RC 19944 (02/21/95)
Engineering and Technology

IBM Research Report

Dynamic Modelling for Re-Engineering Supply Chains

Daniel Connors (corresponding author), Chae An, Stephen Buckley, Gerry Feigin, Anthony Levas, Nitin Nayak, Raja Petrakian, Ramesh Srinivasan

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Ranga Jayaraman

IBM Corporate Finance and Planning
IBM Corporate Headquarters
Old Orchard Road
Armonk, NY 10504

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties).

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Dynamic Modelling for Re-Engineering Supply Chains

Daniel Connors (corresponding author), Chae An, Stephen Buckley, Gerry Feigin,Anthony Levas, Nitin Nayak, Raja Petrakian, Ramesh Srinivasan

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

Ranga Jayaraman

IBM Corporate Finance and Planning
IBM Corporate Headquarters
Old Orchard Road
Armonk, NY 10504

Dynamic Modelling for Re-Engineering Supply Chains

ABSTRACT: In this paper we describe a modelling approach and methodology used to study and design a supply chain for a major IBM product line. A discrete event simulation model encompassing the key business, logistics and physical processes of the supply chain is described. Capabilities of the model are given and experimental results comparing the operation of the supply chain under various parameter settings and policies are provided.

1. Introduction

The management of a supply chain is a critical factor in the success and profitability of a manufacturing company. In the computer industry, where product life cycles are on the order of months, the design and execution of the supply chain is of great importance since manufacturers face challenges on both ends of the supply chain. On the demand side, customer acceptance of new technologies is often unpredictable, such as the explosive growth of customer demand for multimedia personal computers. Market research indicates that the sales of personal computers during the Christmas season of 1994 rose by 33% over that of 1993, 75% of which was for multimedia personal computers [1]. The spurt in the sales of IBM's main-frames in 1994 shows the need to reassess not only the demand in the client-server market place, but also the performance potentials and short-falls of competing technologies. On the supply side, where component procurement lead times are also measured in months, manufacturers of such high-technology products are buffeted by the speculative nature of the availability and price schedules of many commodity components. For example, the supply of dynamic random access memory (DRAM) chips over the past few years vividly illustrates the influence of unpredictable factors such as corporate alliances and third-party licenses for production and rapid introduction of several generations of memory devices [2]. For a manufacturer like IBM the challenges in dealing with the market and supply side forces are aggravated by the existence of multiple product lines requiring many common resources. While economies of scale do result, operational issues such as allocation of capacity to different product lines crop up and need to be addressed dynamically.

Often, it is necessary that a supply chain be re-engineered in order to respond to the external influences of the supply and demand. Many of the changes and redesigns of the business and physical processes in a supply chain required to improve the performance are either self-evident or emerge naturally when a systematic understanding is developed through interviews with people involved in the business. What is usually lacking is an understanding of how different changes interact with each other and the quantitative assessment of impact resulting when these changes are implemented. The latter is particularly useful in prioritizing the changes to be implemented in the face of scarce resources available for non-routine tasks in a business such as re-engineering the supply chain.

In this paper we present the methodology that we employed in responding to the need for re-engineering the supply chain of one of IBM's major product lines to improve its inventory levels and service levels to the market place. We developed a model of the entire supply chain in which the major business processes and physical processes and their interrelationships were captured. The scope of the model developed in IBM may be described as follows. All the significant products produced in this product line were included in the model. The product assembly, electronic card assembly and component supply, both internal and external, warehousing and shipping processes were modelled. Important business processes such as base plan generation, materials requirements and safety-stock planning, and order-entry were approximately modelled. The business process model that we developed was then analyzed using discrete-event simulation. The analysis comprised of measuring the performance of the supply chain under various scenarios of manufacturing, distribution, planning, supply management and product design processes and parameters. Our model was capable of implementing the existing logic of the business processes, or logic based on analytical tools, such as those described in the literature in Section 2, and tools we have developed. Thus, our model also served as a test bed to validate and compare alternative decision support tools. Out of this analysis a set of recommendations made up of enhancements to the existing processes as well as radically new processes was presented to the senior management.

The rest of the paper is organized as follows. In Section 2 we discuss different approaches to supply chain re-engineering considered by others. We also discuss the extents to which various techniques in the inventory literature have been successful in dealing with supply chain management. The methodology employed in modelling the supply chain is discussed in Section 3. A detailed view of the supply chain that was studied, including the business and physical processes involved, is presented in Section 4. In Section 5 we present various experiments dealing with new proposals for the business and physical processes.

2. Literature Review

In response to the importance of supply chain management to businesses, researchers and analysts are investing more resources and time in trying to gain a better understanding of the dynamics of supply chains and in developing models that can be used to support strategic or tactical supply chain decisions. In this section we review some recent work that is both well rooted in scientific knowledge and is applicable to real world problems.

2.1. Managing the Supply Chain as a Whole

Operations researchers have been studying the topic of supply chain management since the early 1960s. They have traditionally modelled supply chains as multi-echelon inventory networks with uncertain demand and have attempted to find solutions that are optimum or nearly optimum across the complete network [3, 4]. Until recently, they had met with limited success in the sense that their results could not be readily transferred and used in industry. Their theoretical analyses were based on assumptions that were often unrealistic: the multi-echelon systems are serial or arborescent, only a single item is produced, replenishment lead times are fixed, etc. [5]. The concept of echelon stock, the inventory of a component at one echelon plus the inventory of the downstream items that use the component, was commonly used for centrally controlling inventory.

One departure from that tradition was that given by Cohen and Lee [6] who developed an analytical approach for modelling supply chains that was based on decentralized inventory control mechanisms. This approach incorporates various elements of the supply chain, including raw materials supply, production, warehousing, and distribution of finished goods to end customers. The approach can be used to either predict the cost and serviceability performance of the supply chain or to determine the appropriate control policies that minimize costs for a given set of serviceability measures associated with the various echelons of the supply chain. In their approach, each echelon in the supply chain is modelled with a different set of stochastic equations that constitutes a sub-model. A decomposition heuristic is used to solve the supply chain optimization problem; it optimizes the individual sub-models while exploiting the linkages between them.

To manage the supply chain as a whole, Hewlett-Packard has developed an analytical framework, [7, 8], that takes into account uncertainties that characterize raw materials supply availability, manufacturing, distribution, transportation, and customer demand. The framework is intended to support strategic decisions, not just operational ones. It helps determine inventory requirements of a given supply chain and compare various supply chain designs. Two case studies are given in [7], where this approach was successfully applied at Hewlett-Packard. The author emphasizes that companies should actively collect valid data on supply chain uncertainty. Such data allows the effective use of powerful analytical methods. The author also shows how the output of quantitative models can help managers make shrewd decisions through greater insight in the dynamics of a supply chain. Lee and Billington [8] provide details on the analytical framework. It is similar to the decomposition based approach developed by [6] except that a single basic model is used for all echelons in the supply chain. A search heuristic is used to find solutions that optimized operations in the supply chain. The model is based on the assumptions that inventory replenishment is controlled by a periodic basestock system, demands at all echelons are normally distributed, and demands at various time periods are independent. Despite the limitations imposed by these assumptions, the modelling framework has its advantages. It is flexible, analytically tractable, and it can be used to engineer the supply chain and to determine the optimal inventory investment required for providing a given service level.

At Digital Equipment Corporation (DEC), a Global Supply Model [9] was developed to evaluate the cost and speed of supply chain networks. The model also analyzes the impact that vendor selection, facilities locations, modes of transportation, duties, differing tax rates, and transfer prices have on the supply chain performance. The model is optimized by using deterministic mixed-integer programming and by applying various solution methods such as elastic formulation, factorization, and cascades. It has a Graphical User Interface (GUI) which displays the output on a world map. The model was used at DEC in at least twenty studies and is credited in helping decrease their supply chain costs by about 10% and cycle time by as much as 50%.

Towill and Del Vecchio applied filter theory to study supply chains [10]. In their work they focus on the phenomenon of demand amplification which is the distortion of real customer demand as the information propagates upstream in the supply chain away from the customer. One key characteristic of demand amplification is that demand variation increases along a supply chain. The authors model the supply chain as a series of amplifiers, where each echelon was described through transfer functions, which are ratios of polynomials in the Laplace operator, s. They propose that alternative supply chain designs be evaluated by comparing their filter characteristics. Through simulation experiments they show that their approach results in supply chains that reduce total inventory in the system for a given service level.

2.2. Managing Facets of the Supply Chain

Not all work relevant to supply chain management studied the chain as a whole. For example, Fuller, O'Conor and Rawlinson [11] focus on the front-end of the supply chain. They contend that while it is becoming cheaper to manufacture goods, logistics and distribution costs have not changed much. They also add that most logistics systems still do not provide enough flexibility to create value for the customer. They determined that this is due to the fact that often a single logistics process is used to satisfy demand for a variety of products by customers with different serviceability criteria. As a result, many customers are either unsatisfied with existing service levels or are being overcharged for services they do not require. To help companies design logistics systems that are less costly and that create value to the customer they propose the use of their "Logistically Distinct Business Methods" which divide customers into separate market segments that can be supported profitably and competitively by distinct logistics processes.

On the other hand, in Fisher *et al* [12], it is argued that most companies have failed to incorporate demand uncertainty into production planning processes that currently treat demand forecast as some true representation of reality. The authors propose a new forecasting, planning, and production approach called "Accurate Response" designed for an environment where demand is highly volatile. Using a probabilistic model of demand, products are divided into two categories. The first category includes products for which demand can be forecasted accurately and that can be built in advance. Products in the second category have more volatile demand and are not manufactured until some real market information becomes available. A mixture of historical information and expert judgment is used to estimate the expected accuracy of forecasts. In addition, information on missed sales opportunities and early market demand information is incorporated to improve the quality of the results. The authors believe that their approach provides a competitive edge to companies having short lead times and flexible manufacturing. They also mention that to be successful, an implementation of their approach should be coupled with changes to the supply chain, such as delaying product differentiation in both the manufacturing process and the bill of materials, and reducing the variety of raw materials used.

Improvements to information systems are widely believed to be a major contributing factor for successful supply chain re-engineering. Evans and Towill [13] conducted simulation studies of the Forrester production-distribution model [14] to investigate the impact that implementation of Electronic Data Interchange (EDI) can have on the cost structure of a supply chain. The authors determined that through lead time reductions, EDI can result in significant financial savings. A major finding of this work is that making information about the actual customer demand immediately available to all players in the supply chain can dramatically improve the performance of the supply chain. The authors showed that such an approach reduces demand amplification. They also argued that to remain competitive, companies will need to adopt "Global Integrated Logistics Information Systems" that coordinate information within a supply chain, providing demand forecasts globally and synchronizing demand with supply quickly and efficiently.

3. Software Architecture and Implementation

This section of the paper describes the requirements, architecture, and implementation of the software system that facilitated the modelling and simulation of the supply chain.

3.1. Requirements

A set of software requirements emerged from our understanding of the domain of the supply chain, and from our desire to apply the model to solve pressing supply chain business problems.

It was clear that the software system would need to provide a powerful modelling methodology enabling the representation of the complex behavior of the process steps and their associated interdependence. A variety of decision support tools needed to be integrated into the process control policies. Simulation was the underlying analytical approach that would execute the dynamic behavior of the models. The various system components would be developed in parallel by a diverse team of programmers and domain experts, thus requiring a high degree of modularity. We intended to reuse the underlying system for many different types of experiments on supply chains so that reuseability, maintainability and extensibility were important. It was apparent that a graphical user interface could not be developed in the allotted development time. Many ideas for a generic business process modelling and analysis tool had been developed in another project. This system developed for the supply chain analysis provided a perfect opportunity to test the ideas. These and other requirements led us to the design of an object-oriented software system.

3.2. Architecture

The software system consists of three major components: Modelling, Simulation and Analysis. Each of these components contains a set of classes that manages a different aspect of the system. The Modelling component is used to define the attributes and behavior of processes, interconnect them into networks and create entities that flow through the processes. It also contains various support classes, such as *bill of material* and *inventory*, that represent relevant model information.

The Simulation component executes the process network based on a set of simulation parameters (the length of time to simulate, various control policy settings, etc.), collects relevant data (cycle time, queue lengths, etc.) and outputs the collected data for analysis. The Analysis component averages data collected from multiple simulation replications, computes confidence intervals and facilitates visualization and presentation of results in the form of graphs and charts.

It was clear that a hierarchical modelling approach would facilitate the composition of process models from reuseable building blocks, the representation of processes at various levels of abstraction and the efficient maintenance of various process alternatives in a single model. Models are constructed from three classes of objects: **Nodes**, **Entities** and **Arcs**. A **Node** captures an executable representation of the structure and behavior of a business process. It encapsulates the relevant attributes and control policies that determine actions of the process. Detailed descriptions of each of the relevant supply chain nodes such as Demand and Supply Planning are presented in the next section. **Arcs** are used to interconnect nodes into networks that define the flow of information and physical items. **Entities** are the actual elements that flow through the process network and either trigger some node behavior or are transformed as a result of the process. They contain attributes that qualify the entity's state and accumulate relevant data, such as cycle time, as they traverse the process network. For example, Orders and Parts (informational and physical entities) flow into the Manufacturing node and are transformed into products that are ultimately delivered to the customers who ordered them.

Nodes are hierarchical and can be recursively decomposed into subnetworks that represent their behavior at a finer level of granularity. In addition, alternative implementations of a node may be defined and associated with the node. This facilitates the representation and storage of alternative process implementations within one model. The Node classes can be partitioned into two groups: Primitive nodes and supply chain nodes. Primitive nodes are very general and provide the basic functionality required to build process networks. They are usually composed to create nodes that exhibit complex behavior. They include nodes for creating and consuming entities (Source/Sink), modelling the passage of time while processing (Delay/RandomDelay), and controlling the flow of entities through the process graph (Branch). Supply chain nodes are nodes that have been constructed out of primitive nodes and capture the behavior of specific key business processes of the supply chain network. Very complex control policies can be constructed through the composition of primitive nodes supplemented with code that calls other programs which provide material requirements planning (MRP), and other decision support capabilities (see Section 5.2). These supply chain nodes have been linked together and capture the sequence of processes that entities undergo in the overall supply chain business process.

The Simulation component contains the classes related to executing the process models and collecting the relevant data for analysis. It includes a Date class which handles calendar dates (year, month, day, hour, minute, second), a TimeStamping class for recording time related statistics, and a RunControl class for specifying simulation related parameters (how long to simulate, number of replications, etc.). Analysis was performed on the output files produced by the simulation runs using a variety of commercial products such as spreadsheets, charting and graphing packages and custom-built routines for averaging results, computing confidence intervals and other analysis techniques.

3.3. Implementation

After evaluating several modelling and simulation programming languages, we selected MODSIM II [15] as our implementation language. We chose it on the basis of its object-oriented programming (OOP) methodology, its process oriented simulation capability and its extensive support for modelling and discrete event simulation (random variables, monitors, events, graphics, etc.). We implemented approximately 80 classes which have about 500 associated methods.

The resulting system evolved over the past several years to accommodate the experiments that are reported in the subsequent sections. The object-oriented design and implementation afforded building, understanding, maintaining, and extending the system through this period. This experience provided us with a testbed to evaluate the efficacy of our modelling and analysis approach and helped us refine our thoughts on the requirements and design of a software system that facilitates this process. As a result we have used what we have learned through this effort to develop a new general purpose Business Process Modelling and Analysis Tool (BPMAT). The advantages of BPMAT are that a user builds a model and defines experiments through a graphical user interface, which requires no programming. In addition, models can be easily understood and shared among users and domain specific libraries of building blocks, such as the supply chain nodes described in the next section, are available for use. This paper focuses on the original prototype system and experiments performed using the system. For more information on our new system BPMAT, see [16].

4. Description of Supply Chain

In this section, we will provide a description of each node in the model of the supply chain we analyzed along with the management policies governing its operations. The scope of business processes we covered were customer, part supply, manufacturing, product distribution, distribution channels, order management, demand planning, and supply planning. We will describe the salient characteristics of the model of each of these processes. A high-level view of the supply chain is given in Figure 1.

4.1. Customer

In our supply chain model, we modelled the demands for products with a Customer node, based on actual sales history. The customers are grouped into regions and orders are generated for each region. An order contains the names and quantities of the desired products, the name of customer region, and the name of the appropriate distribution channel, such as a specific retailer. From the quarterly historical data, the actual demand for each model at each channel for a given week is generated randomly using a multinomial probability distribution, while guaranteeing that the total generated demand for a year is equal to the total actual demand for that year. Optionally, the Customer node can also generate demand in a deterministic manner.

4.2. Order management

The Order Management node replicates the behavior of a company's order entry and scheduling system. It receives orders from customers or distribution channels, and schedules and releases orders for production and/or distribution. This operation also incurs a delay.

After receiving an order, it checks the requested ship date (RSD) to see if the RSD is within the nominal order fulfillment cycle time. If it is, it releases the order to the Product Distribution node or to the Manufacturing node as long as sufficient inventory is available. If the manufacturing policy is Build-To-Order, then the inventory of components are checked to verify if the items in the order can be produced. If there is not enough inventory, the order stays inside the Order Management node until the next day, at which time the inventory levels are checked again to see if the order can be released. In our model, for simplicity we did not

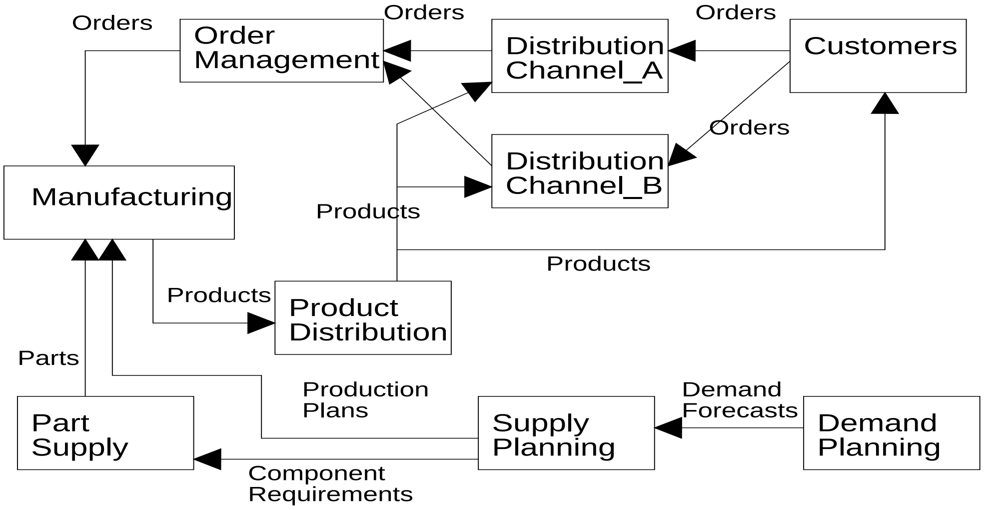

Figure 1: The business processes of the supply chain model

allow any partial shipment of an order; i.e., if an order contains a request for 100 items, the inventory for all 100 items must be available before the order is released.

4.3. Demand planning

For a business, the demand planning process attempts to generate an unbiased forecast of expected shipments of each product in each time period (typically monthly or quarterly). For example, a demand outlook for a product may look like:

1Q:1200 2Q:2400 3Q:3600 4Q:2000

indicating forecasted demands for 1200 units for the 1st quarter, 2400 units for the 2nd quarter, and so on. The forecasts are based on historical sales information, product announcements, market and economic conditions, competition, etc.

This forecasting process is a very complicated and difficult task and we did not attempt to replicate it in our supply chain model. Instead, since the goal of our supply chain model was to analyze the dynamics of interactions in the supply chain, we used the historical demand forecasts to drive the planning process in the model. Hence, the Demand Planning node simply reads in the actual demand forecast for each month and passes it to the next stage in the supply chain, the Supply Planning node.

4.4. Supply planning

In the supply planning node, our model replicates the behavior of a materials requirements planning (MRP) system. The demand forecast for the finished products are translated into two types of detailed plans: the production requirements (base plan) on the production centers and the component requirements on the parts suppliers through a bill of materials explosion. A large number of parameters govern this translation. These include lead times, frozen zones, minimum thresholds for order quantities and changes in orders, safety stock levels, etc. These parameters may be set by specific part numbers or according to part groupings based on the value of annual usage. The capability of the suppliers to produce a feasible, committed production and procurement plan is assessed during this process. The model also captures the delay typical in an actual planning process. A planning cycle may take anywhere from a few days to a few months.

4.5. Part Supply

Following every planning cycle, the Part Supply node receives a procurement plan from the Supply Planning node. The procurement plan specifies shipment requirements for parts to be delivered to a Manufacturing node, in weekly or monthly buckets. The Part Supply node transforms the procurement plan into daily requirements and generates a *starts plan* for each part. Production lots are then started according to this plan. From here, the Part Supply node has several options:

- **Coupled execution, no flexibility**: The simplest option is to produce and deliver parts according to their published *manufacturing lead times*, with no changes to the procurement plan between the lead time and the due date. Under this option, a lot is started when it is a manufacturing lead time away from its due date. Once the manufacturing lead time has elapsed, the lot is shipped to the Manufacturing node, where it is added to the raw parts inventory.

- **Coupled execution with flexibility**: Under this option, the manufacturing lead time is split up into a *frozen zone* and a *flexibility zone*. Lots whose due dates are closer than the frozen zone cannot be changed; they will be shipped as scheduled. Within the flexibility zone, both due dates and lot quantities can be changed from one procurement plan to the next, within limits established by supplier contracts. In practice, this type of flexibility can be achieved at a supplier by keeping inventory or by changing the mix of a lot through late personalization.

- **Decoupled execution**: The procurement plan can be decoupled from the shipment of parts by maintaining a *supplier kanban*. Under this option, the procurement plan represents the maximum level of the kanban during each planning period. Flexibility as described above is available with this option, based on limits established by supplier contracts. The consumption of parts from the kanban by the Manufacturing node triggers pull signals to the Part Supply node. The Part Supply node responds to a pull signal by replenishing the Manufacturing node's inventory with the pull quantity within the *replenishment lead time*, a delay which is usually much less than the manufacturing lead time, since the Part Supply node has been given a series of procurement plans for which to prepare. The pull signals in a period cannot exceed the procurement plan. At the end of each period, the Manufacturing node is liable to the Part Supply node for any parts that were forecasted in the first bucket of the procurement plan but were not pulled.

Whether execution is coupled or decoupled, actual delivery of parts to the Manufacturing node may be on time, early, or late, based on a specified distribution.

4.6. Manufacturing

Manufacturing requires a specified cycle time to convert input parts into output products. The performance of a manufacturing node is constrained by several factors: capacity, availability of components, and randomness in cycle times. Our model of a manufacturing operation focuses only on the above three factors and does not try to capture detailed operation sequences. The production with the Manufacturing node may be regulated by production plans in a Build-To-Plan scenario or on actual orders from customers in a Build-To-Order scenario. Another variant is to regulate the production by a finished goods kanban. Section 5.1 has additional details describing the manufacturing strategies we model.

4.7. Product distribution

Product distribution is the term we use to refer to the process of delivering the components comprising a customer order from the supplier sites to the end customer. From the perspective of product distribution, the suppliers consist of all those locations at which components comprising a finished order are manufactured or assembled. The distribution process begins once the components comprising the order are ready for shipment from the various suppliers, whether internal IBM suppliers or external suppliers, and ends when the order is delivered to the end customer. In terms of customer satisfaction, product distribution is one of the most important steps in supply chain management. Without an efficient mechanism for delivering fully configured orders to customers in a cost effective and timely manner, all the other functions associated with the supply chain will be for naught.

Distribution of orders involves both the physical shipment of components comprising an order to a staging area, or consolidation center, where the order is prepared and then shipped to the end customer, as well as logical management of this process. Logical management includes assigning ship dates to components and configured orders, preparing bills of lading, and determining where components and configured orders should be shipped from and to. Physical shipment may be done in one or more steps and may involve several transshipment points. For example, system units may be first shipped by truck to a transshipment point, then loaded into a larger container and shipped to a second transshipment point, and finally unloaded from the container and shipped by truck to a consolidation center.

In the context of computers, the distribution process is made complicated by several factors. First, a customer order typically comprises more than just a system unit. Usually, orders are placed for a system unit together with a monitor, keyboard, mouse, pre-loaded software, cables, and documentation. Other peripheral devices such as printers and storage backup devices are often included in orders as well. In such orders, several different machine type models may be required and each machine may need additional hardware installed to handle network communications. An additional factor is that customers may have varying service level requirements. In general, service level requirements vary by customer type as well as by geography. For example, purchasers of personal computers for home use may demand delivery of their computers in as little as 24 hours while large corporate customers may be willing to accept longer delivery times as long as a firm delivery date can be met. Also, in the intensely competitive personal computer market, service level requirements are often dictated by what competitors are offering.

Several critical decisions are involved in designing a distribution network. At a strategic level, important considerations include deciding on the number of consolidation centers to have, where they should be located, and the functions they should perform. At an operational level, stipulating the frequency and precise routing of shipments within the network as well as specifying an inventory stocking policy, that is, where finished goods inventory should be kept in the distribution network and how much should be kept, if any, are of central importance. The primary performance measures against which a distribution network will be judged are the total cost of distribution and the service levels provided to customers. The primary costs associated with the delivery network are:

1. Costs of operating and maintaining consolidation centers.

2. Inventory holding costs.

3. Freight costs of shipping components comprising the order to a consolidation center.

4. Freight costs of shipping finished orders from the consolidation center to the end customer.

5. Import duties.

In our supply chain model, the product distribution network for products consists of warehouses and transportation links with associated costs listed above. Warehouses can perform multiple functions. The warehouses receive the finished goods from the manufacturing nodes, places them into inventory and fills orders from this inventory. Since they can receive different product types, they can merge different products into an entity for an order. Bulk shipments of products can also be broken into individual shipments to specific customers. These operations incur delays within the warehouse and then the orders are shipped to their customers via transportation links. The transportation is modelled as a batch operation with a delay and cost.

4.8. Distribution channels

Products may be distributed to end customers through distribution channel partners. Two types of distribution channels are modelled: a Retailer and a Direct-marketer. A Retailer node maintains an inventory of finished goods. Periodically, a Retailer node computes the average demand for each product and orders enough stock to maintain a specified level of inventory based on its specific stocking and ordering policies. If the current inventory exceeds a specified level, it is returned to the manufacturer. In addition to the regular inventory replenishment orders, a Retailer may also submit "unexpected" customer orders for which he does not have inventory. Once per day, a Retailer node ships finished goods to customers from its inventory.

A Direct-marketer node is much simpler than a Retailer node. When a Direct-marketer receives a customer order, it forwards it directly to the Manufacturing node using the Order Management node. Completed orders are received from the Manufacturing node or its warehouses and sent directly to the Customer node. No inventory is kept at a Direct-marketer.

5. Modelling Capabilities and Experimental Results

In this section we describe some of the capabilities of our model in further detail and the experimental analyses that we performed using the supply chain model. The set of experiments are manufacturing strategies, supply planning and contracting, product distribution and channel policies.

5.1. Manufacturing Strategies

The supply chain model is capable of evaluating different manufacturing strategies. We used the supply chain model to evaluate three distinct manufacturing systems: Build-To-Order (BTO), Build-To-Plan (BTP), and Late Customization (LC). We define BTO to mean that a product is built from its subassemblies only after an order for that product has been received.

In the BTP network, the Manufacturing node builds products according to the base plan, pushing the finished products into the warehouses. The base plan must specify the demand for every product.

Under the Late Customization strategy, products are built from "shells". A shell could consist of such assemblies as a power supply, a frame, planars or motherboards, and base memory and base disk storage. The shells are built according to the base plan and are stocked at the manufacturing warehouse. The final assembly for products using the Late Customization strategy is to order. The base plan for this strategy must specify the demand for the underlying shells. The supply chain model is also capable of evaluating the manufacturing strategy in which shells are shipped from the main manufacturing center to remote manufacturing facilities, and the final assembly or customization of products takes place at the remote site. This strategy is similar to the method used by Hewlett-Packard to assemble a printer as described in [8].

The are several advantages of the BTP strategy. For products having large and stable demands, manufacturing production schedules for these products can be implemented to use the manufacturing resources efficiently. In our implementation of BTP, we smoothed the production to a daily going rate. Provided that the right product is on hand in the finished goods warehouse, we can quickly respond to a customer order, by drawing upon that finished goods inventory. The major drawback of the BTP strategy is that it provides no flexibility for meeting the actual customer orders. If the forecasts and subsequent base plans are not accurate, we may end up building the wrong products. We tie up our valuable assets in finished goods that are not wanted by the customers.

The major advantage of the BTO strategy is that assets are not tied up in unwanted finished goods; we only build a product to a real order. The BTO strategy can exploit the commonality of components that may exist across the different products. Rather than committing a common critical component to a product whose demand has been forecasted, as in the BTP case, the BTO strategy does not commit critical components until the order is firm. The major drawbacks to BTO are its effect on manufacturing efficiency and serviceability. With no finished goods inventory to draw upon under the BTO strategy, a product must be built from its subassemblies and tested after the order has been received. For customer orders requiring very short cycle times, for example, one or two days, it may not be possible to build, test and ship a product to a customer within the requested time. Manufacturing efficiency may also be impacted by a BTO strategy if large setup times are incurred switching from the building of one product to a different product. Finally, order skew may also affect the performance of a BTO manufacturing line. If the customer orders do not arrive in a smooth manner, there may be periods of time in which the manufacturing line does not have the capacity to build the products to order within the customer requirement times.

The advantages of the Late Customization approach include a possible smaller final assembly and test time as compared to the full BTO approach and a more efficient use of manufacturing capacity since the shells are built to a plan. The Late Customization approach may also provide a less costly means to meet service requirements in outlying regions. The cost of operating a remote assembly plant which assembles finished products from shells and other components may be much cheaper than having a full BTO capable plant in the outlying region. The LC approach has its disadvantages. Shells are built to a base plan and so assets may be tied up in shells which are not needed. Secondly, the multiple stage build process may incur additional handling and storage costs, and may expose the shells to more handling damages. Also, the LC approach introduces an additional stocking location, which may increase the overall inventory in the network.

In Figure 2, we show the results of running the supply chain model using the BTP, LC and BTO manufacturing strategies for a specific computer production environment. The safety stock planning levels were set in each case so that the three different networks achieved the same service level. The other planning and execution parameters were kept constant for the three scenarios. For the set of parameters chosen, the BTO strategy achieves the same level of service with significantly less inventory in the network.

We note that the choice of manufacturing strategy, BTP, LC or BTO, does not lessen the importance of part planning and procurement. In an industry in which the procurement lead times for critical components are typically several weeks or months, good procurement planning is crucial for the success of manufacturing performance. The BTO, LC and BTP strategies fail if the right components are not available when needed.

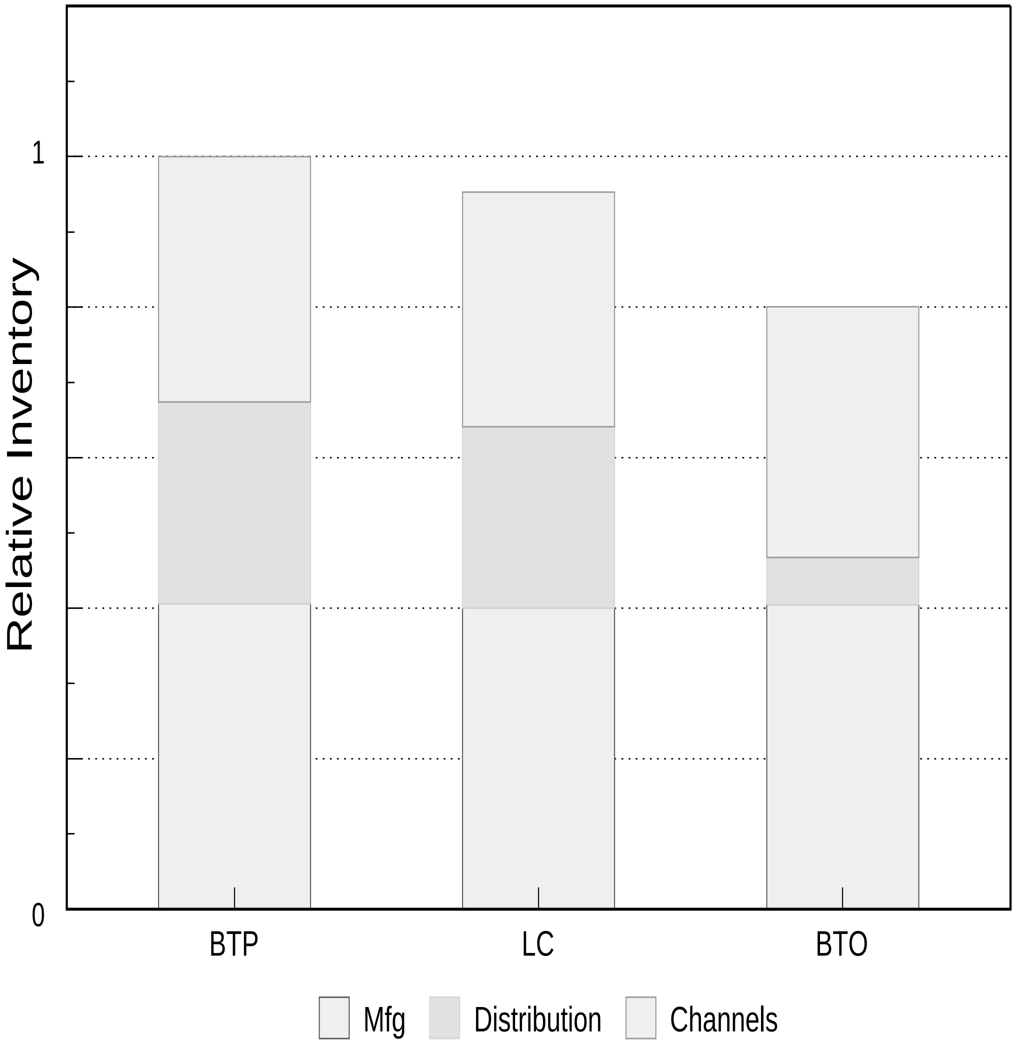

Figure 2: A comparison of the manufacturing strategies BTP, LC and BTO

5.2. Supply Planning and Contracting

As described above in Section 4.4, supply planning takes an unbiased forecast of expected sales and performs a number of computations to obtain a corresponding set of part requirements (see Figure 3). Supply planning is a critical process in the determination of a company's serviceability and inventory. If an insufficient supply of parts is ordered relative to customer orders, serviceability will suffer. If an oversupply of parts is ordered, there may be too much inventory. In manufacturing firms that order multiple part types from suppliers, both of these phenomena can occur, sometimes simultaneously. To improve its situation, a firm can either generate more accurate part and safety-stock requirements, implement closed loop planning, or maintain more flexibility with its suppliers so that part orders can be adjusted as late in the game as possible. Our model examines these possibilities.

5.2.1. Safety-stock planning. Part requirements are currently generated in MRP systems by a time-phased explosion of product demands. However, when product demands are uncertain, safety-stocks are added to average demands (Figure 3, Step 4). MRP systems do not deal with handling the uncertain demands in a satisfactory manner. To plan for safety-stocks practitioners often adjust the product demands to be above or below their average values by a certain fixed percentage and then perform the MRP explosion on these values. In some instances, the materials planning systems set safety-stock levels for parts based on inputs from the planners on the service level for required for parts and assumptions of normal distribution of parts demands. In such cases, however, the correct value of service level for parts cannot be easily determined even if the service required for products are known. Consequently, service level targets for parts are set arbitrarily. In the simulation of the supply chain presented here, we utilized an algorithm to plan the requirements for parts that achieves a specified service level for products while attempting to minimize the safety-stocks of parts required. While benefits of component commonality in reducing overall inventory are well known, this algorithm, referred to here as Flex Planning, was among the first to be developed that addressed industrial size problems. More details on this and other related algorithms developed can be found in [17].

In Figure 4 we present a comparison of the performance of MRP based safety-stock planning method and the Flex Planning method when implemented in the supply chain simulator. The products were built using the BTO strategy in both experiments. It can be inferred from the two graphs in the figure that the Flex

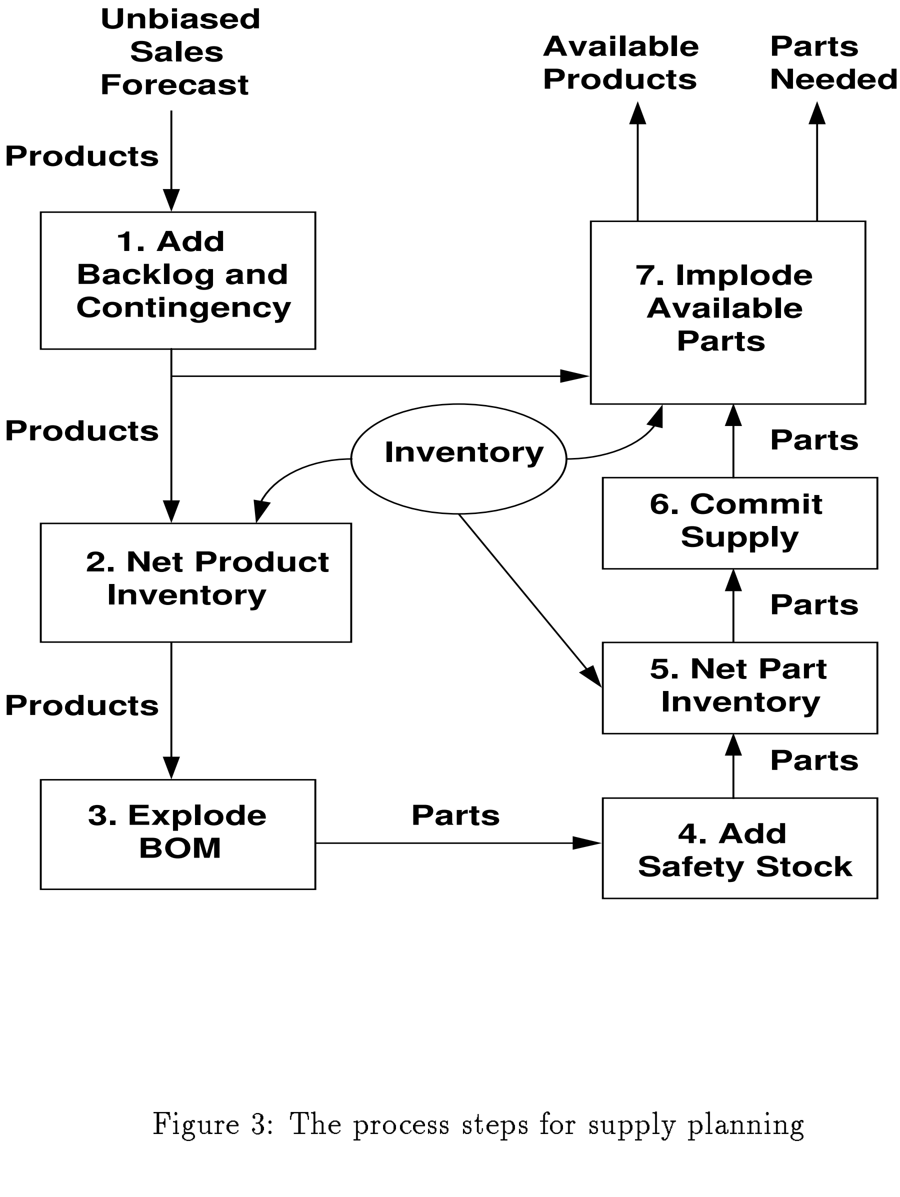

Figure 3: The process steps for supply planning

Planning method required less inventory compared with the MRP based method to achieve the same service level. This effect is attributed to the fact that safety-stocks are set in the Flex Planning method considering component commonality and explicitly recognizing the differences in prices and level of uncertainties among various components.

5.2.2. Closed Loop Planning. After inventory is netted out of the part requirements (Figure 3, Step 5), the suppliers are asked to commit to the remainder of the required supply (Step 6). If most of the parts are 100% committed, but a handful of parts are undercommitted, many of the products cannot be built at the forecasted levels. The quantity of each product that can be built depends upon the allocation of undercommitted parts to the products that need them. If a part is commonly used then this allocation can be tricky. Step 7 represents this allocation, and is referred to as *implosion.* Although implosion can be attempted manually, a much better solution can be obtained by using an optimization tool such as IBM's Implosion Tool [18]. Our simulation deploys this tool. The tool is capable of performing the allocation based on many different objectives, including product priority, fair allocation, and revenue maximization.

Implosion is sometimes referred to as *supply squaring*, because in the process of allocating short parts to products, it identifies parts that should not be ordered because they belong to products that cannot be built. Figure 5 illustrates this concept. *Open loop* refers to a supply planning process which does not perform implosion, and *closed loop* refers to a planning process which does. For this business, it can be seen from the figure that a closed loop supply planning process is marginally useful when there are normal, minor undercommitments, but when there are significant parts shortages, it is critical to perform implosion to avoid inventory problems.

5.2.3. Supplier Flexibility. *Lead time* is the time it takes for a supplier to manufacture and deliver a part order. If a supplier keeps inventory, or its parts have a delayed customization process, it can sometimes change order quantities or mix within lead time. This is referred to as *supplier flexibility.* Outside the lead time, most suppliers will allow changes, unless their capacity is violated.

Frozen zone is the time period relative to the due date in which a part order cannot be changed at all. Minimally, if inventory is not kept by a supplier, its frozen zone is equal to the manufacturing time after final customization plus the delivery time. Maximally, the frozen zone can be equal to or greater than the lead time.

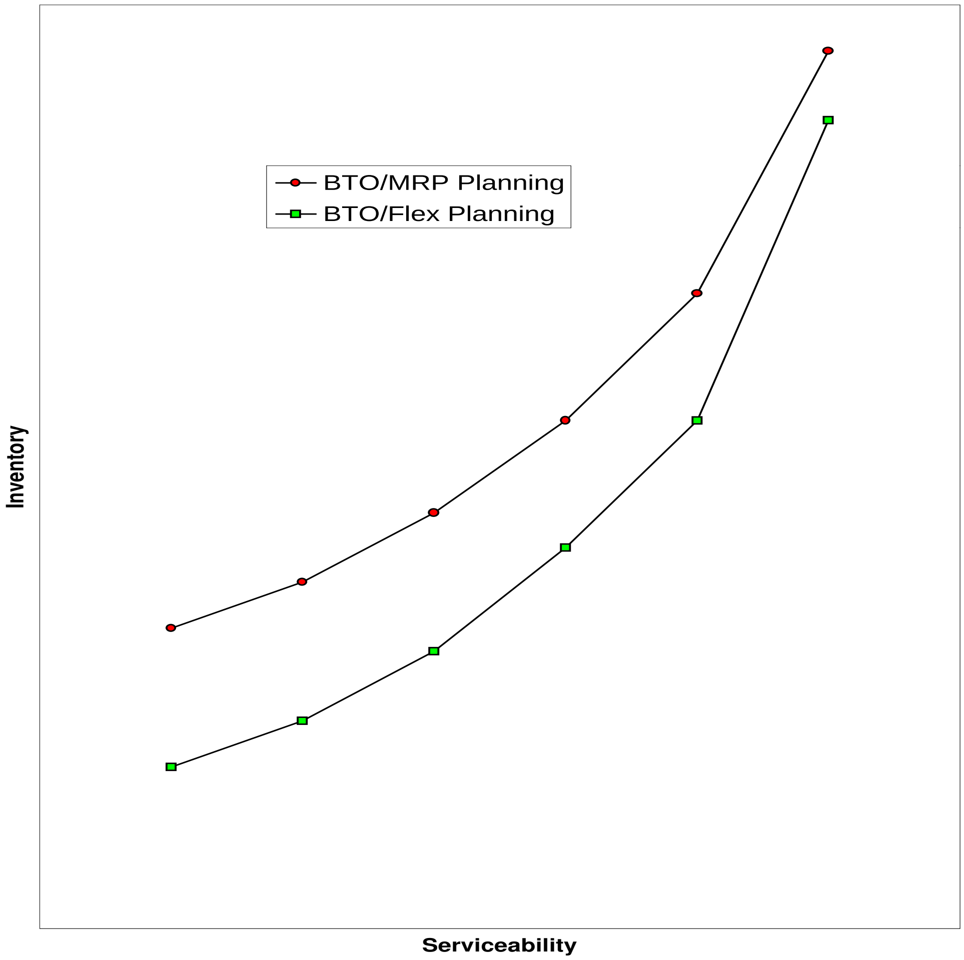

Figure 4: Comparison of safety-stock setting by Flex Planning and MRP based planning

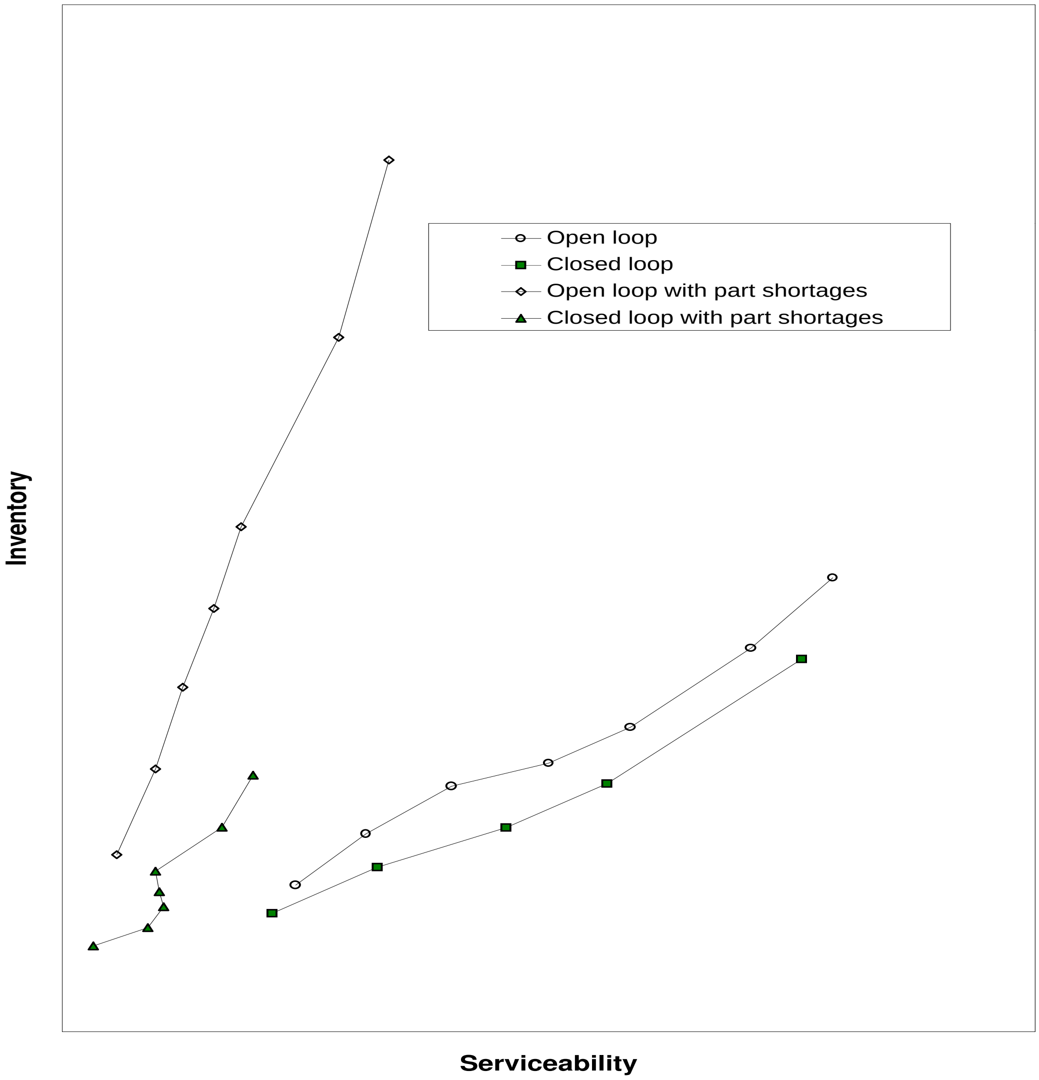

Figure 5: Comparison of open and closed loop planning for various levels of finished goods inventory, with and without severe parts shortages

Figure 6 shows a case study which compares (a) a business in its normal state, (b) the same business with reduced lead times, and (c) the same business with reduced lead times and frozen zones. Clearly, for this business, the forecasts were improving within the original lead times, and thus reducing the lead times paid off. Reducing the frozen zones paid off because some flexibility was allowed between the lead times and frozen zones.

Figure 7 shows a case study which compares (a) a business whose suppliers are inflexible within lead times, (b) the same business with reduced lead times (still inflexible within lead times), and (c) the same business with some flexibility within the original lead times. Case (c) was the best because the flexibility allowed changes to be made very close to execution, when some of the key forecasting adjustments were made.

5.3. Product Distribution

Using the manufacturing network simulator, we were able to create a detailed model of the distribution process that enabled us to test various distribution strategies and to compare their performance in terms of total distribution costs and the achieved service level. We used the simulator to evaluate three different distribution strategies. These strategies can be summarized as follows:

Strategy 1: Customer orders are consolidated at the IBM manufacturing site and then shipped to customers via landing points (transshipment points). Orders are shipped in bulk to these landing points and then, with the orders relatively close to the final customers, they are be shipped individually to the final customers from the landing points. These operations are also referred to as cross-docking.

Strategy 2: Orders are directly shipped to the customers, bypassing the landing points.

Strategy 3: Order consolidation occurs at two locations, at the main manufacturing site and at another late customization center located closer to the final customers. The reason for considering this third strategy is to handle customer orders with stringent service level requirements that cannot not be met at a reasonable cost from the main manufacturing site. The additional consolidation center performs late customization of customer orders, such as adding additional cards or hard disks to the system units, loading language specific software and documentation, and bundling peripheral devices with the system units. The late customization center holds sufficient inventory to meet the customer demand it serves.

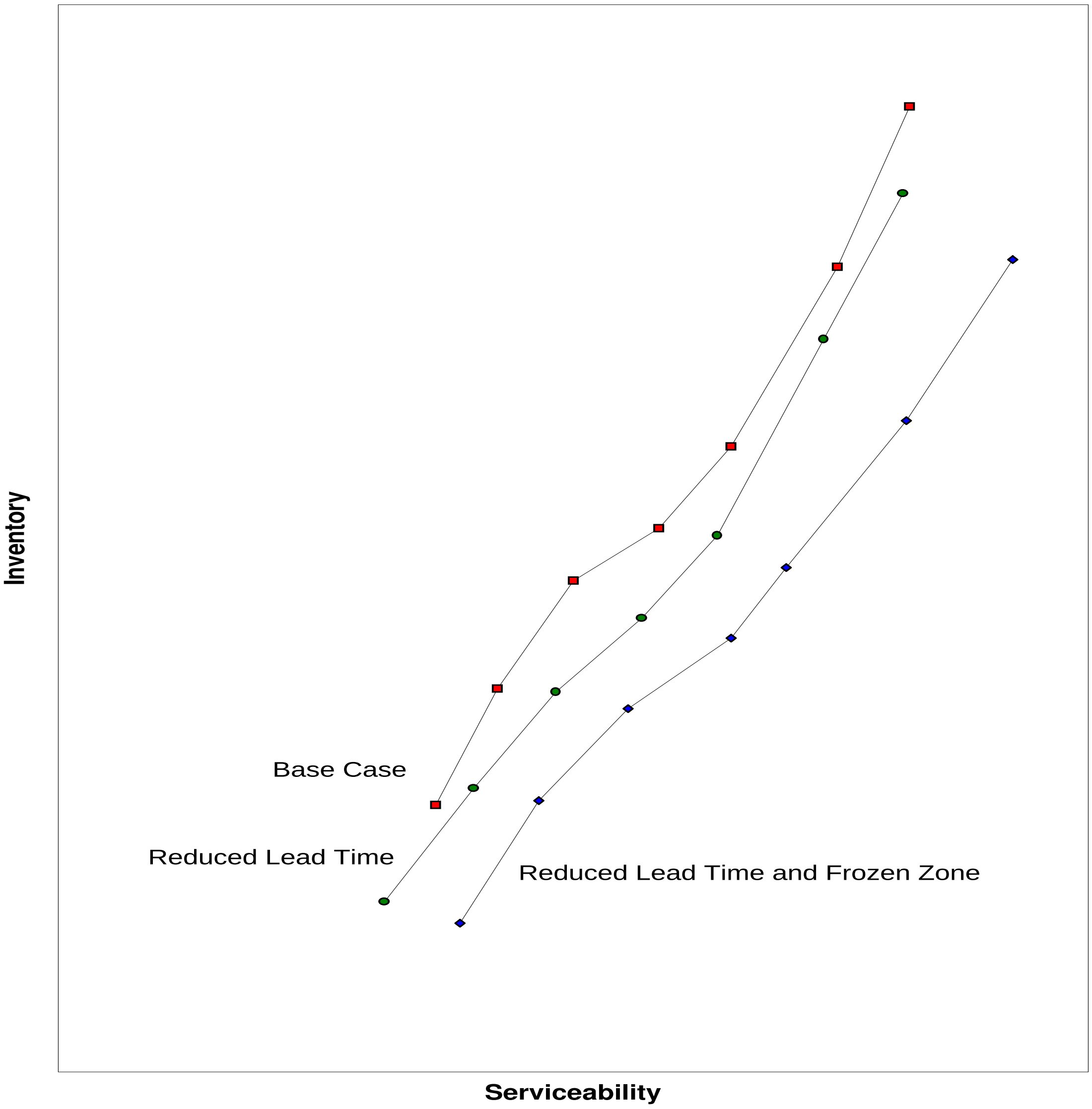

Figure 6: The effect of reducing supplier lead times and frozen zones, for various levels of finished goods inventory

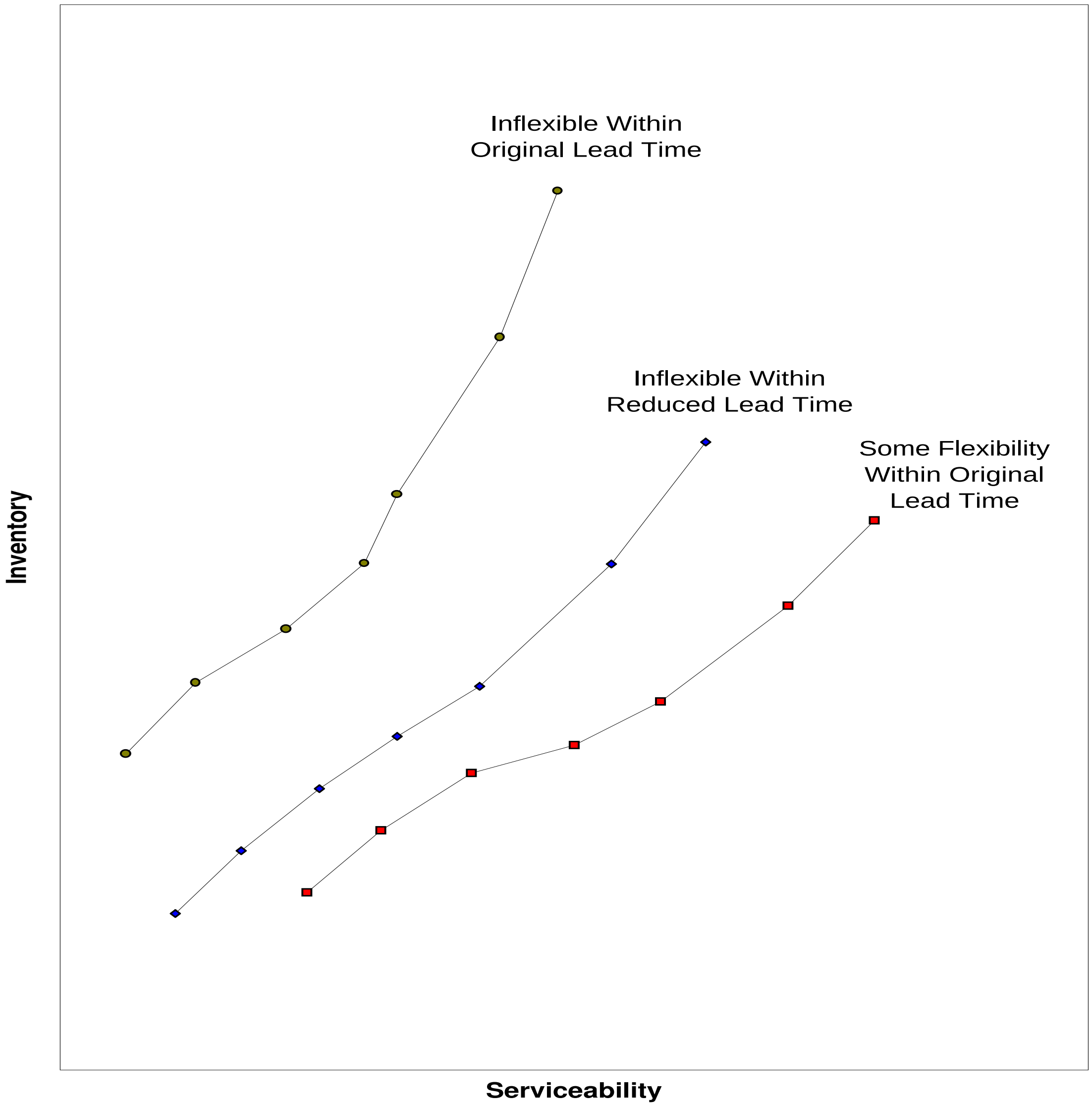

Figure 7: The effect of supplier flexibility within lead time for various levels of finished goods inventory

In our analysis of the first strategy, we considered several cases in which we varied the number of landing points. For the supply chain we modelled, he simulation revealed that the least expensive distribution strategy involved the use of three landing points. When we evaluated cases with as many as eight landing points, we found that the volume of orders shipped to any one landing point was not sufficient to permit frequent shipments from the main manufacturing center. Instead, orders would be held at the shipping dock until a container was full enough to justify a shipment. Thus, significant delays were incurred and inventory holding costs were large. The costs of operating the distribution network with three landing points, instead of eight, were significantly smaller. Cost reductions were due primarily to lower landing point operational costs and a reduction in inventory holding costs. Also, we observed, somewhat unexpectedly, that service levels improved slightly when we reduced the number of landing points because of the increased frequency of shipments to each of the landing points. When we further reduced the number of landing points to fewer than three, overall costs began to increase mainly because of the large inland freight charges incurred when shipping orders from the landing points to the final customers.

When we compared Strategy 1 to Strategy 2, however, we found that distribution costs associated with the least expensive alternative in Strategy 1 were still greater than the cost estimates of having the distribution handled by the direct shipment strategy. Also, the service level provided by this strategy was comparable to Strategy 1.

In our of evaluation of Strategy 3, we were able to clearly quantify the tradeoff in service level and costs associated with the introduction of a late customization center. In particular, the simulation results indicated that introducing a late customization center made sense only if orders requiring the stringent service exceeded 25% of all orders. Since fewer than 25% of all orders would be require the stringent service, based on a pilot study, the introduction of the late customization center did not seem warranted.

As as result of our simulation studies of these different distribution strategies, we recommended that the late customization center was not required and that orders should be shipped directly from the main manufacturing center to the customers.

5.4. Distribution Channels

The distribution channels are often the primary interfaces for the end customers to the supply chain. It is obvious that how well the distribution channel performs is critical to customer satisfaction not only of the channel but of the products they distribute. If the channel or the retailer provides the right products to the customers on time and at low prices, the customers are satisfied. If not, the customers are dissatisfied and may go to a different channel or to a different product.

The effects of the distribution channel policies on the performances of the rest of the supply chain are not as obvious. To understand these effects, we modelled the behavior of the distribution channels and simulated various alternative policies using our simulation model of the supply chain. We characterized the behavior of a distribution channel by several key parameters: level of finished goods inventory, frequency of orders to manufacturers, size of those orders, and return policies. In our model, a distribution channel or retailer receives orders from customers. If there is sufficient inventory at the channel, these orders may be satisfied from the finished goods inventory immediately. If not, the orders are passed on to the manufacturer. In order to maintain a certain level of inventory, the channel places orders to the manufacturers at some regular intervals, typically weekly or monthly. If there is too much inventory whether due to low sales rate or product transition, some or all of the inventory will be returned to the manufacturer.

In our simulation experiments, we varied the average level of inventory at the retailers together with the frequency of orders and order sizes. The rationale for this is that if the average level of inventory is very high, one can place a very large order at infrequent intervals. If the average level of inventory is to be kept low without running out, one must place frequent small orders to the manufacturers. For comparison purposes, safety stock levels in the manufacturing portion of the supply chain were set in each experiment so that the networks operating under the different retail inventory policies achieved the same service level. Figure 8 shows the inventory for the total supply chain, when the retailers' inventory is decreased from two months (Policy 1) in one extreme to three days (Policy 2) in the other extreme. It shows that the total supply chain inventory improves significantly as the retailers' inventory is decreased. The reduction in inventory for the total supply chain is two-fold: one component is accounted for by the direct reduction of the channel inventory, the larger and unintuitive component is the inventory reduction in the rest of the supply chain. In the product line we modelled, the channel inventory originally accounted for about 30% of the total supply chain inventory. If we change the retailer inventory policy in the model to only 3 days, not only is the original 30% belonging to the retailers nearly eliminated, but also there is a significant reduction in the manufacturer's inventory of finished goods and sub-assemblies, while at the same time, improving the serviceability.

There are several reasons for the effect shown on Figure 8. If the retailers maintain large inventories, then they are effectively buffering the actual end-customer behavior from the manufacturers, and the manufacturer does not know the customer buying behavior in a timely manner. When the channel inventory is low, then the buffering effect is reduced and the manufacturer has more direct visibility to the customer buying behavior and therefore, the manufacturer can match their production to what the end-customer wants. Since the orders are more frequent and they are smaller in sizes, the manufacturers can operate their production lines more smoothly and hence make better use of their capacity. With larger, less frequent orders, the manufacturers are challenged with large swings in the production patterns, making it much more difficult to produce and deliver products on-time. These effects have been also demonstrated in the now well publicized story of Proctor & Gamble and WalMart's product replenishment relationship [19].

Another important reason for lower inventory is the reduction of return rates with the reduction of the channel inventory. The lower the returns, the lower the inventory in the total supply chain since the returned goods often cannot be resold. The return rates are lowered as well since the retailers in general would not have much inventory and hence there just isn't much to return. If there are only three days of inventory at the channels and if the product is to be discontinued suddenly, then there would only be three days of inventory being returned.

6. Conclusion

In this paper, we have described a highly detailed operational simulation model of a complex real world supply chain. We have shown the model's utility in the quantification of the impact of redesign alternatives for the supply chain. Based on the experiments performed using the model, we were able to make several key recommendations to the business executives of the supply chain. Our model-based recommendations are being implemented within IBM in a significant re-engineering of its product supply chains. The business results from these changes are expected to be critical to IBM's product competitiveness in the coming years.

An initially unexpected discovery as we did this work is the power of the supply

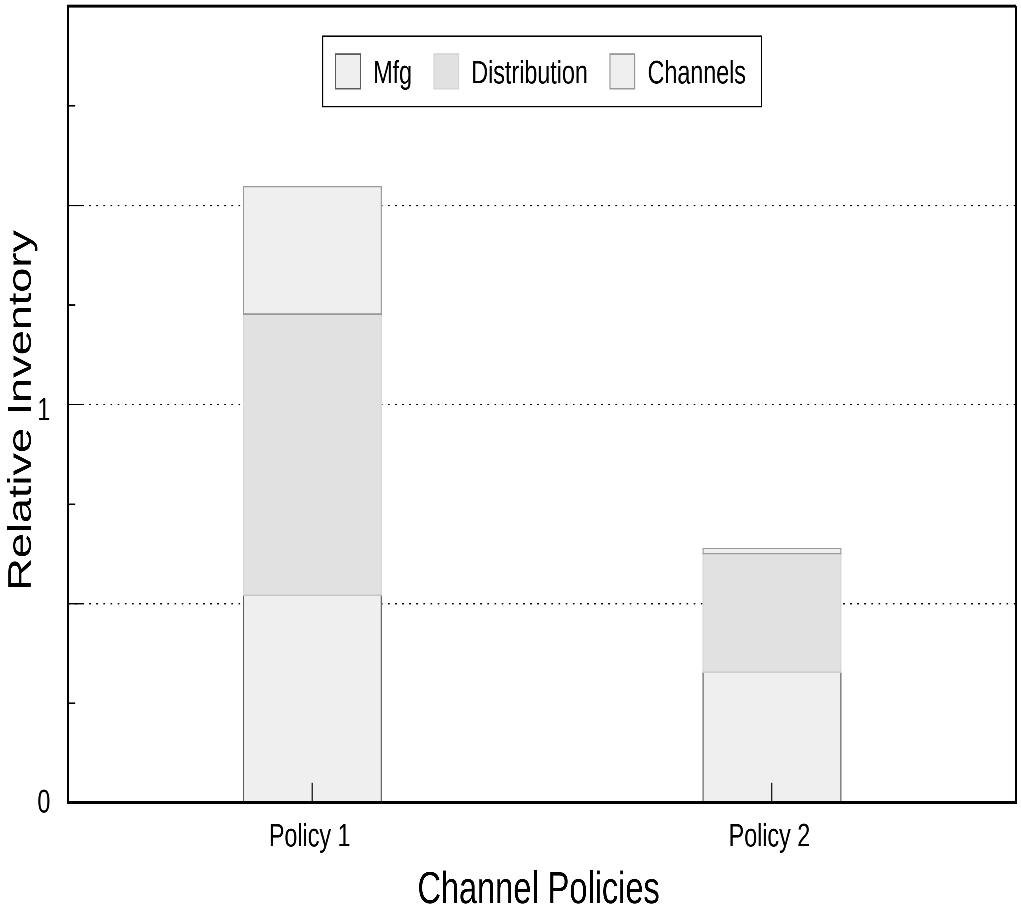

Figure 8: The effects of inventory in the distribution channels while keeping the same customer service level

chain model as a virtual testbed to validate and evaluate the effectiveness of specific decision support tools for individual supply chain processes. Our group has historically developed a number of such tools for the planning and the scheduling processes within the supply chain. Only with the availability of a detailed and validated supply chain model of a real world business have we been able to definitively study their effectiveness in the "laboratory" before actual implementation. Such "laboratory" evaluation has led us to discover the conditions under which a tool is effectively deployable.

A more general conclusion from our work is the role of simulation modelling in business process re-engineering. We have shown that detailed simulation modelling is a powerful approach in business process re-engineering complementary to benchmarking and principles based approaches. Systematic use of simulation modelling can lead to a detailed understanding of the dynamics of the business process, the quantitative impacts of various redesign alternatives, and ultimately serve as a catalyst to generate and experiment with completely new ideas with the potential for breakthrough performance improvements. We believe that as the scope of re-engineering efforts increase in the business world and elsewhere, the use of simulation modelling in business process re-engineering will become more and more widespread to ensure success and reduce risks.

References.

[1] "PC sales over holidays rose 33% from '93 level," *New York Times*, Dec. 27, 1994, p. D5.

[2] "More Mingling with IBM and Siemens: Toshiba Joins 64-Meg DRAM Team", *Electronic News*, May 30, 1994, p. 1.

[3] Clark, A. J. and Scarf, H., "Optimal Policies for a Multi-Echelon Inventory Problem", Management Science, vol. 6, no. 4, 1960, pp. 475-490.

[4] Clark, A. J. and Scarf, H., "Approximate Solutions to a Simple Multi- Echelon Inventory Problem", Studies in Applied Probability and Management Science, Arrow K. J. et al. Editors, Stanford University Press, California, 1962, pp. 88-100.

[5] De Bodt, M. A. and Graves, S. C., "Continuous-Review Policies for a Multi-Echelon Inventory Problem with Stochastic Demand", Management Science, vol. 31, no. 10, 1985, pp. 1286-1299.

[6] Cohen, M. A. and Lee, H. L., “Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods”, Operations Research, vol. 36, no. 2, 1988, pp. 216-228.

[7] Davis, T., “Effective Supply Chain Management”, Sloan Management Review, Summer 1993, pp. 35-46.

[8] Lee, H.L. and Billington, C., “Material Management in Decentralized Supply Chains”, *Operations Research* vol. 41, No. 5, Sep.-Oct. 1993.

[9] Harrison, T. P., Arntzen, B. C., and Brown, G. G., “Global Supply Chain Management”, presentation made on January 10, 1994.

[10] Towill, D. R. and Del Vecchio, A., “The Application of Filter Theory to the Study of Supply Chain Dynamics”, Production Planning and Control, vol. 5, no.1, 1994, pp. 82-96.

[11] Fuller, J. B., O’Conor, J., and Rawlinson, R., “Tailored Logistics: The Next Advantage”, Harvard Business Review, May-June 1993, pp. 87-98.

[12] Fisher, M. L., Hammond, J. H., Obermeyer, W. R., and Raman, A., “Making Supply Meet Demand in an Uncertain World”, Harvard Business Review, May-June 1994, pp. 83-93.

[13] Evans, G. N., Naim, M. M., and Towill, D. R., “Dynamic Supply Chain Performance: Assessing the Impact of Information Systems”, Logistics Information Management, vol. 6, no. 4, 1993, pp. 15-25.

[14] Forrester, J. W., “Industrial Dynamics”, MIT Press, Cambridge, Massachusetts, 1961.

[15] *MODSIM II Reference Manual*, CACI Products Company, La Jolla, California.

[16] Bhaskar, R., Lee, H. S., Levas, A., Petrakian, R., Tsai, F. and Tulskie, W., “Analyzing and Re-Engineering Business Processes Using Simulation”, Proceedings of the 1994 Winter Simulation Conference, pp. 1206-1214.

[17] Srinivasan, R., Jayaraman, R., Roundy R. and Tayur, S., “Procurement of Common Components in a Stochastic Environment”, *IBM Research Report*, RC 18580, 12/11/1992, 45 pages.

[18] *Workstation Implosion Tool: Version 3.3 User Guide and Reference*, IBM T. J. Watson Research Center, April 1994.

[19] Stalk, G. and Hout, T. *Competing Against Time*, The Free Press, 1990.

[20] Dietrich, B., Connors, D., Ervolina, T., Fasano, Jayaraman, R.,J. P., Lin, G., Srinivasan, R., Wittrock, R., "Production and Procurement Under Resource Availability Constraints and Demand Variability," submitted for publication, January 1995.

Copies may be requested from:

IBM Thomas J. Watson Research Center
Publications Office, 16-260
Post Office Box 218
Yorktown Heights, New York 10598