RC 20502 (90865) 7/16/96
Engineering and Technology 32 pages

Research Report

Designing Products and Processes for Supply Chain Management: An Application to the Design of an Electronic Product

Amit Garg
IBM Research Division
T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Designing Products and Processes for Supply Chain Management: An Application to the Design of an Electronics Product*

Amit Garg
IBM, T. J. Watson Research Center,
P. O. Box 218,
Yorktown Heights NY 10598

July 9, 1996

Abstract

In this paper we describe an application of designing products and processes for supply chain management at a large electronics products manufacturer. The objective of our research project was to reduce the costs of complexity resulting from a proliferation of parts and processes in the manufacturer's supply chain. In order to perform this analysis, we developed the Supply Chain Modeling and Analysis Tool (SCMAT). SCMAT models decentralized supply chains and is less data-intensive and yet more general than previous work in this area (Lee and Billington [16]).

1 Introduction

In the last twenty years, many companies have invested heavily in improving their manufacturing, marketing and finance operations. Excellence in these functions has now become a competitive necessity rather than a source of strategic advantage. Companies like HP (Lee and Billington [16]) and IBM have realized the strategic importance of

*This work was done while the author was at Stanford University.

effective supply chain management. Supply chain management takes a more holistic view of the whole enterprise and therefore offers great opportunities for savings.

However, companies have not done extensive analyses of their supply chains because the size of the problems can be daunting even for medium-sized businesses. Lack of modeling tools in this area has also hindered proper analyses of different issues in supply chain management. In this paper we describe the Supply Chain Modeling and Analysis Tool (SCMAT) that we have developed and its application to designing products and processes for supply chain management at a large electronics manufacturer (LEM) (Garg [9]).

Although the SCMAT was developed with our application in mind, it can be used to perform a host of strategic-level analyses of fairly general supply chains. In particular, the types of analyses managers could use the SCMAT to perform include:

- *Inventory-serviceability trade-offs:* This includes the effect of fill rates, customer target response times, and the positioning of safety stocks in the supply chain, etc.
- *Sourcing, Location and Transportation trade-offs:* This includes the effect the sourcing, location and transportation decisions would have on total costs, lead times and their variabilities.
- *Effects of Capacity Limitations:* This includes the effect of capacity on inventories and lead times at each node in the supply chain. Or, on the other hand, determining the capacity required at each node to support the end-customer service requirements.
- *Impact of Lot Sizes:* The effect of lot sizes and their variabilities on inventories and lead times at various stages of the supply chain.
- *Designing Products/Processes for Supply Chain Management:* This includes the effect of different product and process design strategies like postponement, process sequencing (Garg and Lee [10]), and modularization, on inventories and service; the impact of changes in the product mix and product variety on the lead times and inventories in the supply chain.

At the very outset, we would like to define the use of the term "process design". Process design can be performed at an operational and at a tactical/strategic level. At the operational level, it involves the design of production processes required to manufacture the products (Ulrich and Eppinger [20]). At tactical/strategic level, process design involves the design of business processes like distribution, packaging, transportation, order entry and management, etc. (ReVelle et al [18]). In this project, while we studied process design at both operational and tactical/strategic levels, we focused primarily on the tactical/strategic-level process design.

The contributions of this paper are of both theoretical and of practical importance. From a theoretical standpoint, our model is more general in that we model the congestion due to capacity limitations at each site and the interference due to multi-product flows through each site more explicitly than previous research. Our model more robust because it is less data-intensive than existing models of supply chains. These extensions are detailed in Section 2.

From a practical standpoint, as a part of this project, we developed a Differentiation Map and classification schemes to analyze the proliferation of product variety in the supply chain. We found these tools useful in communicating and in analyzing the effect of product differentiations. In addition, it also helped us to map the real processes into the abstractions required by the model. Our paper contributes to the practical aspect of research by outlining a methodology for approaching projects on designing products and processes for supply chain management at a large corporation.

This paper is organized as follows. In Section 2 we review relevant literature. In Section 3 describes the application of the model at the LEM to make strategic-level decisions on designing products for supply chain management. Section 4 describes SCMAT in more detail. Finally, in Section 5 we conclude by describing areas for future research.

2 Literature Review

Researchers have made considerable efforts in modeling multi-site production and inventory networks under different conditions. Multi-echelon systems are very difficult to

model, and therefore very few models yield optimal policies. In addition, even if one obtains an optimal policy, it may be very difficult to compute the values of the policy parameters.

Rather than detail all the papers in this area, we would refer the reader to Graves' [11] excellent review of production planning models for multi-location networks. We will refer only to some of the research that is more closely related to our model.

Multi-echelon production and inventory networks can be operated on a centralized or a decentralized mode. Clark and Scarf [5] obtained a finite-horizon optimal policy for a serial system under centralized control with a base-stock inventory policy. They then presented some approximations for applying their model to distribution networks. Federgruen and Zipkin [8] extended this model to the infinite horizon case. Rosling [19] derived an optimal policy for a general assembly network by reducing it to a serial system under certain initial conditions. One can solve then this serial system by employing Federgruen and Zipkin's results.

The total costs of a system operating under centralized control are usually lower than that of a decentralized system (Axsater and Rosling [1]). However, centralized policies can be very complicated for most general systems. And in practice, most supply chains still operate in a decentralized mode.

Graves [11] developed a model of production networks with various sources of uncertainty. The key to his solution approach is solving the single-site inventory model repeatedly in order to recover the entire network. In this model the production rate at a site is variable and moreover is a decision variable. Although the model can perform different types of supply chain analyses, assuming total flexibility in varying production rates at each site is a limitation of this model.

Cohen and Lee [6] modeled a more general decentralized network. They have solved their problem by decomposing the network into three sub-models: material control, production, and distribution. The material control sub-model is an assembly system that supplies material to the manufacturing sites. The outputs from the manufacturing sites are fed into an arborescent distribution sub-model. The key linkage between manufacturing and distribution is the manufacturing lead time that becomes the replenishment lead time for the distribution sub-model.

Lee and Billington [16], henceforth referred to as LB, developed an analytical model for a decentralized supply chain. This model is very useful for strategic-level analyses of supply chains. It assumes that the replenishment lead time for a product at a node comprises the material lead time, the production lead time, and the delay time. Like Graves, LB have also solved the whole network by repeatedly solving the single-product, single-site inventory problem.

Analysis of the effect of variability due to different sources of uncertainty on the performance of the supply chain is becoming increasingly important. While previous work in this area has considered these sources of uncertainties, it has analyzed supply chain performance based solely on the average values of the performance metrics. To our knowledge, LB's is the first work to have developed expressions for the first two moments for all their performance metrics. This makes their model especially appealing for applications.

Modeling supply chains at large corporations requires enormous amounts of different types of input data. Our research is motivated mainly by problems we encountered in collecting the large number of different types of input data required by the LB model and by some application-specific requirements at the LEM. We have improved upon the LB framework by using fewer types of input data. Since inaccuracies are inherent in input data for most real applications, models using fewer types of input data tend to be more accurate and more robust.

For example, the LB model requires the user to input the delay at the downstream production site due to non-availability of parts. We found that the LEM did not maintain this type of data for its supply chain. Therefore, we used simple analytical approximations to estimate this data.

Another input the LB model requires is the capacity at a node allocated to each product flowing through that node. This input has effectively allowed LB to decompose the node into sub-nodes, one for each product. One can then solve the single-site inventory problem repeatedly for each product-node combination to determine the system performance. In practice, however, many products usually share a common resource: manufacturing lines, storage space, etc. It is very difficult to pre-determine the capacity allocated to each product flowing through a node. In fact, the managers at the LEM

were also interested in capacity requirements at each site in the supply chain and in studying the implications of capacity on lead times and inventories.

This type of analysis is clearly not possible through most previous periodic-review models of inventory networks. In addition, congestion effects due to capacity limitation at each node, and the interference effect because of multi-product flows through each node have not been explicitly captured by any models of inventory networks.

Queueing models allow one to capture the above-mentioned effects explicitly. Our model incorporates queueing approximations within the inventory modeling framework. Considerable research has been done on analytical models of general queueing networks. Broad categories of approaches taken by researchers are as follows:

- Decomposition methods
- Diffusion approximations
- Mean value analysis
- Operational analysis.

We refer the reader to Bitran and Tirupati [2] for a description of and for details on these approaches. Decomposition approaches assume that nodes in the network are stochastically independent, and therefore one can construct the solution for the entire network by solving each node individually. Whitt [21], Buzacott and Shanthikumar [3], and Bitran and Tirupati [2] are some of the authors who use this approach. Since it seems to be consistent with the disaggregation method of solving the production-inventory network, we have employed the queueing approximations used in the decomposition approach in our inventory model. These approximations are described in more detail in Section 4.

3 Application to Designing for Supply Chain Management

In this section we describe the application of SCMAT to designing for supply chain management at the LEM. A new line of products, code-named George, was to be introduced by the LEM. George is a high-technology product in a market that is facing price pressures due to increased competition. The product design team comprising of engineers from software, hardware and mechanical design; manufacturing; marketing; and distribution were aware of the impact of the design on the total cost of the product and were interested in some analyses to select among various product and process design alternatives currently on the table. However, this type of initiative was very new to the LEM. We were part of a corporate manufacturing department team working very closely with the design team to support this initiative.

3.1 Problem Description

One of the main reasons that the design team was interested in our assistance in evaluating different product and process design alternatives was because George was unique within the LEM product portfolio. It shared some common components with the existing products, but the final configuration included several new components and required some additional processes. Therefore, there was a need to determine the best configuration of and the best set of processes required for manufacturing the product.

As result of the numerous discussions we had had with members of the design team, the questions that they wanted us to address can be summarized as follows:

- Where should the main components be manufactured?
- Where and how should the components be integrated?
- Where should the final packaging be performed: at one of the factories or as a part of the distribution operations?
- What kind of packaging is the best for this product?

The project comprised of three main parts: inventory modeling, cost analysis and packaging design. In this paper we will focus mainly on the inventory modeling part of the project.

3.2 Methodology

One of the first tasks for the project team was to develop a *Process Diagram* to map the processes in George's supply chain. This process diagram was used as basis for future communication with the design team, and later was also useful in inventory and cost modeling. However, developing the process diagram for a product that is yet to be introduced presented some challenges. Fortunately for us, all of the components and processes to be used in George were either common with or were similar to some existing product lines within the division. We used the characteristics of the similar or identical processes to construct the process diagram.

In addition to the Process Diagram, we also constructed a *Differentiation Map* as a tool to communicate with engineers from design, manufacturing, and distribution. The Differentiation Map depicted the course of product proliferation along the Process Diagram within the George family. At each point of differentiation, it highlighted

- The source of differentiation—product-driven or process-driven.
- The effect of the differentiation on product—internal or external.
- The number of new sub-families or products that result.

Our first objective was to identify the feasible set of product and process design alternatives to consider. The Differentiation Map was particularly useful for this purpose. Classification of differentiations into *product-* versus *process-*driven and *internal* versus *external* helped us narrow the set of possibilities. For example, we found that changing processes for many process-driven differentiations would not be cost-effective because it would affect many other product lines. Similarly, it was relatively easy to standardize internal differentiations because it would allow us to retain the functionality across product lines without affecting the "look and the feel" of the product. External differentiations were usually cosmetic and were dictated by marketing. These and other

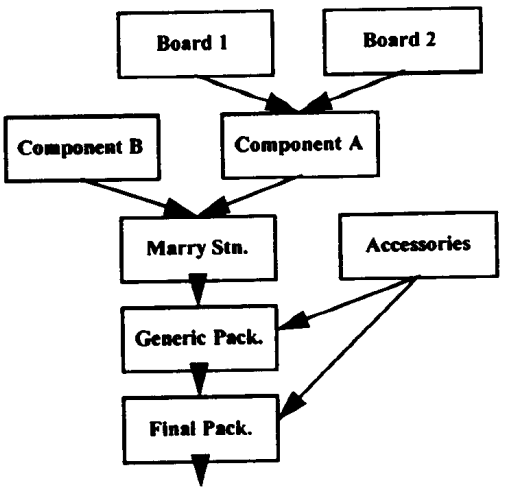

Figure 1: Scenario 1 process diagram

technical considerations were used to determine the set of scenarios we analyzed. Figures 1 – 3 depict a simplified version of the process diagrams for three of the scenarios.

From Figures 1–3 one can see the important stages in the George supply chain. Components A consist of two boards: the Board 1 and the Board 2. The two boards are assembled along with other parts into component A. Components B are manufactured at a different site. In Scenarios 1 and 3, there is a Marry Station. At this station, components A and B are integrated so that they can work together. However, in the long term, this operation will be eliminated, so that any component A can operate with any component B. Figure 2 depicts this long-term scenario.

The LEM produces only some of its accessories; most of them are sourced from outside vendors. The Accessory site represents the internal stockpile of accessories required to support production at the factories. These accessories include batteries, carrying cases, power cords and adaptors, manuals, etc. Some accessories are common to each type of model belonging to George. These accessories are packed along with the component A or component A-B kit in the Generic Packing station. The distribution center performs the final packing where model-specific accessories are included in the package. Some of the nodes considered in these scenarios are located within the same factory, while

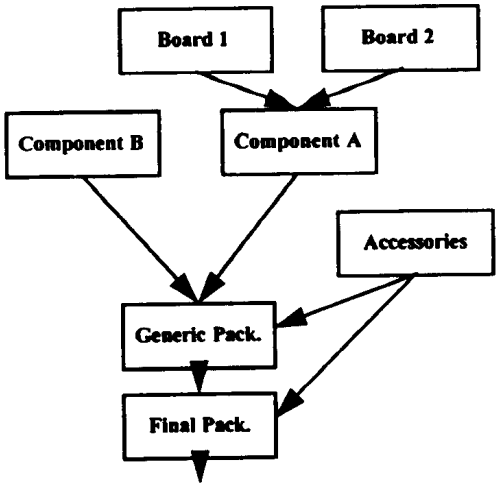

Figure 2: Scenario 2 process diagram

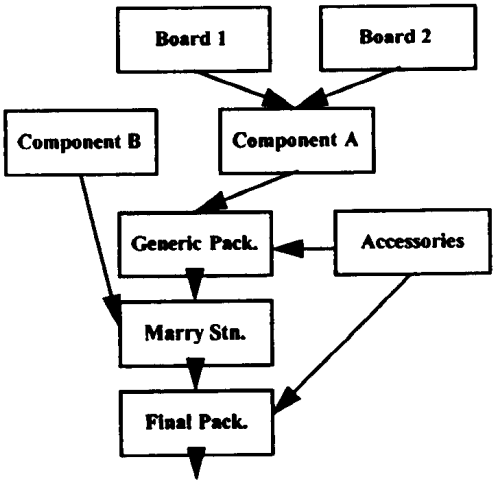

Figure 3: Scenario 3 process diagram

others are at a different site.

Cost, service levels, and inventories were among the criteria for comparing and evaluating the scenarios. As mentioned earlier, there were three main parts to this project: cost analysis, inventory modeling, and packaging design. Packaging design was important because it was an enabler for some of the scenarios, and was also used to reduce the variety within the supply chain. However, we will focus only on the inventory modeling part of this project. The SCMAT described in Section 4 was used for performing the inventory analyses.

3.3 Model Validation

We validated the model in two phases. The first phase included program debugging and verifying the computations and logic through manual calculations. The second phase of validation involved checking the results of the model. Given the short time frame for the project, we used a "reality" check to see if the results corresponded with the actual situation. This also included verifying the accuracy of the input data. In addition, the modeling team had frequent interactions with the clients. This was another means of validating the results of the model. However, we would still need detailed simulations for a more complete validation of some of the approximations made in the model.

3.4 Results

The relative inventory costs at each site under the three scenarios are depicted in Figure 4. The differences in inventory levels among the three scenarios do not appear to be significant. However, this result can be misleading because it does not include the effect of cost buildup as value is added in the supply chain. The differences between scenarios are amplified once the effect of these costs are factored in.

From Figure 4 we can see that Scenario 2 has lower inventory holding costs. A major cause of lower inventories in Scenario 2 is the absence of the Marry Station. Therefore, there are fewer inventory stock-points in Scenario 2 compared to Scenarios 1 and 3. Also notice that Scenario 2 adds cost to the products slower than in Scenario 1. For example, in Scenario 1, the components A-B kit is packed with generic accessories;

while in Scenario 2, generic accessories are first packed with the component A, and this component A-generic-accessories kit is later integrated with the component B. Since generic accessories cost much less than component B, Scenario 2 adds a greater proportion of the total cost than Scenario 1 closer to the point of realization of demand.

Figure 5 compares the effects of service requirements on the inventory costs of the three scenarios. We can see that inventory costs grow exponentially with the service requirements, and that Scenario 2 dominates the other two scenarios. Although the percentage savings across the scenarios do not seem to be very high, these percentages represent significant dollar amounts. The differences among the scenarios are amplified because of differences in the way costs are added at different points of the supply chain in each scenario.

The results of this section only pertain to the inventory implications of each scenario. However, the total cost, which would include the effect of product and process redesign, could make Scenario 1 or 3 preferable to Scenario 2.

3.5 Some Modeling and Implementation Details

We found that this project had several intangible benefits as well. It made the design team more aware of the implications of design beyond the traditional confines of manufacturing. As a part of this project, we first constructed the Differentiation Map to chart the proliferation of products in the George supply chain. This exercise exploded the myth that George does not have much variety. The reason for this misconception was that currently there are only about three different types each of components A and B within the George family. However, the final model that is shipped out includes accessories and packaging that also differentiate the product. In fact, process differences drove a large number of differentiations. In addition, the design team also came to understand the important role played by packaging. This was the first time that packaging design began so early in the product and manufacturing process design cycle.

Gathering data for this project was a great challenge and an immense learning experience for us. The challenge stemmed from the need to communicate the data requirements precisely to the user and to determine the appropriate sources for all the data.

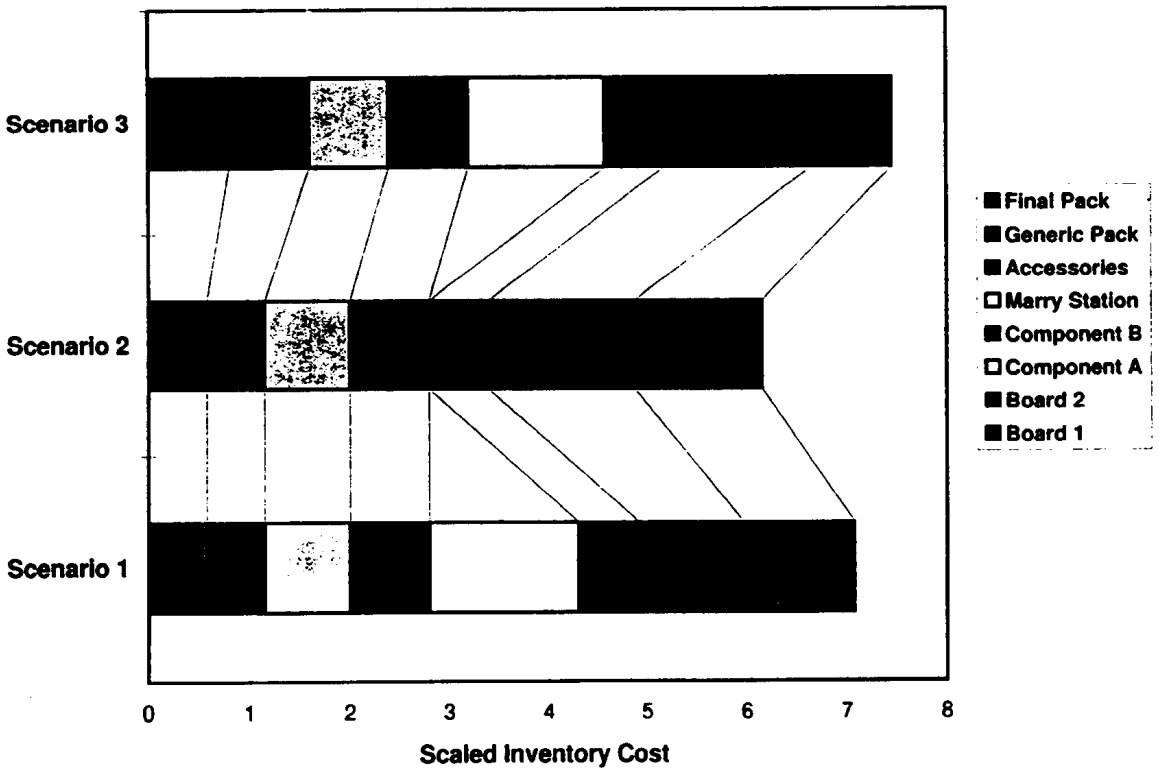

Figure 4: Inventory Costs across the Scenarios

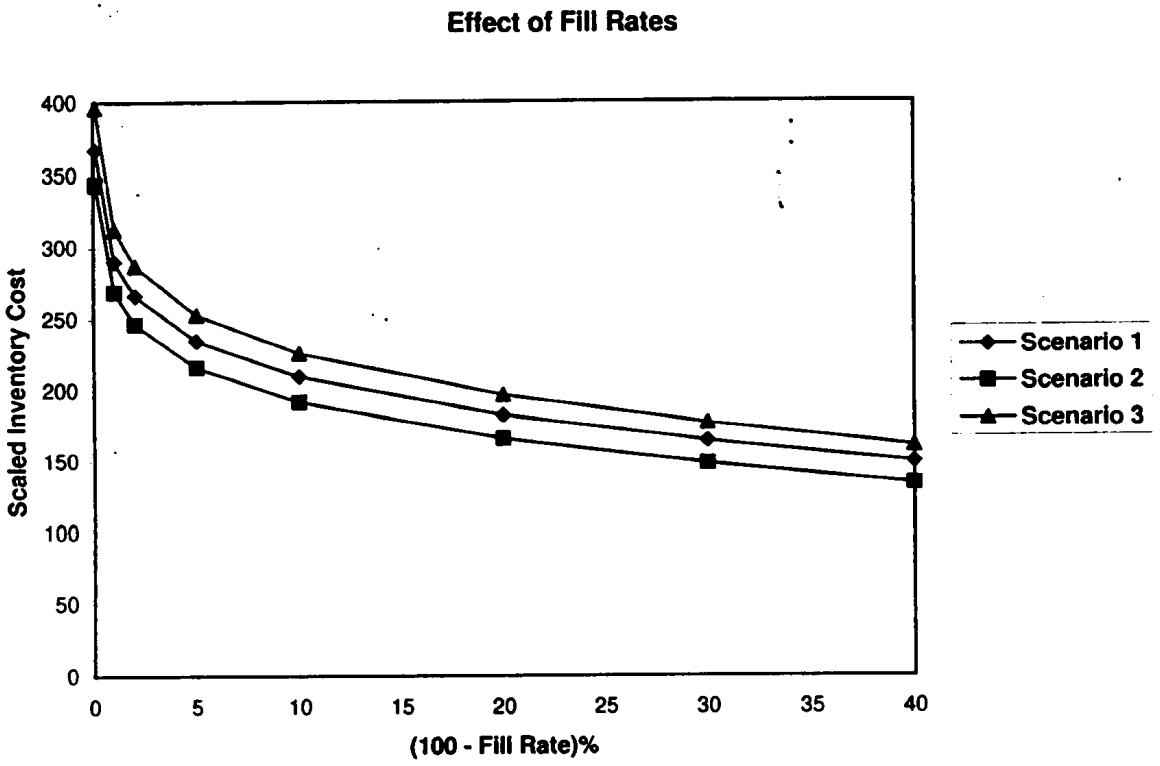

Figure 5: Comparison of the Effect of Service Requirements on Inventory Levels

Therefore, we developed and refined the definitions of each data value we needed. In order to determine the sources, we looked at the commonality and similarity of components across the spectrum of similar products being manufactured within the division. George shares some of its components with existing products. For these components, we used the existing data. However, for the unique components, we used data for similar components being manufactured elsewhere within the division.

4 The Supply Chain Modeling & Analysis Tool

We developed the SCMAT with several implementation-specific criteria in mind:

- Provide answers to decision-makers' questions with a reasonable computation time and with sufficient granularity.
- Implement the model within the project deadlines.
- Keep the number of different types of data required to run the model as small as possible.

Given these criteria and the problem description, we chose a modeling framework similar to that of LB. Like in LB, our framework also uses a periodic-review, base-stock inventory policy at each site. This type of model is sufficiently powerful to address the strategic-level analyses required of our project. A more elaborate model incorporating many operational-level details would not add much to the final decisions while greatly increasing the model development and implementation times due to the increase in the time required to collect different types of input data.

In this model we have made several assumptions and approximations in deriving the analytical expressions for various performance measures. The limitations that these approximations impose on the model are discussed later in this paper.

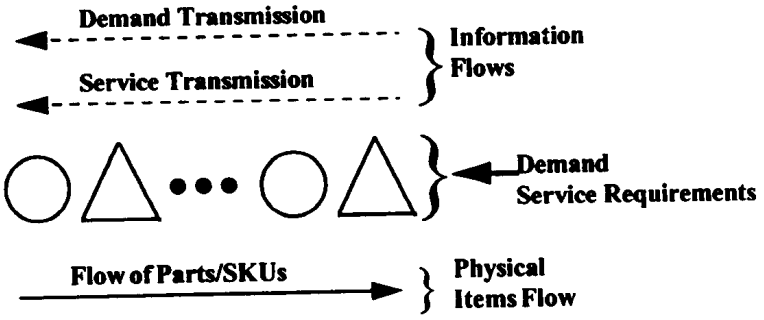

Figure 6: Schematic of Flows in SCMAT

4.1 Model Overview

Nodes in SCMAT can be production sites or distribution centers. At these sites, raw materials, or *parts*, from upstream sites and/or external vendors are transformed into finished products, or *SKUs*. These SKUs may then become parts at some other downstream sites. The Bill of Materials specifies the relationship between the parts and the SKUs in the system. In general, part numbers of inputs at a site are different from those of the outputs. However, if the site is a distribution center, the input and output part numbers are identical because parts simply flow through the facility.

Nodes in SCMAT are connected by two types of flows: the flow of information and the flow of physical items. Both are depicted in Figure 6. There are two types of information flows: Demand Transmission, and Service Transmission. The direction of these flows is from the downstream end-product stages to the upstream raw material stages. Of course, the physical flows originate at the upstream raw material stages and terminate at the downstream end-product stages.

Demands for the SKUs (or the end-products) are translated into demands for each part

and SKU in the system through the Bill of Materials. This is the Demand Transmission Process.

Before explaining the Service Transmission process, we would like to elaborate on the specification of service-level requirements in the SCMAT. Unlike most previous research, we have used a more general specification of the service requirements for an SKU at a given node. Service level is defined as a combination of the SKU *fill rate* and its *target response time*. For example, a service-level requirement of fill rate of 95% with a target response time of 5 days implies that more than 95% of the demands should be satisfied within 5 days of observing them. If the target response time is set to 0, then the service-level requirement reverts to the traditional off-the-shelf fill-rate metric. Therefore, the traditional fill-rate requirement is a special case of the service-level requirement used in SCMAT.

The decision maker specifies service-level requirements for each end-product. The service performance of an end-product is reflected in its fill rate and in its response time, given that it is out of stock. The availability of an SKU at a site depends upon the response times and the fill rates of its input parts from the upstream sites as well as the reorder point at that stage. Therefore, the service-level requirements on end-products drive the service-level requirements for each input part, and therefore for each SKU, at each site. We call this the Service Transmission Process. Since service-level requirement is specified as a combination of the target response time and fill rate at a site, we can see that the Service Transmission Process comprises of Fill-Rate transmission, and Response Time transmission. Note that the Response Time transmission will affect upstream part-level inventories only if an SKU's target response time is greater than its replenishment lead time.

Our model assumes that demands for each end-product are stationary, uncorrelated, and that they are normally distributed random variables. Therefore, from the Demand Transmission process, the demands of each SKU are stationary, uncorrelated, and are normally distributed random variables. We assume that inventory levels for each SKU at each site are governed by a periodic review base-stock policy. For a single-site inventory system, at most two of the following three parameters, the fill rate, the target response time, and the base-stock level, are required to characterize the inventory system. Given any two of these parameters, one can compute the third. As a result of the Service

Transmission process, the specifications of two of the three parameters are available for every SKU at each site. Therefore, we can compute the third parameter value through a search over the vector of possible values for the whole supply chain. The relationship between the three parameters is detailed in Section 4.4.

Base-stock level in the single-site, single-product model depends on the mean and the variance of the replenishment lead time and on the service-level requirement. If the base-stock level is specified, then service performance at a site is a function of the base-stock level and the mean and the variance of the replenishment lead time (Nahmias [17]). The replenishment process at a site includes the production process (if the site is a production site), receiving and storage lead times for inputs, and material lead times for procuring the inputs. If the site is a distribution center, there is no production activity; therefore, the replenishment process includes receiving and storage lead times for inputs and lead times for procuring the parts from their upstream sites.

However, modeling the replenishment process of SKUs that require multiple parts with different lead times is a difficult problem. This complex non-linear programming problem has been studied by Yano [23] and Hopp and Spearman [12]. In the LB model, this data is input by the user. However, in SCMAT, we compute the effect of delay due to non-availability of parts with the help of analytical approximations for the waiting times.

Replenishment lead time in the single-product, single-site inventory model is the sum of the material lead time, the production lead time, and the delay time. This approximation is equivalent to the assumption that these times are independent of one another and are also more or less unaffected by other SKUs flowing through the site. One obtains the solution to the entire network by solving single-product, single-site inventory problems independently. This solution methodology assumes that nodes in the network are stochastically independent.

The material lead time considered in this model is the lead time to procure the part, if it is to be sourced from an external supplier. If the part is manufactured internally, then the material lead time is the lead time to transfer the part from the upstream site to the downstream location. Note that these transfer lead times are not merely the transfer times, because lead times are also driven by lot sizes (Karmarkar [13]).

The production lead time at each site is the only lead time that explicitly reflects the interactions between different products flowing through a site. Like the transit lead times, this lead time is also driven by the production lot sizes of each SKU. Other factors that affect this lead time are capacity constraints, the breakdown frequency and its duration, and the variability of the input process. Variability of the output process at a site is higher than that of its input process because of the effects of variability in the processing time and in machine downtime. Since the output process of one site is the input process into a downstream site, its variability has greater impact on the production lead times downstream. Section 4.5 shows the derivations for the expressions for this lead time.

4.2 Data Specifications

We first define the notation and conventions used in the model. Let indices k and g denote sites, and i and j parts and SKUs.

Site Data

C_k = Capacity in hours per week at site k.
ρ_k = Capacity utilization at site k.
$MTBF_k$ = Mean time between failures in weeks at site k.
$\mu_d(k)$ = Average duration of down time in weeks at site k.
$\nu_d(k)$ = Variance of down time in weeks^2 at site k.

Part Data

F_i = Target fill rate for SKU i.
A_i = Material fill rate for part i.
R_i = Review period in weeks for SKU i.
τ_i = Target response time in weeks for SKU i.
β_i = Effective target response time in weeks for SKU i.

$\mu_M(i)$ = Average material lead time in weeks for part i.
$\nu_M(i)$ = Variance of the material lead time in weeks^2 for part i.
$\mu_D(i)$ = Average delay due to non-availability of part i in weeks given that the part is out of stock.
$\nu_D(i)$ = Variance of delay due to non-availability of part i in weeks^2, given that the part is out of stock.

Bill of Materials Data

$$\delta_{ij} = \begin{cases} 1 & \text{if part } j \text{ is used in SKU } i. \\ 0 & \text{otherwise} \end{cases}$$

n_{ij} = Number of parts i needed to produce one unit of SKU j.

Part-Site Data

$\mu(i,k)$ = Average demand per week in units for SKU i at site k.
$\nu(i,k)$ = Variance of the demand per week in units^2 for SKU i at site k.
$\mu_S(i,k)$ = Average material storage and receiving lead time in weeks for parts required to produce SKU i at site k.
$\nu_S(i,k)$ = Variance of material storage and receiving lead time in weeks^2 for parts required to produce SKU i at site k.
$\mu_{PL}(i,k)$ = Average production lot size in units for SKU i at site k.
$\nu_{PL}(i,k)$ = Variance of production lot size in units^2 for SKU i at site k.
$\mu_{FLOW}(i,k)$ = Average flow time in weeks for SKU i at site k.
$\nu_{FLOW}(i,k)$ = Variance of flow time in weeks^2 for SKU i at site k.
$\mu_{TL}(i,k)$ = Average transfer lot size in units for SKU i at site k to its downstream sites.
$\nu_{TL}(i,k)$ = Variance of transfer lot size in units^2 for SKU i at site k to its downstream sites.
$\mu_P(i,k)$ = Average production lead time in weeks for SKU i at site k.
$\nu_P(i,k)$ = Variance of the production lead time in weeks^2 for SKU i at site k.

$\mu_R(i,k)$ = Average replenishment lead time in weeks for SKU i at site k.
$\nu_R(i,k)$ = Variance of the replenishment lead time in weeks^2 for SKU i at site k.
$S(i,k)$ = Order-up-to point for SKU i at site k.
$I(i,k)$ = Average inventory level of SKU i at site k.
$K(i,k)$ = Safety stock factor for SKU i at site k.
$\mu_r(i,k)$ = Average response time in weeks to demands of SKU i at site k, given that i is backordered.
$\nu_r(i,k)$ = Variance of the response time in weeks^2 to demands of SKU i at site k, given that i is backordered.

Transit Time Data

$\mu_T(k,g)$ = Average transit time in weeks from site k to site g.
$\nu_T(k,g)$ = Variance of transit time in weeks^2 from site k to site g.

In order to construct the solution for the entire network, we need to solve the single-site model for each SKU and site combination. This single-site model requires the demands for the SKU, the required fill rate or order-up-to point, and the replenishment lead time as inputs. Demands and the fill rate requirements are determined as a result of the information flows.

4.3 Demand and Service Transmission

Define $D(k,i)$ to be the set of downstream sites to which site k supplies part i, and $p_j(g,k)$ be the proportion of the requirement of part j at downstream site k that is sourced from site g. Then,

$$\mu(i,g) = \sum_{k \in D(g,i)} \sum_{j} \delta_{ij} n_{ij} \mu(j) p_i(g,k) \tag{1}$$

$$\nu(i,g) = \sum_{k \in D(g,i)} \sum_{j} \delta_{ij} n_{ij}^2 \nu(j) p_i^2(g,k) \tag{2}$$

where g is a site where i is an SKU.

In order to determine the effective target response time at the part level, define $\mathcal{K}_j$ be the set of SKUs that use part j. Therefore, the effective target response time for part j is

$$\beta_j = \max\left(\tau_j, \min_{i \in \mathcal{K}_j} (\beta_i - \mu_R(i,k))^+\right). \tag{3}$$

The service-level requirement at the parts stage is driven by the unit holding costs of the parts and the SKUs, and the service-level requirements at the SKU level. Determination of the optimal fill rates and target response times for all parts and SKUs in the network results in a complex non-linear programming problem (Deuermeyer and Schwarz [7]). In SCMAT, we have used a simple search procedure to determine the vector of inventory parameters for the whole system.

4.4 The Single-Site Inventory Model

If part i is manufactured at site k and is required for manufacturing SKU j at site g, its average material lead time is the transit lead time, computed as follows:

$$\mu_M(i) = \mu_T(k,g) + \frac{1}{2}\mu_{TL}(i,k)\mu_{FLOW}(i,k) \tag{4}$$

$$\nu_M(i) = \nu_T(k,g) + \frac{1}{4}\nu_{TL}(i,k)\mu^2_{FLOW}(i,k) + \frac{1}{4}\mu^2_{TL}(i,k)\nu^2_{FLOW}(i,k) \tag{5}$$

The right-hand side in the above equations is the sum of the time required to transfer items between sites and the effect of time required to build a transfer lot of part i. The mean and the variance of the replenishment lead time can be expressed as

$$\mu_R(i,k) = \frac{\sum_j \delta_{ij}\mu_M(j)}{\sum_j \delta_{ij}} + \max_j \delta_{ij}\left(\beta_j + (1-A_j)\mu_D(j)\right) + \mu_P(i,k) \tag{6}$$

$$\begin{aligned}\nu_R(i,k) = {} & \max_j\{\delta_{ij}\nu_M(j)\} + \sum_j \delta_{ij}(1-A_j)\nu_D(j) + \\ & \sum_j \delta_{ij}A_j(1-A_j)\mu^2_D(j) + \nu_P(i)\end{aligned} \tag{7}$$

In equations (6) and (7) we assume that replenishment lead time is the sum of material lead time, the material delay time, and the production lead time. The expected

material delay time assumes that if multiple parts are delayed, the delay of all parts is masked by the maximum expected delay. The delay, $\mu_D(j)$, is computed as the expected waiting time given that the item is backordered. The probability of a delay due to non-availability part j is $(1 - A_j)$ and the expected delay is $(1 - A_j)\mu_D(j)$. This assumption is valid if the fill rates, A_js, are close to 1. Expressions for the average and the variance of the production lead times are derived in Section 4.5, the computation method of these parameters is an extension of the work in LB.

Define $\mu_{LR}(i,k)$ to be the demand of SKU i during replenishment lead time and review period, $\mu_{LR}(i,k) = \mu(i,k)(\mu_R(i,k) + R_i)$, and $\nu_{LR}(i,k,\tau) = (\mu_R(i,k) - \tau + R_i)\nu(i) + \mu(i)^2\nu_R(i,k)$ to be the variance of demand during the residual lead time. Where τ is a specified response time, therefore, $(\mu_R(i,k) - \tau + R_i)$ is the residual lead time.

We now want to give the expressions for the operating characteristics of the single-site system when a target service performance is specified for the SKU. In order to simplify the exposition of the formulas, in the rest of the section we shall modify the notation by omitting the subscripts i and k SKUs and sites respectively. There are two possible cases:

Case 1: $(F = 0$ **or** $\beta \geq \mu_{LR})$ In this case the SKU is replenished under a Make-to-Order policy. The expressions to the base-stock level, the average inventory, etc. are given below.

$$\begin{aligned} S &= 0; \\ I &= -\mu_{LR} - \mu R/2; \\ \mu_r &= \mu_R - R/2; \\ \nu_r &= \nu_R. \end{aligned}$$

If this SKU is used as a part at a downstream site, the mean and the variance of delay due to its non-availability are $\mu_D = \mu_r$ and $\nu_D = \nu_r$ respectively.

Case 2: $(F \geq 0$ **and** $\beta < \mu_{LR})$ The probability that the waiting time for a customer order in a period is greater than j review periods is the probability that the demand in $\mu_R + R - jR - \beta$ weeks is greater than or equal to S. Demands are assumed to be

normally distributed; therefore, we have

$$1 - F = \left\{ 1 + \exp\left[\frac{2\sqrt{\frac{2}{\pi}}(S - \mu(\mu_R + R - \beta))}{\sqrt{\nu_{LR}(\beta)}} \right] \right\}^{-1} \tag{8}$$

yielding

$$S = \mu_{LR} - \beta\mu + K\sqrt{\nu_{LR}(\beta)} \tag{9}$$

where

$$K = \frac{1}{2}\sqrt{\frac{\pi}{2}} \ln\left(\frac{F}{1-F}\right). \tag{10}$$

The average inventory at hand of SKU i at site k is given by

$$I = \mu\frac{R}{2} + K\sqrt{\nu_{LR}(\beta)}. \tag{11}$$

Let

$$N = \lfloor \frac{\nu_{LR}(0)}{R\nu} \rfloor$$

and for $j \in \{0, \ldots, N\}$ define

$$f_j = \left\{ 1 + \exp\left[\frac{2\sqrt{\frac{2}{\pi}}(S - \mu_{LR} + \beta\mu + \mu j R)}{\sqrt{\nu_{LR}(\beta + jR)}} \right] \right\}^{-1} \tag{12}$$

where the probability the waiting time for an order of SKU i at site k is greater than $j - 1/2$ review periods is $(1 - f_j)$. Therefore, the mean and the variance of the response time given that the SKU is out of stock is

$$\mu_r = \frac{R}{f_0}\sum_{m=0}^{N} f_m - \frac{R}{2} \tag{13}$$

$$\nu_r = \frac{R^2}{4f_0}\sum_{m=1}^{N}(2m+1)\, f_m - \mu_r^2 \tag{14}$$

We omit the details of these derivations here, which are similar to those in the LB paper. In this case, if the SKU is used as a part at a downstream site the mean and the variance of the delay due its non-availability can be expressed as $\mu_D = \mu_r$ and $\nu_D = \nu_r$ respectively.

Now instead of target service levels, if the maximum average units of inventory, I_{max}, that can be carried is specified, the equations would be as follows. In this case, $\beta = 0$, and from equation (11) we have

$$K = \frac{I_{max} - \mu R/2}{\sqrt{\nu_{LR}(0)}}. \tag{15}$$

We can compute the fill rate, $F = 1 - f_0$, and the order-up-to point, S, by substituting the value of K from equation (15) into equation (9). Calculations for μ_r and ν_r remain the same as before. If the maximum units of inventory is specified in terms of weeks of demand, or in terms of total dollar value, they can easily be converted to the form used in equation (15).

4.5 Determination of the Production Lead Time

In order to compute the production lead time at a node, we use the node decomposition technique similar to that implemented in QNA (Whitt [21]). Decomposition techniques extend the product form results of Jackson-type networks to more general systems. Besides assuming nodes to be stochastically independent, this approach also assumes that two-moment approximations provide reasonably good results.

The basic approach to analyzing such systems is to focus on the three processes at each node:

- Superposition or merging,
- Flow through a queue, and
- Splitting or decomposition.

4.5.1 Superposition Process

The superposition process combines jobs from different classes to create an aggregate job that approximately represents all job classes in its first two moments. Therefore, each SKU j processed at a node k is a job of a different class at that node. The arrival

rate for each class at the node is λ_{jk} with a squared coefficient of variation denoted by ca_{jk}. The arrival rate, λ_{jk}, is equal to the mean demand rate for SKU j at site k because the system is assumed to be stable. In this section, since we focus on a single node, we shall drop the subscript k for nodes from all notation to simplify the exposition.

Define $\mathcal{C}$ to be the set of job classes processed at the node, and

$$\alpha = \sum_{j \in \mathcal{C}} \lambda_j, \tag{16}$$

$$ca = \sum_{j \in \mathcal{C}} \frac{\lambda_j}{\alpha} ca_j. \tag{17}$$

The expression for the mean arrival rate for the aggregate job is exact; however, there are no exact expressions for the squared coefficient of arrivals for the aggregate job. The ca in equation (17) has been obtained by the asymptotic method. It can also be obtained via the stationary-interval method. It has been found that considerable improvement in the accuracy is attained by using a convex combination of the approximations due to the two methods. This modified approximation is

$$ca' = wca + 1 - w$$

where $w = [1 + 4(1-\rho)^2(v-1)]^{-1}$, $v^{-1} = \sum_j \lambda_j^2/\alpha^2$, and ρ is the utilization at the node.

4.5.2 Flow Through the Node

We use the same expressions as QNA for the waiting time distributions. Let W be the waiting time in the queue. Then,

$$EW = \frac{\mu_{FLOW}\rho(ca + cs)g(\rho, ca', cs)}{2(1-\rho)} \tag{18}$$

where the function $g(\rho, ca', cs)$ is based on approximations due to Kraemer and Langenbach-Belz [14]) and is given by:

$$g(\rho, ca', cs) = \begin{cases} \exp\left[-\frac{2(1-\rho)}{3\rho}\frac{(1-ca')^2}{ca'+cs}\right] & ca' < 1 \\ 1 & ca' \geq 1 \end{cases}$$

$$\mu_{FLOW} = \sum_j \frac{\mu_{FLOW}(j)\lambda_j}{\alpha}$$

$$cs = \sum_j \frac{cs_j \lambda_j}{\alpha}$$

where $cs_j = \nu_{FLOW}(j)/\mu_{FLOW}(j)^2$, and $\mu_{FLOW} = E\eta$ is the mean processing time for the aggregate job, and ν_{FLOW} its variance.

We now need to compute the variance of the waiting time in the queue. Define D to be the conditional delay given that the server is busy, and cD the squared coefficient of variation of D. QNA uses the M/G/1 approximation for cD for GI/G/1 because D depends more on the service time distribution than on the inter-arrival time distribution.

$$cD = 2\rho - 1 + \frac{4(1-\rho)d_s^3}{3(1+cs)^2}$$

where d_s^3 is the ratio of the third moments of the processing time variable, i.e., $d_s^3 = E(\eta^3)/(E\eta)^3$, where η is the random variable denoting the processing time. Since we are working with two-moment approximations only, we need to approximate d_s^3:

$$d_s^3 = \begin{cases} 3cs(1+cs) & cs \geq 1 \\ (2cs+1)(cs+1) & cs < 1 \end{cases}$$

The squared coefficient of variation of the waiting time in the queue, cW, can be approximated by

$$cW = \frac{cD + 1 - \sigma}{\sigma} \tag{19}$$

where $\sigma \equiv \Pr(W > 0) = \rho + \rho(1-\rho)(ca' - 1)h(\rho, ca', cs)$ and

$$h(\rho, ca', cs) = \begin{cases} \frac{1+ca'+\rho cs}{1+\rho(cs-1)+\rho^2(4ca'+cs)} & ca \leq 1 \\ \frac{4\rho}{ca'+\rho^2(4ca'+cs)} & ca > 1 \end{cases}$$

Now $Var(W) = (EW)^2 cW$; therefore, we can now obtain the expressions for time spent in system, Q, at the node.

$$EQ = EW + \mu_{FLOW}$$
$$Var(Q) = Var(W) + \nu_{FLOW}$$

where $\nu_{FLOW} = cs\mu_{FLOW}^2$.

An issue that needs to be resolved in this model is the effect of lot sizes. Karmarkar [13] showed how lot sizes drive lead times in the system. We have used results of Yao et al. [24] to incorporate these effects.

We also need to incorporate the effect of machine breakdowns and downtimes on production lead times. In our model we have assumed that machine breakdowns occur as a Poisson process. This is quite reasonable since we found that breakdowns are a relatively rare event at the electronic manufacturer's factories. Further, we assume that breakdowns occur only when the machines are busy, and job processing resumes from the same point after an interruption in service. We also assume that downtimes are *iid*, and are independent of the type of job being processed.

Define B to be the number of breakdowns that occur during the processing of a job, and the repair time during each breakdown to be δ, where $E\delta = \mu_d$ and $Var(\delta) = \nu_d$. We can see that B is a stopping time for repair times, therefore, we can apply Wald's identity to determine the average total downtime during the processing of a job, μ_{Down}, and its variance ν_{Down}. We have assumed that $B \sim \text{Poi}(E\eta/MTBF)$. Therefore,

$$\mu_{Down} = EB\mu_d \tag{20}$$

and

$$\nu_{Down} = \nu_d(EB)^2 + EB\mu_d^2 \tag{21}$$

4.5.3 Decomposition

In QNA, the method used to split the output stream from a node is equivalent to the assumption that the routing is Markovian. This is again an approximation because QNA uses deterministic routings. Bitran and Tirupati [2] derive several heuristics which improve upon the results of QNA. Bitran and Tirupati's heuristics perform better than that in QNA because they have explicitly considered interactions between different classes of jobs flowing through the node. Define cd_j to be the squared coefficient of variation of the departure process of job class j, cd to be the squared coefficient of variation of the aggregate job, and $p_j = \lambda_j/\alpha$. Therefore,

$$cd_j = p_j cd + cn_j. \tag{22}$$

The first term in equation (22) reflects the effect of the queueing process, while the second term is independent of the service process. It captures the effect of the distribution of the arrivals of the aggregate job between two arrivals of jobs of class j. Bitran and Tirupati have developed three approximations to estimate cn_j. We have used the Poisson approximation for cn_j, which is less computationally intensive while giving relatively good estimates of the departure-process squared coefficient of variation.

This approximation is motivated by fact that the superposition of a large number of independent renewal processes may be approximated by a Poisson process. We omit the derivations and just present the final expression for cd_j here.

$$cd_j = p_j cd + (1 - p_j)\left[p_j + (1 - p_j)ca_j\right] \tag{23}$$

This approximation performs well when p_j is small, or when there are a large number of products, or when the arrivals of each job class are close to being Poisson. Equation (23) implies that $cd_j \geq ca_j$. $cd_j = ca_j$ will be equal only if the arrivals are Poisson, i.e., $ca_j = 1$. The squared coefficient of variation of the arrival process, ca_j, is a departure process from an upstream node, and is a result of the approximation in equation (23). Therefore, we can see that variability increases progressively as one goes downstream.

5 Conclusions and Future Research

In this paper we describe a tool for modeling and for studying different supply chain issues. We have applied this model to designing products and processes for supply chain management at a large electronics manufacturer. Application of such techniques can result in great benefits both tangible and intangible. Among the intangible benefits were an increased awareness of the implications of product and process design on the entire supply chain. The engineers and the managers involved in the project realized that the product is defined not by what is shipped out of the factory but by what is received by the end-customer. This resulted in the awareness of the proliferation of product variety also due to process differences and to due packaging and other considerations. The effects of these are also felt in upstream stages.

In our application, the differences among scenario inventory levels do not appear to be significant. However, factoring in the effects of cost buildup down the supply chain

showed that the dollar savings were indeed significant. Application of this model to the study of other supply chain issues would be of great interest.

Our model extends previous work in this area by incorporating queueing approximations to model the congestion effects at each site. Like the previous models, ours makes several approximations. These approximations could be tested and improved upon with the help of detailed simulations. This is an area of future research.

Acknowledgments

We would like to thank Prof. Hau Lee of Stanford University for his support and his comments, and the engineers in the product design team and in the corporate manufacturing department of the LEM for their help and their support in this work.

References

[1] Axsater, S. and K. Rosling, Notes: Installation vs. Echelon Stock Policies for Multilevel Inventory Control, *Management Science,* **39**, 10, 1993, 1274-1280.

[2] Bitran, G. R. and D. Tirupati, Multiproduct Queueing Networks with Deterministic Routing: Decomposition Approach and the Notion of Interference, *Management Science,* **34**, 1, 1988, 75-100.

[3] Buzacott, J. A. and J. G. Shanthikumar, Approximate Queueing Models of Dynamic Job Shops, *Management Science,* **31**, 7, 1985, 870-887.

[4] Buzacott, J. A. and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems,* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[5] Clark, A. and H. Scarf, Optimal Policies for a Multi-echelon Inventory Problem, *Management Science,* **6**, 1960, 475-490.

[6] Cohen, M. A. and H. L. Lee, Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods, *Operations Research,* **36**, 2, 1988, 216-228.

[7] Deuermeyer, B. L. and L. B. Schwarz, A Model for the Analysis of System Service Level in Warehouse-Retailer Distribution Systems: The Identical Retailer Case, in: L. B. Schwarz (ed.), *Multi-level Production/Inventory Control Systems: Theory and Practice*, North-Holland, Amsterdam, 1981, 163-193.

[8] Federgruen, A. and P. Zipkin, Computational Issues in an Infinite-Horizon Multi-echelon Inventory Model, *Operations Research,* **32**, 4, 818-836, 1984.

[9] Garg, A., Product and Process Design Strategies for Effective Supply Chain Management, *Unpublished PhD Dissertation,* Stanford University, Stanford, CA 94305, 1995.

[10] Garg, A. and H. Lee, An Analytical Model to Study the Impact of Process Sequencing Decisions in a Supply Chain, *Working Paper*, Department of Industrial Engineering-Engineering Management, Stanford University, Stanford, CA 94305, July 1995.

[11] Graves, S. C., Safety Stocks in Manufacturing Systems, *Journal of Manufacturing and Operations Management,* **1**, 1, 1988, 67-101.

[12] Hopp, W., and M. Spearman, Setting Safety Leadtimes for Purchased Components in Assembly Systems, *IIE Transactions,* **25**, 2, 1993, 2-11.

[13] Karmarkar, U. S., Lot Sizes, Lead Times, and In-Process Inventories, *Management Science,* **33**, 3, 1987, 409-418.

[14] Kraemer, W. and M. Langenbach-Belz, Approximate Formulae for the Delay in the Queueing System GI/G/1, *Congressbook, Eighth International Teletraffic Congress,* Melbourne, Australia, 1976, 235-1–235-8.

[15] Lee, H. L.,*Single and Multiple Location Stochastic Inventory Models,* Lecture Notes, Department of Industrial Engineering-Engineering Management, Stanford University, Stanford CA 94305-4024, 1992.

[16] Lee, H. L. and C. Billington, Material Management in Decentralized Supply Chains, *Operations Research,* **41**, 5, 1993, 835-847.

[17] Nahmias, S., *Production and Operations Analysis, Second Edition,* Irwin, Homewood, IL, 1993.

[18] ReVelle, J. B., N. L. Frigon, H. K. Jackson, *From Concept to Customer: The Practical Guide to Integrated Product and Process Development, and Business Process Reengineering,* Van Nostrand Reinhold, New York, NY, 1995.

[19] Rosling, K., Optimal Inventory Policies for Assembly Systems under Random Demands, *Operations Research,* **37**, 4, 1989, 565-579.

[20] Ulrich, K. T. and S. D. Eppinger, *Product Design and Development,* McGraw-Hill, New York, 1995.

[21] Whitt, W., The Queueing Network Analyzer, *The Bell System Technical Journal,* **62**, 9, 1983, 2779-2815.

[22] Whitt, W., Performance of the Queueing Network Analyzer, *The Bell System Technical Journal,* **62**, 9, 2817-2843.

[23] Yano, C. A., Stochastic Lead Times in Two-Level Assembly Systems, *IIE Transactions,* **19**, 4, 1987, 371-378.

[24] Yao, D. D. W., M. L. Chaudhry, and J. G. C. Templeton, On Bounds for Bulk Arrival Queues, *European Journal of Operational Research,* **15**, 1984, 237-243.