = 5$, the sequential and integrated models yield inventory costs of \$70,000 and \$30,000 respectively. These cost figures are significantly lower than the \$110,000 produced by the traditional ATP. If we further assume that the backorder cost per unit of unfilled demand is $b_m^p = 50$ and the product substitution cost is $c^p = 10$, the sequential model yields a backorder cost of \$175,000, a substitution cost of \$20,000 and a total cost of \$265,000. The integrated model yields a backorder cost of \$75,000, a substitution cost of \$45,000 and a total cost of \$150,000. Given that the total cost arising from the ATP allocation is \$385,000 the percentage cost improvement gained in the sequential model is more than 30 percent, and more than 60 percent in the integrated model. The individual costs pertaining to the different approaches are summarized in Table 6.

	Unit cost	ATP		ATS (sequential model)		ATS (integrated model)	
		Quantity	Cost	Quantity	Cost	Quantity	Cost
Backorders	\$ 50	5,500	\$275,000	3,500	\$ 175,000	1,500	\$ 75,000
Excess inventory	\$ 5	22,000	\$110,000	14,000	\$ 70,000	6,000	\$ 30,000
Substitutions	\$ 10	-	-	2,000	\$ 20,000	4,500	\$ 45,000
Total cost			\$385,000		\$ 265,000		\$ 150,000

Table 6. Detail Cost of ATP and ATS Models

5.2 Impact of Quantity Flexibility on Supply Chain Costs

Next we investigate the impact of quantity flexibility on the performance of ATS-based availability management. Quantity flexibility contracts are widely used by companies that are facing significantly uncertain demand to mitigate supply chain inefficiencies. Quantity flexibility is a supplier's guarantee to deliver up to a certain percentage above the committed supply quantity. It can be specified in a supply contract that allows a manufacturer to adjust its order quantities after an initial purchase order is placed. The manufacturer can purchase options or other derivative instruments for risk management to protect its supply chain against demand risk. Such flexibility enables the manufacturer to reduce its risk in over- or under-stocking, but it generally comes at an extra cost which gives the supplier an incentive to offer flexibility while undertaking more risk. For our purposes, quantity flexibility is a key input to the optimization that limits the amount of additional components available to build alternative configurations.

As before we assume that the top-level demand forecast for configuration m in product family p is $D_m^p = 1,500$ units and the minimum supply commit for each component i , S_i^{\min} , is as shown in the rightmost column of Table 1. To model quantity flexibility, let $S_i^{\max} := (1 + \alpha)S_i^{\min}$ denote the maximum supply quantity, where α is a contingency factor that determines the amount of upside supply flexibility. With quantity flexibility the supplier is committed to provide up to S_i^{\max} units of component i . The manufacturer assumes inventory liabilities only if the allocated supply is less than S_i^{\min} . Figure 6 compares the optimized supply chain cost pertaining to the ATP schedule and the two ATS-based approaches when the contingency factor takes on values $\alpha = 0, 0.1, 0.2,$ and 0.3 .

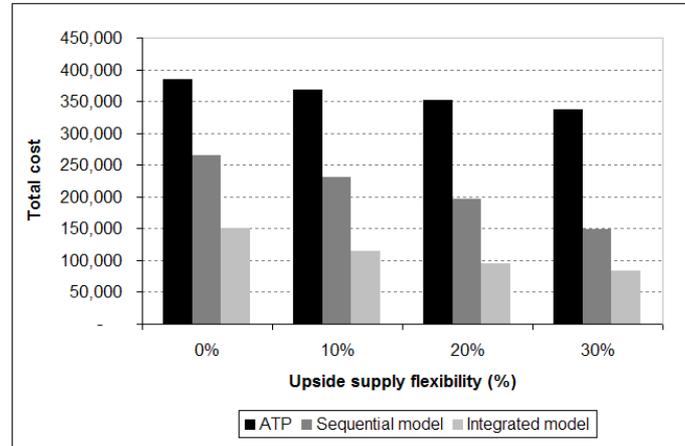


Figure 6. Total Cost of ATP and ATS Models as a Function of Supply Quantity Flexibility

We observe that the supply chain cost of the ATP schedule improves from \$385,000 for the base scenario (no upside flexibility) to \$337,500 for the scenario with 30 percent flexibility for a 12 percent overall cost reduction. This result is intuitive because higher supply flexibility provides more opportunities for building squared set configurations without increasing the manufacturer's inventory liability exposure. For the ATS-based approaches the cost reductions derived from supply flexibility are more dramatic. In the sequential model, the additional supply flexibility translates into 43 percent cost savings (\$265,000 supply chain cost in the base scenario vs. \$148,500 for the scenario with 30 percent flexibility) and in the integrated model the additional flexibility translates into 45 percent savings (\$150,000 vs. \$83,250 supply chain costs). Furthermore the results confirm that even a modest level of upside flexibility can result in significant percentage cost improvements. For example, when $\alpha = 0.1$ the supply chain cost decreases by 13 percent in the sequential model and by 24 percent in the integrated model compared to the base case. Table 7 summa-

izes inventory liability costs, backorder cost, and substitution cost for the different management approaches.

	ATP			ATS (sequential model)			ATS (integrated model)		
	Backorders	Excess Inventory	Substitution	Backorders	Excess Inventory	Substitution	Backorders	Excess Inventory	Substitution
No flexibility	\$ 275,000	\$110,000	-	\$ 175,000	\$70,000	\$ 20,000	\$ 75,000	\$ 30,000	\$ 45,000
10% flexibility	\$ 262,500	\$106,250	-	\$ 145,000	\$62,250	\$ 23,500	\$ 45,000	\$ 24,250	\$ 44,500
20% flexibility	\$ 250,000	\$102,500	-	\$ 115,000	\$54,500	\$ 27,000	\$ 30,000	\$ 21,500	\$ 44,000
30% flexibility	\$ 237,500	\$100,000	-	\$ 72,500	\$43,000	\$ 33,000	\$ 20,000	\$ 19,750	\$ 43,500

Table 7. Detail Cost of ATP and ATS Models as a Function of Supply Quantity Flexibility

Not surprisingly the backorder costs and inventory liability penalties decrease as the supply flexibility increases. However, the relative impact of the additional supply flexibility on cost is substantially higher in the integrated model than in the sequential model. For example, a 10 percent up-side flexibility yields 40 percent less backorders for the integrated model but only 17 percent less backorders in the sequential model compared to the base case. The reason is that the integrated model not only recognizes supply-constrained products, but is capable of finding an appropriate set of saleable product alternatives that can be offered to customers in place of the constrained products at minimal or no additional cost. Notice that the integrated model carries the highest substitution cost which suggests that the optimized conditioned ATP schedules may deviate considerably from the original ATP.

6. Summary

Most firms recognize that supply-demand imbalances differ in magnitude and severity of impacts, but rarely invest resources to develop a systematic approach for resolution. Some leading companies have begun to implement availability management processes that seek to dynamically adjust

product offerings to guide marketing and sales teams. However, these are almost always manual processes that rely on expert's knowledge and partial sets of data. With hundreds or even thousands of product offerings in a typical product portfolio it is impossible to manually assemble the data and reach an optimal conclusion. An automated approach based on optimization not only ensures that the resources invested have the likelihood of producing a successful result, but offers the additional advantage of speed in execution which is critical for a timely resolution of short-term demand-supply imbalances.

In this chapter we have described an optimization-based approach to availability management in assemble-to-order supply chains and have outlined the business processes required to incorporate availability management into supply chain operations. The proposed approach aims at finding marketable product alternatives in a product portfolio that best utilize inventory surplus and replace demand on supply-constrained components. We formulated the problem as a nonlinear program and developed efficient computational procedures to generate optimal solutions. We demonstrated the benefits of product offering conditioning through numerical experiments with realistic production data that were chosen to highlight the advantages of a more sophisticated availability management process.

The principles and capabilities described in this chapter enable manufacturers to determine a financially viable, profitable, and marketable product portfolio, help automate the decisions to resolve supply-demand imbalances and take effective actions to avoid excess and surplus component inventory, and articulate marketable alternate product offerings. They can easily be imbedded into supply chain operations to improve day-to-day flexibility. For example, direct sales businesses that deal with customers directly through their website or telesales system can highlight featured

products on-the-fly based on current component availability and steer customers towards product configurations that they can supply easily and profitably. The models featured in this chapter have already contributed to business improvements in real-world supply chains. IBM has implemented the sequential process described in section 2 in its complex-configured server product line and has executed an ATS business process since 2002. The process has resulted in a \$100M reduction in the first year and over \$20M of inventory reduction in each of the following three years. The diminishing returns were expected because there is typically a “windfall” of excess inventory when first deployed. The inventory reductions only pertain to formally declared excess inventory, i.e. inventory that has no projected demand for over 90 days. Since short-term overages are not declared as excess inventory, the dollar figures exclude additional savings from excess avoidance which can often be substantial.

Acknowledgments

We thank Larry Phillips and Rich Bell of Lenovo and Blair Binney and Dan Peters of IBM for sharing their knowledge and insights on event-driven availability management. We also thank Reha Uzsoy for pointing us to the literature on reverse logistics supply chains.

References

- Balakrishnan, A. and J. Geunes. 2000. Requirements Planning With Substitutions: Exploiting Bill-of-Materials Flexibility in Production Planning. *Manufacturing and Service Operations Management* 2, 2. 166-185.
- Ball, M.O, C.Y. Chen and Z.Y. Zhao. 2004. Available to Promise. In: Simchi-Levi, D., Wu, S.D. and Shen, Z.J. (eds). *Handbook of Quantitative Supply*

- Chain Analysis - Modeling in the e-Business Era*. Kluwer Academic Publishers. 447-480.
- Barnhart, C., E.L. Johnson, G.L. Nemhauser, M.W.P. Savelsbergh and P.H. Vance. 1998. Branch-And-Price: Column Generation for Solving Huge Integer Programs. *Operations Research* **46**, 3. 316 – 329.
- Buckley, S., M. Ettl, G.Y. Lin and K.Y.Wang. 2005. Intelligent Business Performance Management - Sense and Respond Value Net Optimization. In: An, C. and H. Fromm (eds.). *Supply Chain Management on Demand*. Springer. 287-311.
- Cecere, L., D. Hofman, R. Martin and L. Preslan. 2005. The Handbook for Becoming Demand Driven. AMR Research.
- Chen, C.-Y., Z. Zhao, and M.O. Ball. 2002. A Model for Batch Advanced Available-to-Promise. *Production and Operations Management* **11**, 424-440.
- Chen, L. and E. Plambeck. 2005. Dynamic Inventory Management with Learning about the Demand Distribution and Substitution Probability. Working Paper. Graduate School of Business. Stanford University.
- Chen-Ritzo, C.-H. 2006. Availability Management for Configure-to-Order Supply Chain Systems. PhD Dissertation. Pennsylvania State University.
- Dekker, R., M. Fleischmann, K. Inderfurth and L.N. van Wassenhove. 2003. *Reverse Logistics: Quantitative Models for Closed Loop Supply Chains*. Springer, Berlin.
- Dietrich, B., D. Connors, T. Ervolina, J.P. Fasano, R. Lougee-Heimer and R. Witroek. 2005. Applications of Implosion in Manufacturing. In: An, C. and H. Fromm (eds.). *Supply Chain Management on Demand*. Springer. 97-115.
- Ervolina, T. and B. Dietrich. 2001. Moving Toward Dynamic Available-to-Promise. In: Gass, S. and Jones. A.T. (eds.) *Supply Chain Management Practice and Research: Status and Future Directions*. 1-19.
- Ettl, M., P. Huang, K. Sourirajan and F. Cheng. (2006). Product Offering Conditioning in Assemble-to-Order Supply Chains. IBM Research Report RC 23878. <http://domino.watson.ibm.com/library/CyberDig.nsf/Home>

- Fleischmann, M., J. van Nunen, B. Graeve and R. Gapp. 2005. Reverse Logistics – Capturing Value in the Extended Supply Chain. In: An, C. and H. Fromm (eds.). *Supply Chain Management on Demand*. Springer. 167-186.
- Gupta S.M. and K.N. Taleb. 1994. Scheduling Disassembly. *International Journal of Production Research* **32**. 857-866.
- Hale, W., D.F. Pyke and N. Rudi. 2001. An Assemble-to-Order System with Component Substitution. Working Paper. Tuck School of Business at Dartmouth.
- Kapoor, S., K. Bhattacharya, S. Buckley, P. Chowdhary, M. Ettl, K. Katircioglu, E. Mauch and L. Phillips. 2005. A Technical Framework for Sense-and-Respond Business Management. *IBM Systems Journal* **44**, 1. 5-24.
- Lin, G.Y., S. Buckley, H. Cao, N. Caswell, M. Ettl, K. Katircioglu, L. Koenig, A. Nigam, B. Ramachandran and K.Y. Wang. 2002. The New Frontier: Sense and Respond System for Value Chain Optimization. *ORMS Today* **29**, 2. 34-39.
- Lin, G. Y., J.J. Jeng and K.Y.Wang. 2004. Enabling Value Net Collaboration. In: Chang, Y.S., H.C. Makatsoris and H.S. Richards (eds.) *Evaluation of Supply Chain Management*. Kluwer Academic Publishers. 417-430.
- Lee, H. 2004. The Triple-A Supply Chain. *Harvard Business Review* **82**, 10. 102-112.
- Meacham, A., R. Uzsoy and U.Venkatadri. 1999. Optimal Disassembly Configurations for Single and Multiple Products. *Journal of Manufacturing Systems* **18**, 5. 311-322.
- O'Marah, K. and J. Souza. 2004. DDSN: 21st Century Supply on Demand. *AMR Research*.
- Taleb, K.N., S.M. Gupta and L. Brennan. 1997. Disassembly of Complex Product Structures With Parts and Materials Commonality. *Production Planning and Control* **8**, 3. 255-269.
- Xu, S. and Y. Akcay. 2004. Joint Inventory Replenishment and Component Allocation Optimization in an Assemble-to-Order System. *Management Science* **50**, 99-116.