TECHNICAL RESEARCH REPORT

Optimization-Based Available-To-Promise with Multi-Stage Resource Availability

by Zhenying Zhao, Michael O. Ball

TR 2005-4

Optimization-Based Available-To-Promise with Multi-Stage Resource Availability

ZHENYING ZHAO and **MICHAEL O. BALL** mball@rhsmith.umd.edu

Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, MD 20742

Masahiro Kotake

Corporate Manufacturing Engineering Center, Toshiba Corporation, 33, Shin-Isogo-Cho, Isogo-Ku , Yokohama 235-0017, JAPAN

Abstract. Increasingly, customer service, rapid response to customer requirements, and flexibility to handle uncertainties in both demand and supply are becoming strategic differentiators in the marketplace. Organizations that want to achieve these benchmarks require sophisticated approaches to conduct order promising and fulfillment, especially in today's high-mix low-volume production environment. Motivated by these challenges, the Available-to-Promise (ATP) function has migrated from a set of availability records in a Master Production Schedule (MPS) toward an advanced real-time decision support system to enhance decision responsiveness and quality in Assembly To Order (ATO) or Configuration To Order (CTO) environment. Advanced ATP models and systems must directly link customer orders with various forms of available resources, including both material and production capacity. In this paper, we describe a set of enhancements carried out to adapt previously published mixed-integer-programming (MIP) models to the specific requirements posed by an electronic product supply chain within Toshiba Corporation. This model can provide individual order delivery quantities and due dates, together with production schedules, for a batch of customer orders that arrive within a predefined batching interval. The model considers multi-resource availability including manufacturing orders, production capability and production capacity. In addition, the model also takes into account a variety of realistic order promising issues such as order splitting, model decomposition and resource expediting and de-expediting. We conclude this paper with comparison of our model execution results *vs*. actual historical performance of systems currently in place.

Keywords: Available-To-Promise (ATP); Manufacturing Order; Production Capability; Order-Promising and Fulfillment; Mixed-Integer-Programming.

The ability to effectively match demand and supply is fundamental to nearly all supply chain management processes. Under a push-based strategy, demand forecasts are used to match demand and supply, whereas under a pull-based strategy, supply is directly driven by actual customer orders. Increasingly, the "pure" form of each of these strategies is rarely employed, so typical supply chains have an upstream push-portion and a downstream pull-portion, which meet at the push-pull boundary. A key function in such supply chains is the coordination of activities across this boundary. The available-to-promise (ATP) business function can be interpreted as carrying out this role.

The basic purpose of the ATP function is to provide a response to customer order requests based on resource availability. In order to make a reliable response to a customer order, an ATP system must insure that the quantity promised can be delivered on the date promised. Thus, an ATP system must include both order promising and order fulfillment capabilities. In addition, an ATP system should be able to dynamically adapt resource utilization and to prioritize customer orders so as to coordinate supply and demand in a way that maximizes profit. *By its very nature, the ATP system should operate within a short-term operational environment where most resource availability is considered fixed because of raw material procurement lead-time limitations. This distinguishes ATP systems from traditional planning, scheduling and inventory management processes.*

Conventional ATP is associated with a traditional make-to-stock (MTS) production environment associated with long process lead times, relatively standard products and stable demand. In the Materials Resource Planning (MRPII) framework, production decisions are based on the embedded Master Production Schedule (MPS), which takes into account a demand forecast, committed customer orders, existing inventory and production capacity. Hence, APICS defines ATP as "*The uncommitted portion of a company's inventory and planned production, maintained in the master schedule to support customer order promising* (see definition in APICS

(1987)).” Traditionally, the ATP scope includes the on-hand inventory and the planned production at a designated location. The MPS becomes “moderately firm” or even “frozen” once a designated time window is reached. This implies that the planned production quantity becomes static as the planned production time approaches.

Unlike conventional ATP practice, the advanced ATP function studied in this paper refers to a systematic process of making best use of available resources including multi-stage material resource and capacity resource, to commit customers’ orders over a short planning time horizon. Advanced ATP is an execution mechanism. It must take into account uncertainties and changes from exterior suppliers and customers, as well as interior production processes. It must resolve the discrepancy between the push-based forecast-driven planning process and pull-based order-driven execution process. Figure 1 illustrates a simplified framework of ATP practice in an electronic product supply chain at Toshiba Corporation. Because of its central role, the advanced ATP system needs not only to retrieve information (e.g. on the status of inventory, transportation, orders and shipments) from other management modules such as sales and marketing, production planning and procurement, but also to export the results to multiple business modules like sales and production. Certainly, it needs seamless integration with these business modules.

In Figure 1, the forecasted demands are generated based on historical sales and judgment from sales and marketing. Then, various production planning process including aggregate planning, master production scheduling and material requirement planning can be employed to generate the detailed material requirements and capacity requirements plan. Based on these plans, the procurement department will produce the specific procurement orders (POs) to the suppliers, and the production department will allocate production capacity availability. Suppliers deliver raw materials based on the POs and push these to the factory for production. Meanwhile, customers place orders to sales and marketing with a requested product, quantity and due date. Considering internal and external resource availability, sales personnel need to respond to customers about the commitment and delivery schedule either immediately or, within a short time period. Well designed ATP functionality should synchronize and match the push and pull "forces" leading to lower inventory, fewer stock-outs, higher resource utilization and less waste.

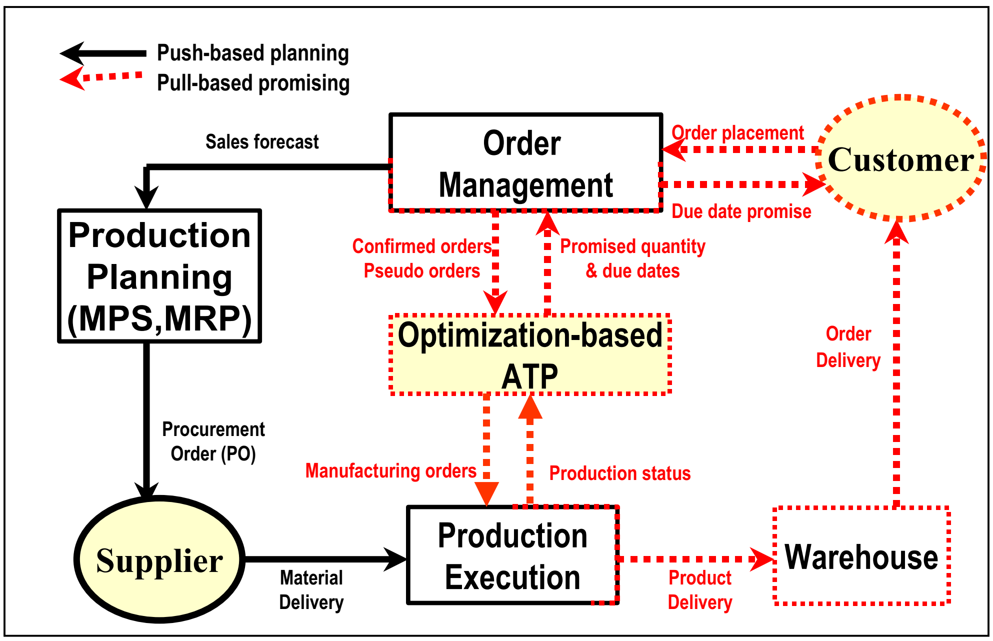

Figure 1 Framework of advanced ATP mechanism in Toshiba electronic product supply chain

The time horizon for advanced ATP is usually short compared with that for planning processes like MPS, MRP, DRP and so on. It ranges from a few days to several weeks. Such time horizons are typically much shorter than the order lead time for some or all required materials. For the shortest time horizons, e.g. a few days, production schedules may be partially or completely fixed. Thus, the advanced ATP problem is one of allocating a fixed set of resources to orders, where both the order requirements and resource availability are time dependent. For a more detailed review of ATP see Ball, Chen and Zhao (2002).

In previous research (see Chen, Zhao and Ball (2000), (2001) and Ball, Chen and Zhao (2002)) we have proposed mixed integer programming models for supporting advanced ATP. For a batch of customer orders that arrive within a predefined batching interval, these models determine which orders to accept or reject, together with individual order delivery quantities and due dates. Production schedules and detailed resource assignments are also output. In this paper, we described adaptations of these models to handle the particular production environment for a Toshiba electronic product. The adaptation required that the model consider multiple levels of resource availability that varied over time. For example, resources in the near-term were defined based on specific manufacturing orders, in the intermediate term based on material availability and in the longer term based only on production capacity. In addition, the adaptation also takes into account a variety of realistic order promising issues such as order splitting, model decomposition and resource expediting and de-expediting. The rest of this paper is organized as follows. In the next section, we provide a literature review. Section 2 describes the ATP problem with multi-stage resources and formulates our mixed-integer-programming model. Section 3 describes our experiments and results. Section 4 gives conclusions and remarks.

1. Literature review

Traditional ATP systems are based on the MPS, which are derived from the aggregate production plan, detailed end item forecasts, and existing inventory and orders (Vollman 1992). Increasingly, one finds articles that either discuss the needs of, or propose the features for, ATP systems (for example, Lee and Billington (1995), Zweben (1996), Fordyce and Sullivan (1999), Robinson and Dilts (1999), and SAP AG (1999)). Among others, eB2X, Inc. (2000), Hill (2000), and Manugistics, Inc. (2000) emphasize the importance of adopting ATP systems to support order promising and fulfillment decisions. However, only a very limited number of papers present quantitative models to support ATP. Hariharan and Zipkin (1995) evaluate the impacts of information for “advance ordering” – customer orders with specified due dates – on inventory policies using stochastic models. Thus, it is not surprising to find several papers (eB2x (2000), Fordyce (1999), Lee (1995), Robinson (1999), and Zweben (1996) that discuss the need for advanced ATP systems, which provide order promising capabilities based on current capacity and inventory conditions within the firm’s supply chain.

Recent research is starting to address ATP issues from order-promising perspective. Taylor and Plenert (1999) introduce a heuristic technique called Finite Capacity Promising (FCP) that keeps track of traditional ATP quantities to generate feasible due dates to promise customer orders. Kilger and Schneeweiss (2000) describe the concept of APS-based “Allocated ATP” motivated by seat-class allocation in airline yield management. Ervolina and Dietrich (2001) describe an approach to carry out efficient order promising in an assemble-to-order setting based on the concept of feature sets. They describe models to allocate available materials to feature sets. The feature set quantities are then used to support the order promising process.

The work in the paper builds on previous models given in Chen, Zhao and Ball (2000) and Ball, Chen and Zhao (2001). These papers develop mixed-integer-programming models for allocating available components and production capacity to competing customer order requests that arrive within a pre-determined batching interval. These models have strong temporal component in that both the component availability and production capacity vary over time and the orders have constraints on possible delivery dates. These models can be viewed as both order-promising and order-fulfillment models, since they specify a schedule for the use of production capacity.

Since the models we present in this paper must consider production capacity and, for the near-term order promising, must assign orders to a fixed production schedule, a part of the functionality of these models can be viewed as production scheduling. There is an extensive literature in the job-shop scheduling area, which quotes manufacturing due dates with various assignment rules, control methods, or analytical models (e.g., McFeely, Simpson, and Simmons (1997), Tsai, Chang, and Li (1997), Hopp and Sturgis (2000), and Duenyas and Hopp (1995)). Cheng and Gupta (1989) offer an earlier survey in this area. They categorize all due date assignment methods into two categories: exogenous (determined by independent external agency) and endogenous (assigned internally by the scheduling model). A paper by Park and Kim (2000) and our previous ATP model consider orders or jobs with exogenous due dates, whereas the due date setting model of Hegedus and Hopp (2001) and quantity-and-due-date-quoting model of Chen, Zhao and Ball (2002) belong to the endogenous category. Although these two models both attempt to optimize due date quoting for customer orders, their model considers stochastic production lead times (in a two-stage production model) and uses a news-vendor-like analytical formulation to obtain minimum-cost due dates for each customer order independently. On the other hand, our model assumes deterministic production lead times and uses mixed-integer-programming to quote due dates for multiple orders within a batch simultaneously. Thus, our model takes into account the current status of the production system, can dynamically allocate and reallocate material and capacity and can trade off the profitability of various orders, whereas their model takes the allocation of materials as static information pre-determined by MRP.

The underlying structure of our optimization-based ATP model is similar to that of many of the production planning and scheduling models in the literature. Thore (1991) summarizes a generic mathematical programming model to maximize a social-welfare-style objective function over logistics networks, which consists of three fundamental dimensions: spatial, vertical, and time. The spatial dimension involves transportation among different geographical locations; it could represent both inbound shipments of raw materials and outbound distribution of finished products. The vertical dimension models the BOM relationship between raw materials and finished products, including intermediate subassemblies. The time dimension simply keeps track of inventory over time. Johnson and Montgomery (1974) describe broader and more in-depth production planning, scheduling, and inventory control models, which, in some cases, include integer (as well as continuous) decision variables.

2. Optimization-based ATP model

We now describe a Mixed-Integer-Programming (MIP) model for the ATP decision problem associated with a particular electronic product (denoted by EP) manufactured by Toshiba Corporation. The order promising process proceeds by iteratively collecting and processing batches of orders. The ATP model is used to determine delivery dates, a decision on whether to split the order and the production schedule for each order. The model must balance available resources relative to a batch of orders requesting multiple products that share certain common components. The objective function criteria include minimization of *due date violation*, *inventory holding cost* and a day-to-day *production smoothness* measure. The due date violation is computed as the sum, over all orders, of the amount delivered late times the number of days late. The holding cost contains both a material holding cost and a finished product inventory holding cost. Production smoothness is based on a measure of day-to-day variation in the production amount of each assembly line at each factory.

2.1 Problem description

The EP supply chain consists of multiple final assembly and testing (FAT) factories all located in Japan, which provide EPs delivered directly to both domestic and international business customers. Due to high product mix, an assembly-to-order (ATO) production framework is employed to increase the degree of product flexibility. The order promising and fulfillment process involves in total several thousand-product models. Order sizes range from a very small number of units to a few hundred. Orders are generated by one of several sales units and are processed by a single central order processing system in Toshiba headquarters. The ATP system collects orders over a 1/4 hour time interval and returns commitments to the sales offices at the end of each ATP run (1/4 hour interval). Order commitments are booked up to approximately ten weeks in advance of delivery.

In the order promising process for EP, Toshiba employs the business practice of never denying an order. If an order cannot be fulfilled before its requested due date, then a promise date beyond the requested date is given, i.e. it is backordered, or the order is split with a portion given an early promise date, e.g. before the due date and a portion given one or more later promise dates. However, an order cannot be split among different factories, namely, one order can only be committed in one factory. In order to emphasize customer satisfaction for EP, Toshiba weights due date violation higher than any holding costs and production smoothness penalty in its order fulfillment decision models. Occasionally, the sales staff will book “pseudo orders” based on enquiry orders from customers to reserve critical resources for anticipated future high priority demands.

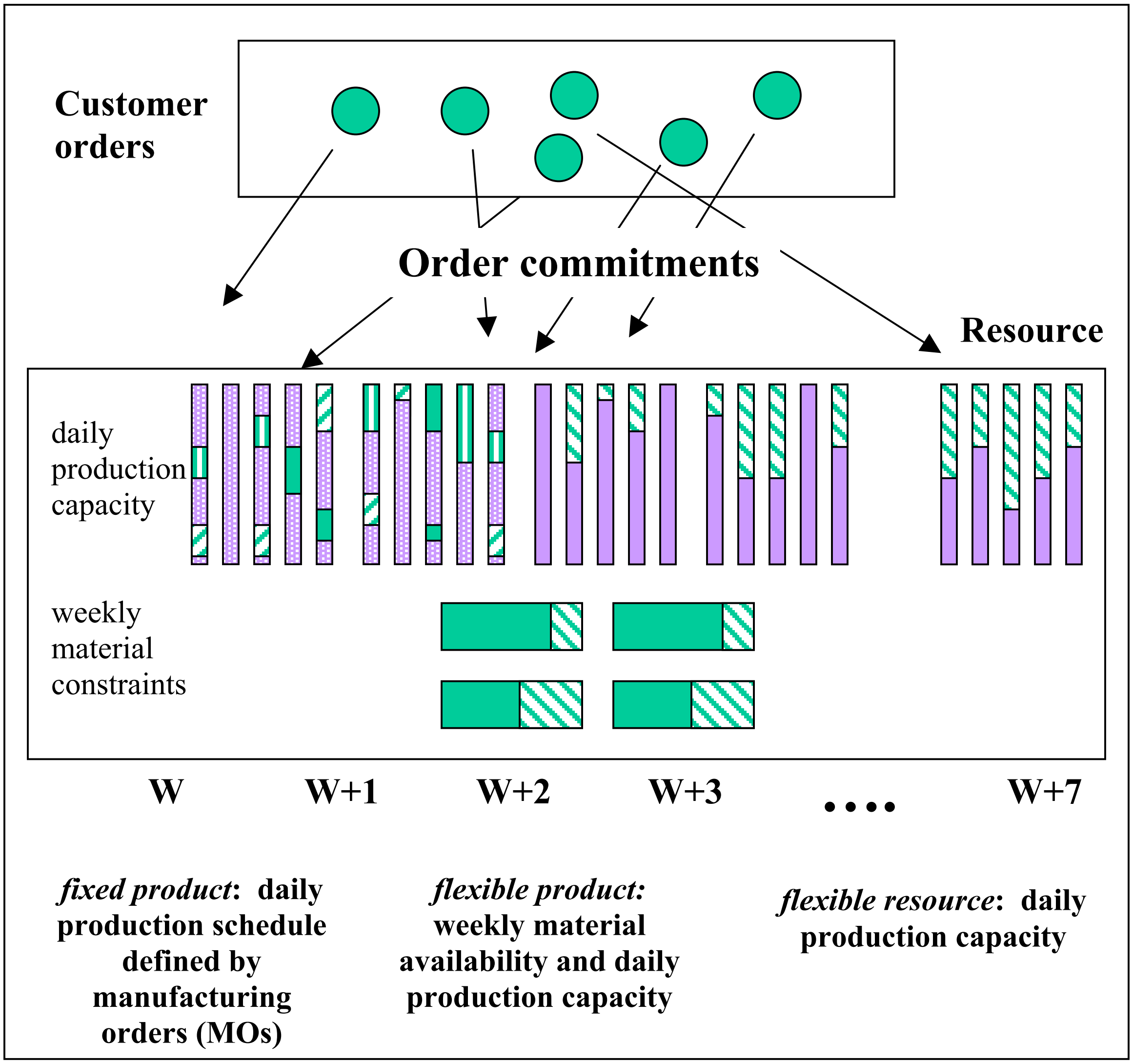

Figure 2: Change in Form of Resources over Time

As illustrated in Figure 2, the nature of the resource constraints varies across the order-promising time horizon. For the *fixed product* interval, which spans from approximately the present time to two weeks into the future, resources, in the form of manufacturing orders (MO) are fixed. An MO specifies the production quantity for each product at each assembly line in each factory. That is, a fixed production schedule is set, which takes into account both production capacity availability and critical material availability. Having a fixed schedule stabilizes production dynamics in the near term and allows for the required materials to be set up and put in place. Any order commitments made for this time interval must fit within the fixed production schedule.

In the *flexible product* interval, two kinds of resources, capacity and material, are considered in order promising. The production capacity is given daily at the factory level in terms of machine-hour and manpower availability. The weekly availability of individual critical materials is aggregated into finished goods level availability grouped based on the bill of material (BoM), balance-on-hand inventory, pipeline inventory and scheduled receipts. It is defined as a *Production Capability* (PC). Any order commitments made for this time interval must satisfy the capacity and material availability constraints. The flexible product interval spans from approximately two weeks to two months into the future.

For the *flexible resource* interval, which covers due dates three weeks into the future, the only constraint considered is production capacity, which is specified daily at factory level in terms of machine-hour and manpower availability. This interval starts beyond the material resource lead times so any resource commitments can be met.

2.2 Notation and Assumptions

Let $\boldsymbol{F}$ be the index set of assembly factories, $\boldsymbol{L}$ the index set of final assembly and testing lines, $\boldsymbol{I}$ the index set of products, $\boldsymbol{J}$ the index set of production capabilities, and $\boldsymbol{M}$ the index set of manufacturing orders. The combination of an assembly line l and a factory f, called a *line instance*, defines a specific line in a factory and is denoted by an ordered pair (f,l), where $f \in \boldsymbol{F}$ and $l \in \boldsymbol{L}$. The combination of a factory f and a production capability j, called a *production capability instance*, defines a particular production capability at a factory and is defined by an ordered pair (f, j), where $f \in \boldsymbol{F}$ and $j \in \boldsymbol{J}$.

Since both weekly and daily resources are involved in the order promising and fulfillment process, an ordered pair (w, d), called a *period instance*, is defined as the index set for a week w and the day-of-the-week d. We assume the current week is week $w^0 + 1$, which is also the starting week for an ATP execution. In this paper to simplify the presentation, we assume the starting day is always Monday (the 1st day of the week) but in the actual implementation we allow any starting day. The planning horizon used in the ATP model is τ weeks, i.e. the scope of the current ATP batch run includes every time period w, such that $w^0 < w \le w^0 + \tau$. Furthermore, suppose that the time lengths of the fixed product period and the fixed product period plus flexible product period are given as τ^m weeks and τ^p weeks, respectively. The week numbers at the end of the fixed product period, the flexible product period and the flexible resource period will be $w^0 + \tau^m$, $w^0 + \tau^p$, and $w^0 + \tau$, respectively. In each week, the working days are defined as $d = 1, 2, .., \omega$, where ω represents the last day of the week.

Under consideration in each ATP batch execution is a set of *newly-arrived customer orders*, which were collected during the most recent batch interval, and a set of *previously promised customer orders*, which arrived in earlier intervals. The previously-promised customer orders consist of all orders that have been accepted, and committed in the preceding time periods, but have not yet been delivered. When considering the newly-arrived orders, all previously-promised orders may be re-scheduled as long as the promised delivery dates are respected. Let $\boldsymbol{K}$ and $\boldsymbol{K'}$ be the index sets of the newly-arrived customer orders and previously-promised customer orders, respectively. Let $\boldsymbol{o}^k$ denote the *k*th newly-arrived order and $\widehat{\boldsymbol{o}}^{k'}$

denote the k'th previously-promised order. A newly-arrived order consists of a product model, due week, day-of-the-week, and requested quantity. It is denoted by the four-tuple: $\boldsymbol{o}^k = \left(p^k, w^k, d^k, q^k\right)$, where $p^k \in \boldsymbol{I}$ is the requested product model, $w^0 < w^k \le w^0 + \tau$ the requested due week, $1 \le d^k \le \omega$ the requested due day of the week, and q^k the requested quantity of the newly-arrived order $k \in \boldsymbol{K}$. A previously-promised customer order consists of a product model, due week, day-of-the-week, requested quantity, and the previously-promised due date violation. It is denoted by the five-tuple: $\hat{\boldsymbol{o}}^{k'} = \left(\hat{p}^{k'}, \hat{w}^{k'}, \hat{d}^{k'}, \hat{q}^{k'}, \hat{v}^{k'}\right)$, where $\hat{p}^{k'} \in \boldsymbol{I}$ is the requested product model, $w^0 < \hat{w}^{k'} \le w^0 + \tau$ the requested due week, $1 \le \hat{d}^{k'} \le \omega$ the requested due day of the week, $\hat{q}^{k'}$ the requested quantity, and $\hat{v}^{k'}$ the previously-promised due date violation (i.e., the corresponding due date violation in the previous ATP execution) of the previously-promised order $k' \in \boldsymbol{K'}$. The per unit per time period due date violation cost is given by c.

In the fixed product period, let $mo_{im}^{fl}(w,d)$ denote the quantity of the mth ($m \in \boldsymbol{M}$) MO to assemble product $i \in \boldsymbol{I}$ at line instance $\left(f,l\right)$ in period instance $\left(w,d\right)$, where $w^0 < w \le w^0 + \tau^m$ (fixed product period). In flexible product period and flexible resource period, the production capacity availability is given as the sum of regular production time and overtime (hours), $t^{fl}(w,d)$, for line instance $\left(f,l\right)$ in period instance $\left(w,d\right)$, where $w^0 + \tau^m < w \le w^0 + \tau$ and $1 \le d \le \omega$. The per unit production costs under regular time and overtime are assumed to be same and given as cp^f (dollars) in the fth factory. The processing time of the ith product in line instance (f,l) is specified as pt_i^{fl} (hours). Meanwhile, production is subjected to the satisfaction of minimum lot size s_i^{fl} for the *i*th product in line instance (f,l).

In the flexible product period, assume that the production capability is pc_j^f for the production capability instance (f,j). One production capability consists of an available quantity, available week period, number of weeks expeditable and de-expeditable, and product candidates. The available quantity refers to the maximum quantity of finished products that can be produced with this production capability, and the available week period denotes the time period that this production capability will be available. The number of weeks expeditable and de-expeditable indicates a fixed number of weeks that this production capability can be expedited by paying an extra cost, or de-expedited by paying a penalty. The product candidates give the set of products that can be produced with the production capability (and combination of the product candidates can be produced so long as the total quantity produced is no greater than the available quantity). We represent a production capability by the five-tuple: $\boldsymbol{pc}_j^f = \left(g_j^f, w_j^f, e_j^f, \overline{e}_j^f, \boldsymbol{u}_j^f\right)$, where $g_j^f, w_j^f, e_j^f, \overline{e}_j^f$, $\boldsymbol{u}_j^f$ are the available quantity, the available week, the number of weeks expeditable, the number of weeks de-expeditable, and the product candidate set of production capability instance (f,j). The product candidate set, $\boldsymbol{u}_j^f$, is denoted by a product candidate vector $\boldsymbol{u}_j^f = \left(p_j^{rf}\right)$, in which $p_j^{rf} \in \boldsymbol{I}$ is the product candidate $r \in \boldsymbol{R}_j^f$, where $\boldsymbol{R}_j^f$ is the product candidate index sets for production capability instance (f,j). The per-unit expedite cost and the de-expedite penalty, which depend on the factory location, are given by ce_j^f and $c\overline{e}_j^f$, respectively. The per-unit holding cost of finished products allocated to specific customer orders and not to any customer orders for factory f are specified by ch^f and $\beta \cdot ch^f$, where $\beta \geq 1$. The initial inventory of the ith product in factory f is given as h_i^f.

To formulate the ATP model, we define the following decision variables.

$C^{kf}(w,d)$: the quantity produced in factory f for newly-arrived order k in period instance (w,d).

$\widehat{C}^{k'f}(w,d)$: the quantity produced in factory f for previously-promised order k' in period instance (w,d).

Z^{kf}: newly-arrived order commitment indicator (1, if the specific newly-arrived order k is committed in factory f; 0, otherwise).

$\widehat{Z}^{k'f}$: previously-promised order commitment indicator (1, if the specific previously-promised order k' is committed in factory f; 0, otherwise).

B^k: the quantity uncommitted for newly-arrived order k in current ATP run.

$\widehat{B}^{k'}$: the quantity uncommitted for previously-promised order k' in current ARP run.

$Q_i^{fl}(w,d)$: the quantity of product i produced at line instance (f,l) in period instance (w,d).

$Y_i^{fl}(w,d)$: 1, if product i is produced on line instance (f,l) in period instance (w,d); 0, otherwise (used to enforce lot size constraints).

$I_i^f(w,d)$: the inventory of product i in the fth factory at the end of period instance (w,d).

$X_{ij}^f, X_{ij}'^f, X_{ij}''^f$: the regular quantity, expedited quantity, and de-expedited quantity of production capability instance (f,j) for producing product i. The quantity will be zero if the ith product is not a candidate of production capability instance (f,j).

V^k: the due date violation of the kth newly-arrived order.

$\widehat{V}^{k'}$: the due date violation of the k' th previously-promised order.

H^f: the total cost in factory *f*.

$\hat{U}^{fl}, \breve{U}^{fl}$: the maximum and minimum production capacity used in one day at line instance (f, l).

N: a large constant.

2.3 Model Formulation

The MIP formulation is given by:

Minimize

$$w' \cdot c \cdot \left(\sum_{k \in \boldsymbol{K}} V^k + \sum_{k' \in \boldsymbol{K}'} \widehat{V}^{k'} \right) + w'' \cdot \sum_{f \in \boldsymbol{F}} H^f + w''' \cdot \alpha \cdot \sum_{f \in \boldsymbol{F}, l \in \boldsymbol{L}} \left(\hat{U}^{fl} - \breve{U}^{fl} \right) \tag{1}$$

Subject to

Order promising and fulfillment:

$$\sum_{f \in \boldsymbol{F}} \sum_{w=w^0+1}^{w^0+\tau} \sum_{d=1}^{\omega} C^{kf}(w,d) = q^k - B^k \quad \text{for all } k \in \boldsymbol{K} \tag{2}$$

$$\sum_{f \in \boldsymbol{F}} \sum_{w=w^0+1}^{w^0+\tau} \sum_{d=1}^{\omega} \widehat{C}^{k'f}(w,d) = \widehat{q}^{k'} - \widehat{B}^{k'} \quad \text{for all } k' \in \boldsymbol{K}' \tag{3}$$

$$\sum_{w=w^0+1}^{w^0+\tau} \sum_{d=1}^{\omega} C^{kf}(w,d) \le Z^{kf} \cdot N \quad \text{for all } k \in \boldsymbol{K}, f \in \boldsymbol{F} \tag{4}$$

$$\sum_{w=w^0+1}^{w^0+\tau} \sum_{d=1}^{\omega} \widehat{C}^{k'f}(w,d) \le \widehat{Z}^{k'f} \cdot N \quad \text{for all } k' \in \boldsymbol{K}', \ f \in \boldsymbol{F} \tag{5}$$

$$\sum_{f \in \boldsymbol{F}} Z^{kf} = 1 \quad \text{for all } k \in \boldsymbol{K} \tag{6}$$

$$\sum_{f \in \boldsymbol{F}} \widehat{Z}^{k'f} = 1 \quad \text{for all } k' \in \boldsymbol{K}' \tag{7}$$

Due date violation calculation:

$$V^k = \sum_{w=w^k+1}^{w^0+\tau} \sum_{d=1}^{\omega} \left(\left(\left(w - w^k \right) \cdot \omega + d - d^k \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{kf}(w,d) \right) \right) + \sum_{w=w^k}^{w^k} \sum_{d=d^k+1}^{\omega} \left(\left(d - d^k \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{kf}(w,d) \right) \right) + B^k \cdot \left(\left(\tau - w^k + 1 \right) \cdot \omega - d^k \right) \quad \text{for all } k \in \boldsymbol{K} \tag{8}$$

$$\widehat{V}^{k'} = \sum_{w=\widehat{w}^{k'}+1}^{\widehat{w}^{k'}+\tau} \sum_{d=1}^{\omega} \left(\left(\left(w - \widehat{w}^{k'} \right) \cdot \omega + d - \widehat{d}^{k'} \right) \cdot \left(\sum_{f \in \boldsymbol{F}} \widehat{C}^{k'f}(w,d) \right) \right) + \sum_{w=\widehat{w}^{k'}}^{\widehat{w}^{k'}} \sum_{d=\widehat{d}^{k'}+1}^{\omega} \left(\left(d - \widehat{d}^{k} \right) \cdot \left(\sum_{f \in \boldsymbol{F}} \widehat{C}^{k'f}(w,d) \right) \right) + B^{k'} \cdot \left(\left(\tau - \widehat{w}^{k'} + 1 \right) \cdot \omega - \widehat{d}^{k'} \right) \quad \text{for all } k' \in \boldsymbol{K'} \tag{9}$$

Due date violation requirements for previously-promised customer orders:

$$\widehat{V}^{k'} \le \widehat{v}^{k'} \qquad \text{for all } k' \in \boldsymbol{K'} \tag{10}$$

Finished-product flow:

$$I_i^f(w,d) = I_i^f(w,d-1) - \sum_{k \in \boldsymbol{K} \,\&\, i=p^k} C^{kf}(w,d) - \sum_{k' \in \boldsymbol{K'} \,\&\, i=\widehat{p}^{k'}} \widehat{C}^{k'f}(w,d) + \sum_{l \in \boldsymbol{L}} Q_i^{fl}(w,d)$$
$$\text{for all } i \in \boldsymbol{I},\ f \in \boldsymbol{F},\ w^0 < w \le w^0 + \tau,\ 1 \le d \le \omega \tag{11}$$

$$I_i^f(w-1,\omega) = I_i^f(w,0) \quad \text{for all } i \in \boldsymbol{I},\ f \in \boldsymbol{F},\ w^0 < w \le w^0 + \tau \tag{12}$$

$$I_i^f(w^0,0) = h_i^f \quad \text{for all } i \in \boldsymbol{I},\ f \in \boldsymbol{F} \tag{13}$$

Manufacturing order requirements:

$$Q_i^{fl}(w,d) = \sum_{m \in \boldsymbol{M}} mo_{im}^{fl}(w,d)$$
$$\text{for all } i \in \boldsymbol{I},\ f \in \boldsymbol{F}, l \in \boldsymbol{L}\ w^0 < w \le w^0 + \tau^m, 1 \le d \le \omega \tag{14}$$

Production capability requirements:

$$\sum_{l\in \boldsymbol{L},d=1}^{\omega} Q_i^{fl}(w,d) = \sum_{j\in \boldsymbol{J},w=w_j^f \,\&\, i\in \boldsymbol{u}_j^f} X_{ij}^f + \sum_{j\in \boldsymbol{J},w=w_j^f-e_j^f \,\&\, i\in \boldsymbol{u}_j^f} X_{ij}'^f$$
$$+ \sum_{j\in \boldsymbol{J},w=w_j^f+\overline{e}_j^f \,\&\, i\in \boldsymbol{u}_j^f} X_{ij}''^f$$

$$\text{for all } i\in \boldsymbol{I},\ f\in \boldsymbol{F},\ w^0+\tau^m < w \le w^0+\tau^p \tag{15}$$

$$\sum_{i\in \boldsymbol{I}} X_{ij}^f + \sum_{i\in \boldsymbol{I}} X_{ij}'^f + \sum_{i\in \boldsymbol{I}} X_{ij}''^f \le g_j^f \quad \text{for all } f\in \boldsymbol{F},\ j\in \boldsymbol{J} \tag{16}$$

$$X_{ij}'^f = 0 \text{ for all } i\in \boldsymbol{I},\ f\in \boldsymbol{F},\ j\in \boldsymbol{J} \text{ and } w^0+\tau^m < w_j^f \le w^0+\tau^m+e_j^f \tag{17}$$

$$X_{ij}''^f = 0 \text{ for all } i\in \boldsymbol{I},\ f\in \boldsymbol{F},\ j\in \boldsymbol{J} \text{ and } w^0+\tau^p-\overline{e}_j^f < w_j^f \le w^0+\tau^p \tag{18}$$

Production capacity requirements:

$$\sum_{i\in \boldsymbol{I}} Q_i^{fl}(w,d)\cdot pt_i^{fl} \le t^{fl}(w,d)$$

$$\text{for all } f\in \boldsymbol{F},\ l\in \boldsymbol{L}\ w^0+\tau^m < w \le w^0+\tau,\ 1\le d\le \omega \tag{19}$$

$$Q_i^{fl}(w,d) \ge s_i^{fl}\cdot Y_i^{fl}(w,d)$$

$$\text{for all } i\in \boldsymbol{I},\ f\in \boldsymbol{F},\ l\in \boldsymbol{L},\ w^0+\tau^m < w \le w^0+\tau,\ 1\le d\le \omega \tag{20}$$

$$Q_i^{fl}(w,d) \le Y_i^{fl}(w,d)\cdot N$$

$$\text{for all } i\in \boldsymbol{I},\ f\in \boldsymbol{F},\ l\in \boldsymbol{L},\ w^0+\tau^m < w \le w^0+\tau,\ 1\le d\le \omega \tag{21}$$

Production smoothness requirements:

$$\hat{U}_l^f \ge Q_l^f(w,d) \quad \text{for all } f\in \boldsymbol{F},\ l\in \boldsymbol{L},\ w^0+\tau^m < w \le w^0+\tau,\ 1\le d\le \omega \tag{22}$$

$$\breve{U}_l^f \le Q_l^f(w,t) \quad \text{for all } f\in \boldsymbol{F},\ l\in \boldsymbol{L},\ w^0+\tau^m < w \le w^0+\tau,\ 1\le d\le \omega \tag{23}$$

Total cost calculation:

$$
H^f = ch^f \cdot \left(\begin{array}{l} \sum_{w=w^0+1}^{w^k-1} \sum_{d=1}^{\omega} \left(\left(w^k - w \right) \cdot \omega + d^k - d \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{kf}(w,d) \right) \\ + \sum_{w=w^k}^{w^k} \sum_{d=1}^{d^k} \left(\left(d^k - d \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{kf}(w,d) \right) \right) \\ \sum_{w=w^0+1}^{\widehat{w}^{k'}-1} \sum_{d=1}^{\omega} \left(\left(\widehat{w}^{k'} - w \right) \cdot \omega + \widehat{d}^{k'} - d \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{k'f}(w,d) \right) \\ + \sum_{w=w^k}^{\widehat{w}^{k'}} \sum_{d=1}^{\widehat{d}^{k'}} \left(\left(\widehat{d}^{k'} - d \right) \cdot \left(\sum_{f \in \boldsymbol{F}} C^{k'f}(w,d) \right) \right) + \beta \cdot \sum_{i,w,d} I^{if}(w,d) \end{array} \right) + ca^f \cdot \left(\sum_{w=w^0+1}^{w^0+\tau} \sum_{d=1}^{\omega} \sum_{i \in \boldsymbol{I}, l \in \boldsymbol{L}} Q_i^{fl}(w,d) \right) + ce^f \cdot \sum_{i \in \boldsymbol{I}, j \in \boldsymbol{J}} X'^f_{ij} + c\overline{e}^f \cdot \sum_{i \in \boldsymbol{I}, j \in \boldsymbol{J}} X''^f_{ij} \tag{24}
$$

Integrality:

$Z^{kf} \in \{0,1\}$ for all $k \in \boldsymbol{K}$, $f \in \boldsymbol{F}$

$\widehat{Z}^{k'f} \in \{0,1\}$ for all $k' \in \boldsymbol{K}$, $f \in \boldsymbol{F}$

$Y_i^{fl}(w,d) \in \{0,1\}$ for all $w^0 + \tau^m < w \le w^0 + \tau$, $1 \le d \le \omega$

Nonnegativity:

$$
\begin{array}{l} C^{kf}(w,d) \ge 0, \widehat{C}^{k'f}(w,d) \ge 0, B^k \ge 0, \widehat{B}^{k'} \ge 0, Q_i^{fl}(w,d) \ge 0, I_i^f(w,d) \ge 0, \\ X_{ij}^f \ge 0, X'^f_{ij} \ge 0, X''^f_{ij} \ge 0, V^k \ge 0, \widehat{V}^{k'} \ge 0, H^f \ge 0, \hat{U}^{fl} \ge 0, \breve{U}^{fl} \ge 0 \end{array}
$$

The objective function (1) includes a due date violation term, a total cost term and a production smoothness term. The due date violation term contains accounts for due date violations for the newly-arrived orders and the previously-promised orders. The total cost term includes inventory holding cost, production cost, cost for production capability expediting, and the penalty (cost) for production capability de-expediting. The production smoothness term indicates a penalty associated with day-to-day production capacity variation for different line instance (f,l). It is chosen to reduce the production capacity variation over time. The weights (business priority) for these three terms are w', w'', w''', where $0 \le w' \le 1, 0 \le w'' \le 1, 0 \le w'' \le 1$ and $w' + w'' + w''' = 1$. α is a factor to make the production smoothness penalty comparable to the other two terms.

There are seven major groups of constraints: 1) order promising and fulfillment, 2) due date violation for previously-promised orders, 3) finished-product flows, 4) manufacturing order requirements, 5) production capability requirements, 6) production smoothness, and 7) production capacity requirements. Constraints (2) and (3) specify the commitment and backlog of customer orders. Constraints (4) and (5) define the feasible production factory for each customer order (orders cannot be split among two factories). Note that the order commitment variables Z^{kf} or $\widehat{Z}^{k'f}$ equal one if the *k*th or the k'th customer order is committed in the *f*th factory, and constraints (6) and (7) ensure that only one factory is used for each customer order. Constraints (10) ensure that previously-promised orders should be fulfilled without a due date violation increase in this ATP execution. Balance of finished-product inventory is provided by constraint (11), and the initial inventory conditions for each product in each factory are enforced in constraints (13). Constraints (12) guarantee that the inventory at the end of one week is equal to the beginning of next week. The manufacturing order requirement is modeled in constraint (14), which ensures that all manufacturing orders must be produced. Constraint (15) defines the production capability quantity in a time period, which equals the sum of the planned production capability adjusted for the expedited and de-expedited production capabilities. For each production capability, constraints (16) ensure that requirement is less than the availability. Constraints (17) and (18) specify that no production capability can be expedited from the flexible product to fixed product period or de-expedited from flexible product period to flexible resource period.

Finally, production capacity and production lot size are modeled in constraints (19), (20), and (21). Both flexible product period and flexible resource period are constrained by the available production capacity. Note that, lot size variable $Y_i^{fl}(w,d)$ equals zero if the production lot is above the desired level s_i^{fl} for any product and line instance $\left(f,l\right)$ in any period instance $\left(w,d\right)$. Constraint (22)-(23) calculate the minimum and maximum production capacity used, and constraint (8)-(9) and (24) keeps track of the due date violation and total cost.

3. Experimental Results

We conducted a series of experiments based on data and business scenarios of the EP product from Toshiba Corporation. An assemble-to-order (ATO) or build-to-order (BTO) strategy is used to manage the EP production, distribution and sales. While the long-term material requirements are planned based on a demand forecast and MRP, the actual EP assembly schedule is based on realized orders in a much shorter time.

Currently, order promising decision support is provided by a legacy system. The ATP decisions are based on business experience and heuristic rules. The presence of a large number of customer orders, product models and the multi-stage resource availability complexity would suggest that efficiency can be gained from the application of optimization methods. In order to make the current ATP execution process manageable, the company has made simplifications such as not considering recommitments of previously-promised orders, limiting the production capability (PC) candidates, etc. Moreover, overall order promising performance can be highly variable as it is dependent on the behavior of the individual decision-makers involved. Overall, these drawbacks and simplifications often result in the occasional, simultaneous occurrence of material shortages, seemingly excess inventory and low order fulfillment rates. The model presented in this paper can improve the efficiency and optimality of order promising and fulfillment process, and achieve more desirable system performance. Note that even though our ATP model is formulated based on Toshiba EP product case, it can address more general problems.

3.1 Experimental Setup

The EP products are produced using 49 assembly lines in three factories all located in Japan. The assembly lines are able to run maximally three shifts per day with eight effective hours per shift. The customers directly place orders to the EP product headquarters and the headquarters needs to promise both a delivery quantity and delivery date for each customer order, and generate a production schedule for all factories. All customer orders have the same priority. The company uses an Oracle ERP system to support production planning and basic business functions like finance and human resources.

The company produces 4355 different EP product models. For our experimental setup, we used historical data and ATP results as a baseline for measuring the effectiveness of our optimization-based ATP model. First of all, we selected one day (Monday, April 1, 2002) as the post date for our experiment. We collected the newly-arrived customer orders with a total number of 1162 orders on the post date, and the previously-promised customer orders with a total number of 3834 orders before the post date. Each previously-promised customer order had an associated due date violation (possibly zero), which was based on the commitment made with the original promise. The total number of customer orders was 4996. Additionally, we collected historical data on balance-on-hand inventory, manufacturing orders, production capacity availability, and production capability availability for the nine weeks time period after the post date. The nine weeks time horizon consisted of two weeks for fixed product ($\tau^m = 2$), seven weeks for flexible product ($\tau^p = 6$), and three weeks for flexible resource ($\tau = 9$).

The due dates for the customer orders covered the entire nine week time horizon. The results we report were based on a single day-long scenario. This required several model executions corresponding to multiple successive 1/4 hour intervals. Each model execution had to consider the set of new and previously committed orders, which had due dates that extended over the nine week time horizon. For resources, there were 765 manufacturing orders and 456 production capabilities with 5 – 20 product candidates in each production capability. For production capacity, we collected the product processing time, the regular working time availability, and overtime availability for each day. Furthermore, we also collected the real-life order promising results and corresponding production schedules from Toshiba EP factories.

The ATP model was implemented and solved with the ILOG OPL Studio (version 3.5) operating on a Pentium IV machine with 1G Hz Intel Processor, and Windows 2000 operating environment. All collected data was stored in MS Access 2000 database. The ATP model received input data directly from the MS Access database, and wrote the solution back to the database. MS Access ODBC driver was used to connect the database with the ATP model. Since the ATP model runs multiple times, OPL Script was used as meta-level process to control the ATP model execution. OPL Script was also used for pre-processing before each time ATP model execution and post-processing after each ATP execution. Each ATP execution was based on the customer orders that arrived in the previous quarter hour time interval together with the previously-promised un-delivered orders. Any resource could be reassigned in subsequent executions as long as commitments were maintained.

3.2 Experimental Model Decomposition

We were unable to solve the initial formulation (it contained several million variables and constraints). To create a more manageable approach, we decomposed the model into "master" model and sub-models, which were solved iteratively. Considering that the production capability resource was given on a weekly basis and that a high level of importance was given to weekly-level order promising and fulfillment in Toshiba Corporation, we decomposed the model into a weekly ATP model and daily ATP model. The weekly ATP model outputs a promised quantity and completion week for each customer order based on weekly production capability availability and a weekly production capacity. The daily ATP model then determines the day or days (for split orders) on which the committed quantity is produced. This model is solved independently for each week in the planning horizon. In the next paragraphs we provide an overview of the decomposed ATP models.

In the weekly ATP model, all customer orders and daily production capacity are aggregated to weekly values. The deliver dates for all customer orders are assumed to be on the last day of the week in which the order is scheduled, namely, $d = \omega$, for the purposes of calculating inventory holding cost and order due date violation. With this policy, the total inventory holding cost will be smaller than and the total due date violation will be greater than, the corresponding value produced by the daily model. Certainly, the order delivery dates can also be assigned in other ways such as the first day of the week.

In the weekly ATP model, the order promising and fulfillment constraints (2) – (7), the due date violation constraints (10), the production capability constraints (15) – (18), production smoothness constraints (22) and (23), production capacity requirement constraints (20) and (21), due date violation calculations (8)-(9), and total cost calculation (24) remain unchanged except that the index d is fixed to the constant value ϖ. All decision variables related to day index also remain unchanged except that the index d is fixed to the constant value ϖ. The finished-product flow constraints (11) - (12) become constraints (25), which give weekly product flow conservation, and the finished product initialization constraints (13) become (26).

$$I_i^f(w,\omega) = I_i^f(w-1,\omega) - \sum_{k\in\boldsymbol{K}\&i=p^k} C^{kf}(w,\omega) - \sum_{k'\in\boldsymbol{K}'\&i=\hat{p}^{k'}} \hat{C}^{k'f}(w,\omega)$$

$$+\sum_{l\in\boldsymbol{L}} Q_i^{fl}(w,\omega)$$

$$\text{for all } i\in\boldsymbol{I},\ f\in\boldsymbol{F},\ w^0 < w \le w^0+\tau \tag{25}$$

$$I_i^f(w^0,\omega) = h_i^f \ \text{ for all } i\in\boldsymbol{I},\ f\in\boldsymbol{F} \tag{26}$$

For the manufacturing order requirements constraints (14), the daily manufacturing orders on the right-hand side are required to be aggregated over days for each given week w; we also specify $d=\omega$; this process yields constraints (27).

$$Q_i^{fl}(w,\omega) = \sum_d mo_i^{fl}(w,d) \quad \text{for all } i\in\boldsymbol{I},\ f\in\boldsymbol{F}, l\in\boldsymbol{L}\ w^0 < w \le w^0+\tau^m \tag{27}$$

Similarly, the production capacity constraints (19) are replaced by constraints (28) after summing the daily production capacity over a week for the right-hand side for each given week w.

$$\sum_i Q_i^{fl}(w,\omega)\cdot pt_i^{fl} \le \sum_d t^{fl}(w,d)$$

$$\text{for all } f\in\boldsymbol{F}, l\in\boldsymbol{L}\ \ w^0+\tau^m < w \le w^0+\tau \tag{28}$$

Based on the execution results of the weekly ATP model, the daily ATP model determines the committed quantity and days for each order. Unlike the weekly ATP model, the daily ATP model doesn't need to be executed once for the whole ATP time horizon. Instead, it is executed multiple times subsequently from week $w = w^0 + 1$ to week $w = w^0 + \tau$. Consequently, the production capability constraints (13) – (17) and inventory balance constraint (12) can be eliminated from the daily ATP model since they are not playing any role in a single week. The order promising and fulfillment constraints (2) – (7), the due date violation constraints (10), manufacturing order requirements (14), finished-product flow constraints (9), due date violation constraints (8) and (9), production capacity requirements (19) – (21), production smoothness constraints (22)-(23), and total cost calculation (24) remain unchanged when fixing the week index w to a constant from $w = w^0 + 1$ to $w = w^0 + \tau$. Meanwhile, the order requested quantity q^k for the newly-arrived orders and $\widehat{q}^{k'}$ for the previously-promised orders are replaced by the quantity actually committed from the weekly model results.

The inventory initialization constraints (13) must be changed to constraints (29) and (30) to guarantee satisfaction of the starting inventory at the beginning of the week and the ending inventory at the end of the week.

$$I_i^f(w,0) = r_i^f \tag{29}$$

$$I_i^f(w,\omega) = \overline{r}_i^f \tag{30}$$

where $r_i^f, \overline{r}_i^f$ are starting inventory and ending inventory of week *w*, which are outputs of weekly ATP model.

When using this decomposition strategy, we cannot guarantee that the complete solution produced by the multiple steps will be optimal.

3.3 Experimental Results Analysis

As we said, we have used the due date violation, inventory holding cost and production smoothness as performance measures to compare our optimization-based ATP results with actual ATP results. Table 1 shows the model execution information including the number of constraints, the number of variables, average execution time and computer memory used for the weekly and daily models. The full problem takes a total of approximately six minutes (322.24 seconds) for a single complete execution. Toshiba feels that this execution speed provides acceptable performance. The execution times given include the data input and output times from/to the MS-Access database. A production system would most-likely make use of a different software architecture, e.g. by replacing OPL script with direct C++ database access. Such alternate approaches should produce substantial performance improvements. In addition, performance enhancements certainly could be gained though the use of servers and/or operating systems with better performance, tuning of Cplex parameters, model enhancements, etc.

Table 1: Experimental execution results

	WeekNo	Periods	Orders	Constraints	Variables	Execution time(s)	Memory used(MB)
Weekly ATP		9 (week)	4994	62,391	99,439	154.00	70.8
Daily ATP in fixed Product period	67	7 (day)	3659	61,831	83,028	130.86	46.6
	68	7 (day)	586	10,429	13,770	10.08	8.1
Daily ATP in flexible product period	69	7 (day)	596	13,228	15,692	11.05	9.9
	70	7 (day)	195	6,823	6,894	5.46	4.8
	71	7 (day)	30	2,533	1,681	2.39	1.4
	72	7 (day)	55	3,538	2,815	3.03	2.2
Daily ATP in flexible resource period	73	7 (day)	27	2,488	1,605	2.42	1.4
	74	7 (day)	4	1,663	671	1.86	0.7
	75	7 (day)	1	1,468	450	1.69	0.5
Total						322.84	

In Table 2, we compare the optimization-based results with the historical performance. The total due date violation is measured (as in the objective function) as the sum over all orders of the quantity that is late times the number of days late. We also considered as another performance measure, the total order quantity committed within fixed product period (manufacturing orders). This performance index provides a measure of the efficiency of the fixed production plan since any products produced that are not committed to an order become inventory.

Table 2. Comparison results of optimization-based ATP and actual ATP

Performance Measures	Real-life ATP Results	Optimization-based ATP results	Improvement
Total due date violation	429016	351521	18.1%
Total inventory holding cost	1614023	1577228	2.3%
Order committed with MOs	18439	15797	14.3%

As shown in Table 2, the optimization-based ATP model gives substantially improved performance over actual ATP practice in terms of due date violation, holding cost and order commitment quantity in the fixed resource period. Of particular note is that the optimization-based ATP model reduces overall due date violation by about 18% compared (from 429016 unit-days to 351521 unit-days). This is a significant improvement in the level of customer service provided. Meanwhile, the total inventory holding cost is reduced slightly, almost 2.3% from 1614023 dollars to 1577228 dollars. Although the inventory holding cost is a relatively small percentage, it is noteworthy that the due date violation is weighted higher than inventory reduction so one might expect inventory costs to increase in order to accommodate due date violation reduction. *Yet, a substantial reduction in due date violation was obtained while simultaneously reducing inventory costs.*

From Table 2, we can observe that the optimization-based ATP model leads to the improvement of the resource utilization in fixed product period by 14% (18439 units to 15797 units). This improvement should translate into a reduction in waste, holding cost and/or variable production costs. These benefits can be especially significant for short-life-cycle products typical in the electronics space.

In carrying out our analysis, in order to gain insight into the tradeoffs between the two most important terms, due date violation and holding cost, we conducted a parametric analysis and produced a (partial) pareto frontier of solutions.

First of all, we fixed a very small weight for production smoothness ($w'' = 0.001$), namely $w' + w'' \approx 1$, and let due date violation and total cost play the dominant role in the model. Then, we solved the model using weights that varied between two extreme values, $(w', w'') = (0.001, 0.999)$ (the due date violation dominated scenario) and $(w', w'') = (0.999, 0.001)$ (the holding cost dominated scenario). The following combinations were used:

$(w, w') = (0.01, 0.99)$, $(w, w') = (0.1, 0.9)$, $(w, w') = (0.3, 0.7)$, $(w, w') = (0.5, 0.5)$.

$(w, w') = (0.99, 0.01)$, $(w, w') = (0.9, 0.1)$, $(w, w') = (0.7, 0.3)$, $(w, w') = (0.5, 0.5)$.

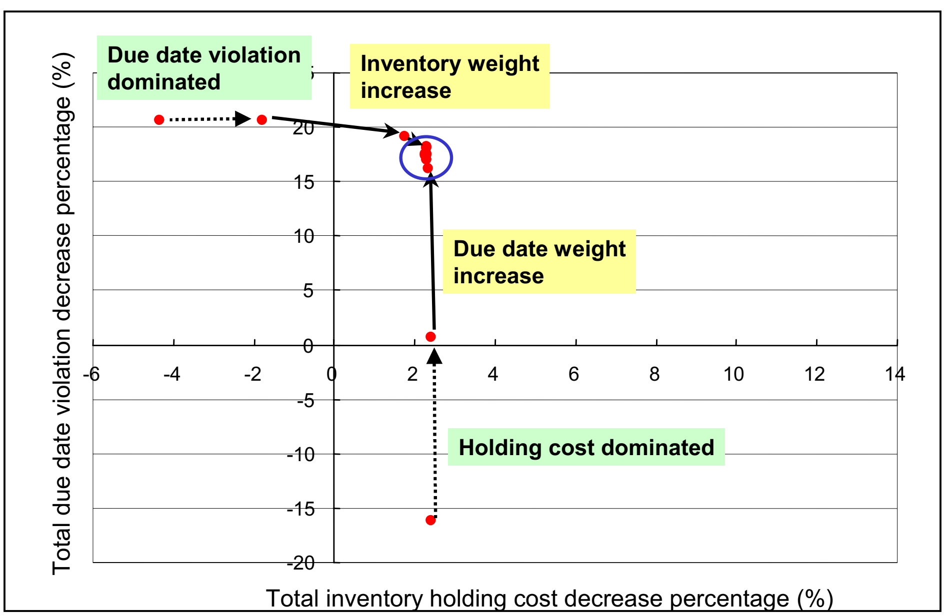

Figure 3. Weight sensitivity analysis in objective function

The corresponding total due date violation value and total holding cost value were obtained by running the optimization-based ATP. Then, the change as a percentage of total due date violation and total holding cost were compared to the actual due date violation and holding cost. These results are plotted in Figure 3.

In Figure 3, a positive percentage value means a decrease in due date violation or holding cost, and a negative number means an increase of due date violation or holding cost. It is particularly noteworthy that a solution can be obtained that simultaneously is very close to both the best due date violation value and the best inventory cost value. In order words, a single solution could be found that is nearly optimal relative to either of the objective functions. We feel that this is a somewhat remarkable property, although we certainly have not shown that this holds in a general setting.

4. Conclusions

Advanced ATP plays an important role in real-time order-promising and fulfillment. The optimization-based ATP model presented in this paper considers multiple products and multi-stage resources including fixed product resources, flexible product resources and flexible resources, as well as production capacities for multiple factories. Some special model features include the ability to split orders and to expedite or de-expedite material shipments. The model specifies a delivery date and committed quantity for each order. We compare our optimization-based ATP results with actual ATP results from Toshiba Corporation. In particular, we illustrate the improvement that optimization-based ATP can provide in terms of total due date violation, total inventory holding cost, and production capacity smoothness. The experiments also provide insight into the tradeoff amongst the various performance parameters. Our results show that system performance can be improved substantially with an effective policy.

5. Acknowledgements

This work was supported by the National Science Foundation (NSF) under grant number DMI9908137 and by Toshiba Corporation. Matching support was provided by the Robert H. Smith School of Business and the University of Maryland Institute for Advanced Computer Studies (UMIACS). The opinions expressed in the paper are those of the authors and do not necessarily reflect the opinions of NSF or Toshiba.

References

APICS (1987), *Dictionary*, 6th Ed. American Production and Inventory Control Society, Falls Church, VA.

BALL, M. O., C.-Y. CHEN and Z.-Y. ZHAO. (2002). "Available to Promise" forthcoming chapter in D. Simchi-Levi, S. David Wu and M. Shen, eds. *Handbook of Supply Chain Analysis in the eBusiness Era.*

BALL, M. O., C.-Y. CHEN, and Z.-Y. ZHAO. (2003). "Material Compatibility Constraints for Make-to-Order Production Planning," *Operations Research Letters*, 31, 420-428.

CHEN, C.-Y., Z.-Y. ZHAO, and M. O. BALL. (2002). "A Model for Batch Advanced Available-to-Promise," *Production and Operations Management*., 11 (4), 424-440.

CHEN, C.-Y., Z.-Y. ZHAO, and M.O. BALL. (2001). "Quantity and Due Date Quoting Available to Promise," *Information Systems Frontiers*, 3, 477-488.

CHENG, T. C. E. and M. C. GUPTA. (1989). "Survey of Scheduling Research Involving Due Date Determination Decisions," *European Journal of Operational Research*, 38(2), 156-166.

DUENYAS, I. and W. J. HOPP. (1995). "Quoting Customer Lead Times," *Management Science*, 41(1), 43-57 eB2x, I. (2000). "Why Promising Orders? How Strategic Order Promising Helps Supply Chains Compete in the New Economy," *Online, http://www.eb2x.com*.

ERVOLINA, TOM AND BRENDA DIETRICH. (2001). "Moving Toward Dynamic Available to Promise," in Saul I. Gass and Albert T. Jones, editors, *Supply Chain Management Practice and Research: Status and Future Directions*, R. H. School of Business, University of Maryland, Manufacturing Engineering Laboratory, National Institute of Standards, and National Science Foundation, 15.1-19.

FORDYCE, K. J. and G. A. SULLIVAN. (1999). "Supply Chain Management, Decision Technology, and e-Business Information Technology at IBM Microelectronics," *MicroNews*, 5, 4, http://chips.ibm.com/micronews/.

HARIHARAN, R. and P. ZIPKIN. (1995). "Customer-Order Information, Leadtimes, and Inventories," *Management Science*, 41 (10) 1599-1607.

HEGEDUS, M. G. and W. J. HOPP. (2001). "Due Date Setting with Supply Constraints in Systems Using MRP," *Computers & Industrial Engineering*, 39(3-4), 293-305.

HILL, S. Jr. (2000). Trade Exchanges & Order Promising: Can Internet Trade Exchanges Deliver Order Promising Capability? [On-line]. Available Internet: www.manufacturingsystems.com.

HOPP, W. J., and M. L. STURGIS.(2000). "Quoting Manufacturing Due Dates Subject to a Service Level Constraint," *IIE Transactions*, 32(9), 771-784.

JOHNSON, L. A. and D. C. MONTGOMERY. (1974). *Operations Research in Production Planning, Scheduling, and Inventory Control*, John Wiley & Sons, Inc., New York, NY.

KILGER, C. and L. SCHNEEWEISS. (2000). "Demand Fulfillment and ATP," in: Stadtler, H., & Kilger, C., eds. *Supply Chain Management and Advanced Planning: Concepts, Models, Software and Case Studies*. Berlin, Germany: Springer, 135-148.

LEE, H. L. and C. BILLINGTON. (1995). "The Evolution of Supply-Chain-Management Models and Practice at Hewlett-Packard," *Interfaces*, 25, 5, 43-63.

MANUGISTICS, INC. (2000). Delivering on Customer Promises – A Critical Component for eBusiness Success. [On-line]. Available Internet: www.manugistics.com.

MCFEELY, D. J., SIMPSON, W. P. III and V. SIMOMS, D. J. Jr. (1997). Scheduling to Achieve Multiple Criteria in an Air Force Depot CNC Machine Shop. Production and Inventory Management Journal, 38(1), 72-78.

PARK, M.-W. and Y.-D. KIM. (2000). "A Branch and Bound Algorithm for a Production Scheduling: Problem in an Assembly System under Due Date Constraints," *European Journal of Operational Research*, 123(3), 504-518.

ROBINSON, A. G. and D. M. DILTS. (1999). "OR & ERP: A Match for the New Millennium?" *OR/MS Today*, 26, 3, 30-35.

SAP AG. (1999). "Functions in Detail – SAP Advanced Planner & Optimizer: Global Available-to-Promise," http://www.sap.com/.

TAYLOR, S. G. AND G. J. PLENERT. (1999). "Finite Capacity Promising," *Production and Inventory Management Journal*, 40 (3) 50-56.

THORE, S. A. O. (1991). *Economic Logistics: The Optimization of Spatial and Sectoral Resource, Production, and Distribution Systems*, Quorum Books, New York, NY.

TSAI, C.-H., CHANG, G.-T. and R.-K. LI. (1997). "Integrating Order Release Control with Due-Date Assignment Rules," *International Journal of Production Research*, 35(12), 3379-3392.

VOLLMANN, THOMAS E., WILLIAM L. BERRI, DAVID C. WHYBARK. (1997). *Manufacturing Planning and Control Systems*, 4th ed., New York : Irwin/McGraw-Hill.

ZWEBEN, M. (1996). "Delivering on Every Promise: Real-Time Sales Order Promising of Supply Chain Resources," *APICS – The Performance Advantage*, 6, 3, http://www.apics.org/magazine/.