Using Forecast Variability and Risk Pooling to Determine Optimal Safety Stock Levels within a Supply Chain

by

Scott A. Roza

B.S. Marine Engineering
United States Naval Academy

Submitted to the Sloan School of Management
and the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

~~Master of Science in Management~~ [M.B.A.]
and
Master of Science in Mechanical Engineering

at the
Massachusetts Institute of Technology
May 1998

Signature of Author..
Department of Mechanical Engineering
May 8, 1998

Certified by..
Donald B. Rosenfield
Senior Lecturer, Thesis Advisor
Sloan School of Management

Certified by..
Stanley B. Gershwin
Senior Research Scientist, Thesis Advisor
Department of Mechanical Engineering

Accepted by..
Anthony T. Patera
Acting Chairman of the Graduate Office
Department of Mechanical Engineering

Accepted by..
Lawrence Abeln, Director of Master's Program
Sloan School of Management

Using Forecast Variability and Risk Pooling to Determine Optimal Safety Stock Levels within a Supply Chain

by

Scott A. Roza

Submitted to the Sloan School of Management
and the Department of Mechanical Engineering in Partial Fulfillment
of the Requirements for the Degrees of
Master of Science in Management
and
Master of Science in Mechanical Engineering

Abstract

Bay Networks, like other high tech companies whose competitive advantage lies in the design of their products, has moved to a manufacturing model that relies heavily on outsourcing. Almost all the components of its products, from metal boxes and power supplies, to electronic components and the printed circuit boards on which they reside, are manufactured and assembled by subcontractors. This strategy has been pursued to reduce costs and to increase flexibility.

To make this system work effectively, considering the long lead times sometimes associated with networking electronic components, Bay Networks forecasts demand to eliminate a portion of the demand uncertainty. Bay then passes its forecast to the subcontractors so they can deliver material as Bay expects to use it. This model makes the forecasting function of extreme importance when determining whether Bay will end up with too much or too little material.

This thesis analyzes the "outsourced" supply chain that Bay Networks currently uses and how forecasting is really the cornerstone of information dissemination within the supply chain. Additionally, this thesis will look very specifically at how demand for the motherboard of a particular networking product is forecast and how it is an example of taking outsourcing too far. Since Bay Networks is forecasting motherboards with a specific memory configuration, before they actually know what the customer wants for memory, they are reworking many boards to effectively meet customer demand. Further analysis will show that forecasting motherboards in the plain vanilla format, i.e. without memory, and configuring with memory only when they receive a customer order, will allow Bay Networks to take advantage of the concept of risk-pooling. With risk-pooling Bay Networks will be able to increase customer service levels with less inventory- a win -win situation considering that the typical tradeoff consists of increasing inventory to improve service levels.

Specifically, it will be shown that the optimal level of inventory safety stock for these and other motherboards can be significantly reduced by holding them in a format which allows them to be easily and flexibly configured to satisfy demand when a customer order arrives rather than holding the material in a specific format.

Thesis Supervisors:
Donald B. Rosenfield, Senior Lecturer
Stanley B. Gershwin, Senior Research Scientist

Acknowledgments

First and foremost, I would like to thank my wife, Marguerite, for her encouragement and endurance throughout the semesters, the internship, my job hunt, and finally, this thesis. Amazingly she supported me while working full time and through ***TWO*** pregnancies and births that have allowed us to grow our family by 100% - two beautiful daughters who could do no better than to emulate their mother. Who was really working hard here? I am forever grateful.

Next, I would like to thank the folks at Bay Networks for making my internship experience enjoyable and extremely rewarding. Specifically, I would like to thank Vah Erdekian and Tony Zacher for supporting the LFM program and making the internship possible. I would also like to thank the core of Manufacturing Business Planning: Brian Keith, Bob Howard, and Anish Johns for their help and wisdom day in and day out. Finally, I would like to thank Jay Stearns, Joe Susienka, and Steve Witham for their insight about forecast accuracy.

At MIT I would like to thank those people associated with the LFM program. This of course includes my advisors, Stanley Gershwin and Don Rosenfield, who let me set my own course through the internship but asked the right questions to keep me from getting "lost at sea." I would also like to thank everyone else associated with LFM: the staff that ensured my life was simple and focused; the professors that shared both wisdom and humor to make learning fun; and finally, all my LFM classmates, from whom I learned more than I ever could have imagined.

Table of Contents

Table of Figures

1. INTRODUCTION

1.1. Problem Description

Bay Networks, like many manufacturing companies, feels constant market pressure to build and deliver its products more quickly with less cost. To reduce cost, Bay has ardently pursued a strategy of outsourcing. Electronic components, like microprocessors and memory, are manufactured and purchased from the traditional powerhouses of those industries. Other worldwide manufacturers, like Solectron, then solder these components to printed circuit boards. Solectron then ships these completed printed circuit boards (PCB's) to Bay Networks when Bay needs them for a customer order.

Using this manufacturing strategy allows different members of the supply chain to become experts in their particular value added process. For example, Solectron, a dominant assembler of printed circuit boards, will make the necessary investments in capacity to offer state of the art material processing in both through-hole and surface mount technology. Since this is the lifeblood of their business and their competitive advantage, Solectron can do this at a cost that is far less than if Bay manufactured PCBs for itself. One of the reasons that Solectron can do this so effectively is that they have become a supplier for multiple customers. As one customer gains market share, another one typically loses that same share. Thus, Solectron experiences a more stable demand stream and is able to more efficiently utilize capacity. Additionally, Solectron experiences some scale advantages by supplying PCBs to multiple customers.

To reduce delivery time to customers, Bay's strategy has been to stock inventory. Since many of Bay's customers demand a specific product format, such as multiple power supplies or different microprocessor and memory configurations, Bay's production floor operates as a "make to order" line. Inventory is thus carried in a format that is usually order non-specific so that it can be easily configured once a customer order has arrived.

Since Bay would like to minimize the amount of inventory carried within its four walls, it tries to have material from subcontractors arrive just when it is needed. Understanding that lead-times for some electronic components are three to four months, Bay uses a planning group within manufacturing to forecast demand. Traditionally, this forecast has been accurate at the aggregate level, such as the total demand in dollars for the quarter, but less accurate at the line-item (specific PCB) level.

This thesis will address my work at Bay Networks. The project description asked for deliverables that would include methods to improve Manufacturing Business Planning's (MBP) forecast. Management stated that if forecast accuracy was improved, Bay could both decrease inventory and more rapidly fill customer orders. This thesis will describe my work to optimize the use of information within the manufacturing supply chain and describe recommended process changes to improve the whole system. To outline the issues involved in improving Bay's supply chain I will describe the current production planning method in section 1.2. Section 1.3 will outline the problem solving approach and finally, section 1.4 will outline the thesis contents.

1.2. Production Planning Methods

The simplified material flow for Bay Networks is shown below in Figure 1.1. As indicated in the figure, Bay only does the final assembly and test portion of the manufacturing process. Component manufacturers or distributors deliver components to a subcontractor. This subcontractor assembles printed circuit boards using the components and then ships PCBs to Bay Networks as requested. To complete the networking hardware ordered by the customer also requires supply chain support for power supplies, power cables, and metal boxes with user interfaces. However, these items in total only make up about 10% of the total cost of the product and therefore were not studied. Also, these items tend to have replenishment lead-times that are far shorter than those for the PCBs and thus are far less expensive to carry in inventory.

Typical replenishment lead-times for PCBs from the subcontractor are on the order of four weeks. Component lead times are more variable. Some commodity-like components can be restocked in days while other more application-specific components can take up to 3-4 months if demand were to change drastically. Typically the subcontractor will order components from manufacturers under the terms of a specific Bay contract. If demand outpaces original forecasts by an enormous amount then component distributors will be used. If the distributor does stock the right component, they will charge a significant premium above that charged by the manufacturer. These costs are then passed on to Bay.

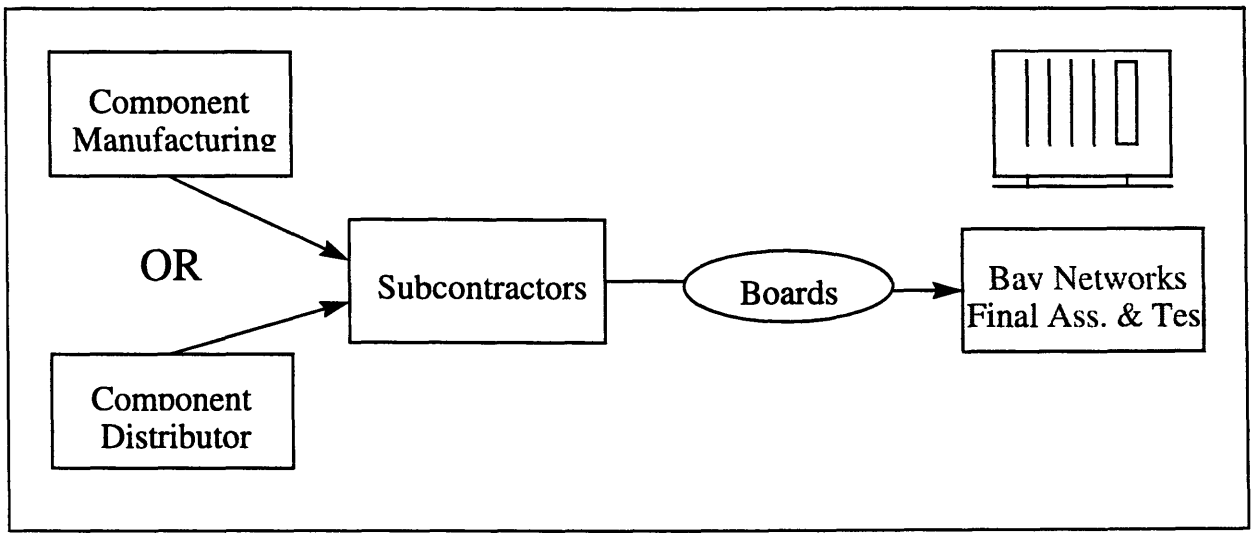

Figure 1-1: Bay Networks' Simplified Supply Chain

To minimize component markup from distributors, in addition to expediting (overtime, extra set-up, etc.) charges from the PCB subcontractor, Bay Networks forecasts demand. The forecasting group aims to use historical demand data, current price-performance characteristics of competitor products, and networking marketplace trends to create a rolling three quarter forecast. This forecast includes a comprehensive one quarter forecast that breaks down demand by individual *line-item*. An example of a *line-item* would be a data processing PCB for a particular small office network hub. Perhaps there are 3 different types of data processing PCBs for this hub product. In addition to the hub box being a forecast *line-item*, each type of PCB would also be a *line-item* since each could be ordered separately.

The current quarter forecast is updated 4 times. The first aggressive look at the quarter is completed about one month before its start. Every 30 days, the forecast is updated using more current information. Thirty days before the end of the quarter, the last forecast is completed along with the first detailed look at the next quarter. MBP attempts to capture the general market trend as opposed to *exact* demand for product families and product line-items in forecast quarters other than the current quarter.

Bay Networks uses an MRP (Materials Requirements Planning) system to plan production. The forecast is a critical element of the MRP system. Master schedulers use the demand forecast for the current quarter and break it up into weekly demand buckets. Using the forecast in conjunction with the product Bill of Materials, the MRP system ensures that the requisite number of supporting parts, in addition to the PCBs, are available for production. Material buyers use the output of the MRP system to make purchases and have material delivered to Bay's shop floor.

Since the relationship between Bay Networks and the PCB subcontractors is dictated in a business contract, Bay has somewhat limited flexibility to change the current quarter forecast. If demand dropped through the floor, Bay could cut its forecast from a revenue perspective, but the subcontractor, who aims to maximize capacity utilization, would still be able to ship some predetermined number of assembled PCBs. On the other hand, if the forecast underestimated demand and Bay had to expedite material, the subcontractor could charge Bay a premium both for components and for extra cost incurred in the PCB assembly. Either way, if the forecast error is excessive, Bay's costs increase. This increase is in the form of increased Cost of Goods sold or in inventory carrying costs. Like many companies, Bay feels that the cost of a lost sale exceeds the cost of inventory. To avoid lost sales, Bay adds between 10 and 20 percent to the quarterly forecast to provide extra capacity in the supply chain and to have extra material on hand to meet customer demand. Higher percents are used for more volatile demand or for longer lead time PCBs. Bay's manufacturing organization feels this extra "flex" as it is called has been crucial in meeting customer demand when the forecast contained large errors. It is also one of the reasons that at times inventory has been excessive. One of the goals of this thesis is to outline how Bay can use less "flex" and still adequately meet customer demand.

1.3. Problem Solving Approach

In determining how much PCB inventory, if any, should be held by Bay Networks one must determine the following: PCB replenishment lead time, customer fulfillment lead time, the variability of demand over the replenishment lead time, and the desired service level (fraction of demand serviced from inventory).

Although not completely deterministic, replenishment and customer lead times can be selected from the stated goals and contracts that bind Bay to its subcontractors. Companies typically will strive to ensure that these goals are indeed met and are therefore quite reliable although not completely deterministic. Although a desired service level can be easily stated, and therefore changed, supply chain process improvement usually focuses on either lead time reduction or demand volatility management through forecasting. Bay's desire in this piece of work was to have its forecast improved so that its current same service level could be achieved with less inventory. Other process improvement teams were working to decrease lead times at PCB supplier sites.

Initial focus included quantifying forecast error variability and designing forecast metrics. Work was then aimed at using forecast variability as a driver in determining adequate safety stock. Final analysis included a redesigned material flow and manufacturing process that would allow Bay to increase its forecast accuracy (because line-items could be aggregated), increase service levels, reduce forecast generation time, and reduce inventory. This panacea would require little if any capital investment or additional manpower.

1.4. Thesis Overview

The most substantial portion of this research was related to the determination of a more efficient process to configure product to customer orders. Desired outcome included improvements in forecast accuracy and reduced inventory. Since Bay makes product to customer order, finished goods inventory was not addressed or analyzed. Instead, the configuration and inventory of assembled PCBs as delivered from Bay's subcontractors was analyzed. This is appropriate given that the PCBs have the longest lead-time and highest percent cost of any product component.

Chapter 2 includes more detailed look at Bay Networks' supply chain and how the Manufacturing Business Planning (MBP) organization develops demand forecasts. This chapter also discusses the output of MBP, MBP's customers, and the assumptions behind the accuracy of the forecast. This chapter shows an example of a forecast for a specific product's motherboard (PCB) and the actual forecast error. Closing out the chapter is a discussion of opportunities for improvement in the forecast through analysis and manufacturing process improvements.

Chapter 3 outlines the mathematical theory behind the concept of Risk-Pooling. This strategy would allow Bay to reduce demand variability and thus inventory through the aggregation of motherboards. Included are examples of two and more than two random demand streams. Additional analysis outlines the increased benefit when negative correlation exists between the demand streams.

Chapter 4 discusses the specific changes required to implement Risk-Pooling at Bay Networks. Also addressed is the resistance that Bay may face both internally and from its subcontractor to such process changes.

Finally, Chapter 5 discusses conclusions and other opportunities.

2. Information Flow within Bay Networks' Supply Chain

Bay Networks uses its manufacturing forecast to reduce the effect of demand variability on both its materials acquisition and production processes. In this capacity the forecast becomes the key information dissemination mechanism within the supply chain. During the generation phase of the forecast key pieces of information are gathered, interpreted, analyzed, and finally synthesized into a demand forecast for the following 3-4 quarters.

Once complete, the forecast is then used to procure materials for production. This is accomplished by two groups: the Master Production Schedulers (MPS), and the materials buyers. The material buyers procure material to meet the quarterly forecast. Typically, they maintain an open purchase order with the supplier and replenish material when in-house stock is insufficient to meet demand over a 2-4 week period. The MPS group directly uses the MBP forecast to run the Materials Requirements Planning (MRP) system. Before data is input into the MRP system they will add between 5-20% of extra "flex" to the forecast as a safety stock.

Since this thesis focuses on minimizing forecast error and reducing Bay inventory, only the MBP and MPS groups, and their corresponding processes, will be addressed in this chapter.

2.1. Demand Planning

Demand planning is the responsibility of the Manufacturing Business Planning (MBP) group. This group consists of six people: a manager, and five planners that develop forecasts for their individual products. The generic process followed by the planners is shown in Figure 2.1.

The first step in the process entails gathering the necessary information about the particular product or line-item within the product line. For example, in a mid range networking product, such as the Advanced Remote Node (ARN) router, there are five different metal boxes that act as supporting frames for the array of power supplies and printed circuit boards that are inserted into the box as purchase options. As for the "guts" of the router, the customer can also choose from four different electronic switching motherboard families. Each family then has five different flavors of motherboards that differ primarily by the amount of memory that the board has. Both the boxes and motherboards mentioned above have demand histories and market information which requires gathering.

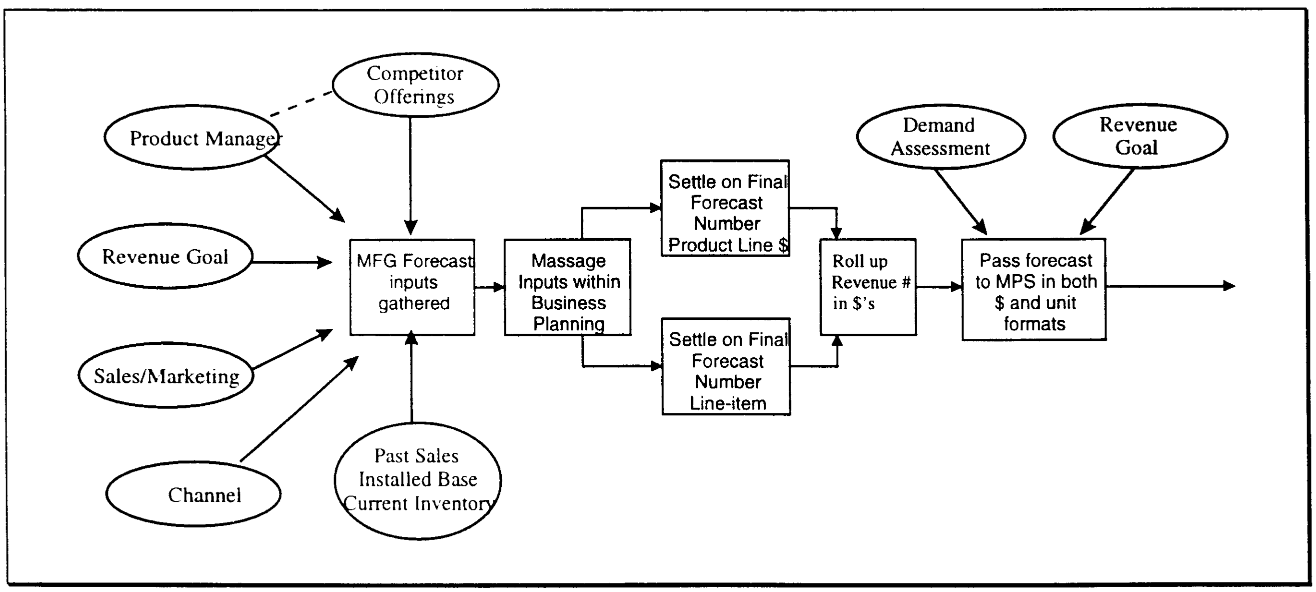

Figure 2-1: Manufacturing Business Planning's Simplified Process

Key information inputs include: product positioning and business plan information from the product manager; the product line corporate revenue goal; sales and marketing incentives or any leads on possible big orders; distribution channel promotions and upcoming orders; current inventory position of the channel and Bay; and finally, historical demand data coupled with the customer installed base information. Using this information, the forecaster in MBP would determine the individual line-item forecasts for the ARN product line and simultaneously roll these numbers up into a gross dollar forecast using average sales price data.

Before this information is passed to the materials planners in the MPS group, senior management in manufacturing would ensure that the total forecast for the quarter adds up to the corporate revenue goal as determined at the monthly demand assessment meeting. The demand assessment meeting is a collaborative forecasting meeting whereby key representatives from finance, manufacturing, product management, and sales determine which products will be pushed by the sales force. Top management also decides on a realistic revenue target and product mix. Usually, adjustments are made to the forecast after the demand assessment meeting. This demand assessment also serves to ensure that all functional groups within the company are in agreement about corporate goals.

While analyzing the forecasting process at both the product planner level and at the Demand Assessment level, we discovered that some simple exponential smoothing is used in the forecasting process. However, it was used far less frequently than the judgmental forecasting method. Product planners within MBP and the higher management represented at the Demand Assessment meeting claimed that the volatility of the marketplace warranted the use of judgmental forecasts.

Research into the use of judgmental forecasting revealed that the method is most effective "in a world characterized by change" (Griggs, 1995) and (Sheerin, 1997). Specifically, "human judgment could be superior to the forecasting models in recognizing changes in the pattern of the data or if it is able to better integrate outside information about the change into the forecasting process" (Griggs, 1995). This is especially true for Bay Networks given that many product life cycles are less than two years and that life cycles are frequently interrupted by new product introductions by both Bay and by its competitors. As part of the work in this thesis, an analysis of 15 randomly selected line-items revealed that Bay's current forecasting process outperformed simple exponential smoothing forecasting methods for 10 of the 15 line-items. Although the statistical analysis could be more rigorous, the sample results certainly do indeed back the claim by Bay's management that judgmental forecasts are effective for Bay's products.

The Advanced Remote Node (ARN) router is an example of product that is forecast by MBP. The ARN router product is called a *product line* and will typically be forecast at the aggregate dollar level. Product lines are forecast at the dollar level and represent the sum of the individual product components that can be optioned by a customer within the product line. This dollar-level forecast informs both production line management and other manufacturing management of product trends (either up or down) in the market and therefore how the production line should be staffed.

To be useful from a material management perspective, the forecast for the ARN product family must be further broken down into its product components. This entails forecasting the demand for each *line-item* in the ARN product family. A line-item would represent one of the individual box or motherboard flavors mentioned above. Depending on market conditions and trends, the mix of demand for the line-items in the ARN family could be vastly different from quarter to quarter, while the top line aggregate dollar forecast remains unchanged. Since each line-item needs to be stocked and ordered from subcontractors separately, each must also be forecast separately.

2.1.1. Output of MBP

An example of the forecast as generated by MBP is presented below in Figure 2-2. Numbers in Figure 2-2 are disguised but representative of actual numbers.

Line-item	Q297 actual	Q397 actual	Q497 actual	Q198 actual	*Q298 Forecast*	*Q298 flex*	*Q298 total*
Enet MTR 0 Meg	324	451	809	1093	730	0	730
Enet MTR 4 Meg	146	277	378	413	500	0	500
Enet MTR 8 Meg	519	1625	2006	5127	6000	0	6000
Enet MTR 16 Meg	577	1003	1342	1580	1800	0	1800
Enet MTR 32 Meg	31	62	99	86	60	0	60

Figure 2-2: Sample Manufacturing Business Planning Forecast

The table is formatted as follows:

- The first column lists the line-item name. In this case each line-item is an Ethernet (data switching method) motherboard with memory (from 0 to 32 megabytes).
- The next four columns are the actual shipment data for last four business quarters.
- The sixth, seventh, and eighth column include the demand forecast for the current quarter, the 'flex', which is extra material added by MBP for potential big orders, and the total.

2.1.2. Customers of MBP

The MPS group is the primary customer of the MBP forecast. This forecast format example, which will exist for about 200 different line-items, will be generated by MBP and electronically handed off to the materials planners in the MPS group. The materials planners will then add additional materials 'flex' to the total demand forecast as a safety stock for the quarter. Finally, they will break the demand into monthly and weekly buckets and use the data to execute the MRP program.

2.1.3. Time-Frame of Forecast

MBP generates a new, comprehensive forecast about once a month. This update will entail both a bottoms-up analysis of the customer order data that is new since the last forecast and a tops-down analysis performed at the monthly demand assessment meeting.

The most detailed analysis will be focused on the current business quarter. From a planning perspective a quarter is considered current from about 4 weeks before it starts until about 3-4 weeks before it ends. This is the time frame that MBP feels it can most accurately predict what will happen. It is also when the supply chain is most likely to be capable of responding without charging Bay excessive expedite and overtime fees. Once the time frame has moved to within 3-4 weeks of the end of the quarter, MBP is already focusing on the next quarter and the supply chain will not accommodate significant forecast changes. Quarters beyond the current quarter are forecast to capture only general trends and are not forecast as rigorously.

If customer order data shows that demand differs significantly for a particular line-item from the original forecast MBP will update this line-item forecast and pass the information to the materials planners. This update will occur outside of the more comprehensive forecast. Thus, line-item forecasts can be managed on an exception basis between comprehensive forecast updates. Both demand planners and materials planners will track their respective line-item order rates and update the forecast as necessary.

2.2. Materials Planning

Materials planning is the responsibility of the master schedulers in the Master Production Scheduling (MPS) group. This group has 5 schedulers that utilize MBP's forecast to ensure that adequate material is on hand to meet customer orders. Unlike the forecasters in the MBP group, whose sole responsibility is to model and forecast demand as accurately as possible, the master schedulers must ensure that adequate material flows into Bay Networks' shop floor to fill customer orders. This is true whether the forecast was accurate or not; therefore, it is the master schedulers that determine safety stock (flex) requirements. The simplified process for the MPS group is shown in Figure 2-3.

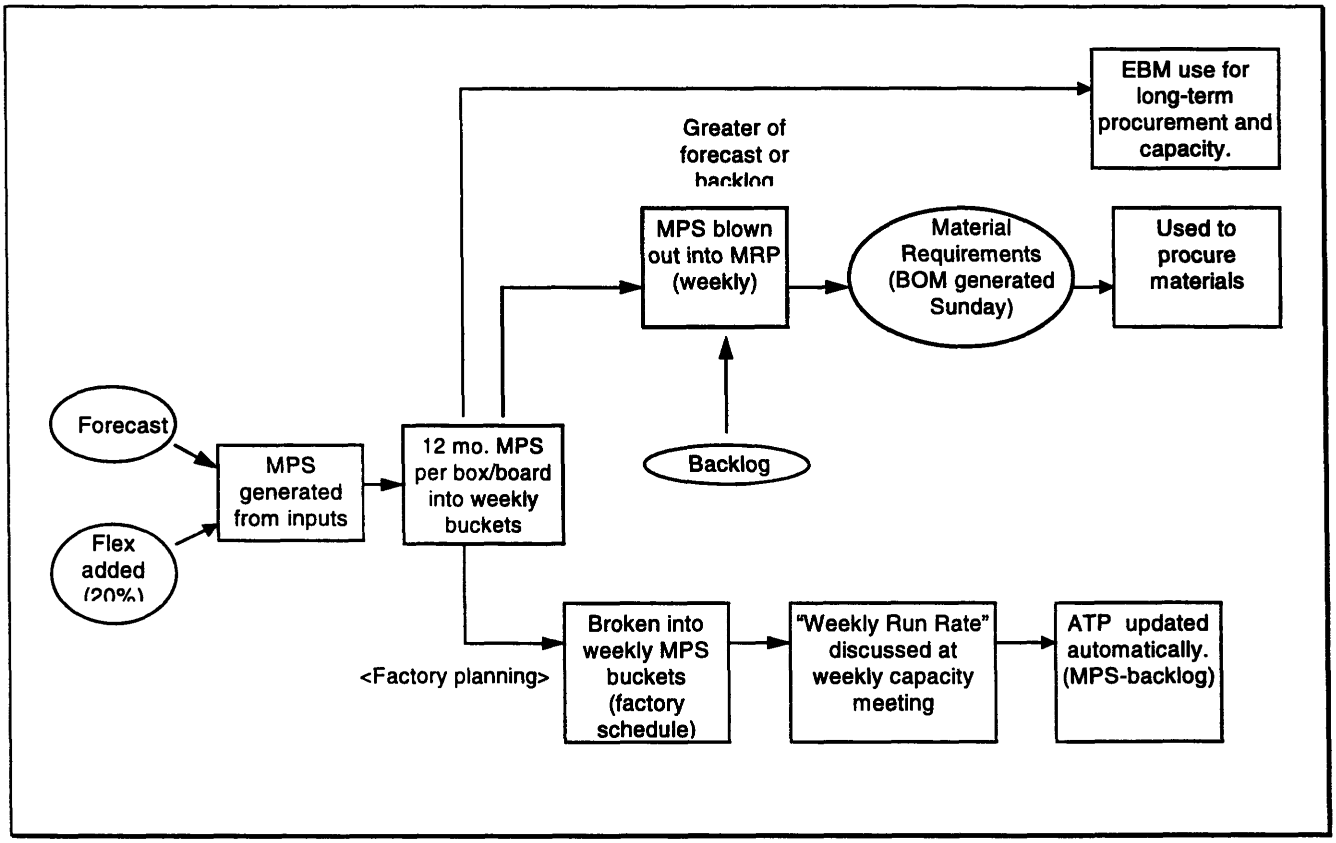

Figure 2-3: Master Production Schedule (MPS) Flow Chart

The first process step for a master scheduler is to obtain the most recent MBP forecast and to add the appropriate amount of flex onto the gross forecast number. This step applies to all line-items in the forecast. A master scheduler will choose a flex percentage based on 'gut-feel' and a combination of other factors. These factors include the lead time of circuit boards from the suppliers, the accuracy of past forecasts, the amount of material on hand from previous quarters, and the expected demand growth rate. In many cases this percentage is set somewhat arbitrarily. The literature (e.g., Krupp, 1997) demonstrates that the flex percentage is really a safety stock and is a function of the factors that the master scheduler is already considering plus the carrying cost of inventory and the cost of a lost sale.

The master scheduler will then break the forecast into weekly buckets for the current quarter. More distant quarters are left in monthly and quarterly buckets. This process is standard for Solectron customers. From this point forward the master production schedule has three main uses.

The first is indicated by the topmost branch in Figure 2-3. This branch represents the long term forecast in monthly and quarterly buckets being used by External Business Managers (EBM) to secure commitments for capacity and components from suppliers. This is only a rough forecast but is sufficient for order of magnitude planning within the supply chain.

The next branch down represents the master scheduler using the current quarter forecast to determine material requirements for the next 16 weeks. In this capacity a master scheduler will determine the greater of the backlog or the forecast and push the resulting larger number through MRP to determine total material requirements. This task is accomplished each week with the result being distributed to the purchasing group to make purchases from suppliers.

The lowest branch represents the demand in weekly buckets being used by the production control group to schedule production within the factory. Usually the schedule is set up to minimize overtime by level loading the quarterly demand within the factory. When demand for a particular week exceeds production capacity, the production control group will pre-build some assemblies in weeks with less than full demand. If this is not possible, overtime and/or temporary work will be hired. The 'weekly run rate' determines the weekly production required to meet the total demand for the remainder of the quarter. If total production exceeds capacity, additional shifts will be scheduled. ATP stands for 'available to promise' and is the difference between what can be built in the rest of the quarter and what has already been scheduled. Basically, ATP represents the amount of product that the sales group can promise to be delivered during the current quarter.

2.2.1. Assumptions about Forecast Accuracy and Flex

As mentioned previously, one of the considerations in determining the amount of flex to push through the MRP system is the forecast accuracy. Prior to this thesis study forecast accuracy was measured for each line-item by determining whether the forecast was within 15% of the actual shipments for the quarter. If the forecast was within plus or minus 15%, the line-item was determined a 'hit' – a 'miss' was any line-item outside of the band. Unfortunately, the MPS group did not look for trends in the accuracy of the forecast from quarter to quarter. Thus, no emphasis was placed on determining exactly which line-items and which product lines were consistently forecast well and required less flex, and which ones were likely to be more volatile and therefore required more flex.

Overall, the MPS group accurately assumed about 30% of the 200 total line-items would fall within this accuracy band. Based on their experience, master schedulers would increase the flex for line-items which they thought had been volatile or had caused lead time problems with customers in the past. Since the process used a 'gut-feel' approach, these selected line-items were not always the worst offenders. As stated by one of the master schedulers, "I will never get fired for too much inventory, maybe yelled at, but not fired. Now if I let customer lead times go through the roof, that is a different story - I could lose my job for that."

2.2.2. Assumptions about Lead-time and Inventory Exposure

In the Bay Networks' supply chain for printed circuit boards, the MPS group assumes that the replenishment lead-time for large changes in the forecast is about 4 weeks. This number is based on experience and by the contract between Bay Networks and their primary subcontractor, Solectron. The contract states that Bay cannot cut its forecast inside of the next 4 weeks without still taking the material at the end of the quarter. Bay must also pay extra fees for additional material (beyond the forecast) that is requested inside of 4 weeks, although this premium is seldom charged as Solectron does not want to discourage Bay from increasing its order. Once the forecast is changed and run through MRP by the MPS group, it will take about 4 weeks for Solectron to deliver the PCBs to Bay Networks under normal conditions. This somewhat long time results from the fact that Solectron tries to maximize its capacity utilization by supplying to multiple PCB customers and therefore must schedule/squeeze Bay's order into the production schedule.

The MPS group also expects that any PCBs not purchased from Solectron and not cut from the forecast in the last 4 weeks of the quarter will still be delivered at the end of the quarter. These PCBs will be held as inventory by Bay Networks until they are consumed in the next quarter. If one assumes that a forecast for a particular line-item is 100% accurate, then the flex that is added by the MPS group will not be consumed and will be the inventory that Bay has on hand at the end of the quarter. This flex can thus be considered to be the safety stock that Bay uses to hedge against abnormal demand surges. If the demand forecast does have some error, then the inventory held by Bay Networks at the end of the quarter will either be less than or more than the flex amount. Recall from section 2.2 that this flex is set somewhat arbitrarily and that the literature (Krupp, 1997) has demonstrated that flex is really a safety stock which can be calculated using standard inventory management models (e.g., Nahmias 1993).

This cycle will repeat each quarter with the MPS group using the forecast and an appropriate flex number. The purchasing department will once again use the MRP output to compare required material to on-hand material and order replenishment as necessary. At the end of the quarter, the difference between the forecast + flex number and the actual shipments will once again remain as inventory.

2.3. Opportunities for a Win-Win: Motherboard Aggregation

Based on Bay Networks' process for forecasting, procuring, and stocking of motherboards such as those for the ARN product family, many employees felt that Bay could realize significant benefits from forecasting only generic motherboards. This would entail forecasting and stocking total motherboard demand for the 0 Megabyte Ethernet motherboard only. Individual memory flavors would be configured as customer orders were received. Total time to configure a board would be a few minutes at most as it was moving down the assembly line.

2.3.1. Likely Forecast Accuracy Improvements

Forecasters within the MBP group stated that they felt they had a much better feel for *total* demand of particular motherboards families (ARN Ethernet, ARN Fiber, ARN Asynchronous Transfer Mode) as compared to the demand for *each memory flavor* within a motherboard family. Since forecasters first calculate total demand for a particular motherboard family (ARN Ethernet) and then break out the demand for the different memory flavors, forecasters felt they were introducing additional error into the process that is both unnecessary and significant. One forecaster claimed that while all five memory flavors of the ARN Ethernet motherboard could fall outside the +/- 15% accuracy band, the total aggregated demand would probably fall within the accuracy band. Chapter 3 will address the mathematical theory behind the expected effect of aggregation on demand variability.

2.3.2. Fewer Memory Conversions

Often, near the end of a business quarter, Bay Networks would have insufficient ARN Ethernet motherboards of a particular memory configuration but excess of another. When this happened, Bay's production staff 'converted' motherboards by removing memory from a motherboard and replacing it with the desired amount of memory. As part of this process the electronic inventory system is updated with each converted motherboard's new bar code identification.

Unfortunately, these conversions occur at the busiest time of the quarter and are completed primarily by temporary workers. Often, the bar coding and system updating is performed incorrectly and later requires reconciliation. Also, since the process is usually performed only two weeks of the quarter, there is a steep and inefficient learning curve each time. In one quarter, 22% of all motherboards shipped required conversion. This also led to over 300 motherboard reconciliation's for just the ARN product family alone.

Without conversions, Bay Networks could reduce the number of temporary workers hired at quarter end, reduce inventory reconciliation, and efficiently integrate motherboard configuration into its process - possibly with no additional full-time personnel.

The next chapter outlines the theory behind risk-pooling multiple demand streams. Chapter 4 details the specifics of applying this concept to Bay Networks and the expected benefits.

3. Risk Pooling: Higher Service Levels with less Inventory

In most circumstances when the decision is made about how much inventory to carry, companies usually must trade off the increased expense of more inventory, in the form of carrying costs, against the company's ability to service more customer demand, in a shorter period of time, from the increased inventory. While many models exist to analyze this tradeoff, the real difficulty is in trying to accurately quantify the cost of inventory or in trying to determine just what a stock-out will cost the company in the long term. One undeniable truth is that if a company could increase its service level while reducing its inventory, without actually needing to quantify carrying and stock-out costs, it would represent a "no brainer" win - win situation. This panacea exists in the properly applied concept of risk pooling. For Bay Networks, the concept of risk-pooling can be leveraged only when the configuration of motherboards with memory is postponed until an actual customer order is received. More detail on the theory of postponement can be found in the literature (Lee & Billington, 1993).

This chapter will mathematically demonstrate how Bay Networks can become more responsive to customers while actually reducing both the unit and dollar value of material it holds. This is accomplished by stocking a more generic 0 Megabyte Memory Motherboard in lieu of the 0, 4, 8, 16, & 32 Megabyte Motherboards that are currently stocked. To service a customer order, Bay manufacturing employees will configure each motherboard with the desired memory chip as they assemble the rest of the product instead of taking a subcontractor pre-configured board from the inventory stock pile.

Since Bay Networks already stocks memory chips, this process change adds no additional stocking costs for memory. This will however allow the more expensive motherboards, at approximately $1000.00 per board, to be stocked in aggregation. This aggregation will serve to pool the demand variability as seen by the stockpile. Since the demand streams for the memory flavored motherboards were not perfectly positively correlated before aggregation, adding their demand streams together actually reduces the variability as a percentage of the total demand. This reduction in variability, or in the standard deviation of demand, allows Bay to carry less inventory while simultaneously increasing service levels.

Section 3.1 will outline the math for the cases of two independent random variables, two dependent random variables, and finally, for the more generic case of many random variables. Section 3.2 will then show how Bay Networks can extract value from the concept of risk pooling.

3.1. Probability Theory of Risk-Pooling

The concept of risk-pooling is most easily understood by analyzing the result of adding two separate demand streams and calculating the resulting mean, μ_T, and standard deviation, σ_T, of the total (T) demand. The component demands of the total demand are modeled as random variables. Each random demand stream has a mean and a variance. Variables for demand stream "X" are (μ_X, σ_X), and for demand stream "Y" are (μ_Y, σ_Y).

Since T represents the total demand for the two demand streams of the Printed Circuit Board family, then T is also a random variable equal to the sum of two other random variables, X and Y. The expected value of the sum of a set of random variables is always equal to the sum of the expected values (Drake, 1988). Thus, the expected value for the sum of demand streams X and Y, regardless of the relationship between X and Y is:

$$E(T) = \mu_T = \mu_X + \mu_Y$$

However, the variance of the sum of two random variables is not the sum of the individual variances. The variance of a random variable T is defined as:

$$Var\ (T) = E\ [(T - \mu_T)^2]$$

Since T= X+Y, and $\mu_T = \mu_X + \mu_Y$, we can substitute and get:

$$Var(T) = E[(X + Y - \mu_X - \mu_Y)^2]$$

Rearrangement and expansion results in:

$$Var(T) = E[((X - \mu_X) + (Y - \mu_Y))^2]$$

$$Var(T) = E[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)]$$

As mentioned previously, the expected value of a sum is equal to the sum of the expected values, therefore:

$$Var(T) = E[(X - \mu_X)^2] + E[(Y - \mu_Y)^2] + E[2(X - \mu_X)(Y - \mu_Y)]$$

The first two expected values are by definition the variances for the individual random variables, X and Y. The final term is defined as the covariance of X with Y and is denoted by the symbol, σ_{XY}. The final result for the variance of T is:

$$Var(T) = \sigma^2_T = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY}$$

This result verifies that the variance for the sum of two random variables is equal to the sum of the variances for the two random variables plus two times an additional term, which is either positive or negative, that will either increase or decrease the total variance beyond the simple sum. The covariance of two random variables reveals, in simple terms, the relationship or relative magnitude of explanatory knowledge gained about one variable from knowing about the other variable. For example, if the value of one variable, X, were to double, knowing the variance of each X and Y, and the covariance between X and Y, would allow one to compute how the value of Y is expected to change.

3.1.1. Special Case of Two Independent Random Variables

Further algebraic manipulation (Drake, 1988) reveals that the covariance σ_{XY} can be expressed as

$$\sigma_{XY} = E(XY) - E(X)E(Y)$$

If X and Y are *independent* random variables, our knowledge of the expectation for the sum of random variables reveals that E(XY) = E(X) E(Y), and that the covariance, $\sigma_{XY} = 0$. If the covariance is 0, then the variance of the sum of two independent variables, or demand streams in this case, is just the sum of the variances. From the perspective of an inventory stocking strategy this means that less inventory is necessary to maintain the same service level (Lee & Billington, 1993). Why is this so? Using a simple safety stocking strategy where the safety stock is proportional to the standard deviation of demand over the replenishment lead-time (Nahmias, 1993), independence between the random demand streams results in the standard deviation of the sum of independent random variables always being less than the sum of the standard deviations. Thus, the total demand variability is less than the sum of the individual demand stream variability and less safety stock is required to protect the company over the lead-time.

The next section outlines how the demand does not necessarily need to be independent for this result to hold.

3.1.2. Special Case of Two Dependent Random Variables

Continuing with the conclusions of the last section, it can be shown (Drake, 1988) that the expression for the covariance between two variables, X and Y, can be represented by

$$\sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y$$

where ρ_{XY} is the correlation coefficient between variables X and Y. The correlation coefficient is a term that describes the degree to which variables X and Y move together. For example, if one were to determine that over a particular period of time the value of X doubled, and that the correlation coefficient was also known, it would be trivial to determine the amount of expected change in the value of variable Y. A simpler method to describe the correlation coefficient is a mathematical definition: an estimate of the correlation coefficient is the slope of the line which best fits (by minimizing the mean square error) a plot of normalized variables X* and Y*. This is only an estimate since we are calculating the correlation coefficient of our population samples and not that of the entire population. The normalized variables are described by

$$X^* = (X-E(X))/\sigma_X \qquad Y^* = (Y-E(Y))/\sigma_Y$$

Physically, ρ_{XY} can be interpreted in the following method: ρ_{XY}^2 is the fraction of the mean square error that disappears if we know the value of variable X before we must guess the value of variable Y. Looking back on section 3.1.1, where X and Y were independent, one can see that knowing the value of X tells you absolutely nothing about the value or Y, therefore $\rho_{XY} = 0$. On the other hand, if X and Y are perfectly correlated, then the knowledge of one variable will tell you the value, without doubt, of the other variable. In this case $\rho_{XY} = 1$. At the other extreme, if X and Y are perfectly negatively correlated, $\rho_{XY} = -1$, knowing the value of one will also tell you the value of the other, but X and Y will move in opposite directions. Thus, the value of ρ_{XY} must fall between -1 and +1, which is the range of correlation of X and Y as they go from being perfectly negatively correlated, to independent at $\rho_{XY} = 0$, to perfectly correlated at +1.

Making the appropriate substitutions it can be shown that:

$$\sigma_T^2 = \sigma^2_X + \sigma^2_Y + 2\rho_{XY}\sigma_X\sigma_Y$$

which is the expression for the variance of the sum random variables, X and Y. Given that X and Y have some degree of variability, quantified by σ_X and σ_Y, the value of the total variance will in large part be determined by the degree to which there is correlation between X and Y. In the next two sections, the magnitude of σ_T will be discussed.

3.1.2.1. Positively Correlated Variables

In the case when the variables X and Y are perfectly correlated, the variance of the sum is equal to the following:

$$\sigma_T^2 = \sigma_X^2 + \sigma_Y^2 + 2(1)\sigma_X\sigma_Y$$

And the standard deviation of the sum can be calculated as follows:

$$\sigma_T^2 = (\sigma_X + \sigma_Y)^2$$

$$\sigma_T = \sigma_X + \sigma_Y$$

which states that the maximum value for the standard deviation of the sum of two random variables is the sum of their standard deviations. In this case, aggregating demand streams from these two different products within a product family would not facilitate safety stock reduction since the standard deviation of the total demand is just the sum of the individual standard deviations. Keep in mind that we are assuming that the safety stock is proportional to the standard deviation of the sum.

Any case where there was some positive correlation between X and Y but not perfect correlation would yield a σ_T that is some value less than the maximum. In this case there is a positive effect of pooling the variability of the random demand streams. Thus we can see that in all cases where $\rho_{XY} < 1$, the standard deviation of the sum will be less than the sum of the standard deviations - and there will be a benefit to pooling the demand variability from individual line-items within a product family into a single demand stream.

3.1.2.2. Negatively Correlated Variables

At times, random variables, or demand streams for various products, will be negatively correlated. This could occur when a consumer would select either one product or the other, but not both. Under these circumstances, the demand for one product will likely increase as the demand for the other decreases. For the case when this relationship occurs in lockstep, there is perfect negative correlation and $\rho_{XY} = -1$. In this case the equation for the variance of the sum

$$\sigma_T^2 = \sigma^2_X + \sigma^2_Y + 2\rho_{XY}\sigma_X\sigma_Y$$

becomes

$$\sigma_T = \sigma_X - \sigma_Y$$

which is the smallest value that the total standard deviation and thus the safety stock can assume. One can conclude that the variance and standard deviation of the sum are at a maximum when there is perfect correlation between X and Y and that these parameters decrease as the value of the correlation coefficient decreases from 1, through the total spectrum, to -1.

3.1.3. The Variance of the Sum of Random Variables

In this section we will expand our expression for the variance of the sum of two random variables to an expression for many random variables. It can be shown (Drake, 1988) that the variance of the sum of many random variables is just the sum of the variances of the random variables plus the sum of two times the covariance of each variable with all the others. The result for the sum of four random variables is shown in the table below, where all the variances are listed on the diagonal from upper left to lower right, and all the covariances are listed in all the other cells. The total variance is given by summing all the terms listed in the table. Note, each of the covariance terms appears twice.

Variable	W	X	Y	Z
W	σ^2_w	σ_{wx}	σ_{wy}	σ_{wz}
X	σ_{wx}	σ^2_x	σ_{xy}	σ_{xz}
Y	σ_{wy}	σ_{xy}	σ^2_y	σ^2_{yz}
Z	σ_{wz}	σ_{xz}	σ_{yz}	σ^2_z

A table of the above format can be expanded in a similar format to as many variables as necessary to describe the sum. Adding all the terms in the table results in:

$$\sigma_T^2 = (\sigma_W^2 + \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2) + 2(\sigma_{WX} + \sigma_{WY} + \sigma_{WZ} + \sigma_{XY} + \sigma_{XZ} + \sigma_{YZ})$$

and upon substitution for the covariance term yields:

$$\sigma_T^2 = (\sigma_W^2 + \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2) + 2(\rho_{WX}\sigma_W\sigma_X + \rho_{WY}\sigma_W\sigma_Y + \rho_{WZ}\sigma_W\sigma_Z + \rho_{XY}\sigma_X\sigma_Y + \rho_{XZ}\sigma_X\sigma_Z + \rho_{YZ}\sigma_Y\sigma_Z)$$

If all of the correlation coefficients are equal to 1, meaning that each variable is perfectly correlated and the variance of the sum is at its maximum value, then the equation above will reduce to

$$\sigma_T^2 = (\sigma_W + \sigma_X + \sigma_Y + \sigma_Z)^2$$

or

$$\sigma_T = (\sigma_W + \sigma_X + \sigma_Y + \sigma_Z)$$

which is the exact result we obtained for the case of two variables where the standard deviation of the sum is the sum of the standard deviations. This result states once again that the standard deviation of the sum of many random variables is at most the sum of the standard deviations. Also, any scenario with less than perfect correlation will yield a smaller standard deviation for the sum. This concept is the theoretical driver for the concept of "risk-pooling." Risk-pooling of multiple demand streams decreases the total demand variability and thus the total required safety stock.

3.2. Extracting value from the concept of Risk-Pooling

Inventory and safety stocks are routinely used to buffer against yield uncertainty in manufacturing systems or to buffer against demand variability over replenishment lead-times. Bay Networks uses safety stock to protect the company from unexpected demand fluctuations, especially those that increase demand, during the time it takes to order and receive material to fill the increased demand. Since the bulk of the long lead time material is printed circuit boards that differ only my memory configuration, Bay Networks could theoretically stock motherboards in a pooled generic format and effectively reduce the demand uncertainty.

Assume that the demand for each of four (excluding the 0 megabyte version) different memory flavored motherboards can be modeled as a standard normal probability distribution, with a mean demand per unit time, μ, and a standard deviation per unit time,

σ. The demand will also be assumed to be independent from time period to time period (in reality there is usually time correlation in demand for an industrial or consumer good, but this does not affect the general nature of the result), and must be filled from inventory over the replenishment lead-time, t. Given this set of parameters, the expected demand for a specific memory flavored motherboard over the time period is μt, and the variance of demand over the same lead time is $\sigma^2 t$ (the variance of the total demand over time, t, is the sum of the individual demand periods' variances). Establishing an initial inventory for this motherboard requires enough inventory for the expected demand and some additional amount, that is some multiple of the standard deviation of demand per unit time. This multiple, K, depends on the probability that we require of the safety stock to meet demand over the replenishment lead-time. Typically, the service level probability is chosen and then K derived from the inverse of the normal probability distribution function based on the chosen probability. For an initial inventory level of

$$I = \mu t + K\,(\sigma t^{1/2})$$

where K is equal to 1.65, the standard normal probability distribution says that there is a 95% chance that total demand will be less than our predetermined inventory level.

If four different memory flavors exist for a motherboard, then the inventory for each flavor can be determined. Consider the case where the different flavors are 4, 8, 16, or 32 megabytes of memory. Demand for each flavor has its own set of normal probability distribution parameters, μ, and σ. For example, for each motherboard flavor, the following set of variables apply:

μ_T = the mean demand for the total set of motherboards per unit time

σ_T = the standard deviation for the total set of motherboards per unit time

μ_i = the mean demand for motherboard flavor, i, per unit time

σ_i = the standard deviation for motherboard flavor, i, per unit time

To ensure that the probability of meeting demand from inventory over the replenishment lead time is .95, inventory levels for each motherboard flavor must be set according to the following equations:

$$I_4 = \mu_4 t + 1.65\ (\sigma_4 t^{1/2})$$

$$I_8 = \mu_8 t + 1.65\ (\sigma_8 t^{1/2})$$

$$I_{16} = \mu_{16} t + 1.65\ (\sigma_{16} t^{1/2})$$

$$I_{32} = \mu_{32} t + 1.65\ (\sigma_{32} t^{1/2})$$

where the inventory levels are different for each motherboard flavor, as determined by the respective mean and standard deviation of demand for each flavor. Total inventory held in memory specific format without risk-pooling will equal:

$$I_T = I_4 + I_8 + I_{16} + I_{32}$$

$$I_T = (\mu_4 + \mu_8 + \mu_{16} + \mu_{32})t + 1.65(\sigma_4 + \sigma_8 + \sigma_{16} + \sigma_{32})t^{1/2}$$

If the motherboards could be held in a format which easily allowed them to service all demand streams, 4 to 32 megabytes, then only one type of motherboard would be held in inventory. When an order is received, the motherboard would be configured to meet the specific characteristics ordered by the customer. The inventory to provide service in this environment of risk-pooling is calculated as follows

$$I_T = \mu_T t + 1.65\sigma_T t^{1/2}$$

where the values for the total mean demand and total standard deviation are calculated as in Section 3.1.

$$\mu_T = \mu_4 + \mu_8 + \mu_{16} + \mu_{32}\ ,$$

$$\sigma^2_T = (\sigma^2_4 + \sigma^2_8 + \sigma^2_{16} + \sigma^2_{32}) + 2(\sigma_{4,8} + \sigma_{4,16} + \sigma_{4,32} + \sigma_{8,16} + \sigma_{8,32} + \sigma_{16,32})$$

which yields

$$I_T = (\mu_4 + \mu_8 + \mu_{16} + \mu_{32})t + 1.65\sigma_T t^{1/2}$$

So the only difference in the required level of inventory between the two stocking strategies resides in the terms associated with the standard deviation of demand. Specifically, if the standard deviation of the risk-pooled demand for the entire set of motherboards, σ_T, is less than the sum of the standard deviations for the individual memory flavors, $\sigma_4 + \sigma_8 + \sigma_{16} + \sigma_{32}$, then the total required inventory will be less. We know from Section 3.1 that in all cases except those in which the different memory flavors are perfectly correlated, that the following equation will hold true.

$$\sigma_T < \sigma_4 + \sigma_8 + \sigma_{16} + \sigma_{32}$$

When this occurs, then

$$(\mu_4 + \mu_8 + \mu_{16} + \mu_{32})t + 1.65\sigma_T t^{1/2} < (\mu_4 + \mu_8 + \mu_{16} + \mu_{32})t + 1.65(\sigma_4 + \sigma_8 + \sigma_{16} + \sigma_{32})t^{1/2}$$

and thus

I_T (with risk-pooling) $< I_T$ (with no risk-pooling)

This section has shown, in conjunction with Section 3.1, that motherboards which are not perfectly correlated require less inventory when they can be pooled to service one demand stream. It will now be shown that this decrease in inventory will occur with a simultaneous increase in overall service level.

For the case where the inventory is pooled, the K factor of 1.65 will provide a service level of .95 to all the products pooled together regardless of the correlation between memory flavors. This is not the case for the larger un-pooled inventory that is specific to each memory flavored motherboard.

With a multiplier factor, K, of 1.65, the service level for an individual memory flavored motherboard is .95. However, this is the service level for the *entire* set of motherboards only in the very unique case in which the demands are perfectly correlated (which is very unlikely). When the demands are perfectly correlated then the different flavors all stock-out or have sufficient inventory. Thus the service level for all flavors, both individually and in total, is .95.

Assume for demonstration purposes that the demands for each individual memory flavor are independent of one another. The probability for no stock-outs can easily be calculated. Note, total demand is satisfied when the demand for each memory flavor is satisfied. The probability for this to occur for each flavor is .95. With four independent memory flavors, the probability that no stock-out will occur is $.95*.95*.95*.95 = .81$. This is significantly less than the pooled value of .95 which is serviced with less inventory!

If the demands were perfectly correlated, then all or none of the inventories of flavored motherboards would either stock-out or be adequate to meet demand. In this case the service level would be the same, .95, for both the pooled and un-pooled inventories. In this special case the probability of satisfying all product line demands is equal to the probability of satisfying each product line demand. Thus, the un-pooled inventory would equally satisfy all demand with probability of .95. Note that this is a special case and that any other set of circumstances where the correlation is less than perfectly positive, there will be benefits to risk-pooling.

4. Risk Pooling within Bay Networks' Supply Chain

As seen in chapter 3, the concept of risk-pooling, when applied to demand streams that are anything but perfectly positively correlated, will decrease the total demand variability and the necessary safety stock to service the total demand. This chapter will outline the current process for memory configured motherboards, an improved process to deliver to the customer these same motherboards, and finally, some obstacles to implementation of the improved process.

4.1. The Current Process for Memory Configured Motherboards

Bay Networks' current process to bring in memory configured motherboards is shown below in Figure 4-1. One can see that Bay Networks receives PCBs already configured with memory according to the forecast that Manufacturing Business Planning generated and that the master schedulers fed into the MRP system. Since there are five different memory flavors for the ARN Ethernet motherboard, each must be safety stocked individually to satisfy its particular demand stream. Assuming the lead-times shown in Figure 4-1 of 4 weeks for the subcontractor and 1 week for flow time through Bay's plant, the required safety stock for each flavor is also shown.

These safety stocks are calculated using the standard safety stock formula, $SS = K\sigma\sqrt{t}$ (Nahmias, 1993), where K is the multiplier factor for the service level, σ is the standard deviation of demand per unit time, and t is the lead-time under consideration. This formula assumes that the demand per unit time is independent from time period to time period and that the service level represents the probability of a stock-out during a cycle. For these calculated safety stocks the K was 1.05 for a service level of .83. This value was chosen because it represented the approximate service level that Bay currently provided from inventory. While this number may seem low, Bay Networks could actually service an increase in demand by pulling material from suppliers more readily than the four weeks indicates and thus could hold less material. The standard deviation of demand was calculated for each memory flavor independently and thus the main driver in the difference between the required safety stock.

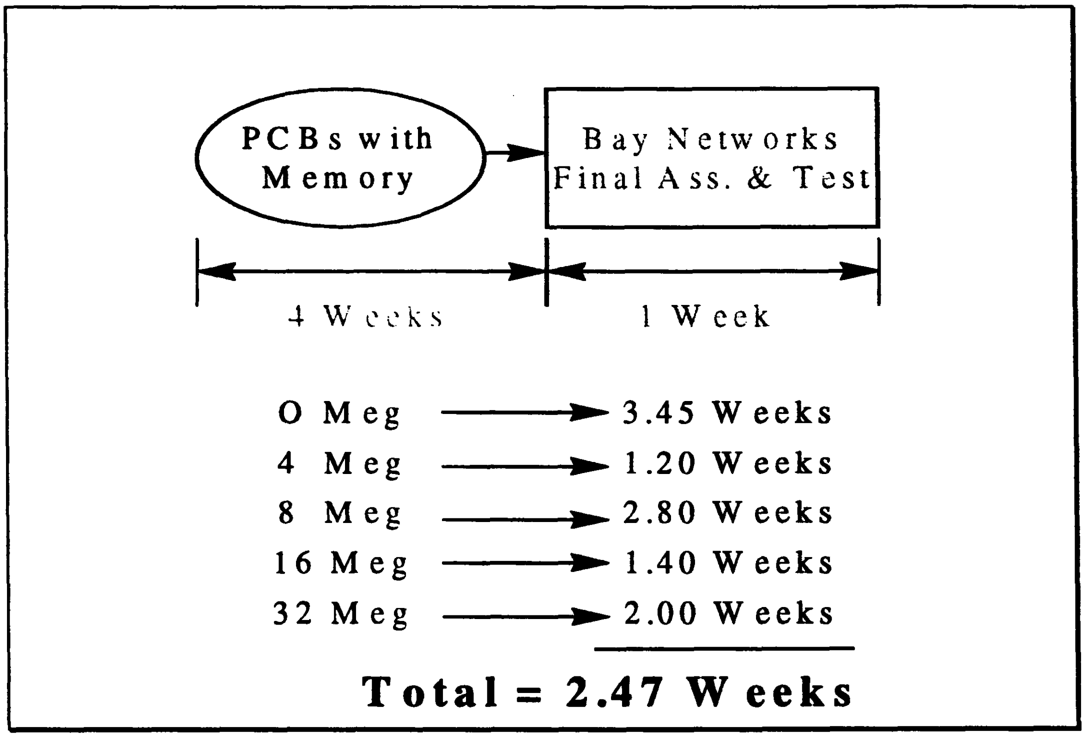

Figure 4-1: The Current Process for Memory Configured Motherboards

While the calculated safety stocks all differ, in practice a master schedule would carry the same stock for each memory flavor of the ARN Ethernet motherboard by arbitrarily placing a 15-20% (2 - 2.5 weeks) flex on each flavor. Although the calculated safety stocks do differ, the total stock required is about 2.5 weeks. This is very close to the *across the board* safety stock that would be placed by a master scheduler on each line-item and thus very close to the total safety stock carried for the ARN Ethernet motherboards.

4.1.1. When the Process works well

Bay Networks' current process effectively services demand when the forecast error is about the same for each memory flavor. When this occurs Bay will over or under consume motherboards at the same rate regardless of memory flavor and will replenish its supply as necessary from suppliers. Given that its safety stock has been set correctly, severe shortages of motherboard flavors will not occur except in the most extreme of demand circumstances.

4.1.2. When the Process falls short

This process of stocking memory flavored motherboards with similar safety stocks does not work when the forecast error is different for each motherboard. Unfortunately for Bay Networks, this is more likely to be the case than not the case. Some of the memory flavors have more data history and much higher demand than other flavors and therefore are far easier to forecast. The wide disparity in average forecast error over a one year period is shown in Figure 4-2 where the average error ranges from 13 to 37 % of the quarterly demand. Also, these average errors only represent the *absolute* (absolute value of (forecast-actual/actual)) error of the forecast. At times throughout the year the demand for one memory flavor far exceeded the forecast while the demand for another flavor fell short of the forecast. When this occurs, Bay's production group tries to minimize the amount of excess material ordered by using conversions. Recall that this is the process whereby memory is removed from a motherboard only to be replaced by the correct memory for the customer order. This is also the process that tends to be extremely inefficient since it occurs for only two weeks of the quarter and is primarily completed by temporary workers.

While a forecast error that is different both in magnitude and direction for each memory flavor makes Bay's current process chaotic, it is just the scenario that would make risk-pooling effective. Risk-pooling under these conditions would allow the negative correlation between demand streams to mitigate the individual forecast errors inherent in the current forecast. For example, if all the motherboards were held in a generic format, the under forecast of one flavor would cancel the over forecast of another flavor. Even if the demand streams were independent of slightly positively correlated, chapter 3 showed how sheer aggregation tends to decrease total variability and thus the required safety stock. Section 4.2 will outline the new Bay Networks process for memory flavored motherboards.

4.2. An Improved Process for Memory Configured Motherboards

This section outlines an improved process to configure motherboards with memory as customer orders are received. This new process replaces the current practice of stocking pre-configured motherboards as delivered by Solectron, Bay's supplier. Issues addressed in this section include: the expected decrease in forecast error; the reduction in required safety stock; and finally, the elimination of inefficient memory conversions performed at the end of the business quarter.

4.2.1. Forecasting at the Aggregate level

Currently, forecasters in the Manufacturing Business Planning group forecast at the specific memory level for ARN Ethernet motherboards. Forecasters get to this specific memory level forecast by first determining the total demand forecast of Ethernet motherboards considering both the installed base of ARN routers and the expected growth rate of the installed base. Next, a forecaster will divide this total forecast into five segments, each of which represents the demand for a memory specific motherboard. This process of segmenting the total demand automatically will create some degree of negative correlation between the segments. This is explained in the following example: Imagine that the total aggregate forecast is perfect – 100% accurate. Next, a forecaster then segments the 0 Megabyte memory flavor from the total. If a forecaster makes this segment too small the total for the other segments will be too big. Thus, the process automatically introduces some negative correlation between the 0 Megabyte segment and the rest of the flavors. This process of segmentation, and the error that is introduced, is repeated for all the memory flavors.

Forecasting at the aggregate level eliminates this extra variability introduction into the forecasting process. Evidence of this is seen in Figure 4-2 below. While the forecast error for the individual memory flavors ranges from 13 – 37 %, the aggregate average forecast error over the last 4 quarters is only 10%.

PCB Model	Ave. Forecast % Error	Next Qtr Forecast (Units)	Service Level fraction from inventory	Material Safety Stock (Units)	Material Safety Stock (%)
ENET MTR 0	37%	730	.83	210	28.7
ENET MTR 4M	13%	500	.83	49	9.7
ENET MTR 8M	30%	6000	.83	1396	23.3
ENET MTR 16M	15%	1800	.83	204	11.4
ENET MTR 32M	21%	60	.83	10	16.4
AGGREGATE	**10%**	9090	.83	**730**	**8.0**

Figure 4-2: Aggregate/ Individual Forecast Errors and Required Safety Stocks

4.2.2. Safety Stocking at the Aggregate level

The primary advantage of forecasting aggregate motherboards is the decreased safety stock resulting from the reduced forecast error. The last row of Figure 4 – 2 shows the new safety stock for the ARN Ethernet motherboard when held only in the generic 0 Megabyte memory flavor. For the total forecast demand of 9090 motherboards, 730 units, or about 8 % of total demand, are required as safety stock for the quarter. Contrast this new safety stock to the 1896 motherboards (over 1139 more) that are required to provide the same service level when individual flavors are safety stocked independently.

This reduction in inventory represents almost a $600,000 savings for just the ARN Ethernet motherboard family. This result is shown in Figure 4 – 3.

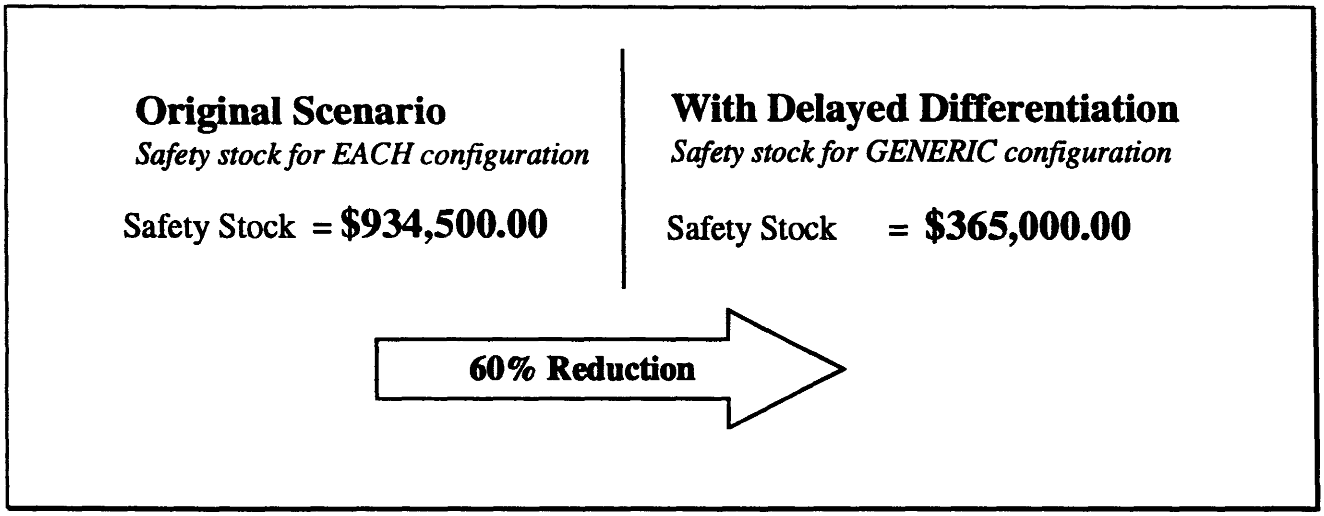

Figure 4-3: Safety Stock Reduction in Dollars

4.2.3. Configuring Motherboards only as needed

Bay Networks' improved process would allow the shop floor to stock only one type of ARN Ethernet motherboard – the generic 0 megabyte memory motherboard. When a customer order was received for an ARN product, a motherboard switching mode and memory configuration would be selected. This is similar to the current process. However, when the 'pick list' is generated on the shop floor a worker will pull a generic motherboard from the stockpile and configure it with the appropriate memory chip prior to sending it through the rest of the final test and assembly process. At the initial assembly station for the ARN product family, memory configuration will replace the stocking of multiple pre-configured motherboards. The new process is shown below in Figure 4-4:

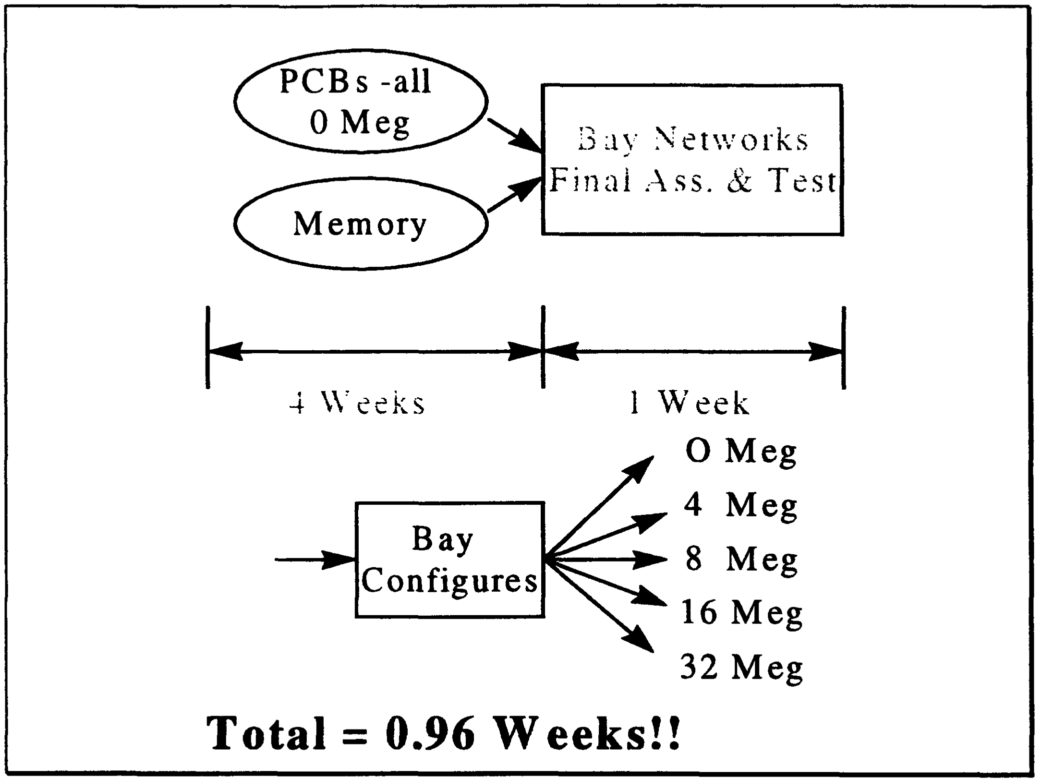

Figure 4-4: Bay Networks' Improved Configuration Process

This simple new process has many advantages over the current process. For one, this new process would reduce the number of storage locations and the number of individual part numbers maintained in the electronic inventory system. Fewer storage locations allows for fewer mistakes in the quarterly inventory cycle-count system that audits inventory on a rolling basis. One motherboard flavor will eliminate the need for inefficient conversions at the end of the quarter, the man-hours wasted performing conversions, and the additional man-hours wasted reconciling the system inventory count created by botched conversions. One motherboard flavor will also reduce the time spent forecasting demand for the individual flavors and will improve the forecast accuracy metric by which forecasters are measured.

Finally, one motherboard will save Bay Networks money. Bay will pay less of a markup to have memory configured and delivered from Solectron and will have to hold far less inventory to adequately meet demand. As shown in Figure 4-4, a total of .96 weeks of inventory will replace the 2.5 weeks currently in place for the ARN Ethernet motherboard. As inventory shrinks, Bay will pull the cash value of the inventory immediately out of the system and then save the opportunity cost of that cash in each subsequent quarter. One of the main reasons that postponement and risk-pooling is so effective in this situation is the vast difference in lead times for the two scenarios. The lead time, especially near the end of the quarter, to obtain memory configured motherboards from Solectron is quite long while the actual time for Bay to configure generic motherboards in-house is just a matter of minutes. This allows Bay to be more responsive to customer orders with less inventory.

4.3. Obstacles to Implementation

Like all process improvement opportunities, there are risks and obstacles to moving towards improvement implementation. For the motherboard aggregation and configuration postponement some resistance was expected by both Bay's primary subcontractor, Solectron, and internally to Bay, by various players.

4.3.1. Subcontractor Resistance

Since Bay Networks had attempted to bring configuration in-house one time previously, Bay expected Solectron to resist the change for the same reasons that they had before but with less fervor. Initial resistance, three years earlier, was primarily based on the large margin that Solectron made marking up memory from its purchase price. Memory prices were as much as 5-10 times what they are today and Solectron made a significant portion of their profit on the memory component of the motherboard. Based on Solectron's pricing model, which was not activity based, the removal of memory would all but remove any profit on motherboards. When faced with the possibility of Bay configuring their own motherboards, Solectron reallocated profit margin on an activity based method and also gave Bay some price concessions on all motherboards. Bay thus decided not to bring configuration in-house.

Now, three years later, Bay Networks is revisiting the decision and expects some resistance based on the reduction of value-add to Solectron -- but far less resistance than last time since memory chips are far less expensive. At a minimum, Bay would hope to convince Solectron to be more flexible in supplying different memory flavored motherboards with a much shorter lead-time.

4.3.2. Bay Networks Resistance

There are likely to be two main reasons that Bay Networks would resist postponed configuration of motherboards: (1) Bay sees themselves as a manufacturing organization that wants to use primarily turn-key vendors, and (2) Bay has as a goal of reducing the manpower related expense to product manufacturing.

Given that Bay is striving to outsource more and more of their manufacturing, the idea of more product manufacturing within its own walls runs counter to the goal of using only turnkey vendors. This goal, to remain useful as a business proposition, must be justified on a product by product basis rather than applied across the board. While some products, especially those that are shrink-wrapped and have fixed configurations, are well suited to turnkey vendors, the ARN product family falls into the category of products that should have order specific features added as late as possible. Bay must seek to balance the often times reduced cost of outsourcing with the improved customer satisfaction, reduced inventory, and improved flexibility of keeping the 'make to order' processes in-house. To outsource such processes as memory configuration, permanently restricts Bay's supply chain to the accuracy by which they can forecast.

The site managers for Bay Networks have manpower related goals that dictate how much can be spent, as a percentage of total product cost, on the shop floor. If the postponed configuration process did require the addition of a full-time employee within Bay Networks it could work against the manpower expense goal. Bay Networks must aggressively account for the savings in temporary workers, inventory reconciliation time savings, and inventory opportunity cost savings to acknowledge the total savings to the company.

5. Conclusions & Recommendations

The last four chapters outlined Bay Networks' supply chain and its processes for forecasting demand and for configuring product. Chapter 1 provided a brief background for this work while Chapter 2 detailed the information flow within Bay Networks' supply chain -- specifically the current process for forecasting motherboard demand and for configuring motherboards prior to their integration into a product. The mathematical theory behind the concept for risk-pooling was shown in chapter 3 while the specifics for the concepts use at Bay Networks was addressed in chapter 4.

This chapter concludes the analysis with recommendations for Bay Networks to both improve forecast performance and to increase supply chain service levels and less inventory. This chapter also suggests some opportunities for further study.

5.1. Recommendations

The following is a list of recommendations based on the analysis in this thesis and based on the author's observations while conducting the research that led to this thesis:

- *Expect uncertainty in forecasts* – Although this may seem like an obvious point, many Bay employees relied on the forecast to eliminate demand uncertainty. In reality, poor forecasting can actually have more variability than the underlying demand or the simplest exponential smoothing method (Rosenfield, 1994). Even great forecasts will still have some degree of variability from the actual demand – the key for Bay Networks is to understand the accuracy of its forecasting group and plan for that level of variability.

- *To improve forecasts you must measure the forecast error* – The first step to improving the forecasting performance of Manufacturing Business Planning begins with the implementation of relevant metrics. If it is MBP's goal to improve its ability to forecast highly aggregated revenue for product families then metrics should be put in place to measure the forecast at this very high level. If MBP's goal is to improve its disaggregated line-item forecast, then metrics must be designed to measure forecasts at this low level. It is critical that Bay realize that to improve its processes it must first understand its current capabilities and weaknesses. This understanding begins with relevant metrics.

- *Inventory can be reduced with improvements on multiple dimensions* – The initial focus of this research at Bay Networks aimed to improve forecast accuracy. I heard many times throughout my research that if forecast error was reduced, current inventory levels could be reduced. In actuality, many other areas of opportunity exist for reducing inventory. One such piece of 'low hanging fruit' is the postponement of memory configuration outlined in Chapter 4 of this thesis. The decreased variability afforded by risk-pooling and the reduction of inventory it provides could be immediately implemented for 5 motherboard families beside the ARN Ethernet motherboard. Additional plans to create new motherboard families was being discussed as this research was concluding. These new motherboards provide even more opportunity for inventory reduction. Finally, equally important as demand variability to reducing safety stocks is the lead-time for material throughput from suppliers and Bay's plant. Parallel process improvements could occur at supplier and Bay plants as the forecast is being improved.

- *Use forecast variability to set safety stock 'flex'* – Rather than arbitrarily setting flex percentages based on the gut feel of a materials planner, Bay should systematically set the number based on the actual forecast variability for a particular line-item. Additional adjustments should then be made based on the supplier and nuances of its particular lead time (Krupp, 1997).

- *Measure inventory over the entire quarter rather than just at the end* – Current inventory measurements occur at an artificially low point at the end of the quarter. This metric does not accurately quantify the exposure to obsolescence and carrying costs that actually exist throughout the quarter. Even worse, the drive to reduce inventory at quarter's end may actually hinder process improvements to reduce inventory throughout the rest of the quarter if end of quarter inventory were to increase. This is another case of measuring the right parameter before attempting to improve it.

- *MBP and the Master Schedulers should prioritize line-items for forecasting* – Because forecasters in MBP have limited time to regenerate forecasts, they must prioritize which line-item forecasts require the most current information. The appropriate dimensions for this prioritization are: (1) the lead-time of the line-item, (2) the strategic significance of the line-item, (3) the cost of a particular line-item, and finally, (4) the unit volume of the line-item. Since some of these dimensions are subjective, both MBP and the Master schedulers should agree on the priorities of each line-item.

- *Manufacturing Business Planning should conduct training within its group* – Currently there is no formal process to transfer knowledge from experienced forecasters to those with less experience. Given that Bay now requires so much mandatory training per employee there is an advantage to conducting training implementing a formal knowledge transfer process within the group.

5.2. Opportunities for Further Study

Some opportunities for further study within Bay Networks' supply chain include:

- *The tracking of actual premiums demanded and paid by Bay Networks to suppliers* – Based on the contracts that dictate the interaction between Bay Networks and its Printed Circuit Board supplier base, Bay should pay significant premiums whenever it under forecasts demand and is required to pull additional material from the supplier base. Although there is little data retained by Bay on this topic, the data analyzed showed a reluctance on the part of the suppliers to actually charge these premiums.

  Perhaps since it is in the interest of the suppliers to increase their own demand, they tend to be very supportive of customer demand increases when Bay's forecast was low. Based on this observation, Bay should actively track premiums demanded and paid, and the circumstances that generated the premium. This information can then be used formulate hypotheses about the *true lead-time for line-item increases* and the material 'cost of underage' in these circumstances.

- *Investigate means to track Major Business Opportunities (MBOs)* – Major Business Opportunities possibly disrupt the Bay Networks supply chain more than any other cause. MBOs represent the unforeseen, at least by manufacturing, large orders that suddenly appear with a very short lead-time to delivery. These large orders can create more than 50% of the demand for quarter. Bay should investigate means to ensure that its sales force communicates the possibility of these large orders to the manufacturing group as soon as possible in the sales prospecting process.

References

Chase, Charles W., "Measuring Forecast Accuracy," The Journal of Business Forecasting, Fall, 1995.

Chase, Charles W., "Developing a Business Strategy for Forecasting," Business Forecasting: Best Practices Conference, October 20&21, 1997.

Colgan, J.A., "A Business Planning Model to Access the Tradeoff Between Inventory and Capacity for a Stage 1 Manufacturing Process," MIT Master's Thesis, 1995.

Drake, A., Fundamentals of Applied Probability Theory, McGraw - Hill Publishing Company, 1988

Gilpin, B.C., "Management of the Supply Chain in a Rapid Product Development Environment," MIT Master's Thesis, 1995.

Graves, Stephen C., "Safety Stocks in Manufacturing Systems," Journal of Operations Management, Volume 1, pp. 67-101, 1988.

Griggs, Kenneth, and Marcus O'Connor and William Remus, "Does Reliable Information Improve the Accuracy of Judgmental Forecasts?" International Journal of Forecasting, Vol. 11, pp. 285-293, 1995.

Kahn, Kenneth B., Mentzer, John T., "Forecasting Technique Familiarity, Satisfaction, Usage, and Application," Journal of Forecasting, Vol. 14, 465-476, 1995.

Krupp, James A.G., "Managing Demand Variations with Safety Stock," The Journal of Business Forecasting, Summer, 1997.

Hogg, R.V., Ledolter, J., Applied Statistics for Engineers and Physical Scientists, New York: Macmillan Publishing Company, 1992.

Lee, H.L., and C. Billington, "Material Management in Decentralized Supply Chains," Operations Research, September-October, 1993, pp. 835-847.

Nahmias, S., Production and Operations Analysis, Burr Ridge, IL: Richard D. Irwin, Inc., 1993.

Price Waterhouse, "High Technology Sales Forecasting Roundtable," Price Waterhouse Benchmarking Study, Sept. 17, 1997.

Price Waterhouse, "Improving Supply Chain Management," Price Waterhouse Benchmarking Study, Dec., 1996.

Richard, Christopher, "Data-Driven Logistics Planning," MIT Master's Thesis, 1996.

Rosenfield, Donald B., "Demand Forecasting," The Logistics Handbook, Chapter 4, pp. 325-462, 1994.

Sheerin, Neil, "Using the Concept of the 'Forecasting Cube' to Benchmark Forecast Error," Business Forecasting: Best Practices Conference., October 20&21, 1997.