Characterization of the Cost of Forecast Error in a Complex Supply Chain

by

Chris Schechter
B.S., Mechanical Engineering
Massachusetts Institute of Technology, 1994

Submitted to the Department of Mechanical Engineering
And the Sloan School of Management

In Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management
And
Master of Science in Mechanical Engineering

In conjunction with the
Leaders for Manufacturing Program

At the
Massachusetts Institute of Technology
June, 1998

Signature of the Author ____________________
Department of Mechanical Engineering
/ Sloan School of Management
May 8th, 1998

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

Certified by ____________________
Yashan Wang, Thesis Advisor
Assistant Professor of Management Science
Sloan School of Management

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

Accepted by ____________________
Prof. Larry Abeln
Director of the Master's Program
Sloan School of Management

Characterization of the Cost of Forecast Error in a Complex Supply Chain

by

Chris Schechter

Submitted to the Sloan School of Management and the
Department of Mechanical Engineering on May 8th 1998 in partial fulfillment of the
requirements for the degrees of Masters of Science in
Management and Masters of Science in Mechanical Engineering

Abstract

The importance of forecast accuracy is increasing in the semiconductor industry due to two compounding factors. The complexity of the manufacturing process is increasing which causes manufacturing cycle times to increase as well. Further, the lifecycle of products is decreasing. Thus manufacturers have to increasingly build ahead of demand while anticipating customer preference with increasing accuracy. This thesis investigates the origins of demand forecast inaccuracies at Digital Semiconductor as well as characterizes the main supply chain costs of forecast inaccuracy.

The forecasting cycle is a complicated chain spanning several functional groups each with their own incentives and interests. For Digital Semiconductor, the process is complicated by having one manufacturing site servicing five product lines which each have very different customers, competitors and product specifications. A successful process results in having the right type and number of parts available to customers at the right time. An over-optimistic forecast results in excess inventory and obsolescence, whereas a forecast that is too conservative will result in stock-outs and loss of customers.

The thesis starts with a detailed analysis of Digital Semiconductor's forecasting process. We then describe the analysis and methodology we used to develop a model of Digital Semiconductor's supply chain. This model allowed us to characterize the supply chain costs of historical forecast errors. The thesis concludes with recommendations to improve the forecasting process at Digital Semiconductor.

Thesis Advisors: Stan Gershwin, Department of Mechanical Engineering
Yashan Wang, Sloan School of Management

Acknowledgments

I gratefully acknowledge the support and resources made available to me through the MIT Leaders for Manufacturing program, a partnership between MIT and U.S. Manufacturing.

I would like to thank my thesis advisors, Stan Gershwin and Yashan Wang, for their guidance and support throughout the past year.

I would like to especially thank my LFM and Sloan classmates for the past two years. Without a doubt, the most valuable part of being at Sloan is having the opportunity to learn from the collective group.

I would like to thank my family for their continuous support along the way.

Most of all I would like to thank my wife, Bridget, for her unending patience, understanding and encouragement throughout my tenure at MIT.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

1.1.1 Background of the Semiconductor industry

The semiconductor industry has grown in the last 10 years at a compound annual growth rate of 20 %. At the same time it has become increasingly competitive and complex with most of the US based companies transitioning away from DRAMs into making integrated circuits (ICs) and Multiple Processing Units (MPUs). These products have followed an exponential growth in performance. This growth in technological performance is being achieved with a combination of decreasing line width; increasing transistor density and increasing die size. Further, the product life cycles throughout the industry have been shrinking as well. Thus in order to keep up with constant technological innovation and decreasing life cycles all semiconductor manufacturers have had to increase their capital costs dramatically. The semiconductor industry gross capital expenditures grew more than four fold from 1990 to 1995. Further, in recent years a larger share of capital assets were retired annually, reflecting the rising share of equipment technological obsolescence in the industry. In addition to the rise in capital and research and development costs, the effective cost of semiconductor devices has fallen dramatically. The rate at which technological advancements enable electrical functions to be placed on individual die far exceeds the rate of manufacturing cost increases. The increase in capital expenditures and rate of innovation combined with falling prices has put semiconductor manufacturing companies under intense pressure to reach critical break even volumes for their product by being at the forefront of technology and marketing. The leader in the industry is Intel with close to 80 percent of the market share of MPUs. Smaller companies are left searching for niches where they can sell enough product to leverage the ever increasing capital and research expediters required to remain in the business. To that end, most electronics companies with semiconductor operations are increasing the emphasis on their external IC business and turning semiconductor operations, traditionally captive to the company, free to pursue external business.

1.1.2 Background of Digital Semiconductor (DS)

Ken Olson created the Digital Equipment Corporation (DEC) in 1957. Under his leadership the company soon grew into a manufacturer of mainframes and workstations. It was also the pioneer in data storage and networking equipment in the late 70s and throughout the eighties. In 1982 Digital Equipment Corporation broke into the personal computer market as well. The semiconductor division of DEC, Digital Semiconductor was created in 1974 as a second source to the industry for DEC's key strategic semiconductor products. In 1992 Digital Semiconductor announced its Alpha program, a totally new, open, 64 bit architecture, when the rest of the industry was still at 32 bits. In order to continue the Alpha chip's technological lead, the Hudson fabrication facility was built. However after a few years of disappointing sales of its flagship product, the division transitioned from a captive supplier to a merchant integrated circuit manufacturer with both an internal customer, DEC, and a variety of external ones. This shift in business strategy was required to reach critical break even sales volumes to leverage the investment in capital equipment of the Hudson fabrication facility. The semiconductor division now engineers and manufactures a variety of products in the Hudson manufacturing facility in addition to the Alpha chip. These new products include network chips, low voltage chips for hand held devices as well as bridge chips. In the last two years the division reorganized to service their new customers. The semiconductor division grew an external sales organization, a distribution organization with a network of distributors, an order fulfillment organization and product line organizations aligned around the variety of products being sold to the external market.

1.2 Motivation

The semiconductor industry is very volatile due to decreasing life cycles and intense competition based on constant technological innovation. Yet the structure of the supply chain in the semiconductor business is inherently rigid. The manufacturing process requires very high capital expenditures on equipment that requires very long lead times. Most of the lead times on equipment exceed the life cycle of the products. Further, the average manufacturing cycle time of the products is three to five times longer than the average lead time quoted to customers. Lastly, in the case of Digital Semiconductor, there are six different product families sharing the same manufacturing facility. In this environment, an accurate customer demand forecast combined with supply chain policies that match the accuracy level of the forecast is necessary to be able to service customers and optimize the use of very expensive capital equipment required for semiconductor manufacturing.

1.3 Thesis Objective

The objective of this thesis is three fold. The first is to quantify the cost of past forecast inaccuracy levels. By using empirical data for past forecasts and demand patterns we will create an awareness of the potential supply chain cost savings that a more accurate forecast would yield. The second objective is to yield an understanding of the tradeoffs between capacity, inventory, expediting, lead times and service levels in the complex supply chain of Digital Semiconductor. A better understanding would allow Digital to structure their supply chain to minimize cost while meeting their strategic customer satisfaction objectives. The third objective is to identify and describe business process changes internal to Digital that would allow for more accurate customer demand forecasts.

1.4 Scope and Limitations

In order to understand the cost of forecast inaccuracy and the trade-offs of these costs throughout the supply chain we developed a model that was used in conjunction with empirical data. The mathematical model is a simplification of the actual supply chain. The supply chain is modeled as a three stage line. Each stage mirrors an actual stage in the supply chain of Digital Semiconductor. Each stage is modeled with deterministic cycle times and yields but variable forecasts and release policies. There are three buffers separating the three stages. Although simplified the model captures the main qualitative and quantitative trade-offs of the actual supply chain by modeling the imperfect flow of information and variable demand. The demand and forecasting history of each product line at Digital Semiconductor was characterized using a probabilistic approach. We then input these mathematical descriptions of past forecasts and demand patterns into the model to obtain the cost of poor forecasts, both long term and short term. We were able to quantify the return on forecast accuracy.

Using the same probabilistic description of past forecasts and demand we then changed the release, inventory and capacity practices to determine an optimal set of supply chain policies given the inherent variability of demand in this industry. These policies remain at the device and supply chain level. The model does not involve plant floor scheduling policies but rather remains at the level of high level supply chain. The model assumes that the whole fabrication facility is one black box. Further, the model only addresses one device throughout the supply chain. The opportunity cost of running once device at the expense of another is not modeled.

Finally, the recommended changes to the current business practices that are required to achieve better forecast accuracy are the result of series of interviews, observations and benchmarking of the industry. We were not able to implement many of these recommendations due to a sudden change of ownership of the division. These recommendations nevertheless apply.

1.5 Deliverables of the Thesis

The project undertaken at Digital Semiconductor had several objectives and deliverables in mind. The project was originally designed to reengineer the forecasting processes of Digital Semiconductor. The model creation was necessary to justify the substantial investment in tools and energy required for the reengineering effort. The model was developed to bring an awareness to the substantial costs of poor forecasting and what supply chain options were available to minimize these costs. The reengineering effort had to be put on hold and remained at the stage of recommendations after the model was complete. The deliverables of the thesis include:

- Description and development of a model of the supply chain of Digital Semiconductor

- Description of the inputs to the model (past forecast accuracy and demand patterns) and the processes used to characterize this data in probabilistic form.
- Description of the outputs of the model
- Results of the model quantifying the costs of poor forecast for two product lines during fiscal year 1997
- Investigation of the effects of demand variability on the supply chain costs
- Conclusions and recommendations on inventory, capacity, lead time and expedite tradeoffs drawn from the analyses
- Recommendations regarding changes in current business practices to improve forecast accuracy

1.6 Thesis overview

The rest of the thesis is broken down into seven sections. The first is chapter 2 and describes the existing forecasting process at digital semiconductor. This section details the purpose behind the existing methodology at Digital Semiconductor and the interrelationships with all the organizations involved in the creation of the monthly forecast. This section also maps out the information flow and feedback between the different groups involved in forecasting as well as the existing accuracy metrics. Chapter 3 describes the creation of the model. It begins with a general overview of the model and the mathematical relationships existing in the model as well as their link to the constraints and information flow of the actual supply chain. Then we characterize the model past forecast accuracy and the role and effect of forecast errors on the supply chain. The next section covers the output of the model. It highlights the main costs of forecast inaccuracy and their relationships and tradeoffs as determined by the model. . These results were obtained by numerical experiments run on the model. Quantitative results of the model are also in chapter 4 in which we report the supply chain costs of past forecast inaccuracies. Chapter 5 covers the use of management processes that do not allow the leveraging of the necessary information resident in the company to achieve an accurate forecast. This chapter recommends some practices that would allow Digital

Semiconductor to achieve better forecast accuracy. Chapter 6 summarize the key learnings and recommendations for the whole thesis as well a the possible future studies building on this one.

2. The forecasting cycle

There are two main forecasting cycles at Digital Semiconductor. The first is done annually and the second is done monthly. The annual forecast is a long term strategic planning document, whereas the monthly forecast drives the manufacturing floor as well as the short to medium term sales and marketing efforts. This thesis focuses on the monthly forecasting process because it is mostly this monthly forecast that drives the business. The thesis will cover the demand planning aspect of the forecasting process, not the plant floor scheduling or supply planning aspect of the forecasting process. The monthly forecast is a rolling 2 year forecast. The monthly forecasting process is designed to begin with the customer signal and ultimately end with a monthly production quota that manufacturing can follow. The length of the forecasting cycle, one month, is dictated by the length of time required to have all the departments involved in the forecasting process complete their work in a mostly linear fashion. The following chapter begins with describing the intended purpose of the monthly forecast in section 2.1. The next section, section 2.2 then has a brief overview of all the different departments involved in forecasting future demand and the process by which they all link up to create a monthly forecast. Section 2.3 then focuses on the information flow and feedback between organizations for the purpose of identifying the key areas for improvement. Lastly, the metrics used to track forecast accuracy are presented and discussed in section 2.4.

2.1 Purposes and nature of the monthly forecast

The monthly forecast at Digital Semiconductor serves many purposes. First it is used to determine what material should be started in the upcoming month to satisfy customer demand at the end the manufacturing cycle time. Second, it is used to determine whether demand is going to outstrip capacity in the upcoming 6 to 12 months. If that is the case, the monthly forecast could be used to build ahead enough surplus stock, above and beyond customer demand within the manufacturing cycle time, to meet demand. Third, the monthly forecast is used to determine capacity expansion if the upcoming demand is thought to be greater than capacity for a significant length of time. In that case the decision to buy new tools could be based upon the demand numbers in the monthly forecast. Lastly, the monthly forecast is also used to set revenue targets and track variances from these targets. Because revenue has to be adjusted for customer and distributor product returns, the monthly forecast could have negative sales in the upcoming month if a large number of returns are expected.

Within each product line, the monthly forecast is broken down by part number and within each part number, by processing speed. Therefore, all transitions from one variation of a part to another variation are also tracked and predicted in the monthly forecast. The monthly forecast is a rolling forecast and spans two full years. The granularity of the forecast is monthly. All products sold by Digital Semiconductor are tracked with the monthly forecast. This applies to products sold through distributors or directly to customers, whether these are external or whether they refer to the mother company of Digital Semiconductor, the Digital Corporation. Further, the monthly forecast tracks chips as well as boards sold out of Digital Semiconductor. Further, the monthly forecast includes prototypes and free samples offered to customers in an effort to obtain design wins from them.

The monthly forecasting process was designed to be broken down in two parts, the demand planning and the supply planning. The demand plan is created by the marketing, sales and distribution organizations. It only encompasses customer demand and the expected customer requirements. The demand plan is referred to as the Demand Requirement Forecast (DRF). The DRF, created once a month, is then transmitted to the organizations responsible for executing and meeting the customer demands. After these organizations have translated customer demands into manufacturing requirements the forecast becomes the build plan. An overall high level description of the process is in figure 1. This thesis will focus on the demand planning aspect of the monthly forecast as this initial stage drives the rest of the forecasting process.

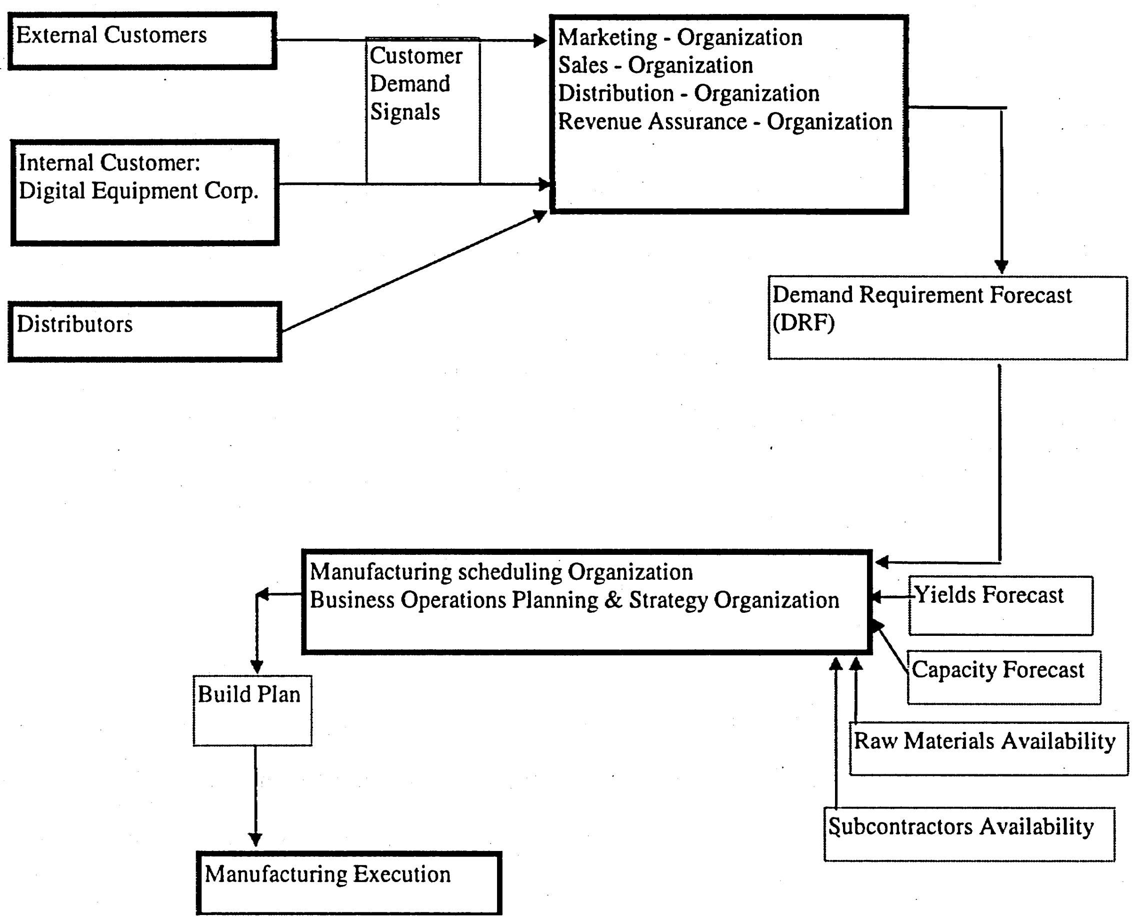

Figure 1: High Level Demand/Supply Planning

2.2 Organizations of DS involved in forecasting

This section contains a brief explanation of the organizations of DS that are involved in the forecasting process as well as their role in the forecast generation.

2.2.1 Sales

The sales organization is divided by geographic regions throughout the world. Each region is headed by an Area Sales Manager (ASM) who is responsible for direct customer contact and sales for all Digital Semiconductor products in his/her area. Working for each

Area Sales manager are a number of sales representatives who will call on clients and establish initial contacts. The sales organization is not aligned along individual products and is given revenue targets on a regional basis, regardless of what products they sell.

At the beginning of every month the Area Sales Managers are responsible for submitting a forecast of customer demand for all products sold in their area for the upcoming 6 to 9 months. The ASMs compile this forecast based on conversations with customers and their sales representatives. The forecast they submit monthly is independent from the revenue targets handed to them annually. The monthly submission is only to allow better customer service as well as guarantee that the demanded products will be available for the customers when they require it. The submission of the ASM forecast is the first formal step in the forecast generation process. The ASM forecast is transmitted to Revenue Operations.

2.2.2 Revenue Operations (ROPS)

Revenue operations is responsible for receiving and recording all incoming orders as well as closing receivables for Digital Semiconductor. Upon receipt of an order, ROPS records the revenue amount and passes the required product, quantity, customer and due date to the Business Operations Planning Strategy group in order for the order to be filled. In addition, Revenue Operations (ROPS) consolidates all the ASM forecasts. The purpose of ROPS involvement in the forecasting process is to transform a forecast divided into regions for all DS products into a forecast divided into products for all regions. ROPS sums up all the forecasted sales for a given product across all regions. Further, ROPS also consolidates the forecast from the regional distributors in a similar fashion. After consolidation into products and aggregating the products into product lines, the ROPS group passes both the distributor forecasts and the ASM forecasts on to the respective product lines. The ROPS group has one week from the receipt of the ASM and distributor forecasts to the handoff of the consolidated forecast to the product lines. The product lines then create their own forecast, the product line forecast, loosely based on the ASM and the distributor forecasts, and return this third forecast back to the ROPS group.

2.2.3 Product Lines

The product lines are the marketing arm of Digital Semiconductor. They are aligned by large product segments: bridges, microprocessors, networks, chips for low voltage hand held devices, custom chips and multimedia chips. Each product line is responsible for the profitability of its product segment. The product lines have a matrix organization that link them to the engineering side of Digital Semiconductor. The product lines also have a customer/supplier relationship with the manufacturing organization of Hudson. The product lines order parts and pay manufacturing competitive transfer prices for them. The product lines set pricing, product promotions and manage customer relationships to secure adoption of their chips. They are also ultimately responsible for the demand forecast that will drive manufacturing. In order to create this forecast, the product lines base their forecast on the forecasts received from the Area Sales Managers and the distributors through the intermediary of the ROPS group. Further, the product lines also may base their forecast on conversations with customers or manufacturing, the run rate of customer orders that are hitting the books or any changes in the market place that they can foresee. The methods used vary widely across product lines as reported in table 1. We obtained these results through extensive interviews. Each product line has one week to compile a forecast for their product. The one week time frame begins upon their receipt of the ASMs' and distributors' forecasts. At the end of one week, the product lines return their forecast to the ROPS organization, which then aggregates all the different product line forecasts together into the DRF.

2.2.4 Business Operation Planning and Strategy (BOPS)

The Business Operation Planning and Strategy group (BOPS) is the link between the product lines and the manufacturing organization. The BOPS group is responsible for balancing the demands placed by all the product lines on the manufacturing organization. The BOPS group is responsible for taking the five sets of demands from the five product lines and supplying the manufacturing organization with a single coherent demand signal. Further, the BOPS group is in charge of filling an order received and logged by the ROPS group. After receiving the DRF from the ROPS group, BOPS translates that demand signal into a schedule of starts that the manufacturing organization can operate by. To do so, the BOPS group uses the latest forecasts of manufacturing yields and capacity as well as raw product availability and machine up-time. The resulting document that the manufacturing organization executes against is the Build Plan. From receipt of the DRF to the finalization of the build plan, the BOPS group has a week.

2.2.5 Distribution

About two thirds of Digital Semiconductor's external clients purchase Digital's products through distributors while the rest purchase directly through the ROPS group. The distribution group is the link between the array of distributors dispersed throughout the world and Digital's sales and product lines organizations. The distribution group at Digital is responsible for the supply of products through the distribution channels. To do so, the group monitors the purchasing and stocking patterns of the distributors as well as their returns and their responsiveness to Digital's customers. The distribution organization creates a monthly forecast as well. This forecast is based upon conversations they have with the distributors as well as their inventory levels. This forecast is passed on to the ROPS organization.

2.3 Forecasting process

The forecasting process at Digital Semiconductor can be characterized as both linear and unsystematic (Figure 2).

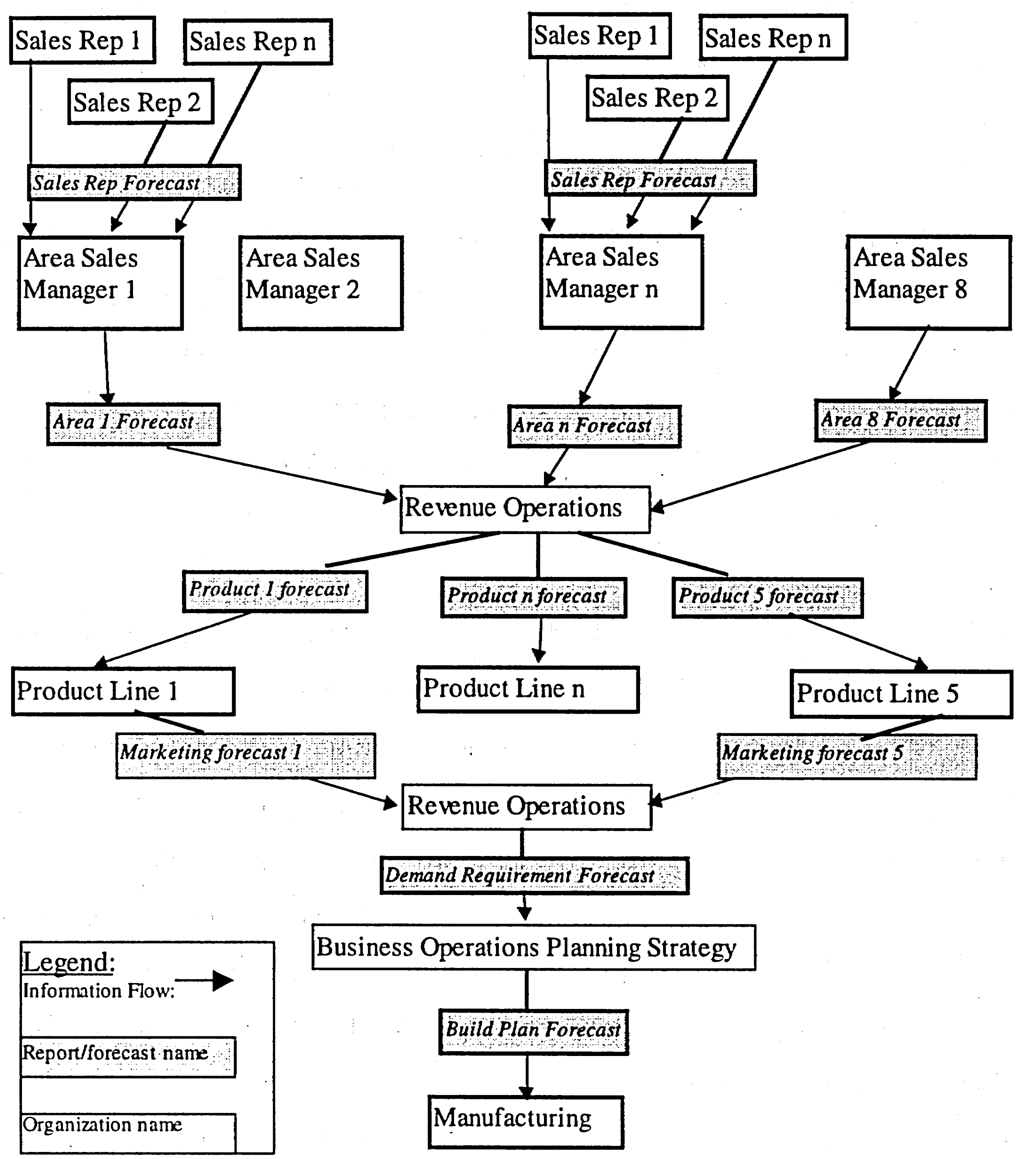

Figure 2: Step by Step Forecasting Process (Distribution forecast omitted)

The forecasting process is a set of sequential steps: At Digital Semiconductor the forecasting process has a cycle of a month. This cycle begins with the sales representatives submitting a forecast for their region to their Area Sales Managers who then approves it and passes it on to the next organization in the forecasting chain. All the forecasts from the different Areas Sales Managers are then disaggregated and reaggregated according to product by the Revenue Operations Group. The later then passes this information under its new form to the product lines who are each responsible for a given product. The product lines approve the forecast and passe it on to the Business Planning and Operations (BOPS) group by way of the Revenue Operations Group. The BOPS group then translates this forecast into an operational form for manufacturing to execute against. Only at this stage in the forecasting process does the first forum for any iterative activity exist: during a one two hour meeting, staff from manufacturing, marketing (the product lines) and BOPS get to discuss and debate the forecast. However, at this very late stage in the process the conversations are geared towards either crafting reaction strategies to either resolve manufacturing bottlenecks that are stopping DS from delivering what is forecasted or understanding how a very low forecast impacts the short term financials of the organization.

The relationships between the groups involved in crafting the forecast are comparable to the traditional "throw it over the wall" interactions between engineering and manufacturing that were prevalent a decade ago in most traditional US manufacturers. These relationships are institutionalized by the overall structure of the process through both the timing of each step and the role played by the Revenue Operations Group (ROPS). The current process is designed so that each group has barely enough time to analyze their section of the forecast during their allotted window. This window is delimited on one side by the day they receive the necessary information and on the other by the day they need to pass it on to the next stage. Also, the exchange of forecast information between group is almost always done through the ROPS group. By serving as intermediary between group, the ROPS organization serves to break down the communication links between the groups generating the forecast (Figure 3).

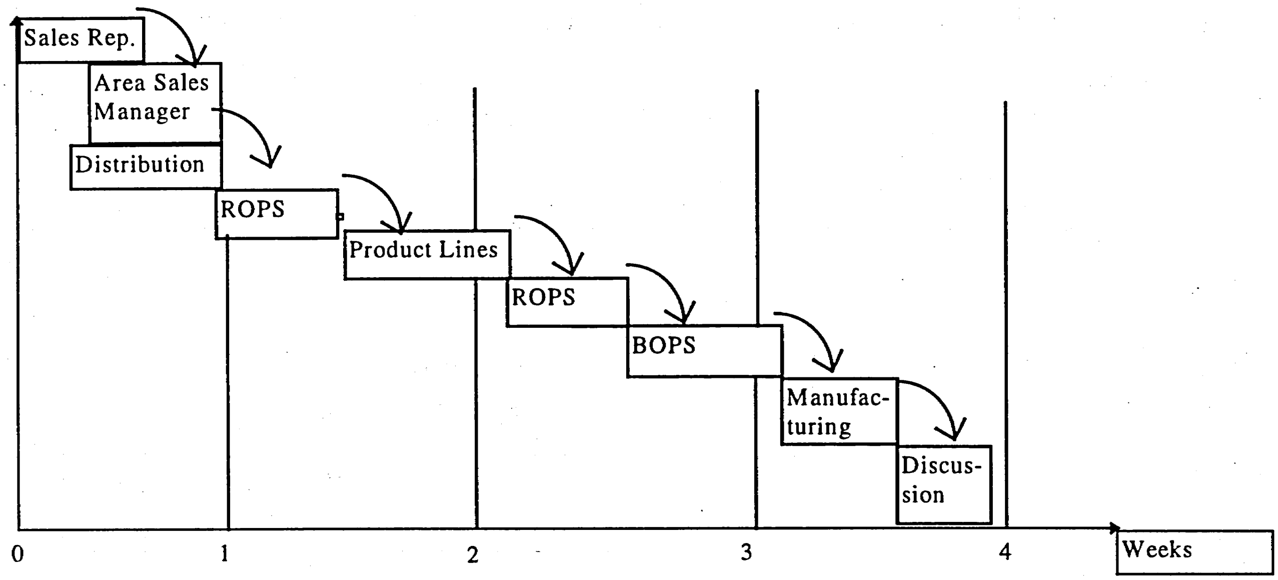

Figure 3: Organizations involved in forecasting

The forecasting process is unsystematic: At Digital Semiconductor the product lines were ultimately responsible for the forecast accuracy and its generation. Each product line followed the same sequence of events but developed its forecast using a variety of methods and sources. Even though all the product lines interfaces with the same organizations such as Sales, BOPS and ROPS, each product line interacted with and trusted the information from these organizations in a variety of ways. Further, internally to each product line the attention paid to forecast accuracy varied dramatically as well. The summary of these differences are in table 1.

As shown in the table, there are vast differences across product lines. For example, even though the Area Sales Manager's forecast is supposed to be the basis for the general forecast, less than half the product lines use it to develop their own forecast. Another interesting metric is the length of time each product line feels comfortable forecasting out to. Keeping in mind, that each month the product lines issue a forecast that spans 2 years, the longest any product line feels comfortable forecasting out to is 6 months, whereas most would not vouch for their forecast further than 3 months out. Lastly, only two of the product lines had a dialogue about the forecast with the Area Sales Managers (ASMs) whereas the other three product lines we surveyed did not. In conclusion, even though officially, all the product lines forecast in a similar manner in the sense that they all go through the same motions, each and every product line uses different sources of information. Each product line then uses judgement and intuition to then create the forecast.

	Product Line 1	Product Line 2	Product Line 3	Product Line 4	Product Line 5
Use ASM forecast	No	Use it as an indication of future trends	Yes, it is the basis for the monthly forecast	No	Occasionally
Dialogue w/ ASM about forecast	No	No	Yes	Sometimes	No
Long term forecasting done	No	Informally	No	Use DEC forecast	Informally
Time horizon for accurate forecast	3 months	3 months	3 months	6 months	6 months
Use current orders to correct forecast	Yes	Yes	No	No	Barely
Update Often (frequency of updates)	Yes (1/week)	No (1/month)	No (1/month)	No (1/month)	No (1/month)
Include BOPS in forecast generation	No	Yes	Yes	No	Yes
Past forecasting accuracy adequate ?	Yes	Yes	No	Not important	Yes
Use feedback to improve forecast accuracy	No	No	Yes	No	Yes

Table 1: Differences in information sources used for forecast generation across product lines

2.4 Information flow and feedback between organizations

The information flow and feedback between the organizations involved in the creation of the forecast is documented in figure 4. The basis for the creation of the forecast is the short term sales forecast issued by the Sales Representative to his/her Area Sales Manager. As the information is progressively passed on to the marketing and to the manufacturing organization, the time horizon of the successive forecasts increase.

The central role of the Revenue Operations (ROPS): the ROPS group has a central role in the forecast generation. It is this group that translates the area specific forecasts from the Sales Managers into product specific forecasts for the Product Lines. The ROPS group also then consolidates the product specific forecasts issued by the product lines into a general forecast that will be passed on to manufacturing. The forecast that is issued by the Sales and Distribution organization is given to the ROPS group as well. Only if the Sales and Distribution forecast is significantly different than expected does Revenue Operation share it with the rest of the organization. Communicating the forecast from the distributors is up to the discretion of the ROPS group. Most often it is not communicated to the product lines who are ultimately responsible for the forecast. The impact of this lack of information sharing is only slowly becoming understood as the Revenue Operations group begins to understand the impact of the ordering patterns of the distributors on the internal supply chain of DS. The marketing forecast that drives manufacturing attempts to predict customer demand, not distributor orders. The later can differ substantially from the former depending on the ordering patterns and inventory strategies of the distributors.

Insufficient feedback: the second striking aspect from this figure is the lack of appropriate feedback to the Area Sales Managers (ASM) even though their forecast should be the basis for the overall feedback. The ASMs receive two types of feedback. The first is through the "Flash report". This report, issued by the Revenue Operations Group, compares the monthly ASM forecast spanning the one month horizon to the actual customer purchases during the forecasted month. Although pertinent, this feedback is focused on the extreme short term. The overall cycle time is longer than one month, making an accurate one month forecast of only limited use. The second feedback available to the ASMs are the "waterfall charts". These charts compare the short and long term forecast of the product lines with the actual number of customer orders for a given product. This feedback mechanism is flawed for two reasons. First, it is focused on the forecast issued by the product line not the ASMs. These two forecast can differ substantially since nearly half of the product lines do not even take the ASM forecast into account when creating their own. (Section 2.3) Second, this feedback mechanism reports how many of a given product were sold for all of the US, Asia and Europe. The ASMs cannot make use of this information because they are responsible for the forecast for their region. A document reporting how many of a given product was sold is of no use to them unless its granularity is at the Regional level. Thus the Area Sales Managers, who most often are offsite in remote locations, cannot improve upon their forecast accuracy because they do not have any feedback as to their current and past accuracy.

Low frequency: lastly all the information flow and feedback occurs only once a month. Since the forecasting process itself is a sequence of consecutive steps and takes a month from the time the Sales Representative issues a forecast to the time manufacturing receives a build plan, the final forecast that drives manufacturing could be as much as two months outdated. Two month in this industry is more than two third of the cycle time and can be as much as a third of the product life cycle.

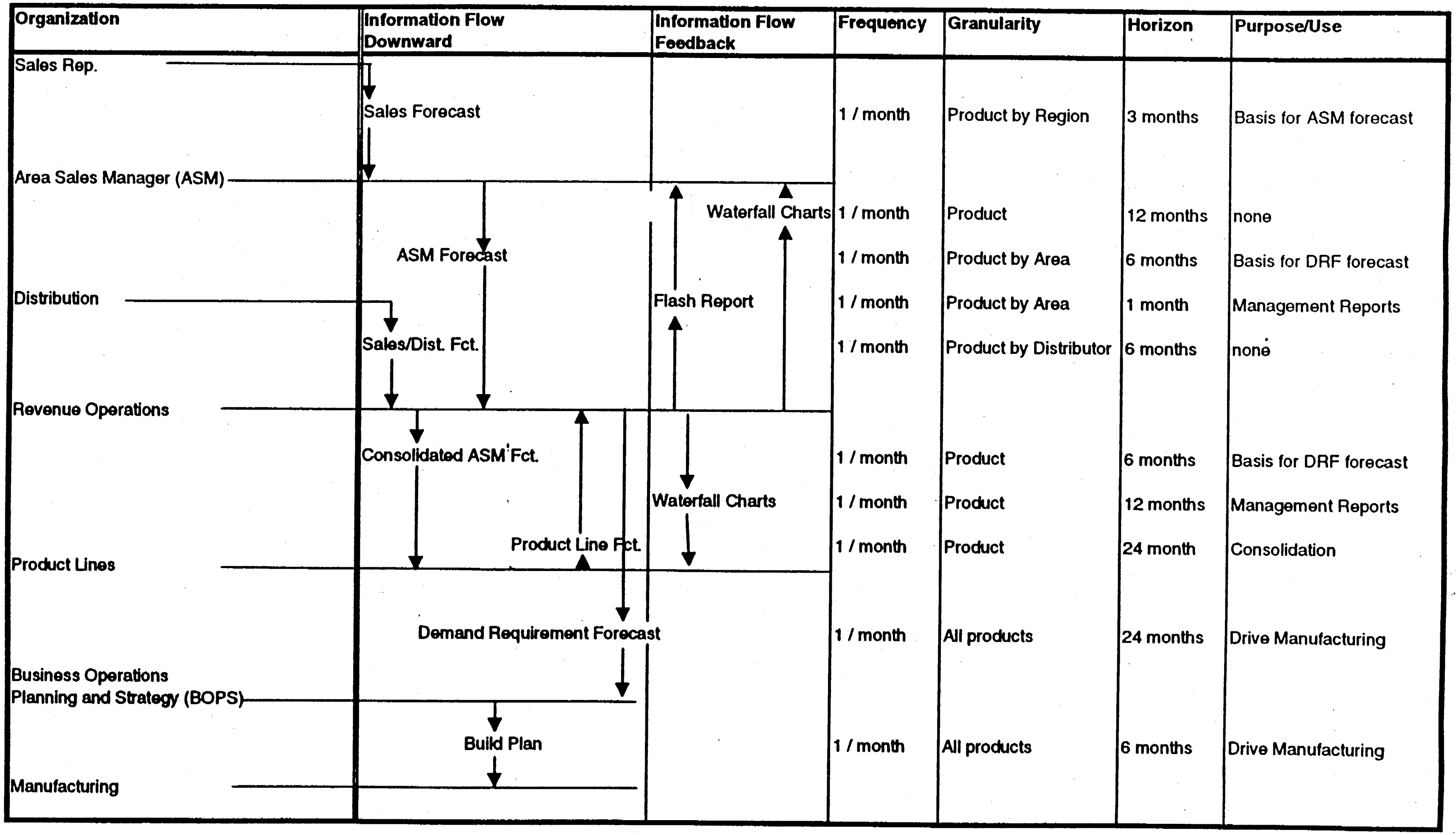

Figure 4: Information flow and Feedback

2.5 Forecast accuracy metrics

The only forecast accuracy metric that was used at Digital Semiconductor was the "waterfall chart". These charts compared the forecast accuracy on a rolling basis as far as 12 month out and were called "waterfall" charts due to their shape.

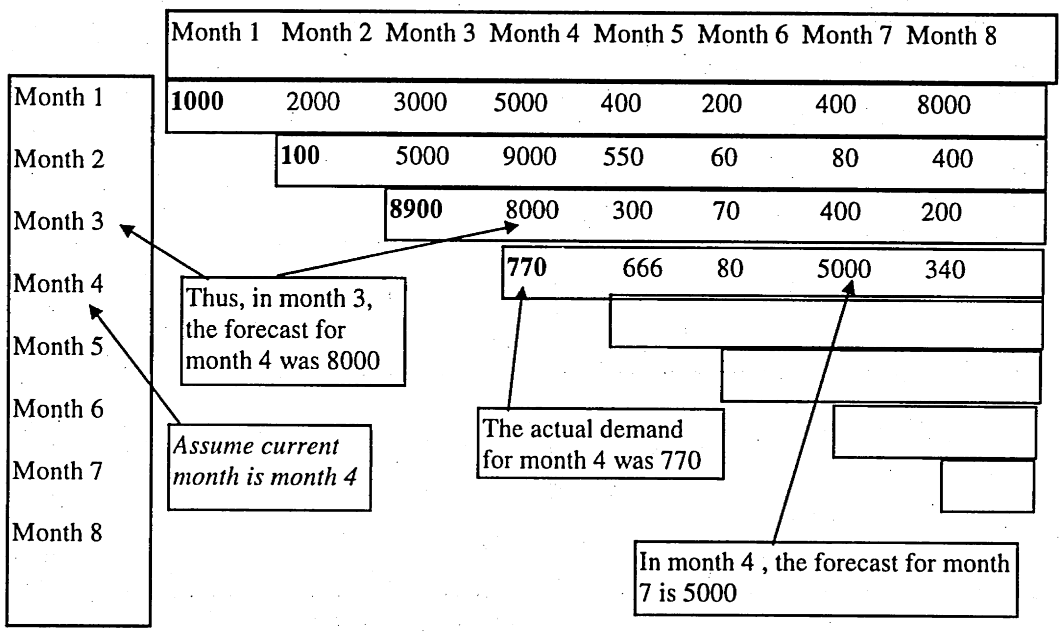

Figure 5: Example Waterfall Chart

The granularity for the "waterfall chart" was one month. Thus the chart documented the forecasted demand for a given month over various time horizons and when the month passed, the chart would compare all the previous forecasts with the actual demand. These charts were used for every device within each product line. Forecast accuracy for each bin or specific speed within a given part was not tracked. The layout of one of these charts is illustrated in figure 5. Further, for every month, the actual inventory was included as well for reporting purposes. Although useful, these waterfall charts were lacking some key elements. First, there were no confidence levels included in the forecast. Ideally, each forecast should have a level of certainty associated with it. Second, each number in the chart should have been given a range to which the confidence level applied. For example instead of simply predicting 8000 units for month 4 in month 4, the correct approach would have been to predict that the actual demand for month 4 was 8000 +/- 2000 with a 85 % certainty. The added range and confidence would allow much more effective planning and optimizing of the supply chain.

3. Constructing a supply chain model

Chapter 2 presented a detailed step by step description of the forecasting process and the information flow across organizations. The focus was on the forecast generation steps. In this chapter we take forecast inaccuracy as given and build a model of the internal supply chain of Digital Semiconductor in order to understand the effect of forecast error. The first section is a brief overview of the model purpose and a description of its capabilities. The second section delves into the dynamics of the supply chain. These occur due to decision heuristics carried out at Digital Semiconductor. This section also describes how the model attempts to replicate them. The next three subsections consist of a brief explanation of the three main stages of the supply chain followed by a description of how we modeled each of their key characteristics and the links between each of them. We then analyze past forecast accuracy and characterize it in a probabilistic form that can be input into the model of the supply chain. The last two sections of this chapter focus on how we incorporated the probabilistic description of forecast error into the model.

3.1 Model purpose and description of inputs and outputs

This section will begin with an overview of the model purpose as well as the general layout of the model. Then a brief description of the general model inputs and outputs will be given. The mathematical relationships behind these inputs and outputs is given in section 3.3.

<u>Model Purpose</u>

The purpose of the model is twofold. The first objective is to determine the cost of forecast inaccuracy using historical demand and historical forecasts. The second is to determine how best to trade off the various costs of forecast inaccuracy within the supply chain. In order to achieve these two purposes, the model needs to be a high level representation of Digital Semiconductor's internal supply chain that will be able to capture the dynamic effects of forecast error as they ripple through the supply chain over time.

Model Overview

The model simplifies the supply chain into the three main manufacturing stages that chips transition through before being shipped to the customer. The three stages are fabrication, assembly and test (Figure 6). The decision rules pertaining to late orders and expediting through each stage are modeled as well. Further each stage also has an associated cycle time and capacity. All stages are linked by inventory buffers which have associated target safety levels that the supply chain constantly strives to maintain.

The model is a discrete time stepping model of Digital Semiconductor's supply chain. The model spans 52 weeks with a time step of a week: each week the model determines how many units to start at each stage of the supply chain. To determine how many units the model desires to starts at each stage and during each week, the model trades off the following quantities:

- amount of material at each inventory storage location
- the forecast 1, 2 and 3 month out
- the backlog of customer orders
- the weekly inventory strategy at each inventory location

After determining the desired amount of material to start at each stage at each week, the model takes capacity and raw material constraints into account as well. Each stage can only start a maximum amount determined by either a capacity constraint or an upstream raw material constraint. The capacity constraints for each stage are variable each week. The model is a fairly accurate, yet simplified, representation of the actual supply chain. See section 4 for a more detailed discussion of the model applicability.

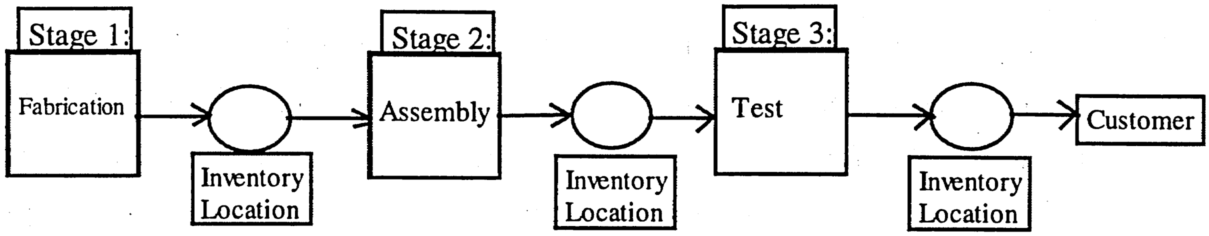

Figure 6: High Level Map of Supply Chain

Model Layout and Structure:

The model is an Excel based model. It is divided into three Excel sheets. Each sheet corresponds to a manufacturing stage and the inventory storage location downstream of it. The three sheets are as follow:

- sheet 1: Fabrication stage and die bank (DB) inventory location
- sheet 2: Assembly stage and ahead of test (AOT) inventory location
- sheet 3: Test stage and finished goods (FG) inventory location

Each sheet is comprised of input cells, output cells and calculation cells that relate the two. On each sheet the columns represent successive weeks whereas the rows represent the relevant inputs and outputs for each week (Figure 7). Many of the rows are calculation cells that relate the input and outputs of each sheet to each other and to the other sheets.

Model Inputs:

Inventory strategies:

For each inventory storage location (DB, AOT or FG) and for every week, the user of the model can enter a multiple of the forecasted demand for that week. This is the amount of material that will be available during that week at that inventory storage location as a safety stock above and beyond what the forecasted demand is. The multiple entered for that week will multiply the forecasted demand to determine the target buffer stock at that inventory location. The model will adjust to attempt to start enough material in the weeks prior to the week in question to guarantee that during that given week there is the requested safety stock.

The model will also react to changes in forecasts or inventory strategies. If the forecast for a given week decreased from one month to the next, causing there to be excess inventory, the model will take that excess into account when assessing the required starts to match the inventory strategy requested by the user for later weeks. The same reasoning applies to changes in inventory strategies. If the model user decreases his/her inventory strategy from one week to the next, the model will assess the extra inventory caused by changes in the inventory strategy when deciding how much to start during a given subsequent week. By following the logic described above, the model reproduces the thought process and actions that the planners at Digital Semiconductor go through to plan and smooth production.

Maximum starts capacity during the week:

Each week the user can specify the maximum number of units possibly started at each stage of the supply chain. The maximum specified by the user constitutes the capacity constraint at that stage for that week.

Actual Demand for Digital Semiconductor:

Each week the user can enter the amount requested by customers during that week. The customer requests come out of the finished goods inventory but are nevertheless communicated to all stages of the supply chain.

Forecast of actual ships from Digital Semiconductor

There are three forecasts that drive the model: forecasts one, two and three months out. Each week the user enters the forecasted amount for that week according to the 1, 2 and 3 months forecast.

The fab stage of the supply chain requires the three forecasts: 1, 2 and 3 months out. The assembly stage only requires the 1 month forecast and the test area does not require a forecast at all, since in the model, the test stage only has a 1 week window into the future. The model assumes that the one week forecast used by test is completely accurate.

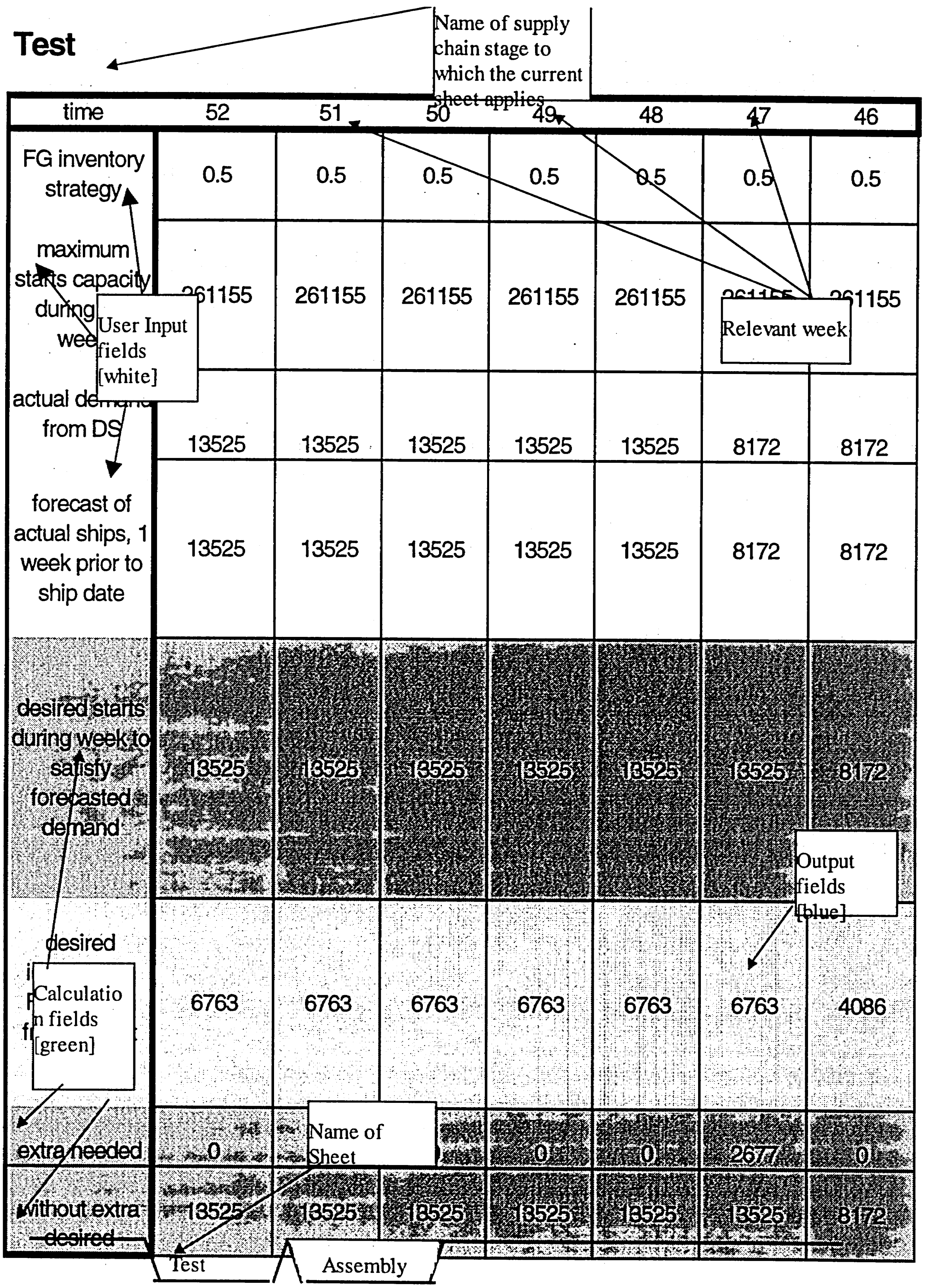

Figure 7: Model Layout

3.Model Outputs:

Desired starts during current week to satisfy forecasted demand, inventory strategy and backlog

This is the number of starts that the model desires during the week. The desired starts is determined by comparing the current WIP and inventory strategy at the given stage of the supply chain with the forecasted demand from that stage. The demand forecast that is relevant is the upcoming forecasted demand within the cycle time of the stage under consideration. (Figure 8) The desired starts do not take capacity and raw material into account. In figure 8 for example, the amount desired started during the current week is A + B, where A is the amount desired started due to past forecast error and B is the amount desired started to meet future demand.

Actual starts during the week

The actual starts during the week at each stage of the supply chain is the desired starts during the week limited by capacity and raw material. The model will start the desired amount provided there is enough capacity at that stage during that week and there is enough raw material in the inventory location just upstream of the given stage of the supply chain.

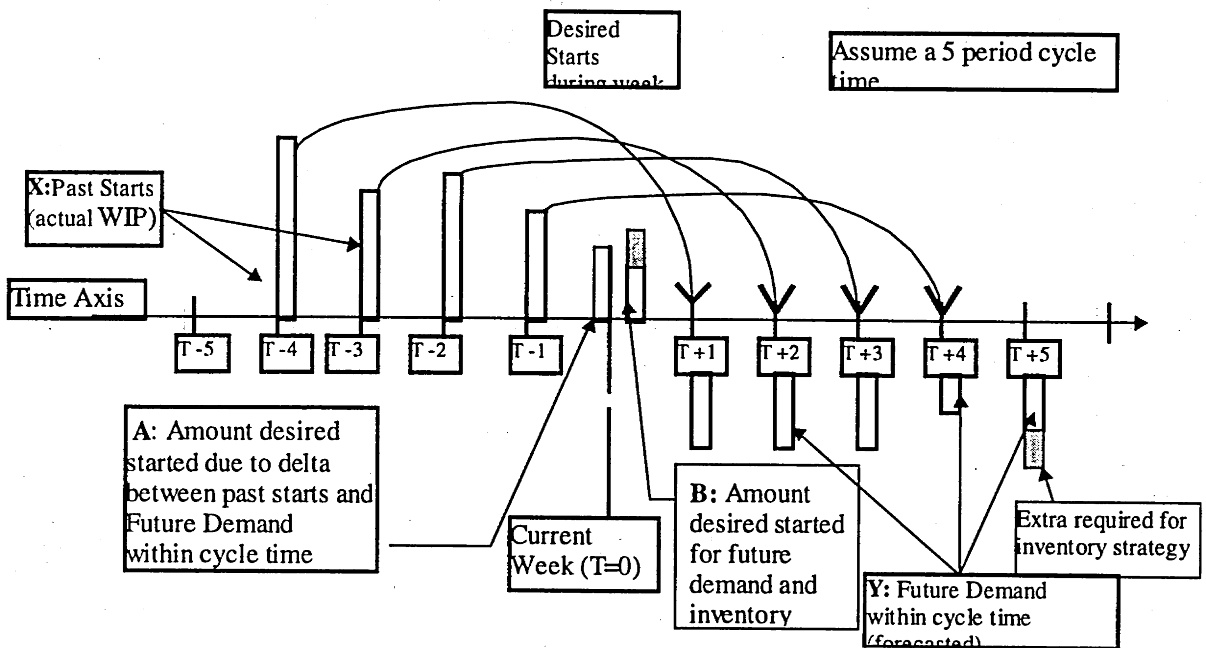

Figure 8: Illustration of the Model's Computation of the Desired Starts during a given week

If the desired amount exceeds one or both of the later quantities, the model will start as many parts as possible gated either by raw material or capacity, whichever is the lowest.

Inventory level at the beginning of the week

The amount of inventory at a stage of the supply chain during a given week is equal to the amount in inventory at the beginning of the previous week minus what was shipped to the downstream stage plus any parts that were started one cycle time prior to the given week.

Backlog that was created during the week

This is the amount of customer orders that were not met by the finished good inventory location during that week. If there was a demand during the week that was larger than the number of parts in the finished goods inventory at the beginning of the week, the deficit will be recorded here.

The quantity recorded in this line can be negative. A negative quantity signifies that there was more shipments out of finished goods than there was demand. In that case, the model over-shipped during the week to decrease the outstanding backlog that resulted from prior weeks.

Outstanding backlog at end of the week

This is the cumulated backlog at any given week. The cumulated backlog is the total number of outstanding orders that have not been met to date. It is the sum of the backlog that was created during the week from the beginning of the year to the current week.

Service level

The service level is the ratio of orders met on time to the total number of orders over the course of the whole year.

Weeks took to resolve backlog

This output indicates each week the amount of time it took the supply chain to resolve any backlog created during the current. The model looks forward in time and calculates how long the model required before the customer order that was only partially filled during that week was fully resolved.

3.2 Dynamics of the model

This section describes the overall dynamics of the supply chain. These dynamics occur because of the very intricate links between the three stages of the supply chain and also because of the inventory and expedite decision criteria prevalent at Digital Semiconductor. The model faithfully replicates these decision heuristics through the use of key variables and logical relationships between them. This section explains how these key variables within the model interact with the main inputs and outputs of the model, namely the number of starts and cumulative backlog each week and at each stage. The section begins with the definition of some key variables used in the model as well as an explanation of how they map onto the decision making heuristic of production planners at Digital Semiconductor. Then a simple example in a single stage supply chain is given and explained in these terms.

3.2.1 Definition and description of key terms and supply chain links

Forecasted shipments from Die Bank or Ahead of Test during current week

In the model there are actually several separate variables under this heading. Each variable corresponds to a different forecast. This variable translates forecasted customer demand from Digital Semiconductor into forecasted demand from a specific inventory holding facility within the supply chain. For example, production planners for the fabrication facility at Digital Semiconductor will use the forecasted demand to then calculate when the fabrication stage will have to have product ready to ship to the downstream stage. There is therefore a shift in time of customer demand. This shift amounts to the cycle time of the downstream stages. (Figure 9)

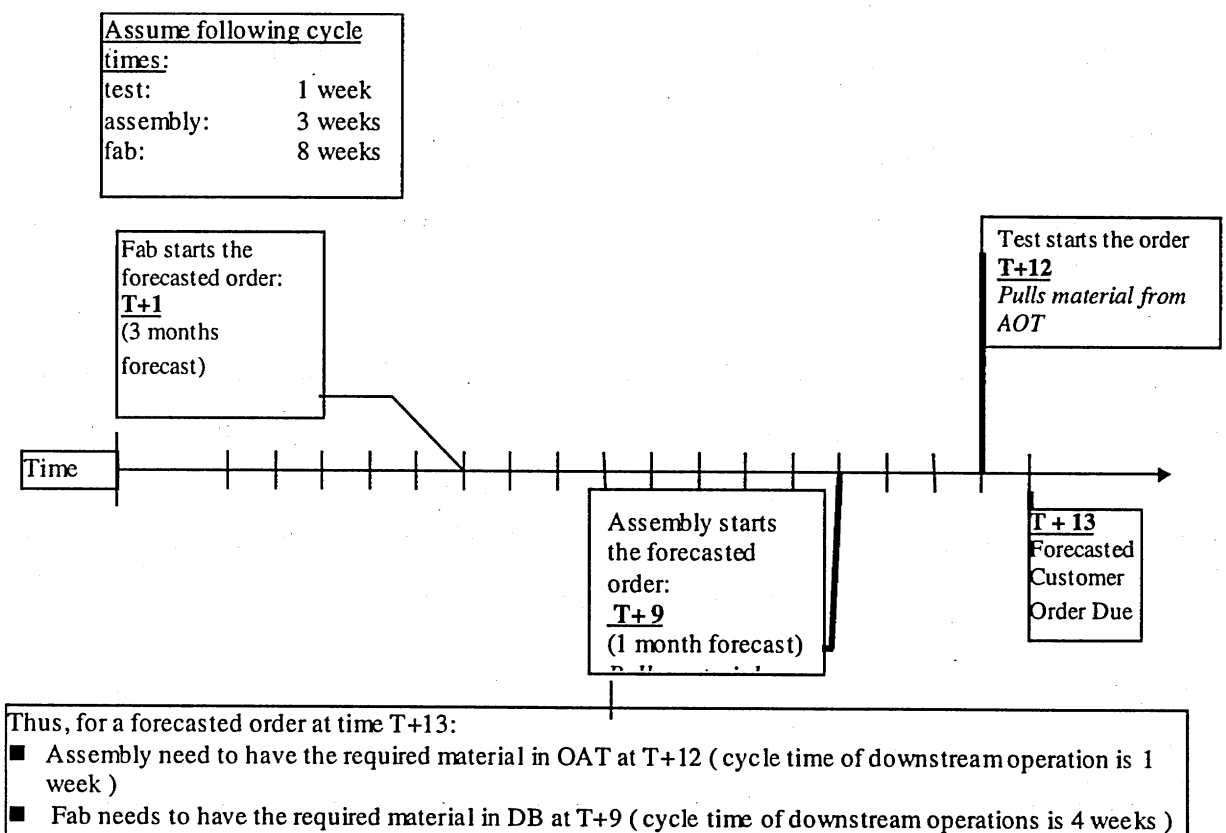

Figure 9:Illustration of the shift in time of the forecasted customer demand in the supply chain

Demand seen by the fabrication facility from assembly during the week

This is the amount of material that the assembly stage desires to pull from the Die bank inventory storage location during the week. If there is not enough material at the beginning of the week to satisfy that demand, the demand from assembly will differ from the actual shipments from die bank to the assembly during the week. If the assembly stage demands are greater than the upstream capacity, the demand will default to the capacity of the assembly stage. The same applies to the interrelationship between the test stage and its upstream inventory storage location, Ahead of Test.

Amount expedited

This output quantifies the amount of material that was expedited through the assembly operation. Expedited material has a cycle time that is two thirds that of a non-expedited material. The assembly operation occurs overseas. Thus an expedited part is simply one that has a shorted transit time.

In the model material, is expedited under the following conditions: if there is spare capacity and raw material and if there is either backlog created during the current week and/or an outstanding customer backlog. Much like planners at Digital Semiconductor, only under these conditions will the model authorize expediting.

Extra needed

This is the number of starts driven by inventory strategy. It is the amount of starts desired above and beyond both the amount required for the forecasted demand and the current backlog. Extra needed captures the number of starts required due to changes in either the actual inventory strategy multiplier and/or changes in the forecasted demand which also lead to changes in the level of safety inventory requirements.

Cumulative backlog of extra at end of period

This is the number of added starts required by the inventory strategy that were not realized because of capacity, backlog or demand constraints. The number can be either positive or negative. A positive number reflects the cumulative number of starts that were required to meet the inventory strategy but that were not actually started. There are several causes for the creation of a positive backlog: not enough capacity or raw material to release the added starts required to fulfill the inventory strategy in the face of actual demand and an actual backlog of orders, both of which take precedence over inventory strategy.

A negative backlog reflects the cumulative number of reductions in starts caused by changes in inventory strategy. As the inventory strategy or the forecasted demand drops, production planners realize that Digital Semiconductor is holding more inventory than its inventory strategy requires. Therefore the planner will want to start less than the forecasted demand to bring its inventory level down. However, if the forecasted amount is smaller than the reduction in starts required to bring inventory levels down, the planner cannot starts negative wafers so he/she will simply register the difference between the desired reduction and the amount he/she would have started without consideration of the inventory strategy as a negative backlog. During the next weeks, the planner will strive to reduce future starts and bring the backlog up. The model exactly duplicates this behavior (Figure 10).

Changes in backlog of extra

This is the amount by which the backlog of extra changes. There are two separate entries in the model for the two distinct cases where extra is positive and extra is negative. The formulation behind those two cases differ.

The model compares the desired amount without the extra starts with the desired amount taking the extra starts into account. The model will also take the amount that the model was capable of starting that week into account. These last two quantities will differ if the model is constrained by capacity or raw material that week. Further, the model will compensate for any Extra starts that have not been addressed in the past few weeks. The change in backlog of extra during the week will be reflected by the Cumulative backlog of extra in the following week. Through these comparisons the model duplicates the behavior of a production planner faced with an erroneous or uncertain forecast (Figure 10).

Extra realized

This is the amount of extra starts that the model could address this week. The amount in this cell can differ from the desired amount of extras depending on whether the model is constrained by raw material or capacity. The amount in this cell is the amount that the model refers to in future weeks to balance the WIP with the forecasted demand. *Extra realized* is the model's way to capture the thought process of a planner who realizes that he/she has released some material into a given stage of the supply chain only to meet inventory buffer targets, NOT to meet actual demand. Thus, the added amount that was released in the supply chain needs to be taken account by the planner when projecting future WIP requirements.

3.2.2 Illustrative example

This example is for illustration purposes only. It applies to a single stage supply chain, whereas the model and Digital Semiconductor's has three stages. Further, the example shown does not incorporate forecast error. It assumes a perfectly accurate forecast. The model developed for Digital Semiconductor does incorporate forecast error. However, even though the example is a simplified one, it illustrates the basic phenomena occurring at Digital Semiconductor's supply chain. The behavior of the key variables defined in section 3.2.1. is illustrated in figure 10. There are four phases in the figure. The first, phase A, has the system in equilibrium. The second, phase B, sees an unchanging customer demand but in increase in the inventory strategy through an increase in the targeted "parts on hand" desired. This increase in inventory strategy, causes positive desired extra starts and an immediate increase in the desired inventory level. This positive number of desired extra causes an increase in the number of wafers released in the supply chain. These starts are not to meet extra demand but rather are caused by the increase in inventory strategy. Further, not all the desired extra starts can be made in the next time period because of a capacity constraint. Because not all the desired extra starts are completed, there is also a positive backlog of extra that is formed. Because there was an increase in the number of parts released there is a gradual increase in the inventory level. After two time periods all the desired extra that was required because of an increase in the inventory strategy has been made, the new higher inventory level has been reached and the number of parts released each time period returns to equilibrium.

During the third phase, phase C, the inventory strategy decreases and the process operates in reverse. There now is a negative Extra starts and there are no more parts released in the supply chain because the inventory level has to be decreased in the face of unchanging demand. However, the drop in parts started is not sufficient to compensate for negative extra starts. There can be no negative starts. Because the decrease in starts was insufficient to meet the negative starts due to the change in inventory strategy, there is a negative Extra Backlog.

Phase D is the only phase during which there is an actual increase in customer demand. The effects throughout the supply chain are identical to an increase in inventory strategy. During these four phases there was no forecast error and no creation of backlog, both of which are present in the model presented in this thesis. Backlog would increase the desired starts in the system until the backlog decreases through shipments above and beyond current demand.

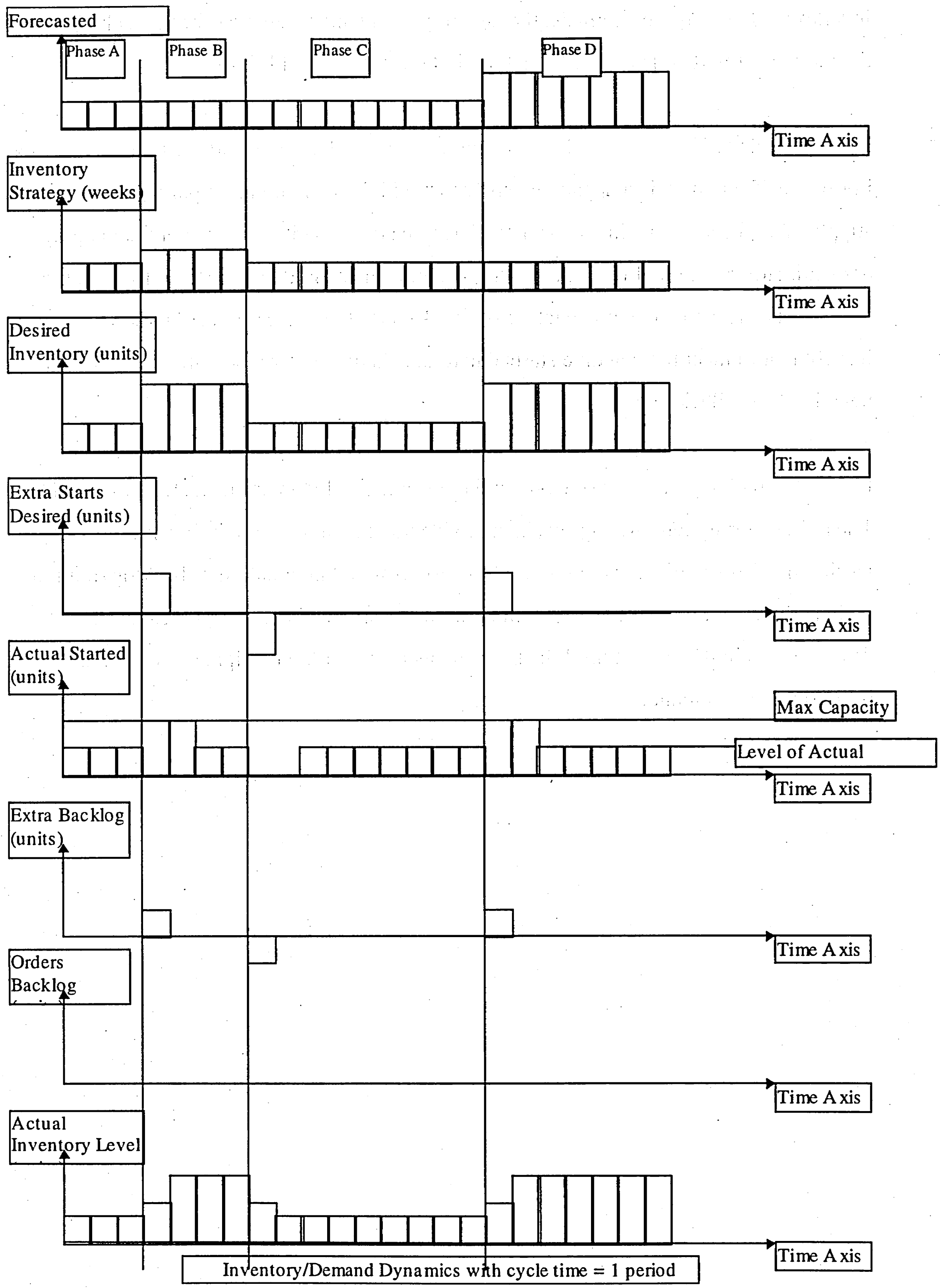

Figure 10: Inventory/Demand Dynamics with cycle time = 1 period

3.3 Mathematical relationships at each stage of the supply chain

This section describes the characteristics of each stage of the supply chain and details the modeling techniques and assumptions that were used. For the mathematical descriptions of the relationships between the different variables in this supply chain model the following convention was used:

- subscripts: time index
- superscripts: stage index
- t: time variable
- X: generic variable
- MAX(A,B): largest of A or B
- If(statement, A, B): formula which will take the value of A if the statement is true and take the value of B if the statement is false
- $\Sigma_{T=1}^{T=t}$ (expression): summation over time of the expression from T=1 to T=t
- MIN(A,B): smallest of A or B

Further, the following abbreviations were used to describe the relationships between the variables of interest:

- Desired starts, or parts released, for a given stage at a given time t: DS_t^{stage}
- Shipments required from a given stage of the supply chain during the current week to meet customer demand according to the 4, 8 or 12 week forecast: Sn_t^{stage} where n is 4 weeks, 8 week or 12 weeks to represent the different forecasts of interests.
- Inventory strategy for a given stage at time t: In_t^{stage} where n is 1 for die bank (DB), 2 for ahead of test (AOT) and 3 for finished goods (FG).
- Actual number of parts released, or started at a given stage during week t: R_t^{stage}
- Number of parts that would be released at each stage without consideration of the backlog of extra starts during week t: RE_t^{stage}

See section 3.2 for an explanation of the terms extra starts and backlog of extra starts.

- Number of additional parts that need to be released because of a backlog of extra parts during week t: RN_t^{stage}
- Changes in the backlog of extra, if extra is negative during week t: CEN_t^{stage}
- Backlog of orders that was created during week t: B_t^{stage}
- Changes in backlog of extra starts during week t: CB_t^{stage}
- Changes in backlog of extra starts, if extra is positive during week t: CBP_t^{stage}
- Changes in backlog of extra starts, if extra is positive during week t: CBN_t^{stage}
- Cumulative backlog of extra starts during week t: SB_t^{stage}
- Amount of extra starts that was released during week t at a given stage: ER_t^{stage}
- Desired amount released in order to meet forecasted inventory and backlog requirements during week t: $DFIB_t^{stage}$
- Amount released in the system during week t, while being capacity limited: RC_t^{stage}
- Amount desired released during week t, including parts started because of extra starts: WD_t^{stage}
- Cumulative backlog of orders at the end of week t: SB_t^{stage}
- Maximum possibly released during week t: MR_t^{stage}
- Desired amount released in order to meet forecasted inventory, backlog and expediting requirements during week t: $RIBE_t^{stage}$
- Inventory level during week t: I_t^{stage}
- Shipment from stage n: Hn_t where n is 1 for fab, 2 for assembly and 3 for test
- New backlog of order created during week t: NB_t^{stage}
- Desired inventory level during week t: DI_t^{stage}
- Forecasted shipment from stage n, during week t and according for a forecast m weeks hence: $FmHn_t$
- Demand from a given stage during time t: D_t^{stage}
- Maximum starts capacity allowable during week t: M_t^{stage}
- Maximum possible started during week t: MR_t^{stage}
- Amount expedited during week t: E_t^{stage}
- Amount desired expedited during week t: DE_t^{stage}

- Customer demand during week t: CD_t
- Number of extra parts needed during week t because of either changes of inventory strategy of changes in forecasts: EN_t^{stage}
- Number of parts desired released without taking extra parts or backlog of extra parts into account: WOD_t^{stage}

For example the formula for the extra starts at time t that are needed for changes in forecasted demand is as follows:

$$= (S4_{t+4}^{assembly} - S4_{t+3}^{assembly}) * (I2_{t+3}^{test} + I2_{t+3}^{assembly})$$

3.3.1 Fabrication/Probe and Die Bank

The fabrication and probe stage is the first stage of the supply chain and has the longest cycle time of the three steps in the supply chain. There is no upstream inventory storage location and the model assumes that this stage is never starved. This assumption is a good one since lack of raw product is rarely an issue at this stage of the supply chain. The main inputs into this stage are the one, two and three month forecast. The model does not allow expediting through this stage of the supply chain. In the actual supply chain there is the possibility to have product expedited. However, the quantities that can be expedited vary tremendously by product line. For some product lines at Digital Semiconductor the amount of die that can be expedited is negligible whereas for others that is not the case. These large difference occur because of variations in die per wafer and yields across the product lines.

Desired starts during the week to satisfy forecasted demand in assembly: $DS_t^{assembly}$

$$DS_t^{assembly} = S4_{t+4}^{assembly}$$

Desired starts during the week to satisfy forecasted demand in fab: DS_t^{fab}

$$DS_t^{fab} = S12_{t+12}^{fab}$$

Desired inventory in DB, 8 weeks from current week: DI_t^{aab}

$DI_t^{fab} = I1_{t+12}^{fab} * DS_t^{fab}$

Shipments from DB during current week (4 week forecast): $F4H1_t$

$F4H1_t = S4^{t+4}_{assembly}$

Shipment from DB during current week at time t (8 week prior): $F8H1_t$

$F8H1_t = S8_{t+4}^{fab}$

Extra needed in fab: RN_t^{fab}

Extra needed at time t is the sum of the added quantity required due to inventory strategy changes and the added quantity required due to changes in forecasted demand. In the fab, these extra starts have to take both the inventory strategy of the assembly stage and the test stage into account. The general structure of the formula has two parts: part 1 accounts for the changes in forecasted demand and part 2 accounts for the changes in inventory strategies.

Extra needed due to inventory strategies = part 1 + part 2

Part 1:

In general:

(changes in forecasted demand)*(inventory strategy of fab + inventory strategy of assembly + inventory strategy of test)

Specifically:

Part 1 of $RN_t^{fab} = (S12_{t+12}^{fab} - S12_{t+11}^{fab}) * (I2_{t+11}^{assembly} + I3_{t+11}^{test} + I1_{t+11}^{fab})$

Part 2:

In general:

(inventory strategy changes from week to week in DB) * (forecasted demand) +

(inventory strategy changes from week to week in OAT) * (forecasted demand) +

(inventory strategy changes from week to week in FG) * (forecasted demand)

Specifically:

$$\text{Part 2 of } RN_t^{fab} = (S12_{t+11}^{fab}) * (I2_{t+12}^{assembly} - I2_{t+11}^{assembly}) + (S12_{t+11}^{fab}) * (I1_{t+12}^{fab} - I1_{t+11}^{fab}) + (S12_{t+11}^{fab}) * (I3_{t+12}^{test} - I3_{t+11}^{test})$$

With extra desired: WD_t^{stage}

$$WD_t^{stage} = RN_t^{fab} + RE_t^{fab}$$

Changes in backlog of extra if extra is positive: CEP_t^{fab}

The formula for changes in backlog of extra has two parts. The first applies if the actual number of starts made during the week is positive while the second applies if the actual number of starts made during the week is zero. The difference stems from the requirement that there cannot be negative starts. Thus when the number of starts is 0, the desired starts is most likely a negative number. This negative number will be reflected in the desired starts and must be truncated to compare the actual starts with the desired starts in a meaningful fashion.

$$CEP_t^{fab} = IF(R_t^{fab} > 0,\ IF((R_t^{fab} - RE_t^{fab}) <= 0, RN_t^{fab}, RN_t^{fab} - (R_t^{fab} - RE_t^{fab})) + IF((R_t^{fab} - MAX(RE_t^{fab}, 0)) <= 0, RN_t^{fab}, RN_t^{fab} - (R_t^{fab} - MAX(RE_t^{fab}, 0))))$$

Changes in backlog of extra if extra is negative: CEN_t^{fab}

The formula is separated in three parts. The first part applies if the Extra starts required is different than 0 and the amount desired without the extra starts is positive, the second part applies if the Extra starts required is different than 0 and the amount desired without the extra starts is negative and the third applies if the extra starts required is 0. (Figure 11)

$$CEN_t^{fab} = IF(AND(RE_t^{fab} > 0, RN_t^{fab} <> 0), IF(RE_t^{fab} >= -RN_t^{fab}, 0, RE_t^{fab} + RN_t^{fab}), 0) + IF(AND(RE_t^{fab} < 0, RN_t^{fab} <> 0), RN_t^{fab}, 0) + IF(RN_t^{fab} = 0, -R_t^{fab} + MAX(RE_t^{fab}, 0) - MAX(0, B_t^{fab}, 0), 0)$$

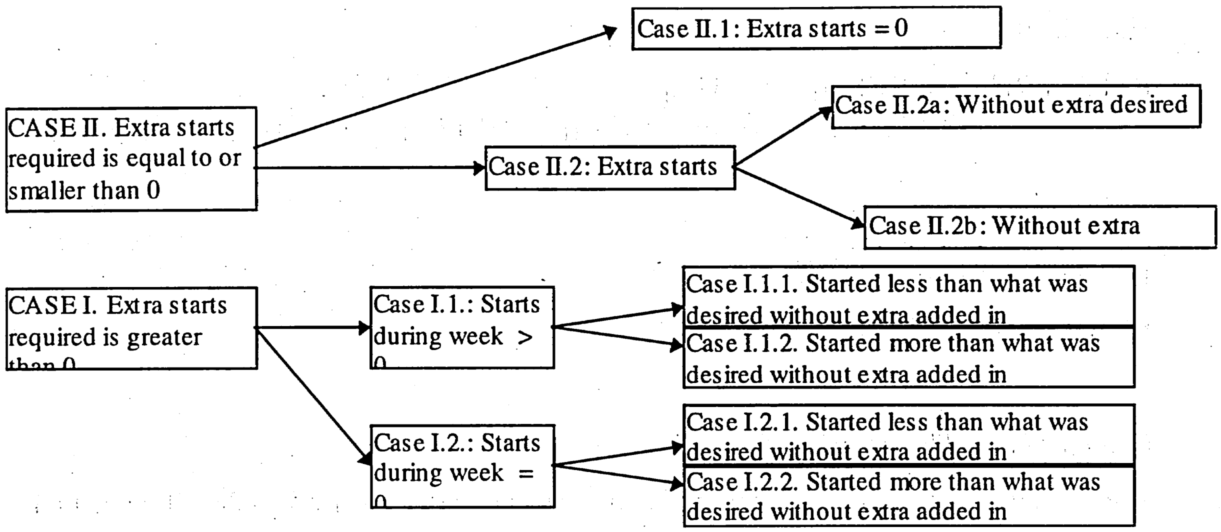

Figure 11: Logic flow of equation for the change of the backlog of extra

Changes in backlog of extra: CB_t^{stage}

This entry simply picks which of the cases in figure 11 to use depending on the sign of the extra starts required:

$CB_t^{fab} = IF(RN_t^{fab} <= 0, CEN_t^{fab}, CEP_t^{fab})$

Cumulative backlog of extra at end of period: SB_t^{stage}

This is simply the sum from week 1 to present of the changes in the backlog of extra:

$SB_t^{fab} = CB_t^{fab} + SB_{t-1}^{fab}$

Extras realized: ER_t^{stage}

$ER_t^{fab} = RN_t^{fab} - CB_t^{fab}$

Desired starts during the week to satisfy forecasted demand and inventory strategy:DFI

This entry attempts to balance the WIP with the forecasted demand. The forecasted demand from die bank at any time t is broken down into three categories: short term (week t +1 and t +2), medium term (week t+3,t+4, t+5) and long term (week t+6, week t+6)

For the short term the model looks at the 4 week forecast, for the medium term the model looks at the 8 week forecast and for the long term the model looks at the 12 week forecast.

The WIP is simply the amount of material that was started in the past 7 weeks minus the amount started because of inventory strategies throughout the supply chain.

Thus in general we have:

amount required started = forecasted demand from DB 8 weeks from now

+ forecasted demand from DB in the next 7 weeks

- amount started in the past 7 weeks

+ amount started because of inventory strategies

Specifically the formula is:

$$DFI_t^{fab} = DS_t^{fab} + \Sigma_{T=t+6}^{T=t+7}(S12_t^{fab}) + \Sigma_{T=t+3}^{T=t+5}(S8_T^{fab}) + \Sigma_{T=t+1}^{T=t+2}(S4_T^{fab}) - \Sigma_{T=t-7}^{T=t-1}(R_T^{fab}) + \Sigma_{T=t-7}^{T=t-1}(ER_T^{fab})$$

Desired starts during the current week to satisfy forecasted demand, inventory strategy and backlog: $DFIB_t^{stage}$

$$DFIB_t^{fab} = DFI_t^{fab} + SB_{t-1}^{fab}$$

Starts made during the week, capacity limited: RC_t^{fab}

At this stage the model adds With extra desired to the cumulative backlog of extra starts desired due to inventory strategy. After adding them the model will start the maximum of the following two quantities: maximum capacity for the week or With extra desired added to the cumulative backlog of extra required for the inventory strategy. Further, the model will also assure that there will be no negative starts even though the inventory strategy or changes in forecasted demand might warrant it. For the fab stage the model assumes that there is no constraint on the raw material available.

Note that With extra desired is the sum of the starts required for the forecasted demand and the starts required for the order backlog.

The specific formula the number of starts made, capacity limited is:

$$RC_t^{fab} = MAX(IF (WD_t^{fab} + SB_{t-1}^{fab} > MR_t^{fab} ,\ MR_t^{fab} , WD_t^{fab} + SB_{t-1}^{fab}) , 0)$$

Demand from assembly operation during the week: $D_t^{assembly}$

This records what the demand from the upstream operation is. However, if the upstream operation demands are greater than the upstream capacity, the demand will default to the upstream capacity. Further, the model will not allow negative demand from the upstream stage.

$$D_t^{assembly} = MAX (MIN (RIBE^t_{assembly}, MR_t^{assembly}) , 0)$$

Inventory level in DB at the beginning of the week: I_t^{stage}

$$I_t^{fab} = I_{t-1}^{fab} - SH1_t^{fab} + R_{t-8}^{fab}$$

Shipments from DB to assembly during week: SH1

$$SH1_t^{fab} = R_t^{assembly}$$

Backlog created during week and cumulative backlog: NB_t^{fab}

This is the same backlog and backlog creation as in the test stage.

$$NB_t^{fab} = NB_t^{test}$$

3.3.2 Assembly and Ahead of Test

The assembly stage of the supply chain is done offshore through a subcontractor. Thus parts can be expedited through this stage by reducing the shipping time. The cycle time of expedited parts is two thirds of that of non expedited parts. The added costs of expediting is significant. The basic mathematical formulas underpinning the modeling of this stage are similar to the fabrication stage. However the assembly interfaces with both the upstream operation and inventory storage location, (Fabrication and Die Bank) as well as the downstream operation, test. This interface causes some of the formulas from the fabrication stage to be altered as well as the creation of some new ones. These new or altered formulas are explained below.

Desired starts during the week to satisfy forecasted demand in assembly: DS_t^{stage}

$DS_t^{assembly} = S4_{t+4}^{assembly}$

Desired inventory in AOT, 3 weeks from current week: DI_t^{stage}

$DI_t^{AOT} = I2_{t+4}^{assembly} * S4_t^{assembly}$

Forecasted shipments from AOT during current week (1 week forecast) $F1H2_t$

$F1H2_t = S1^{t+1}_{test}$

Forecasted Shipment from AOT during current week at time t (4 weeks prior): $F4H2_t$

$F4H2_t = S4_{t+1}^{assembly}$

Extra needed in assembl: $EN_t^{assembly}$

Extra needed at time t is the sum of the added quantity required due to inventory strategy changes and the added quantity required due to changes in forecasted demand. In assembly these extra starts have to take the inventory strategy of the assembly stage and the test stage into account. The general structure of the formula has two parts: part 1 accounts for the changes in forecasted demand and part 2 accounts for the changes in inventory strategies.

Extra needed due to inventory strategies = part 1 + part 2

<u>Part 1</u>:

In general:

(changes in forecasted demand)*(inventory strategy of assembly + inventory strategy of test)

Specifically:

$(S8_{t+8}^{assembly} - S8_{t+7}^{assembly}) * (I2_{t+7}^{assembly} + I3_{t+7}^{test})$

Part 2:

In general:

(inventory strategy changes from week to week in OAT) * (forecasted demand) +

(inventory strategy changes from week to week in FG) * (forecasted demand)

Specifically:

$(S8_{t+8}^{assembly}) * (I2_{t+8}^{assembly} - I2_{t+7}^{assembly}) + (S8_{t+8}^{assembly}) * (I3_{t+8}^{test} - I3_{t+7}^{test})$

Without extra desired: $WOD_t^{assembly}$

This is simply a copy of the Desired starts during the current week to satisfy both forecasted demand, inventory strategy and backlog for that same week.

With extra desired: $WD_t^{assembly}$

$WD_t^{assembly} = EN_t^{assembly} + RE_t^{assembly}$

Changes in backlog of extra if extra is positive: $CBP_t^{assembly}$

The formula for changes in the backlog of extra has two parts. The first applies if the change in the number of parts released during the week is positive while the second applies if the number of parts released during the week is zero. The difference stems from the requirement that there cannot be negative starts. Thus when the number of parts released is 0, the desired starts was most likely a negative number for that week. This negative number will be reflected in the desired starts and must be truncated to compare the actual number of parts released with the desired number of parts released.

$CBP_t^{assembly} = \text{part 1} + \text{part 2}$

Where

Part 1 = $IF(R_t^{assembly} > 0, IF((R_t^{assembly} - WOD_t^{assembly}) <= 0, EN_t^{assembly},$
$EN_t^{assembly} - (R_t^{assembly} - WOD_t^{assembly}))$

Part 2 = $IF((R_t^{assembly} - MAX(WOD_t^{assembly}, 0)) <= 0, EN_t^{assembly},$
$EN_t^{assembly} - (R_t^{assembly} - MAX(WOD_t^{assembly}, 0))))$

Changes in backlog of extra if extra is negative: $CBN_t^{assembly}$

The formula is separated into three parts. The first part applies if the Extra starts required is different than 0 and the amount of desired parts released with the extra starts is positive. The second part applies if the extra starts required is different than 0 and the amount desired without the extra starts is negative. The third applies if the etxra starts required is 0.

$CBN_t^{assembly}$ = part 1 + part 2 + part 3

Where

$$\text{Part 1} = \text{IF (AND(} WOD_t^{assembly} > 0, EN_t^{assembly} <> 0\text{)},$$

$$\text{IF (}WOD_t^{assembly}, \backslash >= -EN_t^{assembly}, 0,$$

$$WOD_t^{assembly} + EN_t^{assembly}\text{)},0)$$

$$\text{Part 2} = \text{IF(AND(} WD_t^{assembly} < 0, EN_t^{assembly} <> 0\text{)}, EN_t^{assembly}, 0\text{)}$$

$$\text{Part 3} = \text{IF (} EN_t^{assembly} = 0, -RC_t^{assembly} + \text{MAX(} WOD_t^{assembly}, 0\text{)},$$

$$-\text{MAX(} 0, B_t^{assembly}, 0\text{)}, 0\text{)}$$

Changes in backlog of extra: $CB_t^{assembly}$

This entry in the model picks which of the cases, extra positive or extra negative, is appropriate for the current week:

$$CB_t^{assembly} = \text{IF(} EN_t^{assembly} <= 0, CBN_t^{assembly}, CBP_t^{assembly}\text{)}$$

Cumulative backlog of extra at the end of current week: $SB_t^{assembly}$

This is simply the sum from week 1 to present week of the changes in the backlog of extra:

$$SB_t^{assembly} = CB_t^{assembly} + SB_{t-1}^{assembly}$$

Extra realized: $ER_t^{assembly}$

This is the amount of parts released due to a backlog of extra.

$$ER_t^{assembly} = EN_t^{assembly} - CB_t^{assembly}$$

Desired starts during the week to satisfy forecasted demand and inventory strategy

This entry attempts to balance the WIP with the forecasted demand. The forecasted demand from AOT at any time t is forecasted demand in the upcoming 3 weeks. The WIP is simply the amount of material that was started and expedited in the past 2 weeks minus the amount started because of inventory strategies throughout the supply chain.

Thus in general we have:

amount required started = forecasted demand from AOT 3 weeks from now

+ forecasted demand from AOT in the next 2 weeks

- amount started in the past 2 weeks

- amount expedited in the past 2 weeks

+ amount started because of inventory strategies in past 2 weeks

Specifically the formula is:

$$DS_t^{assembly} = DF_t^{assembly} + \Sigma_{T=t+1}^{T=t+2}(F4H2_T^{assembly}) - \Sigma_{T=t-2}^{T=t-1}(R_T^{assembly}) - \Sigma_{T=t-2}^{T=t-1}(E_T^{assembly}) + \Sigma_{T=t-2}^{T=t-1}(ER_T^{assembly})$$

Maximum possibly started during the week: $M_t^{assembly}$

The model will compare the amount of raw material with the weekly capacity. The smaller of the two number will be the maximum possibly started during the week:

$$M_t^{assembly} = MIN\,(M_t^{assembly}, I_t^{fab})$$

Demand made of DB during the week: $D_t^{Assembly}$

The model quantifies how much material the assembly stage wants to pull from DB during the week. This amount might differ from the actual amount pulled from DB if there is not enough material in DB at the beginning of the week.

$D_t^{assembly} = MAX(IF(DFIBE_t^{assembly} > M_t^{assembly}, M_t^{assembly}, DIBE_t^{assembly}),0)$

Actual starts during the period: R_t^{stage}

The model will start the desired amount or the *maximum possibly started* depending on which is greater. The model will not allow negative starts.

$R_t^{stage} = MAX(IF(MR_t^{assembly} > DIBE_t^{assembly}, MR_t^{assembly}, DIBE_t^{assembly}),0)$

Production available in 3 weeks: $P3_t$

This is the amount of starts that were not expedited.

$P3_t = IF(DE_t^{assembly}>0, MAX(0,R_t^{assembly} - DE_t^{assembly}), R_t^{assembly})$

Amount expedited: E_t

This is the number of starts released during the week that were expedited

$E_t = IF(DE_t^{assembly}>0, MIN(DE_t^{assembly}, MR_t^{assembly}), 0)$

Demand from test made of AOT during the current week: D_t^{test}

$D_t^{test} = MAX(0, RC_t^{test})$

Backlog requested by test + demand from test made of AOT during the week

$backlog+demand = D_t^{test} + SB_t^{assembly}$

Shipments made from AOT to test operation during the week: $H2_t$

$H2_t = MR_t^{test}$

Inventory in AOT at the beginning of week

$I_t^{assembly} = I_{t-1}^{assembly} - H2_{t-1}^{assembly} + P3_{t-3}^{assembly} + E_{t-2}^{assembly}$

3.3.3 Test and Finished Goods

The test and finished good part of the model is relatively simple. The cycle time of this stage is modeled as one week. There is no forecast error and the only constraint to meeting the customer demand occurring during the week is either lack of material from the upstream inventory storage location or insufficient capacity. When orders are not met, the backlog increases and vice versa when the backorders are filled. This backlog is communicated to the other stages of the model. Most of the equations are similar to the ones found in the fabrication and assembly stages. The more relevant equations are as follows:

Desired inventory in FG, 1 week from current week: DI_{t+1}^{test}

$DI_t^{test} = I3_{t+1}^{test} * F1H3_t$

Desired starts during the week to satisfy forecasted demand: DS_t^{test}

$DS_t^{test} = F1H3_{t+1}$

Extra needed in test: EN_t^{test}

Extra needed at time t is the sum of the added quantity required due to inventory strategy changes and the added quantity required due to changes in forecasted demand. In test these extra starts have to only take the inventory strategy of test stage into account. The general structure of the formula has two parts: part 1 accounts for the changes in forecasted demand and part 2 accounts for the changes in inventory strategies.

Extra needed due to inventory strategies = part 1 + part 2

Part 1:

In general:

(changes in forecasted demand)*(inventory strategy of test)

Specifically:

$(F1H3_{t+2}^{test} - F1H3_{t+1}^{test}) * (I2_{t+1}^{assembly})$

Part 2:

In general:

(inventory strategy changes from week to week in FG) * (forecasted demand)

Specifically:

$(F1H3_{t+1}^{test}) * (I3_{t+1}^{test} - I3_t^{test})$

Cumulative backlog of extra at end of period: SB_t^{test}

This is simply the sum from week 1 to present of the changes in the backlog of extra:
$SB_t^{test} = CB_t^{test} + CB_{t-1}^{test}$

Shipments out of FG during the week: $H3_t$

The model will ship the number demanded during the week, provided there is enough material in FG. Any deficit will be recorded as a creation of backlog for that week.

$H3_t = IF(I_t^{test} > CD_t + SB_t^{test}, CD_t + SB_t^{test}, I_t^{test})$

Backlog created during the week: NB_t^{test}

$NB_t^{test} = CD_t^{test} - H3_t$

Cumulative backlog at end of period: SB_t^{test}

$SB_t^{test} = SB_t^{test} + NB_t^{test}$

3.4 Characterizing past forecast accuracy

The demand forecasts are the key inputs into the model. Each stage of the supply chain will base the number of desired starts during the week on the forecast rather than the actual demand. The full cycle time spans almost a quarter whereas the forecast is updated once a month and the decision as to how many parts to release at each stage is made once a week. Thus the number of starts made each week no only reflects what needs to be started today to meet customer orders at the end of the cycle time but also reflects updates or corrections to the past forecasts. Therefore, the relevant forecasts for our model are not only those spanning the cycle time, but also those forecasts with shorter time horizons. The later serve to correct the former and cause the supply chain to react. The relevant forecasts are the 1, 2 and 3 months forecasts.

In order to characterize the forecast error we plotted past forecasts for the whole fiscal year 1997 (Figure 12) As a measure of error we tracked the ratio of actual demand to the forecasted demand for forecasts with a 1, 2 and 3 months time horizon. Thus unity on the x axis indicates a forecast that was on target. Where a number larger than one indicates a forecast that over predicted what the demand was. We plotted the frequency with which the forecasts fell within each range of error as indicated on the x-axis. The resulting histogram resembles a lognormal distribution heavily skewed towards under-forecasting. Therefore we can approximate the short term forecast accuracy with such a distribution if we can determine appropriate values for the mean and standard deviation. In order to find these defining variables, we need to logarithmically transform the distribution of figure 12. We took the natural logarithm of all data points. The resulting distribution of such a transformation is shown in figure 13. As expected the data now appears more Gaussian and simple calculation reveals the mean and standard deviations. (Table 2) These values will be used in the model to understand the tradeoffs of various costs of forecast inaccuracy.

	Forecast, 1month horizon	Forecast, 2months horizon	Forecast, 3months horizon
Mode in Lognormal Distribution	0.85	0.65	0.6
Mean in Normal Distribution	-0.070	-0.187	-0.221
Standard Deviation in Normal Distribution	0.21	0.18	0.2
Percent forecasts larger than 2X actual demand	12	15	19

Table 2: Statistics of Forecast Error

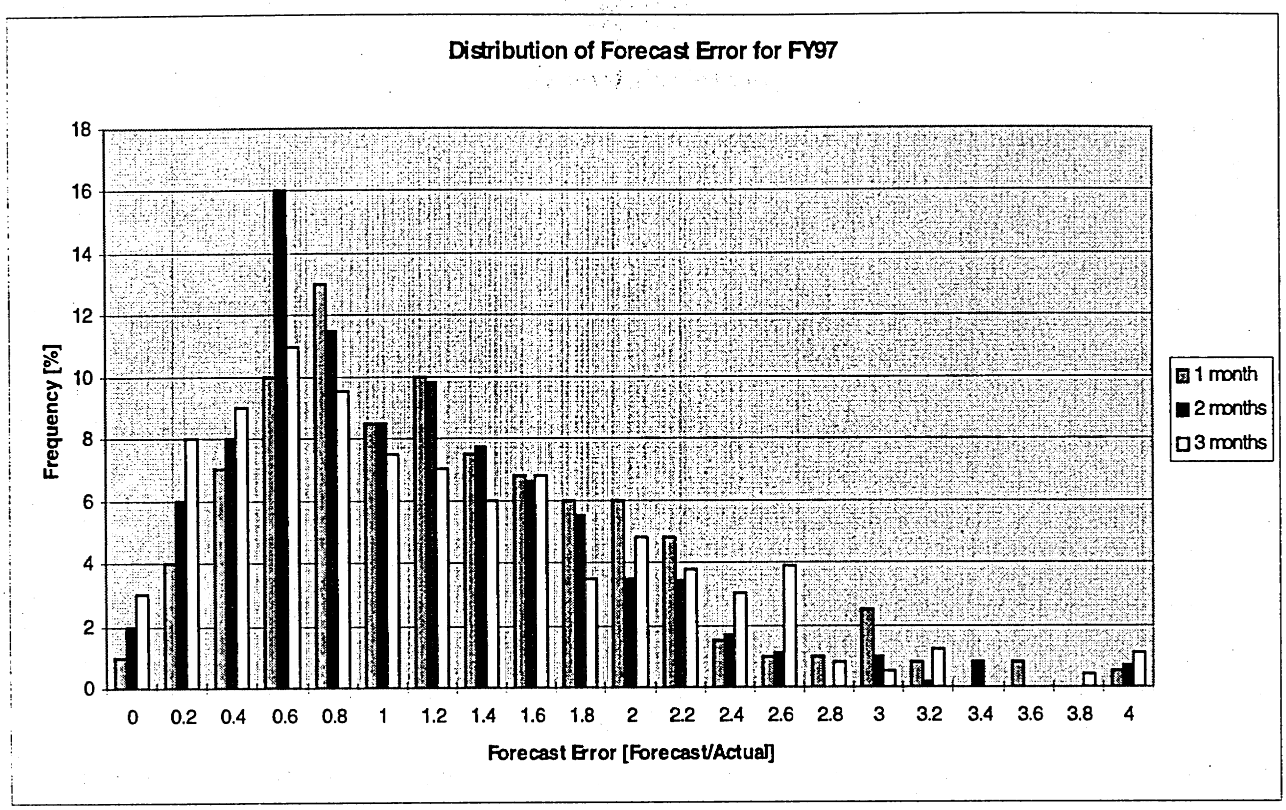

Figure 12: Distribution of Forecast Error for FY97

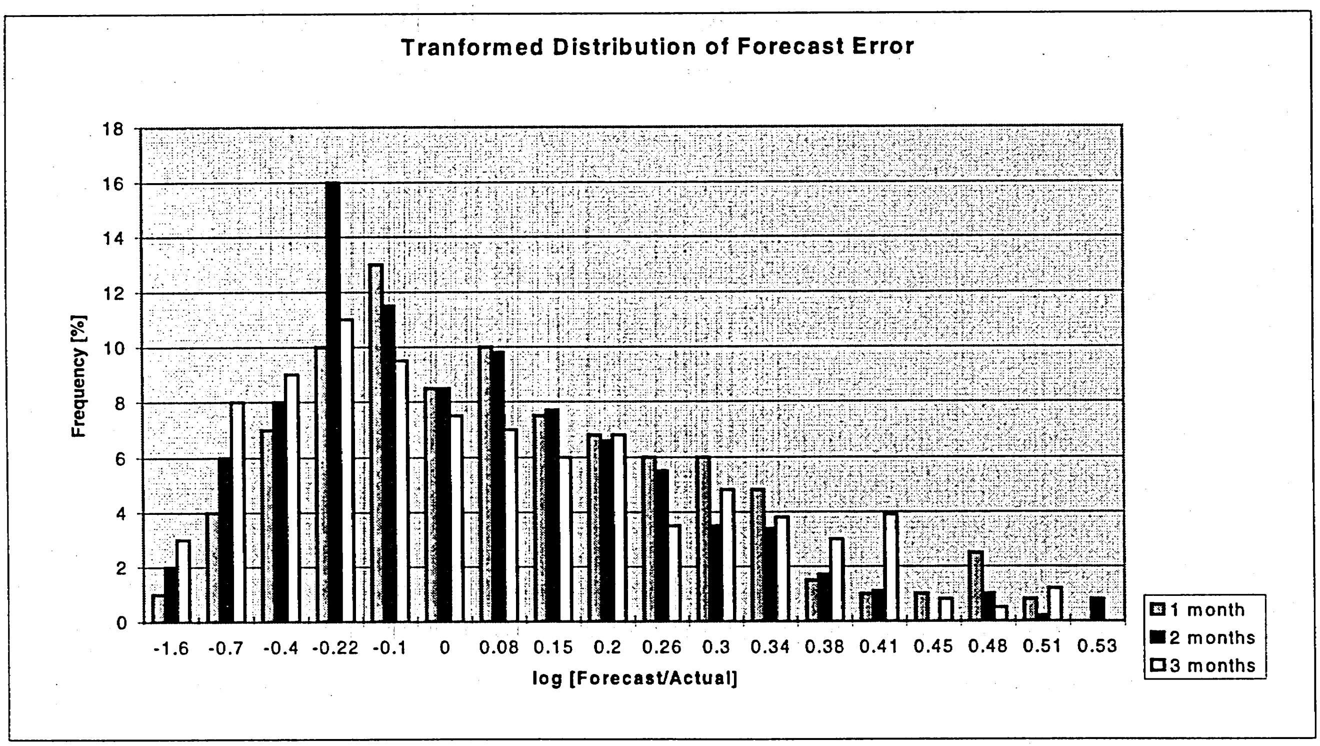

Figure 13: Transformed distribution of forecast error for FY97

3.5 Role and effect of forecasts

There are three forecasts that drive the release of material into the internal supply chain of Digital Semiconductor. Given that the cycle time of most products is 3 months, the 3 months forecast dictates how much material should be started today in order to meet customer demand at the end of the overall cycle time. However, the three months forecast contains errors. These errors are progressively discovered at the time horizon of the forecast decreases. Thus, the two month and one month forecasts allow the internal supply chain of Digital Semiconductor to learn about and react to errors in the three month forecast before the forecasted date occurs. This section details and explains how the one, two and three months forecasts affect the release of material at each stage of the supply chain. In order to detail the effect of each individual forecast, we ran some simulations using simple cases as illustration of the dynamics of the model. Included in the illustrations are outputs from the model.

In all the simulation runs the customer demand is fixed throughout the year at 10000 units per week. For the purpose of simplicity we set the capacity at each stage to be infinite. The inventory strategies that are included in the model are very similar to the ones used at Digital Semiconductor: a target of 6 weeks parts on hand in Finished Goods (FG) and a target of 4 weeks parts on hand in die bank (DB). The inventory strategy targets are based on the forecasted demand. Further, there was no inventory target set for the inventory location Ahead of Test (AOT).

3.5.1 Forecast 1 month out

In this case we set the forecasts one, two and three months out to be equal to the actual demand (10000 units) for all the week except for week 20. For week 20 we set the one month forecast to be only 9500 units for that week (0.95 of actual). We then tracked the resulting change in the number of parts released by each stage of the supply chain: fabrication, assembly and probe. The results are in figure 14.

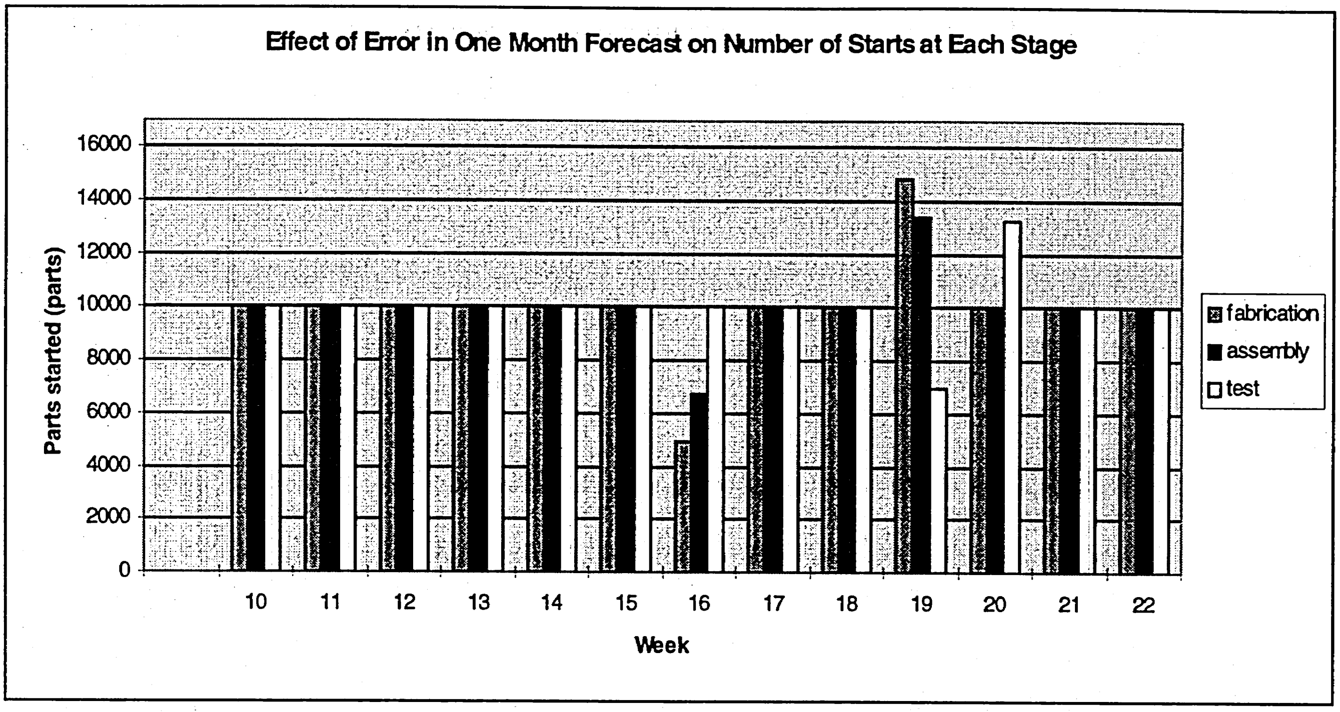

Figure 14: Model output for the case of error in the 1 month forecast.

As illustrated in the figure, in the case of the one month forecast being inaccurate all stage of the supply chain are affected.

- Fabrication stage: up until week 16 the fabrication stage is in steady state producing 10000 units per week. Referring to the three month forecast, the fab stage releases 10000 parts in week 8 to satisfy the forecasted demand for week 20, assuming a 12 week cycle time. As week 12 approaches the fabrication facility now has access to the two month forecast for week 20. The fabrication stage compares this forecast with what has previously been started and stays on track. However, during week 16, the fabrication stage compares the one month forecast with the work in process and realizes that according to the one month forecast it has released too much material into the supply chain. The fabrication stage thus releases less material than what the latest 3 month forecast predicts will be needed during week 28, 12 weeks hence, in order to bring to balance of WIP versus the demand in 4 weeks back into equilibrium.

However, as week 20 approaches, the fabrication facility again has to readjust the WIP level to the actual demand incurred during that week. Thus during week 19, the model now releases 15000 parts because the actual demand is now known. The variation in the number of parts started each week is 10 times greater than 500, which was the original error in the forecast. This factor of 10 is due to the inventory strategies in the supply chain, 4 week parts on hand in DB and 6 weeks parts on hand in FG.

- The Assembly stage: the assembly stage undergoes a similar change in the number of parts released each week. There is a sharp decrease during week 16 as the one month forecast predicts only a demand of 9500 parts for week 20. As week 20 approaches though, and the actual demand is known in week 19 to actually be 10000, the assembly stage attempts to reestablish the balance of what was started in the past and the actual demand by starting 13500 during week 19.
- The Test stage: this stage decides how much material to release each week based on the one week schedule which the model assumes is accurate. Yet even though this stage is driven by an accurate forecast, we still observe a sharp variation in the number of parts started during the weeks 19 and 20. This is due to the fact that the test stage is lacking the sufficient materials to release the required amount of parts during week 19 to meet the demand during week 20, even though the test stage has an accurate forecast. Thus the following week the test stage releases more material than is required to replenish the inventory level in the finished goods area.

Thus, errors in the one month forecast affect all stages of the supply chain, either directly or indirectly.

3.5.2 Forecast 2 months out

In this case we kept all the forecasts at an accurate level of 10000 units except for week 20 in the two month forecast. For week 20, according to the two month forecast, the predicted demand is 95 % of actual. The results are in figure 15.

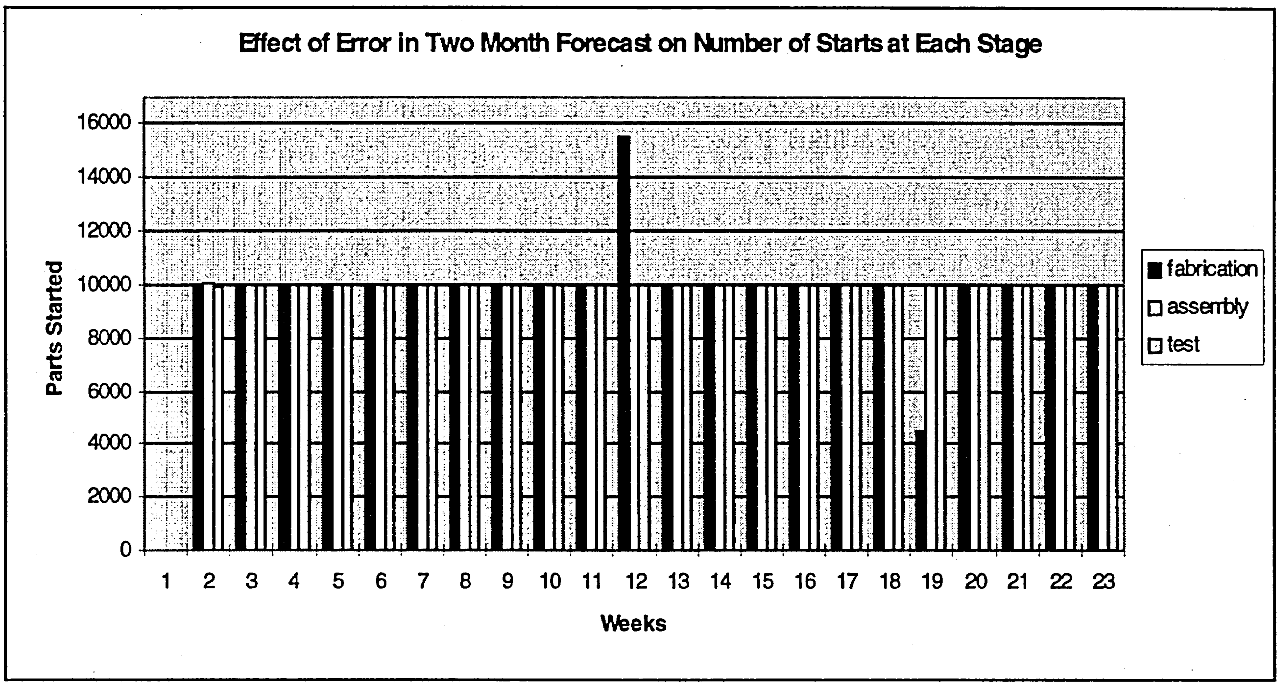

Figure 15: Model output for the case of the error in the 2 months forecast

As is clear by the comparison of figures 14 and 15, the effects of the identical error in the one and two month forecast are dramatically different.

- Fabrication stage: this stage is the only one affected by the error in the two months forecast. During week 12 the fabrication stage decreases the number of parts started in order to balance the forecasted demand with the total number of parts already released into the supply chain. However, as the one month forecast becomes available, there is a correction of equal magnitude in the opposite direction.

- Assembly and test stages: these elements of the supply chain are not affected because they are only driven by the one month forecast and the one week forecast because their respective cycle times are 1 and 3 weeks. Even though the fabrication stage of the supply chain is dramatically varying the number of parts started because of variations in the two months forecast, the test and assembly stage are buffered from these variations by the large inventory target in place for DB.

3.5.3 Forecast 3 months out

In this case we kept the one and two months forecasts accurate while introducing a small error in the three months forecast during week 20 of our simulation. The results were very similar to the case in 3.5.2. The fabrication stage responded to the change in the forecasted demand according to the three month forecast during week 8. However, as the two month forecast became available, the fabrication facility decreased its starts for week 12 in order to bring the balance of WIP and forecasted demand back into equilibrium. Even though the forecast error was only 5 % of actual, the variations in the number of parts released each week was 50 % of actual due to the inventory strategies.

4. Costs of forecast inaccuracy

This section takes forecast inaccuracy as a given. The purpose of this chapter is to both quantify the costs of forecast inaccuracy as well as understand the tradeoffs between them in order to minimize the total cost throughout the supply chain. The chapter begins with a description of the main costs of forecast inaccuracy analyzed in this thesis. Then, beyond a simple description, we use the model of the supply chain developed earlier to understand the complex tradeoffs that exist between these five costs. Finally, we report the added costs incurred throughout the supply chain due to the actual forecast error over FY97.

In our analysis we define forecast error as the variance of the lognormal distribution that characterizes the forecasts one, two and three months out. (Section 3.5) In our characterization of the effect of forecast error on the costs incurred throughout the supply chain, we assumed that the mode of the distribution characterizing the forecasts remained unchanged. We only varied the variance around that mode and mean. Increasing variance was considered equivalent to an increase in forecast inaccuracy.

4.1 Five Costs of forecast inaccuracy

The five main costs of forecast error are: added lead time, lower service level, added capacity, added inventory and the opportunity cost of manufacturing product A when in actuality product B was required to meet demand. We will describe each of these costs in detail.

Added lead time: In our model, lead time is defined as the length of time between the customer request date and customer ship date. This time could be zero in our model, meaning that the customer demand during a given week was met during that same week. However, most often when the forecast is inaccurate the leadtime begins to grow as the supply chain is incapable of meeting unexpected customer demand during the week. There then is the creation of a customer demand backlog, which the supply chain eventually has to ship against. Thus, if all else remains equal an inaccurate forecast can have no effect on any other aspect of the supply chain if the lead times can grow without loss of customers. In our model we did not include any customer balking. Thus, no matter how long the leadtime becomes, customers do no desist to Digital Semiconductor's competitors, rather they patiently wait for their products. At the extreme, the leadtime can become as long as the manufacturing cycle time which would cause short term forecasts to be unnecessary. The model can be used as a management tool to understand the effect of forecast error on lead times. However, The next step of correlating leadtime and loss of customers is beyond the scope of the model.

Lower service level: In our model we define service level as the percent of customer demand that was met during the same week as the demand occurred in. Thus, as the forecast accuracy decreases so does the customer service levels, all else kept equal. The model does not differentiate between partial or complete orders. The model simply aggregates total demand and calculates for a given week, how much of that total demand was the supply chain able to meet.

Added inventory: The single largest costs of forecast inaccuracy is the added inventory that the supply chain must carry is order to maintain high service levels and low lead times despite inaccurate forecasts. In order to maintain these high customer satisfaction metrics while having a forecast that is inaccurate, it is necessary to build buffers between the many stages of the supply chain. As the size of these buffers increase, so does the margin for error that the forecast can have while the supply chain still meets customer demand on time.

Added capacity: This is the added capacity that was required at each stage of the supply chain because of forecast error. In our model we assume a constant capacity throughout the year. Although the model has the capability to do so, we set the capacity at a chosen constant level throughout the period of study (FY97). The added capacity required because of forecast error could have many sources depending on the nature of the forecast error and the demand pattern:

- Added capacity was required to expedite material
- Material that consumed capacity was released according to a forecast that over predicted demand. This effect is compounded by the current inventory strategies in place.
- After under predicting demand, an increase in the backlog occurs, which causes stages of the supply chain to push through more material in a shorter period a time than if the forecast was accurate and smoothing of production could occur over a longer period of time.

Opportunity cost: The opportunity cost is defined as the capacity that was used to manufacture part A when the actual demand called for part B. We did not consider this cost in our analysis. (Chapter 6)

4.2 Tradeoffs between four costs of forecast inaccuracy

In this section we report the result of numerical experiments we ran using the model of the supply chain. Even though the results presented in this section are theoretical, they do shed some light onto the problem because the mathematical model on which they are based accurately describes the constraints and general characteristics of the actual supply chain of Digital Semiconductor. These numerical experiments help us understand the tradeoffs between the various costs of forecast inaccuracy. Thus, with these results, managers can determine what tactics are best suited to minimize the total costs of poor forecasts as well as determine the costs associated with achieving their strategic goals. For example, a manager might decide that customer service levels cannot drop below 90% for the business to be viable. With this model, he/she can understand the various costs that he/she will have to bear to achieve this target.

4.2.1 Treatment of demand and forecasts in our numerical experiments

Treatment of Demand:

The results that follow are not based on a stochastic description of demand. Rather, we chose a demand pattern that was fairly typical of the ones facing Digital Semiconductor.

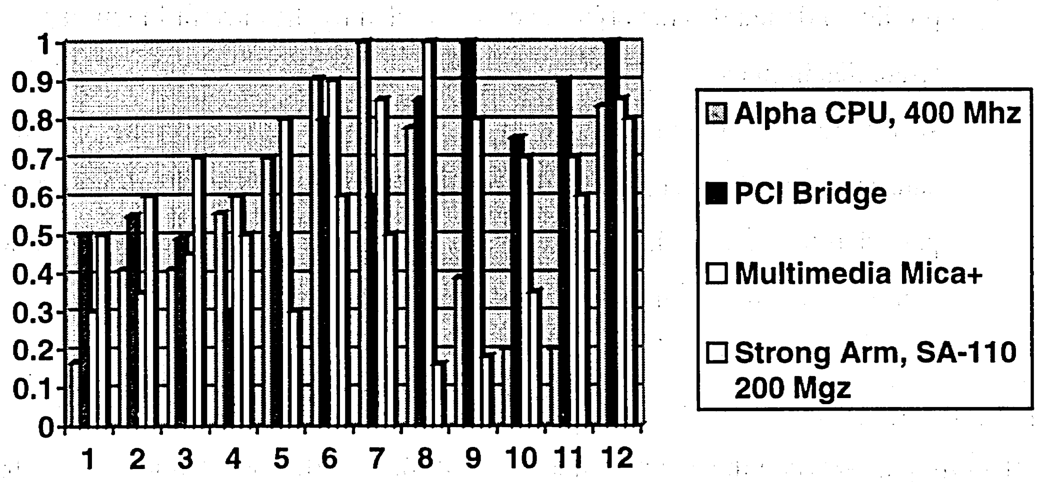

Figure 16: Normalized Monthly Demand Patterns for 4 Typical D.S. Parts

This demand pattern was then used as an input to our model and an invariant throughout our numerical experiments. In figure 16 the demand pattern we input into the model was normalized and the outputs made relative to the input signal. Thus, the results can be generalized to most demand patterns faced by Digital Semiconductor. The specific demand pattern we chose to base our calculations off of was the demand pattern of the Alpha CPU, 400 Mhz.

Treatment of Forecasts:

In order to input the forecasts into our model we used the probabilistic distributions described in section 3.4. As a proxi for a measure of forecast inaccuracy we varied the variance of these forecast distributions while maintaining the same mean and mode. In

the following descriptions "forecast error" will refer to the variance in the lognormal distribution of the forecasts. For each data point we ran the model 1000 times and thus obtained a second variance, this one associated with each output. We varied all three forecast together in a uniform fashion for most of the experiments.
All other inputs such as capacity and inventory strategy were considered deterministic.

4.2.2 Effect of Error on Service Level

We varied the error from half of what it was during FY97, to one and a quarter times the actual forecast error that occurred in the past year. We observed the effect of varying the error in this way had on the average service level. Further, the standard deviation of our output is included in figure 17 as well in the form of error bars representing two standard deviations from the mean. For this experiment we kept inventory strategies and capacity constant, at 6 weeks and infinity respectively. As error decreases the average service level increases, as expected. Further, the variance in our output decreases as well. The difference in service level from an error of three quarters historical error to an error of half historical level is statistically significant to 95 % confidence level. As a rule of thumb, a decrease in the forecast error from historical levels to half of that will increase the service level 10 %, without added inventory or capacity.

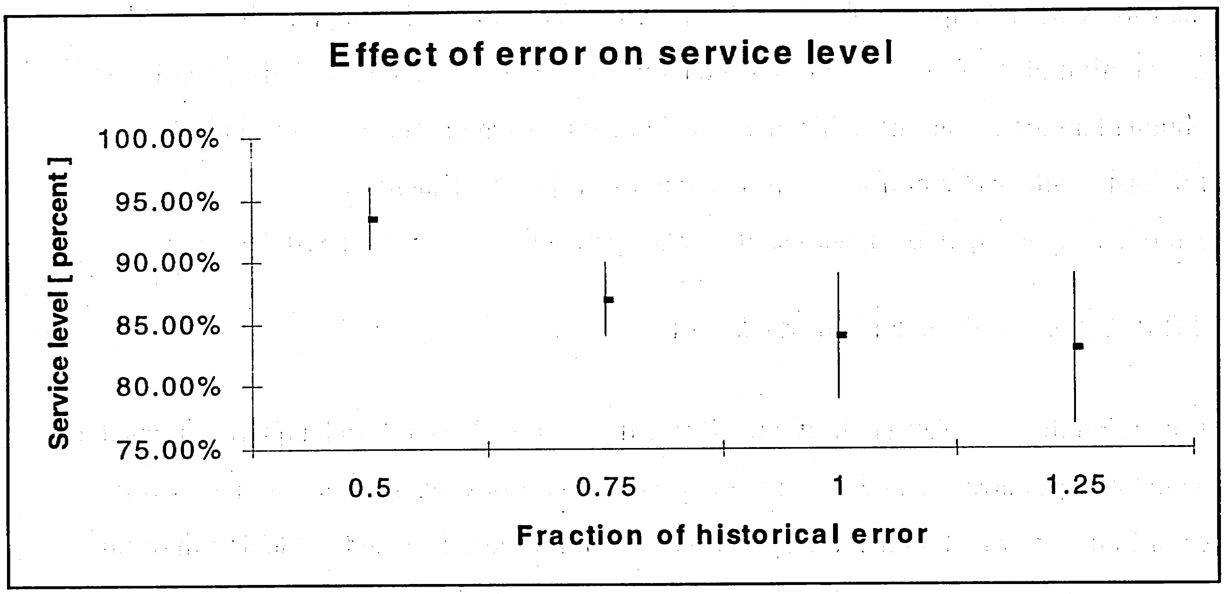

Figure 17: Effect of Forecast Error on Service Level and Output Variance

The results in figure 18 yields two conclusions:

- First, as indicated earlier, as error decreases, average service level increases. However, we learn here that the previously reported gain of 10 % service level that comes with the halving of the error still holds for inventory strategies varying from 4 to 8 weeks parts on hand.
- Second, keeping the error constant, an increase of inventory strategy from 4 to 8 weeks, increases the average service level by approximately 12 %, as long as the error is not larger than the historical error. The marginal improvement of service levels with increasing inventory strategy decreases quickly if the error is larger than five fourth historical error.

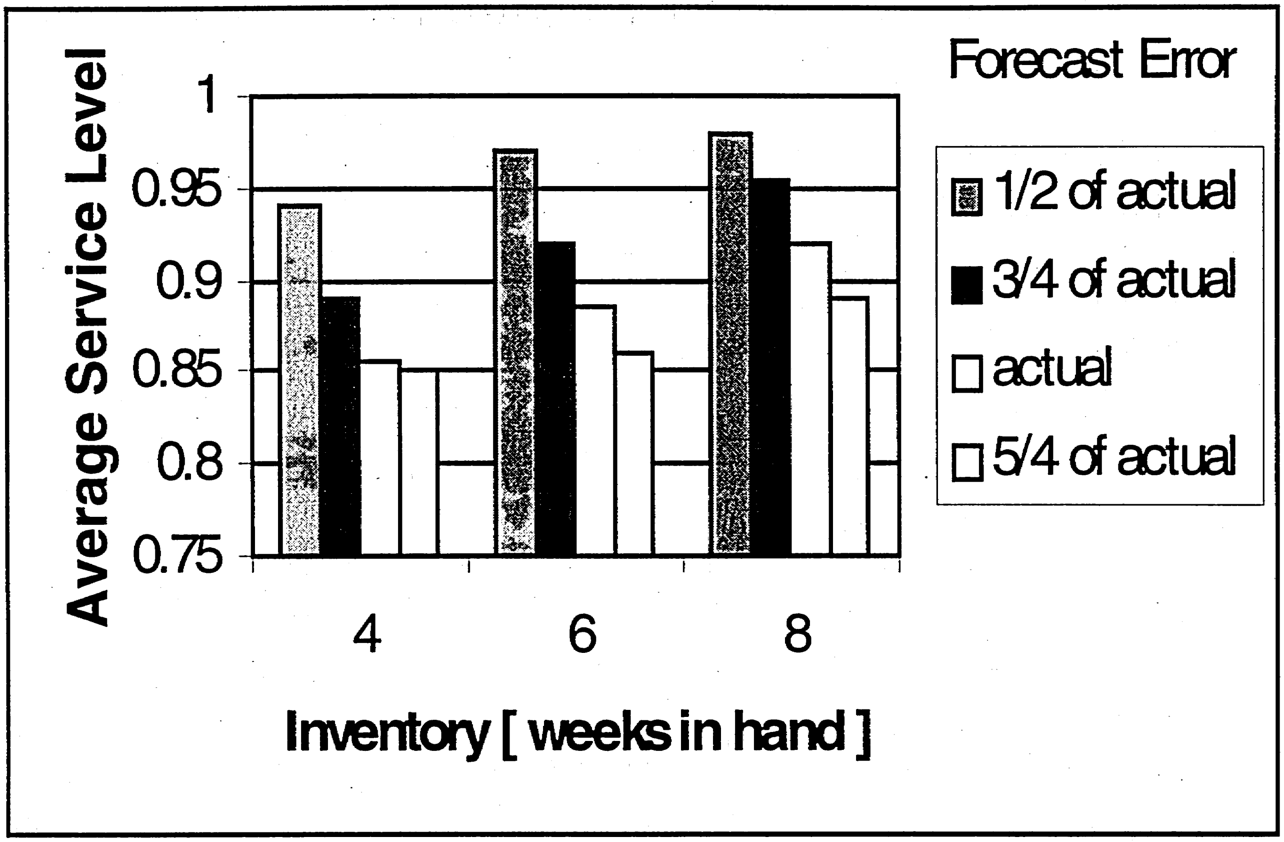

Figure 18: Effect of Error and Inventory on Service Level

4.2.3 Effect of error and inventory on service level

This numerical experiment had two purposes, the first was to capture the tradeoff between inventory strategy and capacity while keeping service level constant. The second was to understand the effect of decreasing error on this tradeoff. We chose to focus on the capacity at the fabrication stage of the supply chain. This stage involves the highest

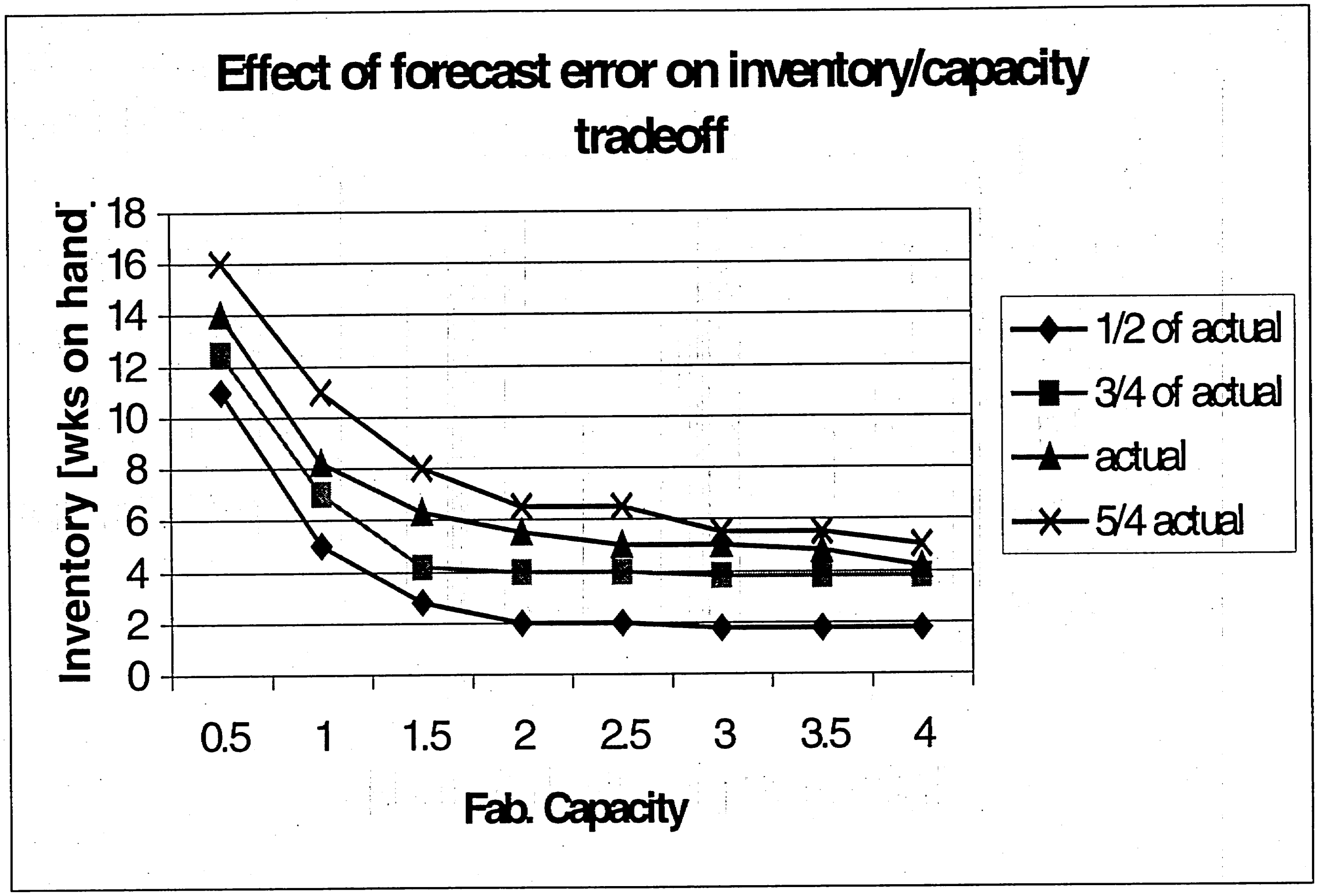

Figure 19: Effect of Forecast Error on the Inventory and Capacity Tradeoff

capital expenses and therefore its capacity is the most valuable in the total supply chain. We varied the capacity from half to four times its current capacity. For the inventory strategy we varied the inventory strategies throughout the supply chain while keeping the ratios of inventory buffer in finished goods and die bank equal. As expected, for a given error, there is a nonlinear tradeoff between levels of inventory strategy and capacity. As expected the rate of substitution of inventory for capacity increases as the capacity decreases, and vice versa. At the current level of capacity and error, the fabrication stage could decrease its capacity by half if the inventory strategy would be raised from 8 to 14 weeks on hand in finished goods, while still maintaining the same service level. Further, as forecast error decreases the shape of the tradeoff curve remains the same. Also, the ratio of inventory and capacity that is released because of a decrease in forecast error seems to remain constant as the error decreases from 5/4 to half of the historical error.

4.2.4 Effect of error on lead times

We observed no statistically significant effects on the average leadtime of decreasing or increasing error within the bounds explored earlier. However, the variance around the average leadtime did increase with the variance of the error in the forecast, as expected.

4.3 Empirical results

In this section we describe the added costs of the actual forecast errors that occurred in FY97. We input the actual demand, inventory strategies and forecasts that drove the internal supply chain of Digital Semiconductor during the past year. Our results assume that the capacity at each stage was infinite while the cycle times were constant. While in reality this is not the case, the assumption is not a bad one, since the facilities of Digital Semiconductor were extremely under-utilized for most of the year. Rocket lots, or material that is rushed through the fabrication facility with a very low cycle time, was an option available to Digital Semiconductor and was often used. However, only limited amounts of material could be processed in that fashion.

Secondly, we assumed that an average of 90 % service level was achieved. This level is close to the corporate targets that are set for Digital Semiconductor and is in line with both the rest of the industry and this division's goal to adequately serve its external customer base. In conclusion, the following results are an indication of how much past forecast errors would have costs the division if they wanted to achieve their target service level given both the poor forecasts they issued as well as the demand patterns that they faced. The following are averages across the twenty highest revenue generating parts at Digital Semiconductor

The results were as follows:

Production Stage / Inventory location	**Excess Capacity Consumed (%)**	**Added Holding & Obsolescence Cost (%)**
Fabrication & Probe / Die Bank Inv.	20 %	15 %
Assembly / Ahead of Test	30 %	20 %
Test / Finished Goods Inv.	25 %	20 %

Table 3: Costs Due ro Forecast Inaccuracy for FY97

The above results are averages across all product lines. In total 20 products were averaged and the standard deviation was less than 3 % for all six results.

The variation across production stage and inventory location is due to the specifics of the errors in the one, two or three months forecast. The holding and obsolescence costs were calculated from as a ratio of the monthly ending inventory

The results are striking by their shear magnitude. The opportunity for savings both on inventory holding and obsolescence cost as well as decreased capacity utilization is tremendous.

5. Business process changes required to improve forecast

The following chapter describes the key recommendation uncovered after conducting more than 40 interviews across all the functional groups of the organization (marketing, manufacturing, sales, distribution and manufacturing) as well as observing the forecast generation process. These recommendations address the root cause of forecast error at Digital Semiconductor.

5.1 Redefinition of forecast purpose

The monthly forecast generation process as described in figure 2 serves many often conflicting purposes. In essence, every month the Digital Semiconductor debates and creates a two year forecast which involves many organizations and which attempts to serve all of them. This section begins with a list of the numerous goals of the monthly forecast and an explanation of why they conflict or are detrimental to achieving an adequate accuracy level for the short term production forecast. Then, we issue a set of recommendations of how to organize the forecast generation process is a way more conducive to achieving forecast accuracy.

The current forecast has many purposes:

- Production planning: the monthly forecast is first and foremost used to determine how much the manufacturing operation needs to start today in the supply chain in order to meet their forecasted customer demand at the end of their cycle time. In order to do this the forecast is broken down to the stock keeping unit (SKU) which details not only the part numbers but also the specific speed required (bin). Production planning is focused on a three to four month time horizon and could benefit from as frequent updates as possible to allow more time for manufacturing to respond to any new information about customer demand that sales representative might receive weekly of daily. However under its current format the production planning forecast is bundled with a monthly forecast that spans two years. The main cause for this apparent disconnect is that the same forecast used for production planning is also used for:

- Marketing and medium range planning: the same monthly forecast also serves to determine and document major marketing decisions or events. These could include the growth of new customers, the entrance of new competitors in existing markets or the targeting of new markets. These projections are more marketing targets rather than actual short term production forecasts that manufacturing should build against. The current forecasting process does not incorporate confidence levels into the projected demand, making it impossible to differentiate what is a broad brush marketing goal from what is a forecasted customer demand that the sales organization is 90 percent sure they are going to sell. Further, the level of granularity that is needed for making marketing projections is much less than for production planning. Yet the monthly forecasting process requires the marketing organization to specify once a month the monthly demand for each and every product down to the speed bin level. Although this level of granularity is required for production planning which has a two to four month horizon, it only adds confusion and creative guesswork when the process is carried out for forecasts spanning several years in an industry where the typical life cycle of a products is between 6 and 18 months.
- Revenue planning: the monthly forecast is also used to benchmark the monthly forecasted revenue against yearly budget. Within a tight range, the bottom line revenue that results from the monthly forecast must correlates directly with the number of parts the forecast claims will be sold. There are two negative consequences to tying the monthly production planning forecast to revenue planning. The first is that the Product Lines (marketing) who are both responsible for meeting their budgets and creating the forecast, have an incentive to change both the quantity and mix of the products on the monthly forecast to reflect an even and steady growth of earnings against their targets regardless of the actual demand. The second consequence of tying revenue forecasting to the same forecast that drives manufacturing, is that customer returns and other revenue altering activities will be reflected in the forecast. Thus it is possible to have a monthly forecast that will have no demand for a product because

the expected number of returns is approximately the same as the expected number of sales. Even though it might not be possible to sell the returned merchandise from customer A to customer B because of branding or timing issue, still the monthly forecast will reflect no demand for that month.

- Long term capacity planning: another purpose of the monthly forecast is for long term capacity and utilization planning. Every month, the marketing, revenue operations and manufacturing organization debate the long term utilization and capacity planning of the internal supply chain. To address these issues every month is unnecessary and diverts attention and energy away from the short term production planning issue. By including the purpose of capacity planning with the purpose of achieving an accurate short term forecast that drives manufacturing the current forecasting process at DS confuse demand planning with production planning. The latter consists of understanding how the manufacturing organization will meet the forecast and the former is understanding whether the forecast is accurate. By merging them, most of the attention at DS is spent debating production planning strategies instead of focusing on improving and understanding the forecast.

All these different and often conflicting goals of the monthly forecast detract from achieving an accurate short term forecast that can be used to drive manufacturing. The recommended solution to these issues is to separate the long term revenue planning from the short term production planning and forecasting. Thus the forecasting process with a horizon of 6 months or less should be done monthly by the sales, manufacturing and to a lessor extent by marketing at the SKU level. The long term (6 months to 2 years) forecasting process would have a quarterly cycle and would involve mainly the sales, marketing and BOPS organizations. The long term forecasting would be at the product level, not the device or SKU level and would be reconciled with the short term forecasts once a quarter.

5.2 Information and assumption sharing across organizations

Currently the forecasting process at Digital Semiconductor is very linear with each organization involved completing their task and passing it to the next organization in the forecasting chain. Further, the linkages between the organizations are through the Revenue Operations group. This very fragmented process causes the many organizations involved in the process to not share or know the assumptions behind the forecast they are given by the organization upstream of them. Consequently, many of the organizations who develop the forecast sequentially will duplicate the same assumptions as the others. For example, an Area Sales Managers will make an assumption as to the effect of an increase in the price on the upcoming demand for a product, create a forecast and pass it on to the Product Lines through the Revenue Operations Group. The Product Lines will then further decrease the forecast, not knowing to what extent the Area Sales Manager took the increase in price into account, if at all. Further, as described in Chapter 2, there is no adequate feedback loop to the Area Sales Managers as to the accuracy of their product specific long term forecast. In order to resolve this lack of communication the forecasting process needs to be changed in two ways. Lastly, because the process is sequential and linear, new information received by a Sales Representative about the forecast could possibly take as long as 2 months before it is included in a formal feedback.

There are several initiatives that need to be taken in order to resolve these issues. First, the process needs to transition from a linear to a parallel process. The main participants in the forecast generation, the Product Lines, Area Sales Managers, BOPS, Revenue Operations and Sales Distributions, need to simply meet twice a month do discuss their assumptions behind the forecast and share any new information that is pertinent to it. These meetings would occur separately for each product line and could be as short as two hours and still fulfill their purpose. The current lack of communication is increased by the intermediary step of having to channel the forecast through the Revenue Operations (ROPS) Group at almost every step in the forecast generation process. Thus the information technology tools need to be put in place to allow the organization to bypass the involvement of the ROPS group in the forecasting process. All these tools would have to do is allow the product lines to receive and consolidate, through a simple filter, the forecasts from the Area Sales Managers.

5.3 Information and assumption documentation for root cause analysis

As described earlier in section 6.2, there is little to no assumption sharing across organizations involved in the creation of the forecast. Once the sharing of product and market assumptions does occur according to the recommendations outlined in the previous section, these assumptions should be agreed upon by all parties involved and recorded. The recording of these assumptions will allow some root cause analysis to be performed once the forecast comes due and there can be comparisons drawn between the forecast and the actual demand. If there exists a significant difference between the two, the recorded assumptions can be reviewed and an understanding of which assumptions are erroneous can be derived. In this way, there will be some organizational learning taking place and progressively the forecast accuracy will improve as all the organizations involved in the forecasting generation understand what is reasonable to assume and what is not. For example, the ramp up of new products was often underestimated in terms of the magnitude and the date of the ramp up. In order to avoid the reoccurrence of the same error month after month as occurs now at DS, it is important for the product lines and sales organization to understand what assumption caused them to repeatedly misjudged the market.

5.4 Implementation of realistic metrics

The product lines are ultimately responsible for the accuracy of the monthly forecast. However, currently they are evaluated using the waterfall charts described in section 2.5. These waterfall charts show the forecasted demand for various time periods as compared to the actual demand. This metric is very binary, the forecasts are either right of wrong. By their very nature forecasts are almost always going to be wrong. Thus, any kind of metric against which forecast accuracy should be measured, must allow for a range of error. The question then becomes what range of error is acceptable. In our case, we used the model described in chapter 3 to define a range of acceptable error in the 1, 2 and 3 months forecast.

First we broke down the existing products sold by Digital Semiconductor into three types of demand patterns. The three generic categories were growth, steady/decline and cyclical demand patterns. Using the model of the supply chain developed earlier and the historical demand pattern of several parts that fit into one of the three categories, we could plot the range of error that could exist in the one, two and three months forecasts while still meeting the target service levels for a given inventory strategy. Further, this range of error varies increases depending on whether the forecast improves from month to month.

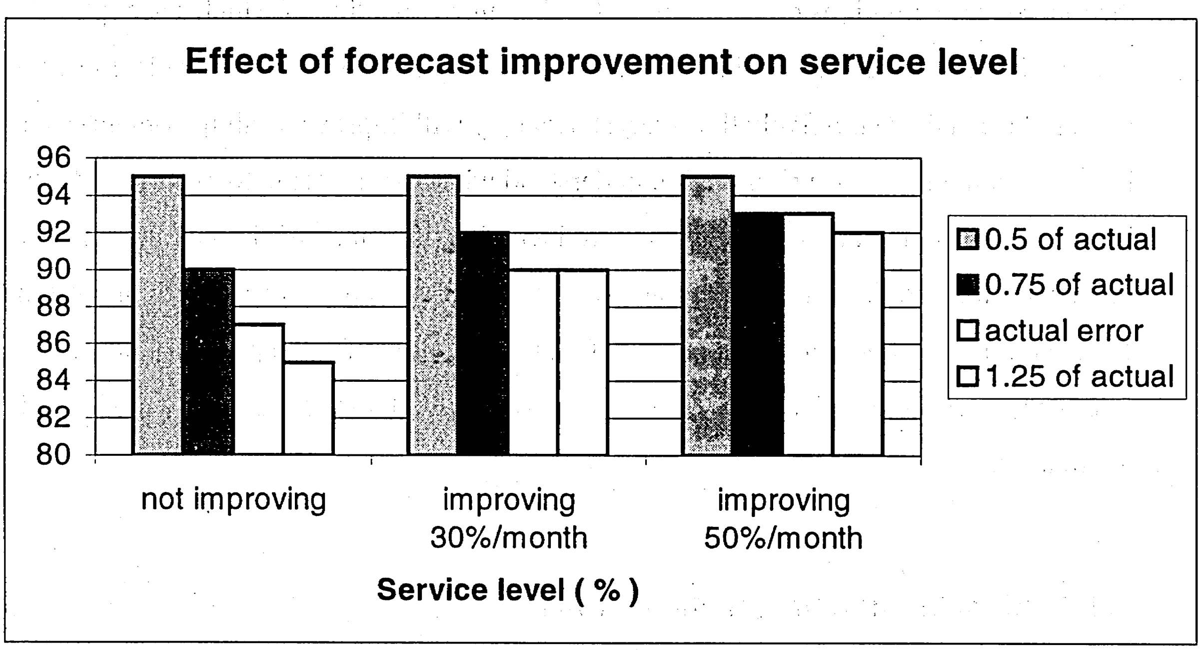

Figure 20: Graph to determine the acceptable range of error in forecast depending on rate of improvement of forecast accuracy.

As is clear in figure 20, if the forecast accuracy increases from the 3 to the 2 month and finally to the 1 month forecast, the starting range acceptable range of error is far larger than if the forecast did not improve. Further, as the target inventory levels increase so does the possible forecast error for a given service level. These results are averages and varied depending on the demand pattern we were analyzing. Further, in all these cases we assumed that there was infinite capacity available.

Thus, with these tools, the product lines could focus on achieving a realistic target accuracy range instead of just trying to either be right or be wrong. If their forecast accuracy was within the target range then, given the target inventory strategies, there should be enough flexibility in the supply chain because of inventory and capacity buffers to meet the desired customer service levels.

6. Conclusion and Recommendations

This section is meant to complement chapter 5. The later outlined some specific changes in the forecasting process required to improve forecast accuracy, whereas the following section discusses some key high level recommendations to both increase forecast accuracy and well as reduce the cost of forecast inaccuracy. Second, the key learnings I have gained from this internship are also included in this section. Lastly, we finish with some suggestions for future study at Digital Semiconductor, and pertaining to supply chain in particular.

6.1 Recommendations

Inventory strategies:

The first key recommendations pertains to the use of inventory to buffer against poor forecasts. At Digital Semiconductor, the approach taken to inventory buffers was one of setting inventory strategies. Thus, instead of setting absolute inventory levels, the target buffers are set as factors of the demand itself. The reason behind this choice is the very short life cycle and uncertain demand patterns of the semiconductor industry. Because most of DS's product lines supplied chips to a very concentrated customer base, the demand patterns were inherently lumpy. An unexpected large order from one of the few larger customers could double the order run rate. To protect against this, the standard inventory strategy was as high as six times the forecasted demand.

While setting absolute targets for inventory level is not appropriate for such a market and customer base, neither is setting an inventory strategy based on forecasted demand which often is two to three times the actual. By setting inventory strategies based on forecasted demand, the supply chain of Digital Semiconductor was essentially institutionalizing the "beer" game. Further, these inventory strategies were set at several locations within the supply chain, which just compounded the problem. In total, Digital Semiconductor was targeting to have as much as 10 weeks of the forecasted demand of added inventory as

WIP in the supply chain at any point in time. Consequently, any error in the three and two months forecast was multiplied 10 times before it reached the beginning of the supply chain.

The solution is to set inventory strategies based on the current, known demand. For example instead of an inventory buffer of six times the expected demand 12 weeks from the current week, the inventory strategy should be a buffer of four times the current demand. By doing so, the safety buffer will be preserved, without amplifying the forecast error.

Rigorous analysis of information resident in the organization:

Currently, the staff generating the forecasts do not have the expertise or the necessary tools to rigorously analyze the information available to them. There is little possibility to conduct any statistical analysis of past demand run rates and forecasts. Neither is there any software in house that would allow the marketing group to conduct correlation studies relating marketing events to changes in the demand and ordering patterns of their customers. However, commercial software with the capabilities to not only easily aggregate all the demand information disseminated throughout the organization but also to analyze it, is readily available in the market. After some benchmarking and analysis we determined that the best supplier of such a software is either i2 technologies or Manugistics.

Further, the expertise to conduct analysis of information is not resident in the marketing, sales or BOPS group. Many organizations in industries with short life cycles and uncertain demand patterns have a forecasting group which supports marketing and manufacturing. In the forecasting group, the necessary expertise to develop accurate forecasting models and to use the software required to analyze the information resident in the organization is developed. We recommend the creation of such an independent forecasting group to allow better forecasting to take place within Digital Semiconductor. The advantage of creating such a group is twofold. First, it would allow the concentration of knowledge and skills as well as continuous improvement. Second, the group's mandate to create an accurate forecast would not be in conflict with the many other incentives that the current marketing organization faces, such as reaching marketing and revenue targets. In conclusion, currently the forecasting decisions made by the product lines are not based on analysis and data, but rather on knee-jerk reactions which causes the inherent volatility of the demand to be amplified.

Alignment of incentives:

The sales organization's lack of incentive to achieve an accurate forecast was one of the main impediments to improving the short term forecast accuracy. The sales force was almost exclusively rewarded with growing the business and increasing revenue not the accuracy of their forecast. An order from a customer meant a commission for them, regardless of whether the supply chain could actually deliver on that order. Because Digital Semiconductor's external business was still in its infancy and most of the production facilities underutilized, the negative impact caused by the practice of promising products and not being able to delivering on them due to poor forecasting had not yet been felt by the sales force. The later did not see a link between an accurate forecast and having the material ready to ship to the customer on their request date. Had the sales force incentives been linked with those of marketing and manufacturing, the sales organization would have been more responsive to implementing initiatives that increased forecast accuracy. It is only a question of time before the loop linking forecasts and service levels will close and sales will be lost because inaccurate forecasts.

6.2 Key Learnings

The main learning in this project was the importance of formal communication channels across the different organizations of Digital Semiconductor. Most of the root causes behind the poor forecasting performance of Digital Semiconductor were related to the key decision makers, the product lines, not being able to access the information resident in the organization in a timely and rigorous manner. This lack of access was prevalent despite the fact that there was open communication between them and the variety of groups that possessed the information relevant to making an accurate forecast. However, no formal, rigorous business processes were in place to capture, document and keep track of the information that was communicated. The outcome was a piecemeal, inconsistent flow of demand forecast information between the Sales Organization, the Product Lines, the Sales Distribution organization and BOPS.

Further, the little formal communication that did occur was through the Revenue Operations Group (ROPS). The later did not add any new information to the forecasting process, but rather only consolidated the information it received from various sources before passing it on to the next organization in the forecast generation chain. However, by having the ROPS consolidate the information as it traveled from one group to another, the clear communication between these groups was broken and delayed. The involvement of the ROPS group also increased the overall length of time required to create a forecast. This added time then increases the required time horizon of the forecast while decreasing the time available to respond to a forecast error. The function of the ROPS group in the forecasting generation could almost be entirely replaced by a commercial demand planning software. (Section 6.2)

Further, the importance of incentives was driven home when we attempted to implement many of the initiatives described in chapter 5. Although almost all the organizations believed and acknowledged the validity of the analysis and the description of the problem, very few responded and only after extensive prodding. The main issue was that the metrics by which they were measured were only indirectly related to forecast accuracy. Even though all the groups involved in the forecast generation were partially responsible for it, they were hardly accountable for the forecast accuracy in any formal way. Rather, it was the manufacturing organization that ultimately was responsible for meeting customer requests, irrespective of the original forecast error.

Lastly, the importance of clearly separating demand planning and supply planning was made clear on numerous occasions. In a fast growing organization such as DS, the distinction between creating a forecast, demand planning and meeting the forecast, supply planning, is critical to understanding the root cause of poor forecasting.

6.3 Future Study

This last section describes the possible next steps to continue and conclude this research.

Model Validation:

In order to understand fundamental trends and tradeoffs the model was intentionally kept at a relatively high level with most of the details of the supply chain omitted. However, to truly optimize the supply chain in terms of minimizing inventory and capacity in the face of a given forecast error, the model would have to be extended. The key addition would be to model multiple products sharing the same manufacturing facility. With this added complexity the model could actually be validated against empirical observations of inventory levels, customer service levels and lead times across the six product lines.

Opportunity Cost:

Because the model only handles one part being processed at a time, the opportunity cost of a poor forecast was not included in our treatment of the costs of a poor forecast. With an extended and validated model, the opportunity cost of making one product at the expense of another could be determined. Thus, the manager would be able to make tactical decisions that would minimize the overall cost of a poor forecast by trading off low margin products for high margin products when capacity is an issue or poor forecasting drives those choices.

Effect of various reaction strategies:

In our analysis the reaction strategies now practiced at Digital Semiconductor were taken as a given. Thus, for example, when a backlog of part was created because of a poor forecast, new parts were released immediately. We took these material release rules as given because we wanted to understand the current costs of poor forecasts. However, a rigorous analysis of the effect of these various reaction strategies would yield an understanding of which one is most appropriate for Digital Semiconductor to its strategic objective while minimizing the overall cost of doing so. A general understanding of which strategies would be most cost effective can be carried out with the model developed in this thesis. Specifically, a study of added inventory vs. more frequent expediting would be appropriate.

<u>Cost modeling:</u>

Ideally, a model of the cost of capacity, inventory and service level should be superimposed upon the model developed for this thesis. Such a cost model would yield some insights as to the accounting costs of many of the practices and tradeoff presented in this thesis. The current analysis remained at the level of parts and percent capacity, not at the accounting level.

6.4 Conclusions

In conclusion, this thesis showed that the costs of current poor forecasting are significant both in terms of the added inventory and the added capacity that is required to meet customer service levels and lead times. Further, these costs can be significantly reduced through the reengineering of the forecasting process to make it simpler, more rigorous and more focused. Further, the model developed in this paper was able to prove that there are many complex tradeoffs among the various costs of forecast inaccuracy that are possible. These tradeoffs can limit the total costs of forecast error. Further, understanding these costs enables managers to make tactical choices and weigh the effect of these choices on the various elements of the supply chain.

Bibliography

Altsrom, P., and P. Madsen, "Simulation of Inventory Control Systems," International Journal of Production Economics. 1992, pp. 124-134

Billington, C., "Strategic Supply Chain Management," OR/MS Today. April, 1994, pp. 20-27.

Gershwin, S., Manufacturing Systems Engineering, New Jersey: PTR Prentice Hall, 1994.

Kanet, J. J., "Toward Understanding Lead Times in MRP Systems," Production and Inventory Management, Third Quarter, 1982, pp. 1-13

Kelsch, M., "Using Supply Chain Flexibility to Account for Uncertainty in Demand." MIT Master's Thesis, 1996.