

MATERIAL MANAGEMENT IN DECENTRALIZED SUPPLY CHAINS

HAU L. LEE

Stanford University, Stanford, California

COREY BILLINGTON

Hewlett-Packard Company, Palo Alto, California

(Received September 1991; revisions received August 1992, January 1993; accepted February 1993)

A supply chain is a network of facilities that performs the functions of procurement of material, transformation of material to intermediate and finished products, and distribution of finished products to customers. Often, organizational barriers between these facilities exist, and information flows can be restricted such that complete centralized control of material flows in a supply chain may not be feasible or desirable. Consequently, most companies use decentralized control in managing the different facilities at a supply chain. In this paper, we describe what manufacturing managers at Hewlett-Packard Company (HP) see as the needs for model support in managing material flows in their supply chains. These needs motivate our initial development of such a model for supply chains that are not under complete centralized control. We report on our experiences of applying such a model in a new product development project of the DeskJet printer supply chain at HP. Finally, we discuss avenues to develop better models, as well as to fully exploit the power of such models in application.

A supply chain is a network of facilities that performs the functions of procurement of material, transformation of material to intermediate and finished products, and distribution of finished products to customers. One can always identify the supply chain of a manufacturing enterprise, although the complexity of the chain may differ from company to company. As companies become more vertically integrated, their supply chains consist of many sites organized in multiple levels, and the flows between sites form a complex network.

Different sources of uncertainties exist along a supply chain. They include demand (volume and mix), process (yield, machine downtimes, transportation reliabilities), and supply (part quality, delivery reliabilities). Inventories are often used to protect the chain from these uncertainties.

Inventories stored at different points of the supply chain have differing impacts on the cost and service performance of the chain. For example, inventories at various points have different values (a higher value for finished goods and a lower value for raw materials). Also, inventories at various points have differing degrees of flexibility (more flexibility in the form of raw materials because they can be turned into different alternative finished products without incurring a significant penalty). Finally, inventories at various points have differing levels of responsiveness (finished goods

can be shipped to customers without delay, whereas some lead time is needed to transform materials into finished goods before shipments can be made). A major challenge to supply chain managers is how to control inventories and costs along the chain while maximizing customer service performance.

An additional complexity in the management of material flows through a supply chain is introduced by the organizational structure and barriers within the chain. Today's global business environment often results in multiple sites worldwide working together to deliver product, while reporting to different organizational functions or units within the corporation. For example, manufacturing sites may report to different business units based on process or technology similarities, or customer needs. Distribution sites often report to sales or marketing, while overseas sites may report to an international business unit.

If all these sites belong to the same company the organizational boundaries might seem artificial; however, many companies have intentionally decentralized operational control of their business units or functions. Each of these groups has its unique culture, constraints, and objectives. In particular, information flows can be restricted or costly so that complete centralized control of material flows may not be feasible.

Centralized control means that decisions on how

Subject classifications: Inventory/production: inventory applications and multistage systems. Manufacturing: supply chain management.
Area of review: OR, PRACTICE.

much and when to produce are made centrally, based on material and demand status of the entire system. Decentralized control, on the other hand, refers to cases where each individual unit in the supply chain makes decisions based on local information.

We will begin by describing, in general, the decision support that manufacturing managers at Hewlett-Packard Company (HP) require in managing material flows in their supply chains. These needs motivate the development of a model for the management of material flows within a noncentrally controlled supply chain. In our application work at HP, we have focused only on a few of the problems that an integrated supply chain model can address. While the model was applied directly to these problems, we do not claim that it can address all the problems that were cited by management. A literature review on models for supply chain management will precede a description of the characteristics and environment of one supply chain at HP, that of the DeskJet printer. This is followed by a description of the initial model built to match the needs of the DeskJet printer supply chain. We will show how this model was used in a new product introduction project. Finally, we will discuss avenues to develop better models and methods to exploit fully the power of such models in application, as well as to address the full range of problems raised by management that were described in Section 1.

1. CHALLENGES AND NEEDS FOR MODEL SUPPORT IN SUPPLY CHAIN MANAGEMENT

An internal survey of HP manufacturing managers indicated that incoming part availability and part delivery performance are the most important problems they face today. Since incoming parts at one site are often supplied from another site within the company, one can characterize the problem as one of managing lead-time uncertainties throughout the supply chain. The sources of lateness can be traced back to the three types of uncertainty described earlier: demand, process, and (external) supply. In light of these uncertainties and the resulting challenges, we have found four capabilities that manufacturing managers would like addressed with decision support models.

Inventory and Service Benchmarking. The amount of inventory kept within a supply chain is becoming an important measure of corporate manufacturing managers' performance. Because supply chains vary in customer service performance objectives, it is often inappropriate to compare the amounts of inventories

at separate chains. Moreover, these supply chains may differ in the network structure, product structure, transportation times, and degree of uncertainty they face. Therefore, two weeks supply of inventory may be too high for one supply chain, while six weeks supply may be just right for another. To properly assess a supply chain in terms of its inventory investment, one needs to determine the optimal level of inventory needed to support a specific service target given the nature of the supply chain, and then compare it with the current level in that chain. This assessment can only be possible with a model that can relate inventory investments throughout a supply chain with its customer service performance, given the unique characteristics and environment of the supply chain. An inventory/service benchmark also allows manufacturing to rationalize their service targets (i.e., assessing the inventory investment to support a given service target) and facilitates negotiation between manufacturing and marketing on the right level of service targets.

Operational Planning and Control. Material managers have a difficult time determining how much safety stock to hold and when to initiate orders for material from upstream sites, given the various degrees of uncertainty in supply and demand that impact their stock-keeping units (SKUs). Model-based support systems are necessary to anticipate variability and optimize investment.

What-If Analyses. In managing a dynamic supply chain, manufacturing managers are often faced with alternative opportunities. To properly assess the cost and service impact of these alternatives, model-based support systems are extremely useful, because they can provide valuable input in support of management's decision process. Armed with the capability to explore and evaluate multiple alternatives, management may also identify new opportunities that were not evident before. This was indeed the case in our case study (to be described later).

Some examples of such what-if analyses are:

- changes in market demand induced by competitive behaviors, product maturation in its life cycle, or new entrants to the market;
- changes in the design of the supply chain network, e.g., an additional distribution center (DC) or consolidation of factories;
- changes in transportation mode, e.g., air versus surface;
- changes in production capacity.

Design for Supply Chain Management. A more visionary way to use an analytical model is to evaluate alternative product/process designs using inventory investment and service performance as a key measure in manufacturing and distributing the product. In addition to the usual criteria of functionality and performance, operational performance can make or break a new product introduction in today's global marketplace. Such an operational view for design evaluation is a natural extension of the "design for manufacturability" concept.

2. LITERATURE REVIEW

To this point, we have discussed the need for models for decision support. Before moving into a discussion of the application of such models at HP, we review the relevant literature that forms the foundation and motivation for our work. We refer the reader to the excellent review given by Graves (1988) on production planning models in a multisite network. In this section, we will only highlight the specific readings mentioned in his review that are most closely related to the current supply chain problem in addition to highlighting some more recent additions to the literature.

The analysis of multisite inventory networks can be categorized in two ways: those that are managed with complete centralized control, and those that are managed through decentralized control. Clark and Scarf's (1960) paper on centralized control forms the basis of most subsequent work. They considered a series system. Under periodic review inventory control with no setup costs, they were able to show that an order-up-to policy at each node is optimal. The continuous review version of this problem has been addressed by DeBodt and Graves (1985), where a reorder-point, fixed-lot size inventory control mechanism is used. Reorders are triggered on the echelon inventory position; therefore, the system does not operate in a decentralized mode, as more than local inventory status is needed.

Extensions of centralized control systems from a pure serial network to other structures are not straightforward. Federgruen and Zipkin (1984) were able to obtain tractable approximate results for the optimal centralized control policy when the network structure was of the arborescent type, i.e., each node in the network can have at most one supplying node. Such a network structure would fit the distribution portion of a supply chain. To model a manufacturing setting, where multiple inputs are often needed to make a product, an assembly type of network structure is needed.

Schmidt and Nahmias (1985) first considered such an assembly network structure with two inputs, and the resulting optimal policy was found to be a function of the inventory status of both the inputs as well as the finished product. By careful redefinition of the products' lead times, Rosling (1989) was able to show that, under certain initial conditions, the assembly network structure can be reduced to a serial one, such that the main results of Clark and Scarf apply. However, Rosling's results are true when there is only one end product, leaving issues like part commonalities (in the case of multiple end products) impossible to address using his approach.

Graves develops a different approach to the above problem; he models the production rate (the decision variable) at a site as the depletion rate of input inventory. He starts by describing the operating characteristics of a single site, but indicates that the full network can be recovered by repeated use of the single site model. Simple and tractable results can then be obtained. Graves' approach can be used to analyze the value of flexibility in the manufacturing process as well as parts commonality. This model has significant advantages. First, it can address most of the issues raised in Section 1; in particular, the design for the supply chain issue can be handled. Second, the model can be used in a centralized control setting (when all the decision variables are optimized simultaneously), or in a decentralized control mode (when the decision variables are optimized one site at a time). The approach of developing a single site model as a building block for the whole network is an idea that we have also used. There are some limitations to Graves' method, however; one may not be able to vary production rates at will, and it is not clear how his model handles an assembly type network (i.e., where there are multiple inputs to a site).

There is also a significant literature on decentralized control systems. Sherbrooke (1968) describes one such model, known as METRIC, where an arborescent structure is assumed, and where the inventory replenishment policy is of the one-for-one type. There are numerous extensions to METRIC and we refer the readers to the review by Axsäter (1991). Muckstadt and Thomas (1980) compare the performance of two such systems: one with order-up-to points at different sites determined independently, and one where such points are optimized jointly. The latter system was found to perform significantly better in certain cases.

Cohen and Lee (1988) describe a model for networks that are of a more general nature and decentralized. In their network there are multiple inputs to a manufacturing site with outputs feeding

an arborescent distribution network; however, the network is limited to a single manufacturing site. The key linkage between manufacturing and distribution here, is the manufacturing lead time that becomes replenishment lead time for the distribution part of the network. This idea is based on the original work of Karmarkar (1987) and Zipkin (1986). Cohen and Lee assume, as an approximation, that the replenishment lead time is a constant, equal to the mean manufacturing lead time. Using heavy traffic queueing theory, Arreola-Risa (1989) shows that the distribution of the mean manufacturing lead time is shifted exponential, hence improving on the accuracy of the operating characteristics of the model by Cohen and Lee. Pyke (1987) considers a variant of Cohen and Lee's model in which expedited production is allowed at the manufacturing facility.

In general, centralized control mechanisms should perform better than decentralized ones. One possible implementation of centralized control is the use of echelon stock (see DeBodt and Graves) in monitoring inventory at a site. Decentralized control would naturally involve local inventory status, known as installation stock. An exploration of the comparative efficiency of the two systems can be found in Axsäter and Rosling (1990). Federgruen (1989) also evaluates these two types of control mechanisms.

There are two papers that are closely related to the supply chain problems described here. Lee and Billington (1992) describe common problems observed in supply chain inventory management in a variety of industries. Lee, Billington and Carter (1993) describe a particular application case for the redesign of a product to reduce inventory investment at HP. The current paper builds on these two papers, provides the basic model framework from which general supply chain inventory problems can be tackled, and describes an application case that is related to new product introduction.

3. THE DESKJET PRINTER SUPPLY CHAIN

In this section, we will give an overview of the DeskJet printer manufacturing and distribution process. The DeskJet printer is one of several product families manufactured by the Vancouver Division at HP. Other product families include the ThinkJet, QuietJet, Impact, and RuggedWriter printers. Among these product families, the DeskJet family is by far the fastest growing. At the time when the authors were working with the division there were three major product classes within the DeskJet printer family: the

DeskJet, the DeskJetPlus (an upgraded version of the DeskJet), and the DeskWriter (for Macintosh personal computers). These were and still are distributed worldwide.

One has to exercise care when defining the scope as well as the level of aggregation of the supply chain to be considered. A comprehensive supply chain (both in terms of facilities and SKUs represented), while accurate, may be too complex for analysis and the collection of data needed to support the model can be prohibitive. The supply chain of the DeskJet printer is a very complex one, and we will only focus on key facilities as well as some key SKUs, as illustrated in Figure 1.

There are two main component manufacturing stages. One division within HP manufactures ASICs (application-specific integrated circuits) which serve as an input to the board assembly manufacturing stage in Vancouver. Another division manufactures printheads that are needed in the printer's final assembly stage (in Vancouver). Each of these two divisions is a complex supply chain by itself; however, for the purpose of analyzing the Vancouver supply chain, they can be treated as single nodes within the network.

There are two stages in the manufacturing process in Vancouver: printed circuit board assembly and test (PCAT), and final assembly and test (FAT). PCAT involves the assembly and testing of electronic components like ASICs, ROM (read-only memory), and raw printed circuit boards to make logic boards and print head driver boards. FAT involves the assembly of other components like motors, cables, key pads, printheads, plastic chassis and "skins," gears, and the printed circuit boards from PCAT to produce a working printer, as well as the final testing of the printer. The localization of the DeskJet printers for different countries involves packaging the appropriate power supply module, which contains the correct voltage requirements and power cord terminators (plugs), and

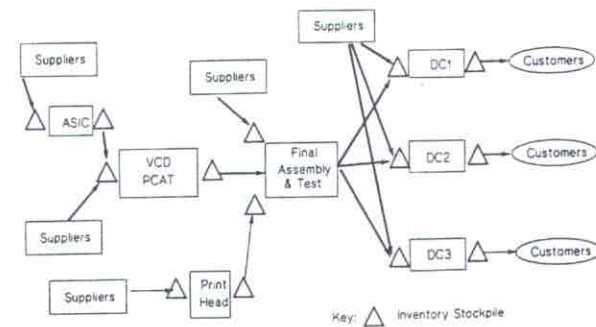


Figure 1. DeskJet printer supply chain.

the manual with the printer. The components needed for PCAT and FAT are sourced from other HP divisions as well as from external suppliers worldwide. In total, there are about 300 suppliers to Vancouver. The number of parts used exceeds 200. Nevertheless, one can focus on the major parts that constitute the bulk of the cost of incoming material, or parts that have the highest degree of supply uncertainty. Figure 2 gives the bill of materials for the DeskJet printers.

Printers made in Vancouver are sent by ship to the distribution centers (DCs) in Europe and the Far East, and by truck to the DC in the U.S. These DCs ship the printers either directly to the customers or to HP sales offices.

The printer industry is highly competitive. Customers of HP's computer peripherals (dealers) would like to carry as little inventory as possible; yet a high level of availability to end-users (consumers) is critical. Consequently, there has been increasing pressure for HP to provide high levels of availability at the DCs. HP management has responded by operating the DCs with large safety stocks.

Manufacturing in Vancouver operates in a pull mode. Production plans are set to replenish the DCs "just-in-time" to maintain the target safety stocks. To ensure material availability, safety stocks are also set up for incoming materials at the factory.

The environment facing the Vancouver division is both uncertain and dynamic. Uncertainties in supply, process, and demand (see the Introduction) coexist. At the same time, the division faces tremendous pressure from external competition, as well as internal competition from the LaserJet and PaintJet printers. The introduction of new products (entirely new platforms or upgrades of existing products) to remain competitive is a periodic event. The manufacturing management team in Vancouver were eager to develop a tool to help them address the four issues raised in Section 1 in light of such an uncertain and dynamic setting.

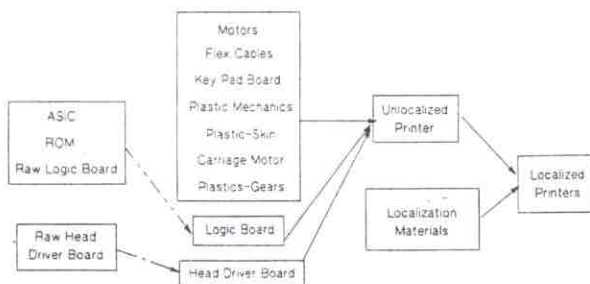


Figure 2. DeskJet printer bill of materials.

4. A MODEL FOR MATERIAL MANAGEMENT OF A SUPPLY CHAIN

In reviewing the literature, with the needs of manufacturing and material managers as described in Section 1 in mind, we identified the need for a decentralized supply chain model that allows for: 1) a generalized network structure, 2) uncertainties in supply, demand, and internal processes (as opposed to demand only), 3) simplicity and tractability for computation, and 4) capacitated production systems. In this section, we will present an overview of a model that was developed to address the needs specified by HP management (see Section 1). To do so, we have employed various approximations. These approximations and the limitations of the resulting model will be discussed later.

4.1. Modeling the Supply Chain

For all sites in a supply chain, there are typically two types of operations: material receiving, and production. A material receiving operation is one that receives input material from some other source(s) and then puts away the material as a stockpile to be used for production. A production operation is one in which manufacturing activities occur, transforming input materials to output materials. In most cases, a stockpile of output material is kept as well. For modeling purposes we will treat these two types of operations as one. Thus, we treat a material receiving operation as a production operation, where the material receipt, inspection and put away activities are the "production" activities of the material receiving operation. Consequently, all operations (material receiving or true production) can be treated as production operations. For representation, we split a site into two separate elements: receiving and production.

We can thus model the complete supply chain as a network of single production sites, each being an inventory system stocking SKUs to supply to customers/downstream sites. Replenishment of SKUs is a production process that transforms parts to SKUs. All demands are assumed to be backordered when shortages occur.

All the sites in the supply chain are linked together by the supply and demand process. The demand placed on SKUs at a downstream site translates to demand for parts at that site via the bill of materials. These incoming parts constitute the SKUs of the supplying site. Thus, downstream demand, in turn, creates demand at the supplying site. Hence, the whole network acts as a "pull" system in terms of material

requirements. We call this the *demand transmission process*.

The availability of input materials (parts) at a site depends on the service performance for these materials at the supplying site where they are known as SKUs. Three measures are needed to characterize the availability level: fill rate (fraction of requirements met without delay), mean delay, and variance of delay when shortages occur. Together, these measures indicate the extent of material shortage delays that can occur at the current site. The transformation of output product service performance measures at a supplying site to input material availability measures at a downstream site is called the *availability transfer process*.

4.2. Basic Model for a Single Site in the Network

The basic single site model assumes that the stockpiles at a single site are controlled by a periodic basestock (order-up-to) inventory system. Such a system approximates the actual control mechanism at the Vancouver division.

Replenishment of stock is a production process for a production site and is a procurement process for a material stocking site. Demands are assumed to be normally distributed. Again, this assumption has been found to be appropriate for the Vancouver HP division that we worked with, whose products are of a very high volume nature.

We assume that for each SKU at every site in a supply chain either a target service level (fill rate) or a target basestock level has been specified. The purpose of the model is to determine the required basestock level to support the target service level or the resulting service performance given the basestock level. In both cases, we want to compute the response times for the demand placed on an SKU at the site. As we will see later, this response time is important as it impacts material supply times of downstream sites.

For a stockpile of SKU i at a site, we need first to compute the mean and variance of the replenishment lead time for this SKU at the stockpile. This replenishment lead time consists of the standard lead time for the transfer of materials from an upstream site to the current site in the production of SKU i , the delays incurred in this transfer, and the actual production times needed for the transformation of materials into SKU i .

One difficulty encountered in modeling this lead time is the requirement for multiple input materials in the production of SKU i . These input materials have different standard lead times, and delays resulting from the availability of materials at the supplying sites differ. The linkage of material stocking levels and

material availability levels at the supplying site and the subsequent delays of material supply to the current site, will be described in the section on material availability process. Thus, one would have captured the combined effect of these multiple input parts on the resupply lead time of SKU i . As indicated in the literature review, exact theories for this problem are either just emerging or are computationally difficult (see Hopp and Spearman 1989, Ernst and Pyl 1992, and Hausman, Lee and Zhang 1992). We have developed an approximate model, the full details of which are described in the Appendix.

The production lead time itself is a function of the throughput rate and flow time in the manufacturing of the product, the occurrence and duration of production downtimes, and the capacity of the manufacturing process. We assume that downtime occurrences are Poisson, and obtain the mean and variance of the elapsed time in the production of a batch of SKU i . The assumption of Poisson occurrences of downtime simplifies the model. At the Vancouver division, unscheduled downtimes are very infrequent, and the effect of this assumption is likely to be minimal. We admit that the assumption may not be applicable in more general settings. Capacity of the manufacturing process is represented as a fixed unit of time (e.g., the number of hours per period, where a period is typically a week). Division of the elapsed time by the capacity yields the production lead time in periods.

Once the mean and variance of resupply lead times are computed, we determine the mean and variance of the demand on SKU i during the resupply lead time plus a review period. This is a standard step to find the target inventory level in standard period review basestock inventory systems (see Silver and Peterson 1985). The basic relationships are as follows. Let:

R = the review period (in weeks);

L = the lead time (in weeks);

D = the mean demand per week;

X = the demand in lead time plus one review period.

Denote also $E(\cdot)$ and $\text{Var}(\cdot)$ as the mean and variance of a random variable, respectively. Then, assuming independence of demand between periods:

$$E(X) = [E(L) + R]E(D).$$

$$\text{Var}(X) = [E(L) + R]\text{Var}(D) + [E(D)]^2\text{Var}(L).$$

The assumption of independence of demand between periods has been found to be appropriate for the DeskJet products in Vancouver, where a period is a month. (Finer demand data such as weekly were not available.)

Given the mean and variance of resupply lead time, we compute either the required basestock level to support a given target service level (fill rate) for an SKU at a site, or the achieved service level for a given basestock policy. In both cases, the response time to the demands on the SKU, given that the SKU is out of stock, can also be determined. The Appendix gives details of these derivations.

4.3. The Demand Transmission Process

At a downstream site the demand for SKU i generates demand for input parts that are needed to manufacture the SKU. This, in turn, becomes the demand for parts at upstream sites that supply the current site. Note that parts are referred to as SKUs at upstream sites.

The demand transmission process develops the first two moments of the demand on the SKUs at an upstream site. Consider an upstream site g . Suppose there are multiple downstream sites that receive the part j from this upstream site. Let:

- k = the index of the downstream sites;
- $K(g, j)$ = the set of downstream sites that are supplied by this upstream site for part j ;
- $u_i(k)$ = the mean demand per week for SKU i at downstream site k ;
- $s_i(k)$ = the standard deviation of demand per week for SKU i at site k ;
- $p_j(g, k)$ = the proportion of the requirement of part j at downstream site k that is to be sourced from the current upstream site g ;
- $u_j(g)$ = the mean demand per week for part j at the upstream site g (known as SKU there);
- $s_j(g)$ = the standard deviation of demand per week for part j at the upstream site g (known as SKU there);
- v_{ij} = the indicator variable: equals 1 if part j is needed for SKU i , 0 otherwise;
- b_{ij} = the number of units of part j needed in one unit of SKU i .

Then:

$$u_j(g) = \sum_{k \in K(g, j)} \sum_i v_{ij} u_i(k) b_{ij} p_j(g, k).$$

$$s_j^2(g) = \sum_{k \in K(g, j)} \sum_i v_{ij} s_i^2(k) b_{ij}^2 p_j^2(g, k).$$

4.4. The Availability Transfer Process

Consider now a downstream site that requires an input material for its production. In our single site model, the supply uncertainties of an input material are characterized by the immediate availability level, and the mean and variance of the delay due to material short-

ages. If the input material is sourced from an external supplier, then the supply uncertainty characteristics can be obtained as data inputs on the delivery performance of the supplier.

Suppose that input material at a site k is sourced from upstream site g . Then, the immediate availability level of the material at site k is the fill rate of the material at site g , where the material is known as an SKU. Delays due to material shortage at site g are the response time of the material at site g , given that the SKU is out of stock there.

4.5. Specification of Target Levels Throughout the Supply Chain

As described above, the basic single site model requires the specification of either the target service level or a basestock policy used by the site. For sites that face end-customer demands there is usually a target service level (or a range of service levels) desired by management. For other sites, the specification of their respective service or basestock levels is actually a decision variable itself. In the current implementation of the model, we have used a simple search heuristic to look for the "best" combination of service level targets at the different sites of the supply chain to support desired end-customer service targets. This is clearly an area for further research.

5. APPLICATION OF MODEL TO THE DESKJET SUPPLY CHAIN

The above model was applied to represent the supply chain of the DeskJet printer. The DeskJet printer supply chain model has been used by the Vancouver division in the following projects:

1. Evaluation of current inventory stocking level effectiveness, Pareto analysis of the contributions to inventory investment from each source of uncertainty, impact analysis of supplier improvement programs, and evaluation of alternative ways of shipping finished products to Europe and the Far East.
2. Analysis of alternative manufacturing/distribution strategies for a new product introduction (code named Maxim).
3. Cost and benefit analysis of product design changes for the DeskJet Plus family to facilitate the localization of product in Europe and the Far East.

In this paper, we will concentrate on describing the new product introduction project, because the issues considered in item 1 are also considered in this project. Project 3 highlights the importance of the "Design for Localization" concept and is described in detail in

Lee, Billington and Carter. In that paper, the validity of the DeskJet supply chain model, which is based on a number of simplifying modeling assumptions, has been tested against real performance data, and the model has been found to be very effective. In our presentation below, the actual data and project name have been disguised to preserve confidentiality.

In developing the basic DeskJet printer supply chain model, extensive data collection were needed. These data include:

- supplier lead time and delivery performance for inputs to each site in the supply chain;
- transit times between all sites of the supply chain;
- manufacturing cycle and flow times, capacity, and downtime frequency and duration (mean and variance);
- mean and variance of demand for all products at each DC;
- internal value of all SKUs at different sites of the chain.

The Vancouver division had been monitoring accurate data on supplier delivery performance and transit times. It had also implemented a tracking system to record production cycle time, flow time, and line downtime frequencies and durations. Nevertheless, the data collection phase was time consuming, as the data resided in different computer systems. Cleaning up the data files to ensure consistency and accuracy is a tedious and manually intensive task.

5.1. New Product Introduction: The Maxim Supply Chain

Maxim is a product that requires Vancouver to make the main engine, ship to a partner company in Japan for integration into the final product, and finally ship to the HP DCs for distribution. The same engine will also be used by the Japanese partner to build the final product under its own brand name, which would be distributed to the customers through its distribution channels. The supply chain for Maxim resembles that of the DeskJet printer with additional links to and from the partner company (see Figure 3). The manufacturing process and bill of materials for the main engine resemble that of the DeskJet printer. Hence, the description of manufacturing and distribution processes and issues for the DeskJet printer in Section 3 are equally applicable to Maxim. Historical data on supplier lead time and performance, and manufacturing times for the DeskJet printer thus apply to Maxim as well.

Since Maxim is a new product, no historical demand exists as a basis to estimate the variability of

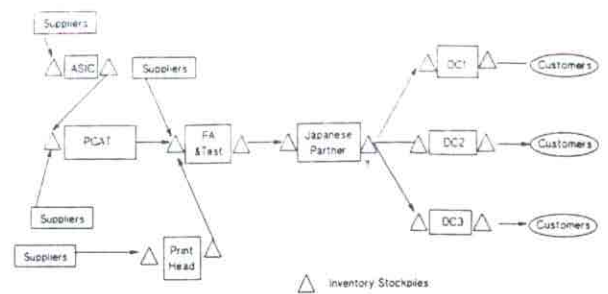


Figure 3. Maxim supply chain.

demand at the three DCs. Vancouver marketing personnel developed forecasts that serve as the mean demand inputs to the model. In the analysis, we used the assumption that Maxim would have the same coefficient of variations as the DeskJet printers.

The original plan developed by the Maxim product development team called for Vancouver to hold no finished engine stock and to ship engines to the Japanese partner on a weekly basis. The plan also called for shipping engines from Vancouver to the Japanese partner, and the finished Maxims from Japan to HP DCs by sea. The flow time at the Japanese factory was quoted (by the Japanese partner) as four weeks. To fully utilize a container size for shipment of finished Maxims out of Japan, the Japanese partner proposed shipping Maxims out on a monthly basis.

To analyze the impact of inventory investment at different points of the Maxim supply chain, which has multiple stocking points, a model such as the one described in Section 4 was needed. We began with the analysis of inventory requirements to meet varying service targets under the original logistics plan. Figure 4 shows two inventory/service tradeoffs, one based on the DeskJet family derived from the DeskJet supply

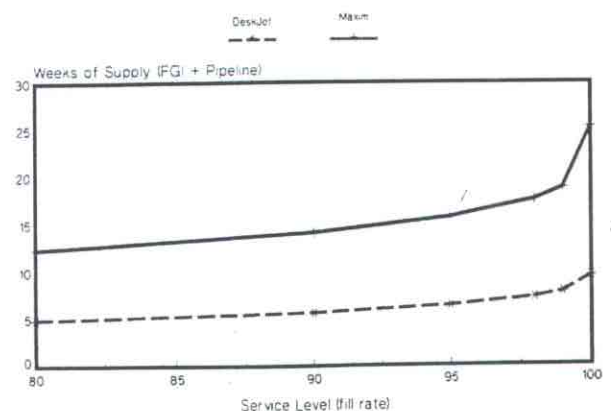


Figure 4. Inventory service tradeoffs for Maxim.

chain model, and one based on Maxim. To compare inventory investments of the two different products, we have shown finished goods inventory (both inventory at the DCs and in transit from Japan to the DCs) in weeks of supply. Note that inventory investments in WIP (work-in-process) and input material for DeskJet and Maxim are similar, because we used the assumption that Maxim has the same coefficients of variation of demand at the DCs as those of the DeskJet printers, and the supply chains for both products up to FAT in Vancouver are identical. Figure 5 shows the composition of inventory investment throughout the Maxim supply chain.

As can be seen from Figure 4, there is substantially higher finished goods inventory investment for Maxim relative to DeskJet printers to meet the same service requirements. In addition to higher Maxim inventory in transit from Japan to the DCs (versus DeskJet inventory in transit from Vancouver to the DCs), the safety stock levels at the DCs were significantly higher. Figure 5 shows that the biggest opportunity for inventory reduction lies in reducing Maxim finished goods at the DCs and inventory in transit.

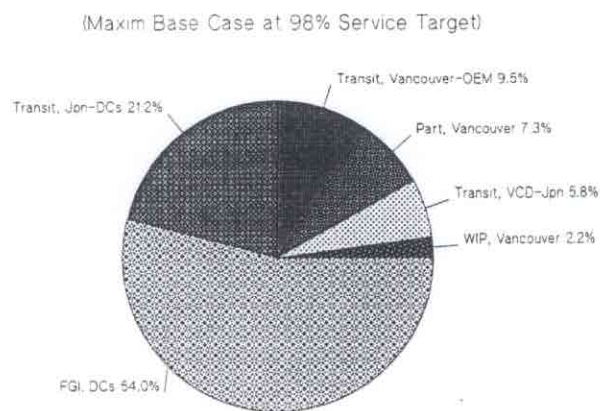


Figure 5. Composition of inventory in supply chain.

It was a bit surprising to the development team that the additional link to the Japanese partner would result in such a substantial increase in finished goods inventory.

The above analysis indicates that either the division will have substantially higher levels of finished goods inventory to support the target service level for Maxim than for its current products, or something has to be done to the logistic structure of the Maxim supply chain to improve its effectiveness.

5.2. Opportunities for Maxim

The Vancouver division has instituted a policy of not stocking engines in Vancouver. One possibility for improving performance of the Maxim supply chain was to use engine inventory in Vancouver to buffer against demand uncertainties at the three DCs. This alternative was found not to be very effective (see Table I).

Another way to shorten the length of time that materials flow through the Maxim supply chain was to shorten the transit time for engines from HP to the Japanese manufacturer. This could be accomplished by air shipment of the engines from Vancouver to Japan. This triggered the manufacturing team at Vancouver to begin exploring the use of HP's Singapore plant to manufacture the engines, because the country is closer to Japan than Vancouver. Figure 6 shows that significant improvement in the inventory-service tradeoff relationship could be achieved through the use of this strategy. Nevertheless, the improved Maxim supply chain was still far from the DeskJet family in terms of performance.

As the analysis of Maxim's supply chain continued to unfold, new potential alternatives were generated. One attractive possibility was the "co-location" strategy. The team discovered that this Japanese manufacturer has a plant in Singapore that could manufacture Maxim. The challenge to HP focused on convincing

Table I
Comparing the Opportunities for Maxim
(at 98% Service Goal)

Item	Base Case	Engine Buffer	Air Engine	Engine at Singapore	Co-location
Inventory ^a (weeks supply)	17.6	20.26	13.7	15.0	10.1
Inventory ^a (\$)	22.0M	21.6M	19.6M	20.4M	15.0M
Inventory reduction		0.4M	2.6M	1.6M	7.0M
Savings/yr. ^b		100K	600K	396K	1,680K
Freight savings/yr. (\$)		0K	-382K	120K	200K
Potential savings/yr. (\$)		100K	218K	516K	1,880K

^a FGI, engine + pipeline.

^b Based on 24% annually.

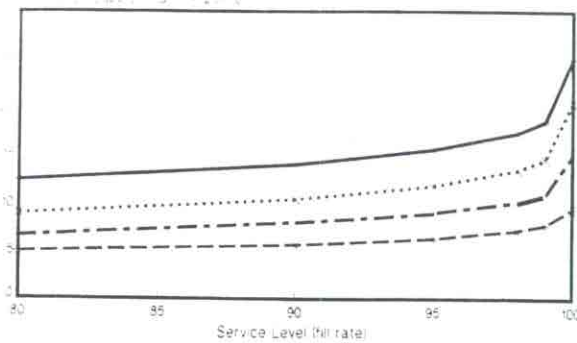


Figure 6. Inventory service tradeoffs for Maxim alternatives.

the Japanese manufacturer to build Maxim in Singapore. Combined with HP's Singapore plant making the engine, potential existed to further shorten material flow time in the supply chain. The volume of HP's traffic flow from Singapore to the U.S. and Europe was also large enough to justify economically more frequent shipments of Maxim from the plant to the DCs (weekly instead of monthly). Moreover, having the two phases of manufacturing close together facilitates coordination and planning operations. This strategy turned out to be the most effective. Figure 6 shows the extent to which the inventory-service tradeoff curve was improved with the co-location strategy.

Besides inventory-service tradeoffs, the previous strategies also resulted in different transportation costs. Table I shows the comparison of costs under the different strategies at the desired 98% target fill rate level. The co-location strategy again provided the best opportunity for cost reduction.

5.3. Discussion

The analysis of the Maxim supply chain uncovered and evaluated some alternative supply chain designs that were not considered in the original logistics plan. It also documented the quantitative costs/benefits of the different strategies. These results were not only useful for HP in identifying the best design to introduce the product, but they also provided a rational basis and strategic direction for negotiation with the Japanese manufacturer.

6. CONCLUSION

This paper described the challenges faced by supply chain managers, and the model-based decision support systems that they desire. The model developed at

HP is one step toward this end. We recognize that there are still many open questions on the theoretical support of the model.

While we are still conducting ongoing research, we note the following important avenues.

Modeling Correlated Demands. Correlation in demands for different end products, or different versions of the end products, may exist at sites that face end-customers. At other upstream sites, demands for the different SKUs are often correlated due to commonality. Models that explicitly capture such correlations are needed.

Modeling Material Delay. The current version develops expected value and variance of material shortage delays by assuming that there is only one material causing a delay. In general, more than one material input can be simultaneously out of stock. In short, we need better models that more accurately capture the effect of multiple (possibly correlated) input delays.

Modeling Congestion Effects of Capacity. The effects of finite capacity on production lead times can be highly nonlinear, for example, congestion or queueing effects. One possibility is to incorporate queueing type models to compute production lead time in the basic single site inventory model.

Extensive Testing of Approximations. The current model utilizes a number of approximations to retain tractability. These approximations are supported by our limited simulations of the model, but more comprehensive testings are needed to ascertain the robustness and accuracy of the approximations.

Optimizing Inventory Levels Across the Supply Chain. In the current application, the model is used as a performance evaluator, and we have used a simple search heuristic to find the optimal service levels at different sites across the supply chain. Better search routines are needed to find the jointly optimal levels at different sites and different SKUs.

Increasing Application Base. We are continually applying (and refining) the model for other products at Vancouver, as well as for other divisions at HP. Such applications provide real tests for the validity and adequacy of the model. They also enhance our understanding of the challenges and issues in the decision making process of supply chain managers.

Developing Build-to-Order Models. Our continuing application of the model at other HP divisions indicates that the service level concept (defined as the immediate fill rate) of the current model is inadequate. There are other divisions which are in a build-to-order

environment, where service targets are specified not in terms of fill rates, but in the reliability of meeting a target lead time. Adaptation of the current model to address such targets is an ongoing task.

APPENDIX: BASIC MODEL DESCRIPTION

Definition

Operational Characteristics

R_i = the review period (in weeks) for SKU i ;

B_i = the target fill rate for SKU i ;

S_i = the order up to point for SKU i .

Demand Characteristics

u_i = the mean demand per week of SKU i ;

s_i = the standard deviation of demand per week of SKU i .

Production Characteristics

x_i = the average cycle time (in hours) for SKU i (cycle time is defined as the reciprocal of the throughput rate);

z_i = the average flow time (in hours) for SKU i ;

n_i = the mean number of occurrences (per hour) of downtime on the production line for SKU i during a production run;

μ_{di} = the mean duration of downtime (in hours) on the production line for SKU i ;

σ_{di} = the standard deviation of duration of downtime (in hours) on the production line for SKU i ;

C_i = the production capacity (in hours per week) for SKU i .

Bill of Materials

v_{ij} = the indicator variable, equals 1 if part j is needed for SKU i ; 0 otherwise;

b_{ij} = the number of units of part j needed in one unit of SKU i .

Supply Characteristics

μ_{Lj} = the average lead time (in weeks) for part j (if the site is a stockpile, then this is the transit time from the upstream site);

σ_{Lj} = the standard deviation of lead time (in weeks) for part j ;

A_j = the material availability (fill rate) for part j ;

μ_{mj} = the average material delay time (in weeks) conditional on material unavailability due to the shortage of part j ;

ν_{mj} = the variance of material delay time (in weeks) conditional on material unavailability due to the shortage of part j .

Operating Characteristics

At each site, the parameters as given in the previous section constitute the input data, from which we can compute the operating characteristics of the site. To do so, we need to define the variables of interest:

μ_{P_i} = the expected production time (in weeks) for SKU i ;

ν_{P_i} = the variance of production time (in weeks²) for SKU i ;

μ_{R_i} = the expected replenishment lead time (in weeks) for SKU i ;

ν_{R_i} = the variance of replenishment lead time (in weeks²) for SKU i ;

μ_{T_i} = the expected demand for SKU i in lead time + review period;

$\nu_{T_i}(z)$ = the variance of demand for SKU i in lead time + review - z periods;

μ_{S_i} = the average response time (delay) (in weeks) to demands of SKU i , given that there is a shortage;

ν_{S_i} = the variance of response time (delay) (in weeks) to demands of SKU i , given that there is a shortage;

I_i = the average inventory level of SKU i (as finished products) in the stockpile;

k_i = the safety factor for SKU i .

The average production lead time is given by:

$$\mu_{P_i} = [(u_i R_i - 1)^+ x_i + z_i](1 + n_i \mu_{di}) / C_i, \quad (1)$$

when $X^+ = \max(x, 0)$.

The rationale for (1) is as follows. The average production requirement per week for SKU i is $\mu_i R_i$. The production time for this requirement is z_i for the first unit, and x_i for the rest. The total real production time is thus $(u_i R_i - 1)^+ x_i + z_i$ hours. During the production run, the process is liable to go down. In fact, for each hour of the production run, an expected downtime of $n_i \mu_{di}$ would be incurred. Hence, the total production time is $[(u_i R_i - 1)^+ x_i + z_i](1 + n_i \mu_{di})$. Dividing by C_i gives the production time in weeks.

Next, the total replenishment time is composed of material lead time + material delay time + production time. Assume that, at most, one material input is delayed at each replenishment occasion. Hence, part j is delayed with probability $1 - A_j$ and duration μ_{mj} . This gives:

$$\mu_{R_i} = \sum_j v_{ij} \mu_{Lj} / \left(\sum_j v_{ij} \right) + \sum_j v_{ij} (1 - A_j) \mu_{mj} + \mu_{P_i}.$$

Note that materials are usually managed such that material availability levels are very high. The

probability of more than one material input delayed is very small.

Consequently, we compute the expected demand during replenishment lead time plus the review period: $\mu_{Ti} = u_i(\mu_{Ri} + R_i)$. Similarly, the variance of replenishment lead time is:

$$\begin{aligned} \nu_{Ri} = & \text{Max}_j \{v_{ij}\nu_{Lj}\} + \sum_j v_{ij}(1 - A_j)\nu_{mj} \\ & + \sum_j v_{ij}A_j(1 - A_j)\mu_{mj}^2 + \nu_{Pi}, \end{aligned}$$

where

$$\begin{aligned} \nu_{Pi} = & \{(\mu_{di}^2 + \sigma_{di}^2)[(u_i R_i - 1)^+ x_i + z_i]n_i \\ & + x_i^2(1 + n_i \mu_{di})^2 s_i^2 R_i\}/C_i^2. \end{aligned} \quad (2)$$

The derivation of (2) is as follows. Without loss of generality, we drop the subscript i . Let P , V , and N denote the random variables for production time (in hours), production volume per week, and the number of occurrences of downtime during the production run, respectively. Recall that the occurrence of downtime is modeled as a Poisson process. Consider ν_P in hours² (division by C^2 would yield ν_P in weeks²).

$$\nu_P = E_V[\text{Var}(P|V)] + \text{Var}_V[E(P|V)]. \quad (3)$$

$$\begin{aligned} \text{Var}(P|V) = & E_N[\text{Var}(P|V, N)] \\ & + \text{Var}_N[E(P|V, N)]. \end{aligned} \quad (4)$$

$$E(P|V, N) = [(V - 1)x + z] + N\mu_d. \quad (5)$$

$$\text{Var}(P|V, N) = N\sigma_d^2.$$

$$E(N|V) = n[(V - 1)x + z].$$

$$\text{Var}(N|V) = E(N|V) = n[(V - 1)x + z]. \quad (6)$$

Equation 6 is due to the Poisson assumption on the occurrence of downtime. Using (5) and (6) in (4), we have:

$$\text{Var}(P|V) = n[(V - 1)x + z](\sigma_d^2 + \mu_d^2). \quad (7)$$

Next:

$$\begin{aligned} E(P|V) = & E_N[E(P|V, N)] \\ = & [(V - 1)x + z](1 + n\mu_d). \end{aligned} \quad (8)$$

Combining (7) and (8) into (3), we get (2). Finally, we define:

$$\nu_{Ti}(z) = [(\mu_{Ri} - z + R_i)s_i^2 + u_i^2 \nu_{Ri}].$$

Service Levels and Response Time Measures

Given the operating characteristics of an SKU at a site, we can then compute the service performance of the SKU, which, in turn, will affect how the downstream sites can reliably receive the SKU from this site.

Case 1: $B_i = 0$

$$S_i = 0.$$

$$I_i = -\mu_{Ti} + u_i R_i / 2.$$

$$\mu_{Si} = \mu_{Ri}.$$

$$\nu_{Si} = \nu_{Ri}.$$

Case 2: $B_i > 0$

The probability that the waiting time for a customer order coming in a period is greater than j review periods is given by the probability that the demand in $\mu_{Ri} + R_i - jR_i$ is greater than or equal to S_i . We use the approximation that the probability of a normal variable being greater than z as $1/\{1 + \exp[2\sqrt{(2/\pi)}z]\}$. Hence, we require

$$\begin{aligned} 1/\{1 + \exp[2\sqrt{(2/\pi)}(S_i - u_i(\mu_{Ri} + R_i))/\sqrt{\nu_{Ti}(0)}]\} \\ = 1 - B_i. \end{aligned}$$

From the above equality, we obtain:

$$S_i = \mu_{Ti} + k_i \sqrt{\nu_{Ti}(0)},$$

where

$$k_i = [1/2\sqrt{(2/\pi)}]\ln[B_i/(1 - B_i)].$$

The average inventory level just prior to the receipt of a replenishment batch is the safety stock. The average inventory level after the receipt of a replenishment batch is the average replenishment size, i.e., $u_i R_i$, plus the safety stock. Hence, the overall average inventory is $u_i R_i / 2 + \text{safety stock}$, so that:

$$I_i = k_i \sqrt{\nu_{Ti}(0)} + u_i R_i / 2.$$

Let:

$$N_i = \text{the largest integer smaller than } \nu_{Ti}(0)/R_i s_i^2.$$

Define, for $j = 0, 1, \dots, N_i$:

$$\begin{aligned} f_i(j) = & 1/\{1 + \exp[2\sqrt{(2/\pi)}(S_i - \mu_{Ti} \\ & + u_i j R_i)/\sqrt{\nu_{Ti}(jR_i)}]\}, \end{aligned}$$

We can then derive:

$$\mu_{Si} = \{[f_i(0) + f_i(1) + \dots + f_i(N_i)]/f_i(0)\}R_i. \quad (9)$$

$$\nu_{Si} = R_i^2 \sum_{j=1}^{N_i} (2j + 1)f_i(j)/f_i(0) - \mu_{Si}^2. \quad (10)$$

To derive (9), we proceed as follows. It is easy to verify that the probabilities that the waiting time being zero, 1, 2, \dots review periods are $1 - f_i(0)$, $f_i(0) - f_i(1)$, $f_i(1) - f_i(2)$, \dots , respectively. The value of N_i is the maximum j so that the square root part in $f_i(j)$ is positive. The probability that the waiting time is

$N_i + 1$ review periods is $f_i(N_i)$. Hence, the cumulative probabilities of the waiting time being zero period, one period, . . . , and so on, are given by $1 - f_i(0)$, $1 - f_i(1)$, . . . , $1 - f_i(N)$, 1. Then μ_{Si} is obtained by noting that the expectation of a random variable is the sum of $1 - [1 - f_i(0)]$, $1 - [1 - f_i(1)]$, . . . , and $1 - [1 - f_i(N)]$, which, upon division by the conditional probability of positive waiting time, or $1 - [1 - f_i(0)] = f_i(0)$, gives (9).

To see (10), let us drop the subscript i for ease of exposition. Suppose that W denotes the random variable of waiting time. The variance of waiting time conditional on positive waiting time equals: $E(W^2)/f(0) - \mu_s^2$. Now, to compute $E(W^2)$, we note that it is equivalent to:

$$\left\{ \sum_{j=1}^N j^2 [f(j-1) - f(j)] + (N+1)^2 f(N) \right\} R^2$$

$$= \sum_{j=1}^N (2j+1) f(j) R^2.$$

The last equality above is easy to verify by induction.

ACKNOWLEDGMENT

The development and implementation of the supply chain model at HP has benefited from the inputs of the following individuals: David Archambault, Brent Carter, Tom Davis, Paul Gibson, Allan Gross, and Steve Rockhold.

REFERENCES

- ARREOLA-RISA, A. 1989. Integrated Production-Distribution Systems With Capacitated Manufacturing Facilities. Unpublished PhD Dissertation, Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, Calif.
- AXSÄTER, S. 1993. Continuous Review Policies for Multi-Level Inventory Systems With Stochastic Demand. In *Handbook in Operations Research and Management Science, Volume 4: Logistics of Production and Inventory*, S. Graves, A. Rinnooy Kan and P. Zipkin (eds.). North-Holland, Amsterdam, 175-198.
- AXSÄTER, S., AND K. ROSLING. 1990. Installation Versus Echelon Inventory Stocking Policies for Multi-Level Inventory Control. Working Paper, Lulea University of Technology, Lulea, Sweden.
- CLARK, A., AND H. SCARF. 1960. Optimal Policies for a Multi-Echelon Inventory Problem. *Mgmt. Sci.* **6**, 474-490.
- COHEN, M. A., AND H. L. LEE. 1988. Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods. *Opns. Res.* **36**, 216-228.
- DEBODT, M., AND S. C. GRAVES. 1985. Continuous Review Policies for a Multi-Echelon Inventory Problem With Stochastic Demand. *Mgmt. Sci.* **31**, 1286-1299.
- ERNST, R., AND D. PYKE. 1992. Component Part Stocking Policies. *Naval Res. Logist.* **39**, 509-529.
- FEDERGRUEN, A. 1989. Methodologies for the Evaluation and Control of Large Scale Production/Distribution Systems Under Uncertainty. In *Logistics: Where Ends Have to Meet*, C. F. H. van Rijn (ed.). Pergamon Press, New York, 143-157.
- FEDERGRUEN, A., AND P. ZIPKIN. 1984. Approximations of Dynamic Multilocation Production and Inventory Problems. *Mgmt. Sci.* **30**, 69-84.
- GRAVES, S. C. 1988. Safety Stocks in Manufacturing Systems. *J. Mfg. and Opns. Mgmt.* **1**, 67-101.
- HAUSMAN, W. H., H. L. LEE AND A. X. ZHANG. 1992. Order Response Time Reliability in Multi-Item Inventory Systems. Working Paper, Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, Calif.
- HOPP, W., AND M. SPEARMAN. 1989. Setting Safety Lead Times for Purchased Components in Assembly Systems. Working Paper, Northwestern University, Evanston, Ill.
- KARMARKAR, U. 1987. Lot Sizes, Lead Times and In-Process Inventories. *Mgmt. Sci.* **33**, 409-419.
- LEE, H. L., AND C. BILLINGTON. 1992. Supply Chain Management: Pitfalls and Opportunities. *Sloan Mgmt. Rev.* **33**, Spring, 65-73.
- LEE, H. L., C. BILLINGTON AND B. CARTER. 1993. Hewlett Packard Gains Control of Inventory and Service Through Design for Localization. *Interfaces* (to appear).
- MUCKSTADT, J. A., AND L. J. THOMAS. 1980. Are Multi-Echelon Inventory Methods Worth Implementing in Systems With Low-Demand Rate Items? *Mgmt. Sci.* **26**, 483-494.
- PYKE, D. F. 1987. Push and Pull in Integrated Production-Distribution Systems Under Stochastic Demand. Unpublished PhD Dissertation, The Wharton School, University of Pennsylvania, Philadelphia.
- ROSLING, K. 1989. Optimal Inventory Policies for Assembly Systems Under Random Demands. *Opns. Res.* **37**, 565-579.
- SCHMIDT, C., AND S. NAHMAS. 1985. Optimal Policy for a Two-Stage Assembly System Under Random Demand. *Opns. Res.* **33**, 1130-1145.
- SHERBROOKE, C. C. 1968. METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Opns. Res.* **16**, 122-141.
- SILVER, E. A., AND R. PETERSON. 1985. *Decision Systems for Inventory Management and Production Planning*. John Wiley, New York.
- ZIPKIN, P. 1986. Models for Design and Control of Stochastic, Multi-Item Batch Production Systems. *Opns. Res.* **34**, 91-104.