# MAKING REAL-TIME PRECISION ADJUSTMENTS TO WORLD-WIDE CHIP PRODUCTION

By
Neelesh Pai

B.E. Electronics and Telecommunications, University of Mumbai 1996
M.S. Electrical Engineering, Pennsylvania State University 1998

Submitted to the MIT Sloan School of Management and the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degrees of

**Master of Business Administration**
**and**
**Master of Science in Electrical Engineering and Computer Science**

In conjunction with the Leaders for Manufacturing Program at the
**Massachusetts Institute of Technology**
**June 2009**



Signature of Author ______________________
Electrical Engineering and Computer Science, MIT Sloan School of Management
May 8, 2009

Certified by ______________________
David Simchi-Levi, Thesis Supervisor
Professor of Engineering and Systems Division and Civil and Environmental Engineering

Certified by ______________________
Don Rosenfield, Thesis Supervisor
Senior Lecturer/Director LFM Fellows Program, MIT Sloan School of Management

Accepted by ______________________
Terry P. Orlando, Graduate Officer
Professor of Department of Electrical Engineering and Computer Science

Accepted by ______________________
Debbie Berechman
Executive Director of MBA Program, MIT Sloan School of Management

*This page has been intentionally left blank.*

**Making Real-Time Precision Adjustments To World-Wide Chip Production**

**Neelesh Pai**

Submitted to the MIT Sloan School of Management and the Department of Electrical Engineering and Computer Science on May 8, 2009 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Electrical Engineering and Computer Science

# ABSTRACT

Intel has recently embarked on a mission to improve its supply chain responsiveness. Currently production lead times are around 4 months requiring a forecast a quarter out. Most customer demand changes happen within lead time since customers only know their demand a few weeks before shipment. While stable production plans help maintain factory utilization rates their inflexibility can also lead to missed revenue opportunities or unneeded inventory. The challenge then is to make planning processes agile enough to react to late demand changes.

The FAB has a 2-3 month throughput time or latency. The subsequent Assembly-Test (ATM) operation has a 1-2 month latency. Increasing competition requires the striking of a balance between competitive service levels and excess inventory. This Thesis looks to develop ways of making more real-time tactical demand updates to production plans used by the global factory network to improve Supply Chain Responsiveness. Using business analytics and organizational processes analysis, ways of making late demand changes to the production plan are evaluated.

The project focuses on Intel's global ATM network due to its proximity to end customer demand. A holistic solution to use available intelligence is proposed. The focus is on creating data visibility across the supply chain and on putting feedback loops in planning processes to intercept planning processes at various points with new information as and when it becomes available. Issues examined include demand signal generation, the choice of different demand signals, solver algorithms to convert demand inputs to a global production plan, inventory target setting and implementation in production plan and finally ATM processes such as SDD (delayed product differentiation at the semi-finished goods warehouse) for Product Mix and volume determination. The hypothesis is that this will lead to a better understanding of the interaction between various planning processes.

Thesis Supervisor: Don Rosenfield
Title: Senior Lecturer/Director LFM Fellows Program, MIT Sloan School of Management
Thesis Supervisor: David Simchi-Levi
Title: Professor of Engineering and Systems Division and Civil and Environmental Engineering

*This page has been intentionally been left blank.*

# ACKNOWLEDGMENTS

The author gratefully acknowledges the support and resources made available to him by the MIT Leaders For Manufacturing Program and the Intel Corporation for this work.

I wish to thank my Intel supervisors Amy Reyner and Jeremy Weinstein for their help and contributions throughout the internship process. I also want to thank Karl Kempf for championing the project and spending a great many hours with me and helping to simplify the complex. Thanks also, to all my Intel colleagues from MMBP, Division Planning and ATM's who shared their knowledge so patiently with me. I wish to thank my thesis advisors Don Rosenfield and David Simchi-Levi for their time and guidance throughout the process.

I especially want to thank my wife Sheetal for her unconditional love, support and patience throughout my time at MIT. I also want to thank my son Raj, who was born during the course of this project and whose playful nature provided welcome relief during the grueling process of writing a thesis. I would finally like to thank my brothers for their help and guidance throughout my academics and my parents without whom none of this would have been possible.

*This page has been left blank intentionally.*

# TABLE OF CONTENTS

*This page has been left blank intentionally.*

# Table of Figures

# Glossary of Terms

**FAB:** Semiconductor Fabrication Plant

**FSM:** Fab-Sort-Manufacturing Intel term for a FAB.

**ATM:** Assembly Test and Manufacturing Plant

**MMBP:** Microprocessor Marketing and Business Planning, Intel Sales and Operations group

**Judged Demand (JD):** MMBP Forecast projected out 4 Quarters, by month, by forecast item

**Billings:** Previous, Weekly Data on what has been actually bought by and billed to Customers

**Backlog (BL):** Customer Projected Demand Forecast out over next 4 months on weekly basis

**CPR:** An internal Business Process Re-engineering effort

**Customer Judged Forecast:** Post-CPR Customer Projected weekly Forecast over next month

**Commits:** Intel Commit to Customer Order Request projected out weekly for next 4 months

**Allocation:** Post-CPR, Intel Commitment response to Customer Judged Forecast

**Geos:** World Wide Sales Geographies

**Geo Forecast:** Forecast by individual Geographies for their regions

**COR:** Weekly Change Order Requests by customers updating prior order Backlog estimates

**Reset:** Intel's monthly production planning process

**OFG:** Outside Factory Guidelines, an ATM capacity planning process

**IFG:** Inside Factory Guidelines, an ATM planning process following OFG for die distribution

**Division Adjusted Demand (DAD):** Aggregate signal used in OFG-IFG process.

**SFGI:** Semi-Finished Goods Inventory

**SDD:** SFGI Delayed Differentiation a Fusing based postponement process

**ASRS:** Automated Storage Retrieval System used for SFGI

**CW:** Component Warehouse, Intel terminology for Finished Goods Warehouse

**ADI/TRDI**: Available Die Inventory/Tape reel Die Inventory

**VMI**: Vendor Managed Inventory

**WOI/DOI**: Weeks of Inventory/Days of Inventory

**OEM**: Original Equipment Manufacturer

**LOEM:** Local OEM, Intel Term for smaller domestic OEM's

**BOH:** Beginning On Hand Inventory relative to planning cycle

**EOH:** Ending On Hand Inventory relative to planning cycle

**Qx Weekly Schedule:** Intel ATM Production Planning Schedule

**STR:** Storage Transfer Request, inter warehouse request

**Roadmap:** Quarterly or Yearly Product or Price Rollout strategy

**EOL:** End of Life of a product

**NPI:** New Product Introduction

**WSPW:** Wafer starts per week

**DSS:** Decision Support System, a tool for business analytics

## 1. INTRODUCTION

In the early era of the PC market, technology innovation and capability were the only things that mattered. As the market has increasingly matured, with technology capability going a far ways to meeting the needs of all but the heaviest workloads, customer service has become equally if not more important as a key competitive differentiator. Increased competition has also meant that the speed of technological innovation has increased rather than decreased in the battle for higher end higher margin market share. It has also meant that the market has been segmented finely with an array of products to fit every niche on the price-performance curve. Thus the industry has increasingly become a high clock-speed industry requiring the servicing of a dizzying array of sku's in an environment of higher demand volatility and competitive pressures.

The main challenge for Intel then has become increasing its customer responsiveness. Improving customer service while reducing inventory is particularly tricky given heightened demand uncertainty. To exacerbate this, the company along with the rest of the industry has a zero penalty cancellation policy till the very last minute before order fulfillment. This means there is no great incentive for customers to forecast accurately or improve forecast accuracy. The entire throughput time for producing chips is around 3-4 months from wafer starts in the Fab and packaging in Assembly Test operations through to stocking in finished goods warehouses. This means the ideal heads up required from customers is 3-4 months before delivery. Given this, even if customers tried their best to forecast well, it is arguable to what extent customers could accurately predict their microprocessor requirements months ahead of end demand. Most demand only crystallizes in the last few weeks before delivery so the real question is how does Intel become more responsive to ever changing market demand, especially when the majority of the changes occur within lead times? While there are many trends that become evident over the 3-4 month lead time from order forecast how can it separate the noise from signal and respond only to meaningful signals? And once the meaningful signals are identified analytically, what can be done within lead time to adjust the signal such that demand is met, without building excess inventory.

The Intel Fab factory network is driven by a demand forecast called "Judged Demand" that is published monthly while the Assembly Test factory network runs to fortnightly plans. Each month, production volumes and product mix for various product families are re-allocated throughout Intel's worldwide network of factories based on this forecast. However as discussed these planning processes that are great for stability and capacity utilization are being buffeted by the need to respond more real time to demand swings resulting from competitive pressures and their impact on product pricing and strategies. These frequent and immediate changes have increased the need for a system to make more frequent forecast updates with which to intercept various points in the supply chain depending on where in the 3-4 month time line new and actionable information is available. The ability to do this well to influence production would drive a more agile supply chain capable of meeting improved customer service levels at leaner inventory levels with the ability to capture more last minute customer upside or change requests all of which are top priorities for the company today. Challenges in driving this change are equally technical and organizational.

This thesis evaluates opportunities for making late demand changes to the production plan using business analytics and organizational process analysis. Some of the key issues are how demand and inventory inputs for current production planning processes get decided. How are they communicated across organizations? How is this tracked? What are the key incentives driving behavior both within and outside the organization? What are the key barriers to making tactical changes quickly and efficiently?

The current project focuses on suggesting ways of improving the supply chain responsiveness by focusing on Intel's global ATM factory network due to its proximity to end customer demand. The project consists of understanding all the production planning processes including demand signal generation processes, the choice of different demand signals, solver algorithms to convert demand inputs to a global production plan, inventory target setting and implementation in production plan and finally ATM processes such as SDD (Semi Finished Goods Delayed Differentiation) and Range Planning for Product Mix and Volume determination. Initially approaches to improving demand forecast accuracy were considered. But these approaches posed challenges to the project. Intel was shifting to a VMI strategy (Vendor Managed Inventory) as well as implementing a SAP on a large scale for its supply-demand planning systems. As this was happening, the names of various demand signals as well as their compositions were changing on a frequent basis. Hence any demand signal accuracy analysis proved very difficult to do given the changing meanings of demand signals at different points in time historically. In addition, due to the multitude of ongoing internal projects on this topic as well as the fact that different forecasts were used at different points in the supply chain making a single analysis less valuable, this approach was eventually discarded.

Intel is already pursuing various standard ways of improving responsiveness including reducing lead times by increasing pipeline inventory (with VMI hubs), reducing forecast accuracy error (several internal efforts strive to do this but there are limits to this) and increased frequency of planning cycles (these cycles are costly, effort-intensive and there is an organizational limit to how often they can be done). Hence this thesis explores other ways such as –

i) Increasing the number of production processes using pull rather than less accurate push forecasts

ii) Increased information transparency and decentralized decision-making

iii) Improved communication and standardizing for consistency

iv) Quicker learning rates based on micro-feedback loops

v) Improving bottlenecks such as scarce ASRS space at SDD a key push-pull interface and information bottle-necks

vi) Improving clarity on ambiguous push-pull boundary with SDD and its shift in a post-VMI world.

The project has concentrated on identifying available intelligence about factors causing demand changes and possible process interception points for delivering this information. The thesis explored ways of making the intelligence actionable within the framework of existing planning processes. A pilot project was conducted to demonstrate the concept and the results of this are explored in detail. The criteria for success is longer term ROI of aligning supply with demand specifically attributable to this project, helping to meet previously unmet demand changes as well as avoiding excess inventory.

In addition to the process of making tactical updates, understanding how to communicate this clearly and measuring the resulting ROI of these changes are also key considerations for the project. Different demand signals used across different processes and the use of different types of product groupings between supply and demand processes, with many-to-many mappings at each level, posed the biggest challenges in meeting project objectives. What this means is that once a wafer level and die-package combination level production plan is generated based on a marketing sku forecast, any changes to respond to customer demand shifts require this complex translation. This makes any swift communication almost impossible, while increasing reliance on solvers with embedded product mappings for translation. The immense complexity this generates makes responsiveness a challenge

This project resulted in a better understanding of the interaction between various processes across the chain, aided standardization efforts of inventory measurements which is used as a key tactical control signal, improved product mix decisions to help grant late customer upside change order requests and reduced costs related to excess inventory buildup. The project increased the ATM revenue under the purview of SDD by up to 10% increasing the flexibility of Intel to respond to changes in volatile demand. In addition a method of swapping demand based on trends in excess finished goods inventory and semi-finished WIP (work in progress) related inventory bubbles was developed. This method was projected to have an ROI of about 500K a year if implemented. Other measurements suggested to capture the benefits will take longer to pinpoint, including improved ROA on ATM assets such as ASRS, increased asset utilization by improved velocity as measured by increased SFGI inventory turns, lower inventory costs through reduction in aged inventory at SFGI and CW due to improved SDD process.

# 2. Background and Context

Intel is the world's largest maker of chips which in its case are primarily microprocessors and chipsets for PC's but also include Flash memories, communications chips and embedded controllers. The thesis primarily focuses on Intel Architecture based microprocessors which account for more than 80% of its revenue.

The microprocessor business serves 3 main vertical customer segments- Desktop (computers), Mobile (laptops) and Server (workstations). Intel sells the bulk of its chips through large OEM's (Original Equipment Manufacturers) like HP and Dell who then sell to retailers like Best Buy etc. It also sells its chips to ODM's (Original Design Manufacturers) like Quanta which make designs for sale by OEM's. Lastly it sells chips through the distributor to increase its reach, with a network of primary and secondary distributors, resellers etc. called the "channel" or "Disti". To most of the above described customers it sells chips in bulk in what is called the "Tray " format. But it also sells individual chips in individual boxes to retail customers through the channel in what is called the "Box" format.

## 2.1. The Intel Supply Chain

**Intel Supply Chain Overview:**

The first and main part of the Intel supply chain is the FSM (FAB-Sort-Manufacturing) global factory network. The semiconductor wafers are made here and they are diced and good chips are sorted from the bad chips based on preliminary testing. These unpackaged good chips are called "die". These available die are then shipped to a second set of factories called the ATM (Assembly-Test-Manufacturing) factory network. Typically (but not always) the ATM's have a receiving warehouse called the ADI (available die inventory). The ADI is used to store the sorted die shipped from the FAB's. These die are then packaged and tested post-packaging in the A/T (Assembly-Test) factories. The only step remaining to convert these chips into finished products is called the "Finish" step. Since pre-Finish the chips haven't been completely processed, they are called SFGI (Semi Finished Goods Inventory). The chips in the SFGI are stored in storage systems called ASRS (automatic storage retrieval systems). The Finish process essentially is an electrical fuse blow process that runs according to an algorithm or process called SDD (SFGI delayed differentiation). The algorithm essentially makes a down binning decision such as rating a higher speed part as a lower speed one (e.g. a 3 GHz part as a2GHz part etc.) or a part that can run at a lower power as one that needs a higher power dissipation (e.g. a 40W part as 50W one). This is done to hard fuse what the customer wants so that no black market can exist where parts with higher ratings are retested and resold for more money. The SDD algorithm essentially has both Intel specified business process guidelines as well as customer specified demand requirements that it tries to meet. When the parts are Finished they are stored in Finished Goods warehouses called CW (component warehouses).

**Strategic View:**

The figure below shows a strategic view of the Intel supply chain i.e. a view showing the 14 strategic decisions that need to be made at different locations in the chain. A decision on which wafer starts to make and how many to make is required as an input to FSM. Within the FSM (and similarly in the ATM) decisions on which WIP (work in progress) to prioritize are required. Since the FSM network is located in a different set of global locations from the ATM factory network, a decision on where to ship the different types of die coming out of FSM is required. At each of the inventory locations such as the ADI, and later the SFGI, the CW and the VMI hub locations inventory decisions on quantity and makeup of inventory as well as relative positioning of quantities across these locations has to be made. At the input of the Finish step decisions have to be made on what to Finish and how much to hold in SFGI. Lastly in the VMI Hubs, decisions on what to reserve for the OEM customer to whom the hub is dedicated and what to share across the hubs is required.

**Tactical View:**

The tactical view of the Supply chain shows how the different supply-demand planning processes interact with the supply chain to make the different strategic decisions required. The Build Plan Reset process is used monthly at the start of each month to make the wafer starts and product mix decisions. A Build Plan solver is used to make this decision. For ATM a two week build plan is generated every two weeks through a twin process Intel calls OFG (outside Factory Guidelines) and IFG (Inside Factory Guidelines). The OFG process is divided into OFG-Request and OFG

Response processes. The OFG Request is a production capacity request based on customer demand made by Division Planners to ATM planners and assumes unlimited capacity. The ATM Planners give a response back taking their capacity constraints into consideration. The IFG process is carried out by Division planners and is essentially a reordering of granted product capacity to make the daily production more in line time-wise with customer demand. The hotlist or CORS process is used to factor in customer change order requests and is done weekly and affects OFG-IFG planning as well as shipments from CW or VMI hubs.

The figures below show two views of the Intel supply chain- A strategic and a tactical view of the supply chain. The strategic view of the Intel supply chain shows the key decisions being made in the supply chain. The tactical view of the supply chain shows the actual supply-demand related processes and where they intercept the supply chain.

**Figure 2-1. Strategic View of Decision Making in the Intel Supply Chain:**

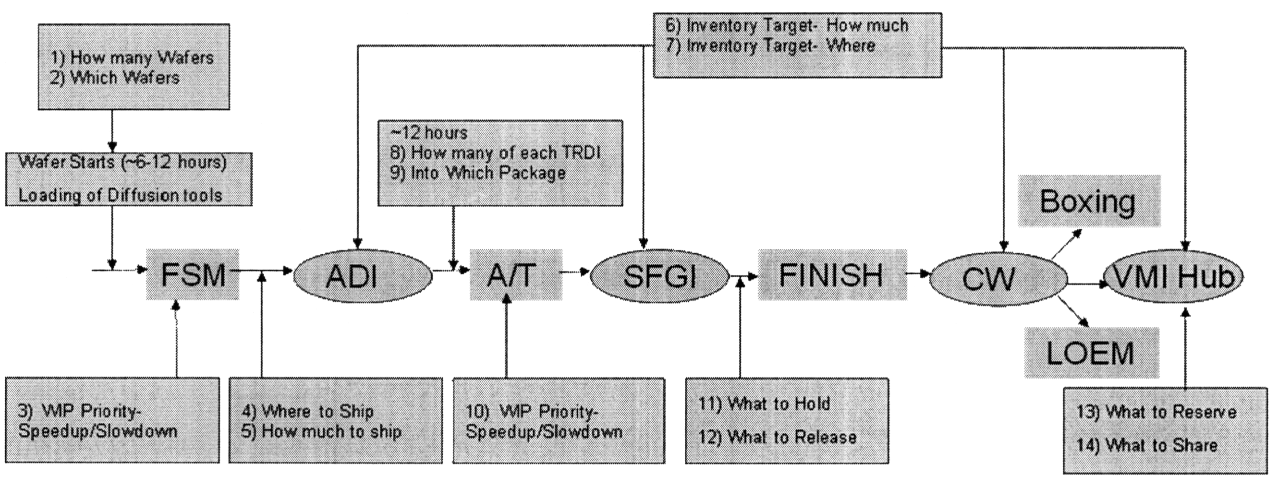


**Figure 2-2. Tactical View of Demand Planning Processes in the Intel Supply Chain:**

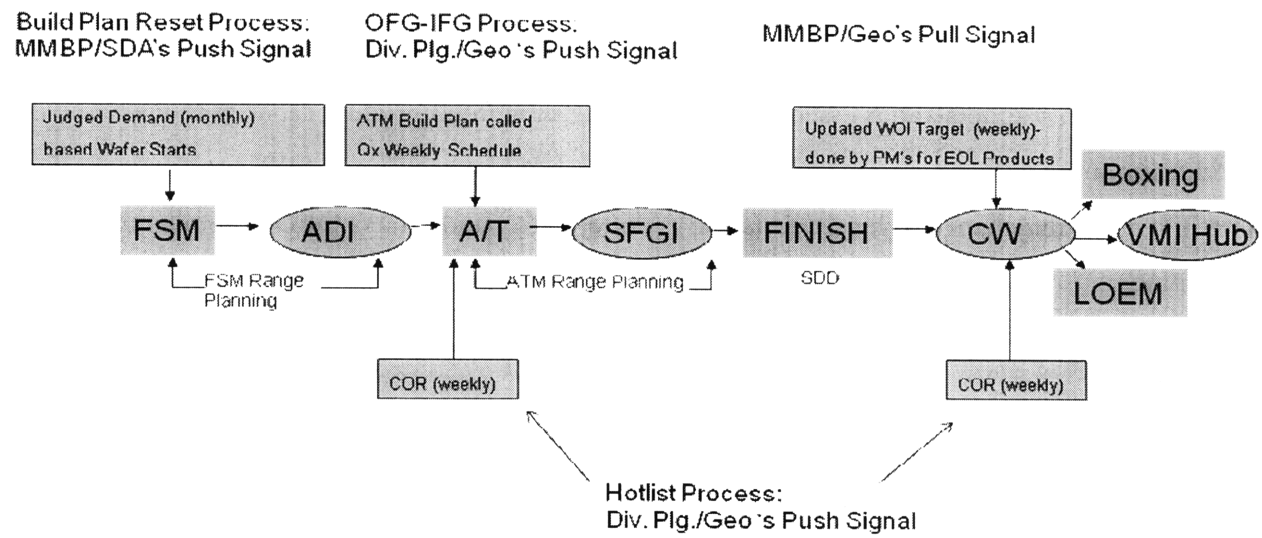
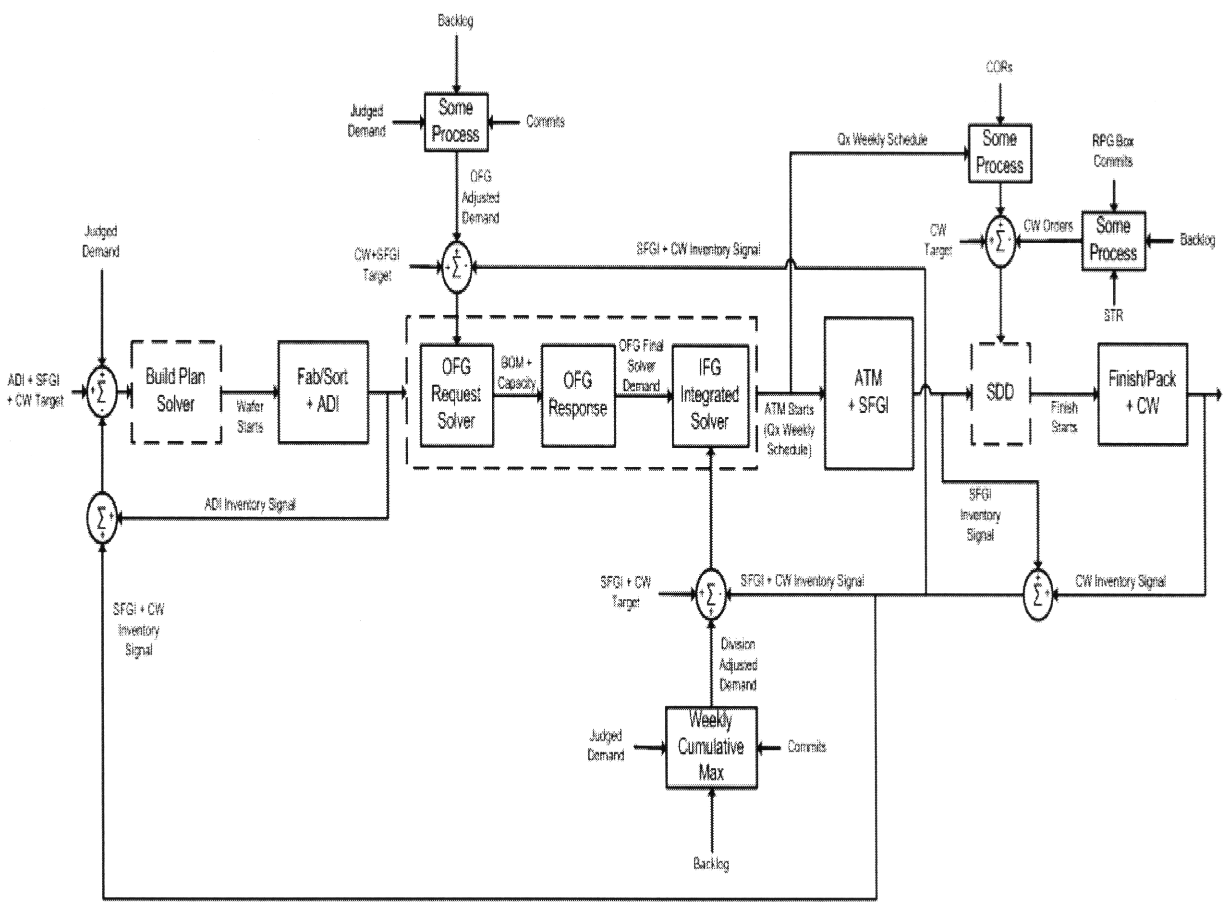



**Figure 2-3. Control View of Production Planning Processes:**



**Control View:**

The above diagram represents the control system view of the Demand and Production Planning processes in the Intel Supply chain. In the horizontal supply chain actual facilities are shown in solid line and planning processes such as Build Plan Reset (using the build plan solver), OFG-IFG (Outside Factory Guidelines and Inside Factory Guidelines which have their own OFG and IFG Solvers) and SDD (SFGI Delayed Differentiation) are shown in dotted lines.

To understand the process and solvers used along with the feedback loops affecting the next cycle or other processes, we proceed from left to right. The Wafer Starts are decided at the start of each month in a process called the Reset using the Build Plan solver. This solver has as its input, the forecasted Judged Demand from MMBP(a quarter out), the Inventory targets from MMBP both for product mix and volume to be kept at different locations such as ADI and (CW+ SFGI), and the beginning on hand inventory amounts from these locations. The BOH forms the feedback loop from the last cycle of planning and stochastic demand pulls from CW and ADI. This then starts in wafer starts for the FAB which will eventually after 3-4 months result in die sitting at ADI.

Every fortnight (beginning and middle of each month) the weekly plan for ATM's is done in a process called OFG-IFG. While a simple description of this has been provided before a more detailed discussion of this follows in the next section. Given that the ATM's typically have a frozen period of about a month, this weekly plan typically results in a plan for 2 weeks of ATM builds a month out (i.e. week 5 and 6). Note that while both Reset and OFG-IFG happen every month, the horizon being affected for Reset is 3-4 month out given FAB lead times and for ATM's is a month or so out given ATM lead times. Also the OFG-IFG process is broken into 3 solvers one for OFG (capacity unconstrained but die constrained) request from Division planning to ATM planners, one for capacity constrained Master ATM planner OFG Response to Division Planning and, finally, one for IFG integrated solver to remix the time line for granted product requests. A thing to note is that since the ATM demand is only a month or so out from true end customer demand; there are now more signals that are available. They are the Customer (order on) Backlog signal, the Intel Commits signal to customer and the Judged Demand(forecast for a month out).These are combined and reconciled to create the demand and inventory signal for the OFG-IFG solvers. As with the Reset, the BOH amounts from CW+SFGI cause results from last month's cycle to affect current month cycle.

The SDD process or Algorithm which will be described in much more detail later is an algorithm that runs the Finish operation to decide how to align push based starts from the ATM resulting in SFGI inventory with customer demand based pulls from CW. This runs twice a day and helps correct some of the forecast accuracy errors that result from the ATM starts being planned a month and a half ahead of true customer demand pulls. The SDD algorithm looks at a composite of the Backlog signals, Request for the Boxing group and VMI Hub demand measures such as STR (Storage Transfer Requests) for demand pull and aligns it with a composite signal of ATM Qx weekly scheduled starts and COR's (change order requests) from customers.

### 2.2. Sales Process

**Customer Commits Process**

The Intel Sales process is essentially an exercise in supply allocation after all demand has been aggregated globally and across sales channels. Since a big part of the microprocessor market is serviced by Intel, its supply situation as well as customer perceptions of its supply situation both influence buyer decisions. The customer supply commitment process is designed to allocate products that may be supply constrained such as top of the line products that typically are in short supply when a new process is introduced. The allocation is designed to prevent gaming behavior and ensure that all customers across geographies and channels get an equitable supply of products, though it doesn't completely eliminate it. This means that customers have to project demand and Intel has to project supply availability. A central organization called MMBP then has to match supply with demand.

If supply outstrips demand every customer gets what they ordered. If demand outstrips supply then MMBP has to make a judgment call based on customer and Intel forecasts of what each geography and customer will probably end up requiring a quarter from now. This is because the factory network has a 3-4 month lead time and current orders are essentially forecasts for what will be needed a quarter out. Intel has a zero penalty order cancellation policy that customers enjoy till essentially the very last minute. This means that if customers cancel orders, Intel is left with unneeded inventory that it may have to take a charge for. Similarly if each sales geography thinks it will grow faster than the expected worldwide growth then judgment is required at a global level. Hence MMBP really focuses on making its own forecast (called Judged Demand) based on the geography forecasts but with adjustments to reflect its own judgment as to how much delivery customers and geographies will actually end up taking.

This judged demand is what is then allocated or "committed" to customers and geographies. These commitments are the right to order products or "book" orders up to the committed amount. After the few weeks to the booking deadline, the remaining commitments are redacted. Geographies do have the right to ask for what's called "allocation" which is a certain amount not directly attributable to a certain order but reserved for expected future orders. The Planning process through "Build Plan Reset" then uses this Judged Demand to do production planning for deciding on wafer starts.

**Hotlist or COR (Change Order Request) Process**

Intel uses a process called the Hotlist or Change Order Request Process for responding to customer demand changes after the regular forecast and planning process is over. This is the key way tactical adjustments are made and as such is very important to understand for purposes of improving customer responsiveness.

Cancelling orders or pushing them out in time is easy. Swapping order quantities from one product line to another related one or pulling the order times in may or may not be easy to honor based on product availability. The most common request is new requests for product called upside requests. Again these require the most review and a statement of supply assurance. Typically customer field sales communicate requests to their geographic business analysts (GBA) who try to meet requests that they can and escalate the unfulfilled request to product supply planners (PSP) in Division Planning. Product Supply Planners have more visibility into the pipeline globally and try to meet the requests they can escalating unmet needs at a weekly Séance meeting with MMBP. Similar to the concept of commits, granted hotlist requests can be underutilized and unused grants are pulled back in. For more detailed discussion of these sales processes see Chow 2004.

### 2.3. Demand Planning and Determination of Wafer Starts

Judged Demand (JD) is the key forecast published by MMBP which is used as the basis for deciding how many wafers to start in the Fab. It is published monthly and used for determining wafer starts through a process called "Build Plan Reset". It is a monthly breakout of projected demand for sku's extending out 4 quarters out. The current quarter or first quarter is used by ATM for Test/Die management and for deciding on inventory strategies. The second quarter projection is the key one that is used for the wafer starts decision since the Fab throughput time is 3-4 months. The third and fourth quarter projections are used for capacity planning and form the starting basis for the long range planning group's work.

The end to end process for getting to a JD takes a little more than a month. The process is done once for all the sku's across vertical business segments and typically involves hundreds of sku's per vertical. Key inputs are obtained from customer sales projections from Geographic sales forces, from macro-economic models, from billing trends and industry dynamics for all market participants, from third party analysis and from various channel information.

While JD forms the basis for Demand inputs into the Wafer Starts determination process, Inventory Strategy and Inputs are key to providing the necessary safety stock to support business and operations strategy. These two sets of inputs by product family are then used in the Build Plan Reset process to determine the Wafer starts and Mix needed for a quarter out demand.

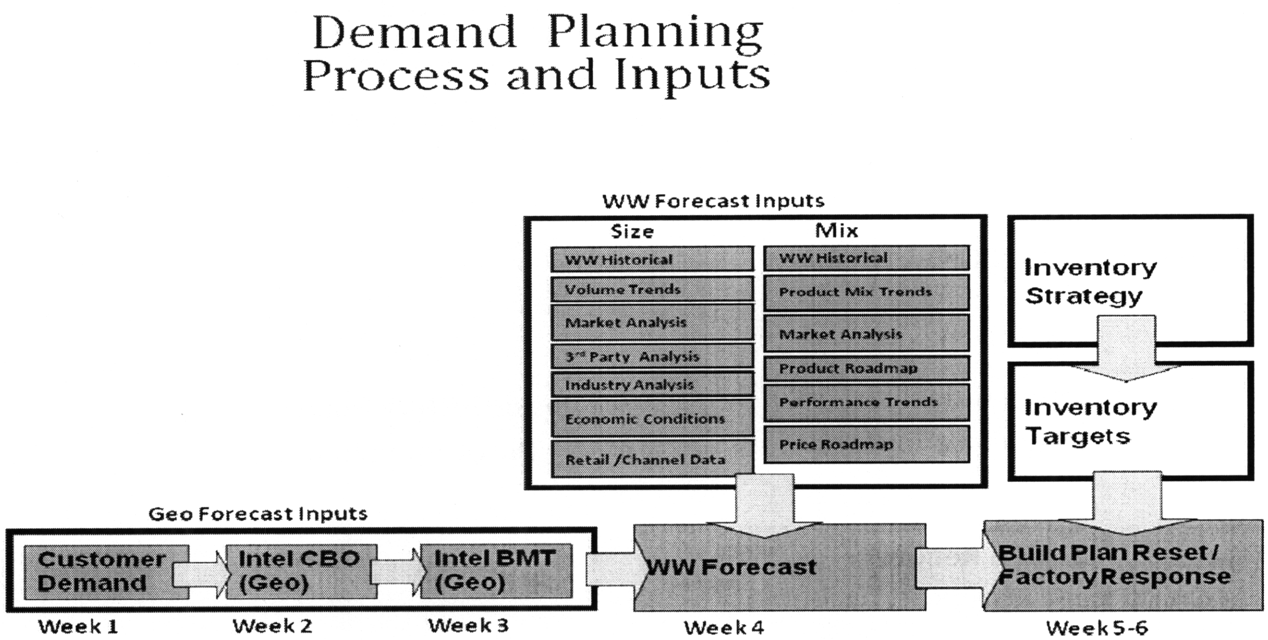



**Figure 2-4. Demand Planning Process and Inputs.**

(courtesy: A. Reyner)

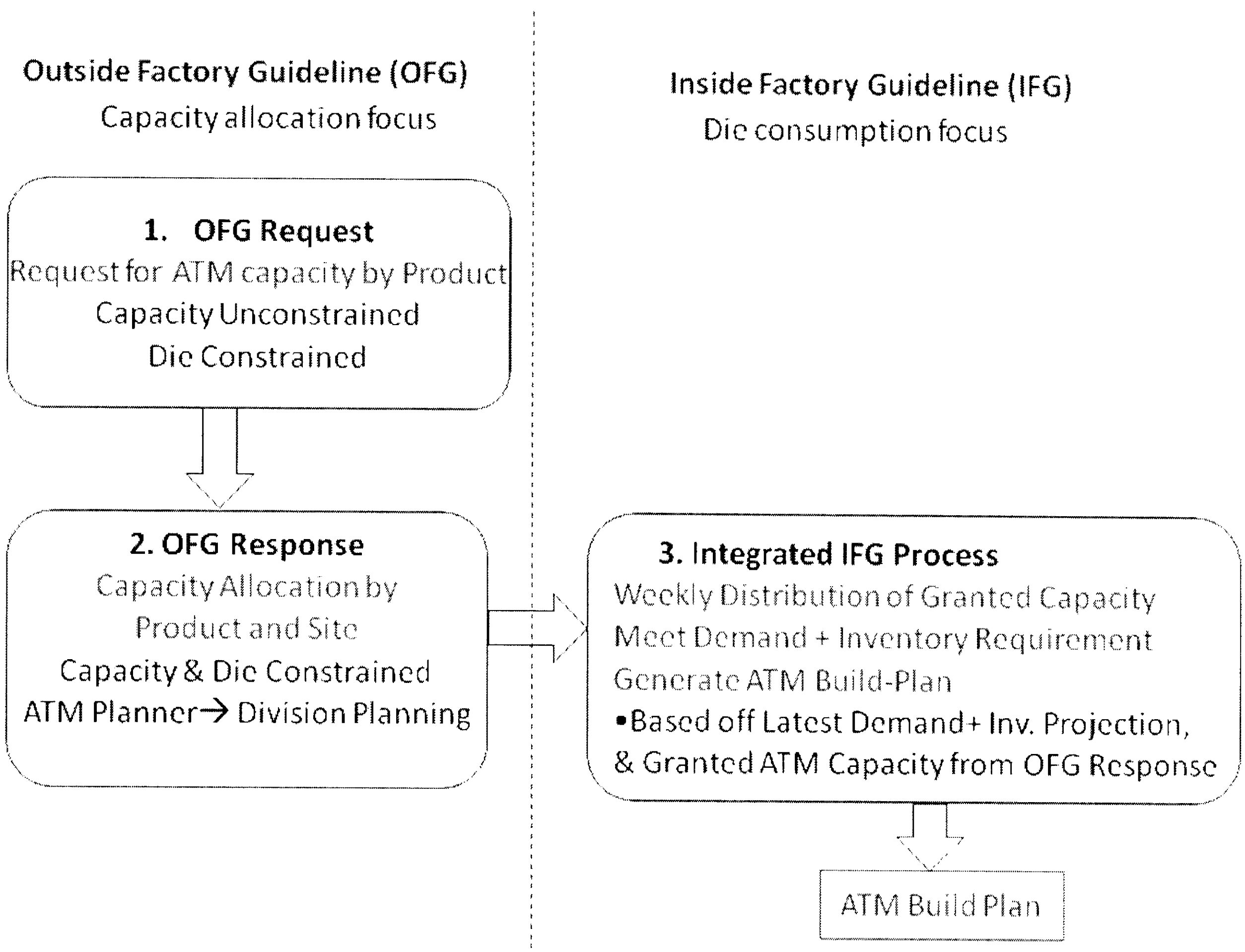
### 2.4 ATM Production Planning for Supply Assurance

**OFG-IFG Overview:**

The key Production planning processes used for ATM planning are the twin OFG-IFG processes. OFG or Outside Factory Guidelines is a process where a capacity-unconstrained (but die-feasible) demand request is made by Division Planners from the ATM planners. The ATM planners as coordinated by the ATM Master Planner respond giving a yes, no or partial yes answer to the request. This not only depends on their available manufacturing capacity but also on the BOM (or Bill of Materials) and tool/facility availability such as for die prep. This response to the OFG Request is called the OFG Response. A mix of manual planning and solvers are used for both these processes. The other Process called the IFG or Inside Factory Guidelines then is carried out by Division Planners with the agreement of ATM's where they rebalance the granted product mix and volume across time to meet the actual demand linearity. The first chart below shows this high level view of the OFG-IFG process.

The second chart below shows the simplified timeline for the OFG-IFG process. The OFG-IFG process is carried out in the first and third week of the month and affects ATM builds post the frozen horizon a month later for weeks 5 and 6 or 7 and 8 depending on the cycle.

**Figure 2-5. OFG-IFG Overview**



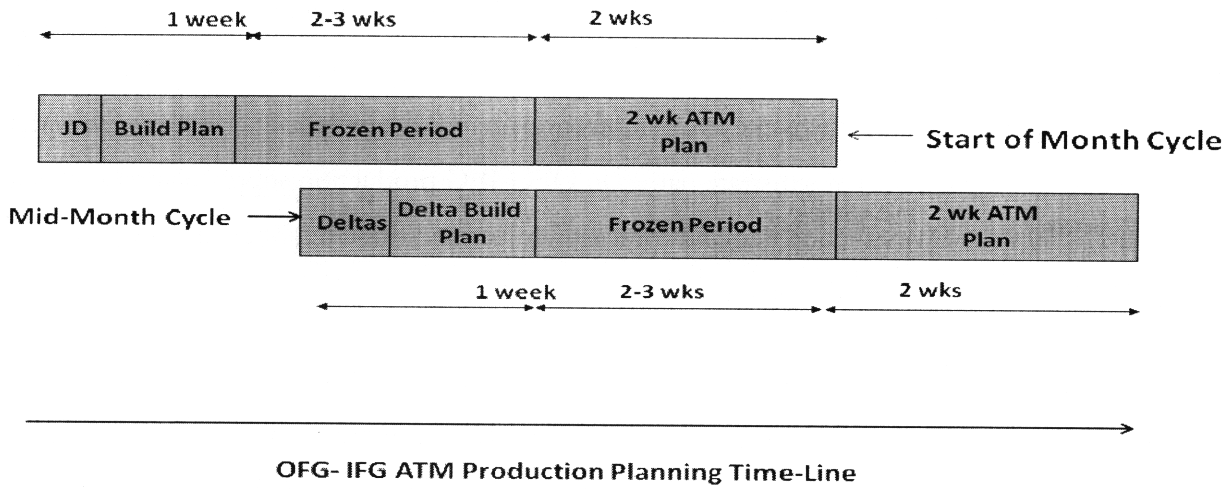



**Figure 2-6. OFG-IFG Timeline**

**<u>OFG Request:</u>**

The OFG Request is Division Planning's communication to ATM on the build targets. The request is unconstrained by ATM capacity but constrained by die availability. A key aim is to meet EOH Inventory Target post projected demand while keeping tool utilization high. A combination of manual effort and solvers is used.

OFG request typically matters only for the next two months though the farther out projections are used for capacity planning. Also the actual build plan that matters is only a two week plan about a month out after the frozen period of about 2-3 weeks to keep some level of predictability in the factory. This does hurt responsiveness though and later on we will see some processes for affecting change here. OFG-IFG cycle is done twice a month both at the start and middle of the month. The demand signal used is some combination of available signals such as JD, customer commitments and customer order backlog. The inventory targets are set by MMBP keeping die availability and other yield parameters in mind.

**<u>OFG Response:</u>**

The OFG Response is ATM's communication to Division Panning on the best possible build capacity for each product family requested given constraints of ATM tooling and die prep capacity along with die availability. The key function of this step is to do a Site Split or Capacity allocation by

Product and ATM Site. OFG response follows the same schedule for both planning cycles and for build plans as the OFG request as it is essentially a yes or no response to the OFG request. A thing to note in terms of the complexity is the mapping from demand to supply language. A single wafer might have for example 20 ATM substrate names in OFG-IFG production supply planning which might map to 50 sku's by marketing demand forecast families and these might eventually map to 100 SAP MM's (Material Master: SAP version of sku).

As far as the process goes ATM provides Division planning with OFG Response in an excel file called the Databank wherein the figures in Databank are ATM's committed build plan for each product. This response is given in weekly detail for the three quarter horizon, by wafer stepping, package and additional levels of detail as needed to ensure capacity utilization. The OFG response incorporates test capacity constraints, product prioritization & die/package availability at each ATM site.

**<u>IFG Integrated Request & Response:</u>**

The IFG integrated Request Response Process uses an IFG solver algorithm that attempts to take the granted capacity post OFG Request-Response and spreads requested builds out across time in such a way that at any time the cumulative total requested builds covers the highest cumulative demand from among a multitude of available demand and inventory signals (such as JD, Backlog and Commits) including that week. As a result the demand and inventory signals in the IFG are different than the one used in OFG. Given that all the capacity requested in OFG might not have been granted the solver has built in intelligence to make both product priority calls. It tries to meet demand on all products according to their priorities followed by their minimum requested inventory followed by the requested Target inventory. Any additional products are assigned based on discretion of planners.

# 3. RESPONSIVENESS

To be able to make Real-Time precision adjustments to world-wide chip production we need to define the problem of improving supply chain responsiveness since production and production planning form a control system with feedback loops that need to continually self-correct. This chapter looks at what responsiveness entails and the most expedient ways of approaching this issue. It ends by prescribing certain operative principles to guide policy actions and recommendations as well as a pilot idea to test some of the ideas described herein.

## 3.1 Complexity and its Implications for Responsiveness

### 3.1.1 Types of Complexities:

The inherent complexities in the chip business are many and varied. They include the following varieties all of which invariably reduce the ability to respond quickly because of the need to navigate through the complexity before crafting a response-

1) Product Complexity – The number of sku's has increased tremendously both in response to the increased competitiveness of the market as well as the proliferation of consumer electronic device types that run on silicon today. This has added all sorts of complexity in terms of processing variation to deal with different performance requirements, size and packaging variations to handle different heat requirements, etc.

2) Complexities arising out of Customer Fragmentation: As the consumer section of the market has gotten increasingly vibrant and comparable with the corporate segment the two segments with their different needs and different demand volatilities have added pressure on demand forecasting and responsiveness. This is seen in the widening OEM customer base and the widening spreads in their forecast accuracies. In addition the globalization of this customer base and demand and the increasing fraction of sales coming from newer emerging markets have also made this task harder.

3) Supply Chain Complexity: As the number of Intel facilities around the world have increased they themselves have added to the supply chain and logistics complexity. In addition since these different factories can make only certain types of technologies or products that makes it difficult to redo a production plan once it has been created and disaggregated around the world to respond to changing demand at the last minute.

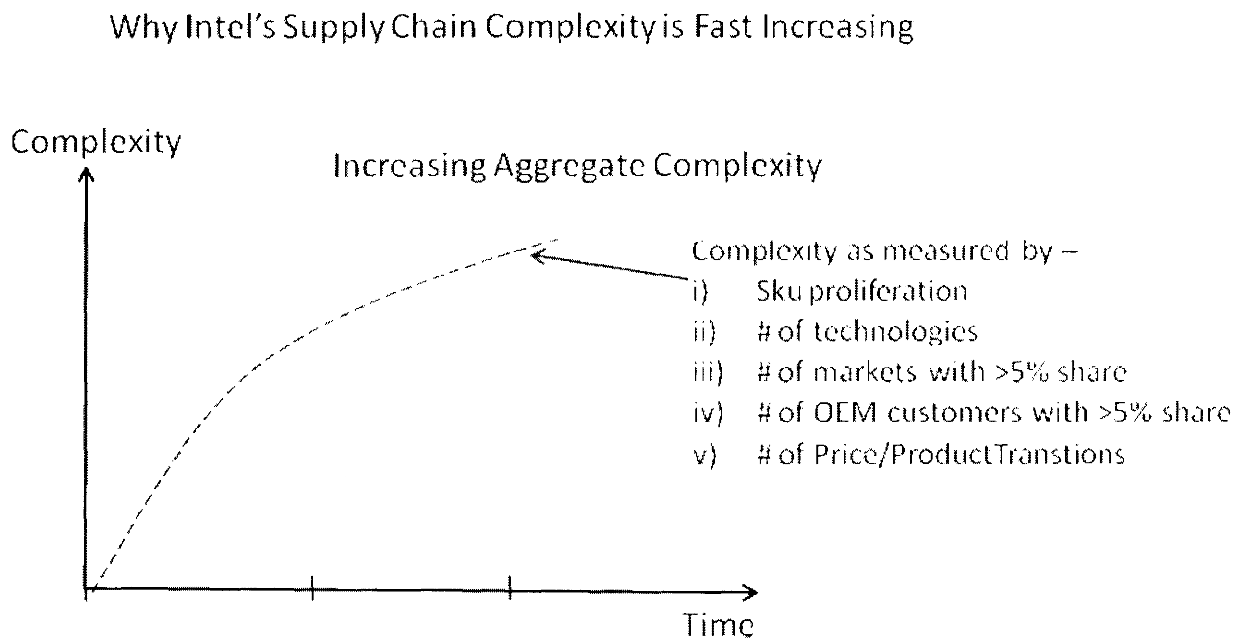



**Figure 3-1. Supply Chain Complexity at Intel**

### 3.1.2 Why Focus on ATM's for Improved Responsiveness:

The timeline for the entire delivery response from customer projection based JD to customer delivery is very long on the order of 3-4 months and is shown below.

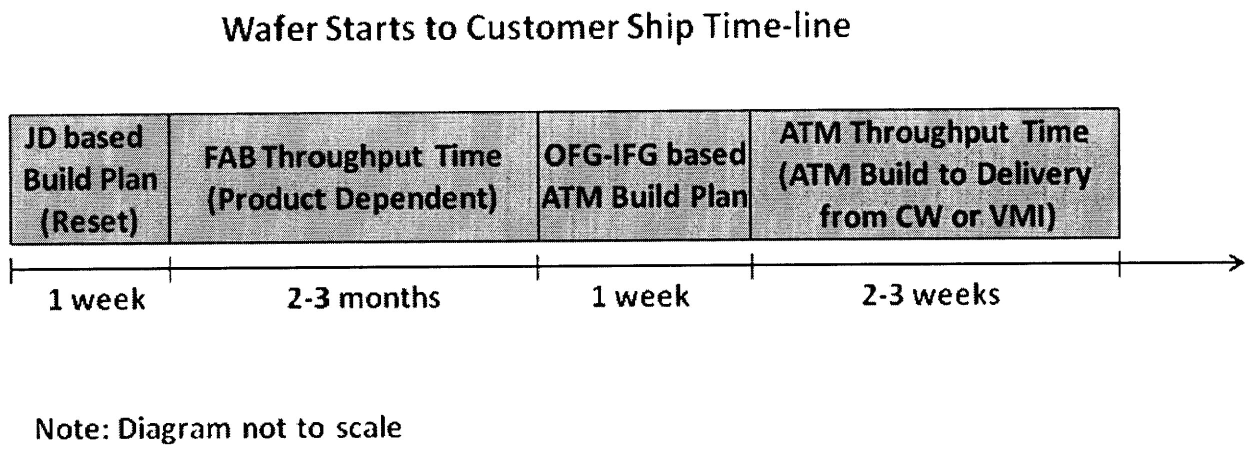
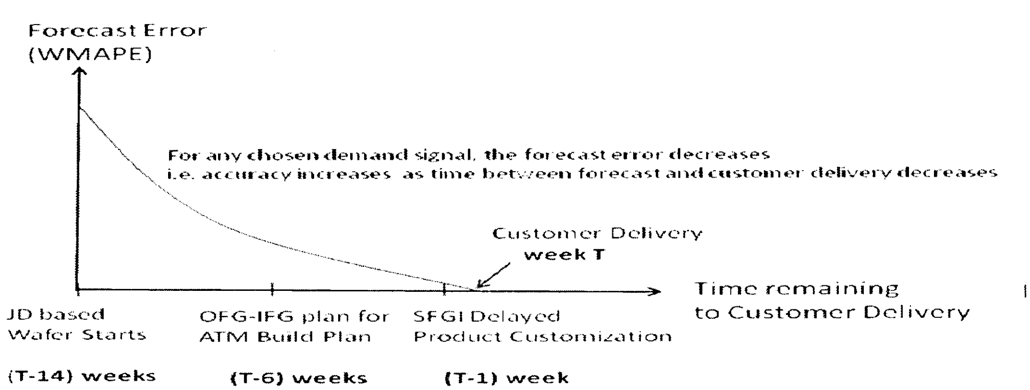



**Figure 3-2. Timeline for Production Processes**

While the JD projection based Wafer starts in the FAB to have the starting die for meeting customer needs a quarter out is a necessity, the challenge is the forecast error is very poor when looking out a quarter. After 3 months when the wafers are processed and the time for making an assembly and test production plan through OFG-IFG nears the accuracy is greatly improved however there is still some significant uncertainty surrounding final sku requirement and the demand only firms up in the last week or so. The final postponed customization done to Semi-finished goods and called SDD for SFGI delayed differentiation is only about a week or so away from Finished goods warehouse. The forecast at this point currently has the best accuracy. This is shown in the chart below. Given this the best point to focus on responsiveness is the ATM side rather than the FAB side. This includes both the OFG-IFG production planning and processes in the last 2-3 weeks of delivery such as SDD and ATM range planning, both of which are discussed in the next chapter.

**Figure 3-3. Degradation of Forecast Accuracy with Time**



## 3.2 Approaches to Help Responsiveness

### 3.2.1 Identifying Opportunities

When trying to identify the opportunities for Responsiveness it becomes imperative that a high level direction is set to help this. Two aspects of the direction are key. One is whether the approach should focus on coordination while involving as many people in the chain as possible or whether it should concentrate the decision making in the central planning group to prevent issues with strategy coordination. The other issue is how decision making should be organized i.e. at what level should decisions be made.

A) Approaches: Command and Control or "Wisdom of the Crowds":

While choosing an approach between centralized command and control and the "wisdom of the crowds" it is worth remembering that it has to be implemented by various groups of people in multiple globally dispersed organizations. It is difficult for centralized decision making to manage globally dispersed supply chains where responsibility for handling different processes lies with different organizations. So a key part of the chosen approach should be the ease of implementation and the ease with which the recommendations can be effectively sold to help facilitate coordination amongst the multitude of organizations. Since buy-in is so crucial to coordination and effective execution of strategy, a more decentralized approach is advocated.

A second reason for this is provided by the market itself. External changes such as economic crises, changing conditions in emerging markets, technological changes such as with atom, netbook, etc., are accelerating the industry clock speed. The current planning system works well when Intel's actions are driving most of the changes but not fast enough when the changes are driving Intel to act. Requiring all information flows to bubble up, pass through MMBP and then trickle down while other organizations wait for a decision to be taken can create bottlenecks. Trying to avoid this bottleneck increases the need for coordinated planning with more devolved decision making, clearer communication and more visibility to information across the organizations.

A third reason is the nature of knowledge about the drivers of change. Knowledge is increasingly disaggregated and local or product specific. Only in some cases does it make sense to bubble it to the very top and wait for a decision. In most cases where speed is of the essence and knowledge is perishable, trends provided by micro-feedback loops should be made actionable by individuals and organizations closest to the process. E.g. SDD is handled almost entirely by Division planning with very little information transmission. The key objective for MMBP then should be to make relevant information visible to decision makers and set business processes or checks based on volumes or product strategy to set trip wire or an "andon cord" calling for their attention.

B) Devolution of Decision Making: Close to the Top or the Bottom:

If we agree that an approach trying to use the distributed intelligence and knowledge in an organization is preferred than the role of MMBP may evolve to the point where it –

i) collects information from every stakeholder (external and internal),

ii) sets operational objectives in line with its strategy,

iii) ensures that incentives, accountability and responsibilities are well defined, well aligned and well understood,

iv) disseminates information making it visible to all(e.g. EOL, NPI, off roadmap moves in price, strategic build-aheads),

v) does more coordination among organizations(Geographies, division planning, etc.) and finally

vi) leaves management of individual pieces e.g. inventory management past ADI to division planning

Hence reducing the center of gravity for decision making and devolving it to the lowest level possible where it is as close to the actual entity in the field making the decision is preferable for increasing supply chain responsiveness.

### 3.2.2 Decision Making Choices:

After speaking with members of the various planning groups across the world, various ideas for improving responsiveness were explored and analyzed. Some of these ideas with their cost-benefit analysis are described in the table below. There were 5 main types of ideas-

1) Making Intelligence Visible: Improving information visibility across all organizations to improve the quality of decision making

2) Improving Communication: Since the setting of WOI targets is used across the board as a key way of communication, making the WOI definitions consistent would help improve communication tremendously. Inventory targets primarily provide directional information for the rest of the chain and so consolidating the number of definitions helps communications without limiting flexibility.

3) Adding Micro-Feedback Loops: Since some of the feedback loops in the planning cycle are pretty long, the goal would be to put in place shorter feedback loops to improve the rate of learning.

4) Improving ATM Planning Processes: Currently there is a lot of opportunity in improving ATM planning processes such as OFG-IFG, SDD etc. since they are closer to the end customer demand in time when the forecast accuracy is higher.

5) Reconciling different demand signals: Planners have many Demand signals to choose from without a consistent business process on what to choose when they widely diverge. Giving guidance here to reconcile the top spreads would help improve planning.

**Figure 3-4. Possibilities for Improving Responsiveness**

| No. | Possibility | Description | Benefits | Challenges |
|---|---|---|---|---|
| 1 | Making Intelligence Visible i.e. using information visibility to improve decision making | Making Added Intelligence Over JD, Backlog, Commits, DAD Visible Downstream through Flags in Database | 1. Using Existing Intelligence<br>2. Delivery Platform for all future Intelligence<br>3. Devolves power to decision-makers downstream | 1. Misinterpreting of Signal such as EOL |
| 2 | Improving Communication through WOI Alignment | Synchronizing WOI as MMBP Tactical Knob | 1. Currently used MMBP tactical knob<br>2. Excess Inventory Reduction<br>3. Frees up Capacity | 1. Multiple Orgs.<br>2. GEO EB target<br>3. Different Incentives |
| 3 | Increasing the learning rate by adding Micro-Feedback loops | Add loops from end of supply chain to earlier points preventing disconnects in planning processes. Improve the velocity of learning through short loop learning turns | 1. Eliminates disconnects<br>2. Disseminates information and learning across the chain faster<br>3.Prevents possibly amplifying bull-whip effects | 1. Separating signal from noise.<br>2. Identifying proper points to intercept and send signal to. |
| 4 | Improving ATM planning processes (OFG-IFG, SDD etc.) | Identify areas for improvement in current planning processes to improve efficiency and responsiveness | 1. Focused on ATM<br>2. SDD is key Push-Pull boundary<br>3. Unexplored with scope for improvement | 1. Identifying required change |
| 5 | Reconciling different demand signals (Commits-BL-JD-Geo tactical forecast) to reduce confusion | Guidance on aligning signals for OFG-IFG planning for improved co-ordination | 1. Improved Demand Input into OFG-IFG Solver<br>2. Excess inventory Reduction<br>3. Frees up Capacity | 1. Geo Bonus Target<br>2. Unsure NPI/EOL Supply Transitions<br>3. Defeats Purpose of clear un-judged customer signal. |

**Figure 3-5. Responsiveness Improvement Possibilities-Impact vs. Implementation Ease**

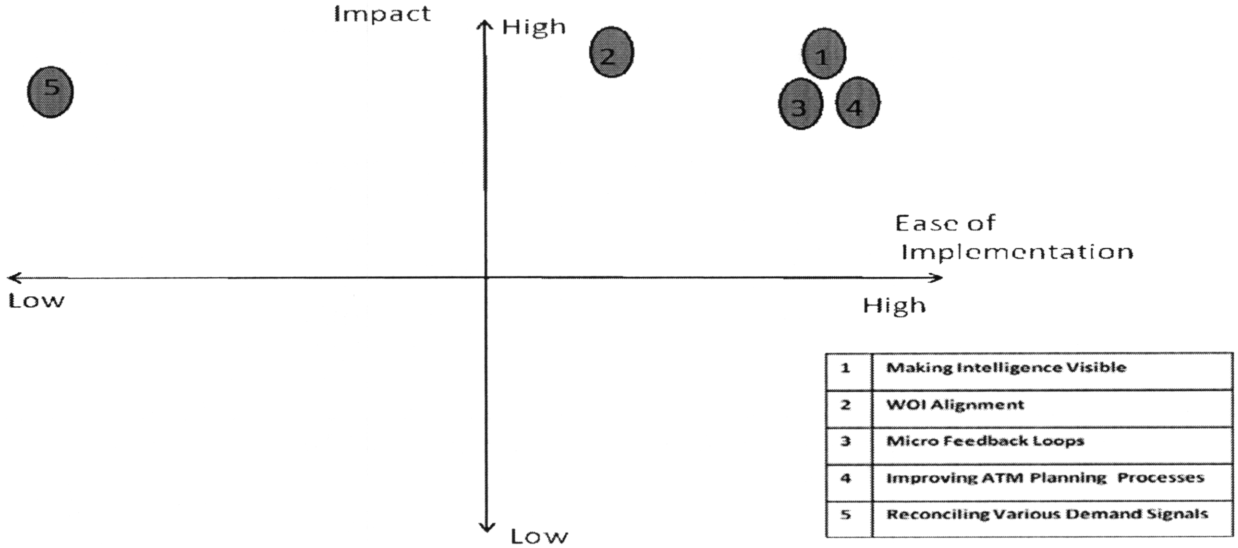
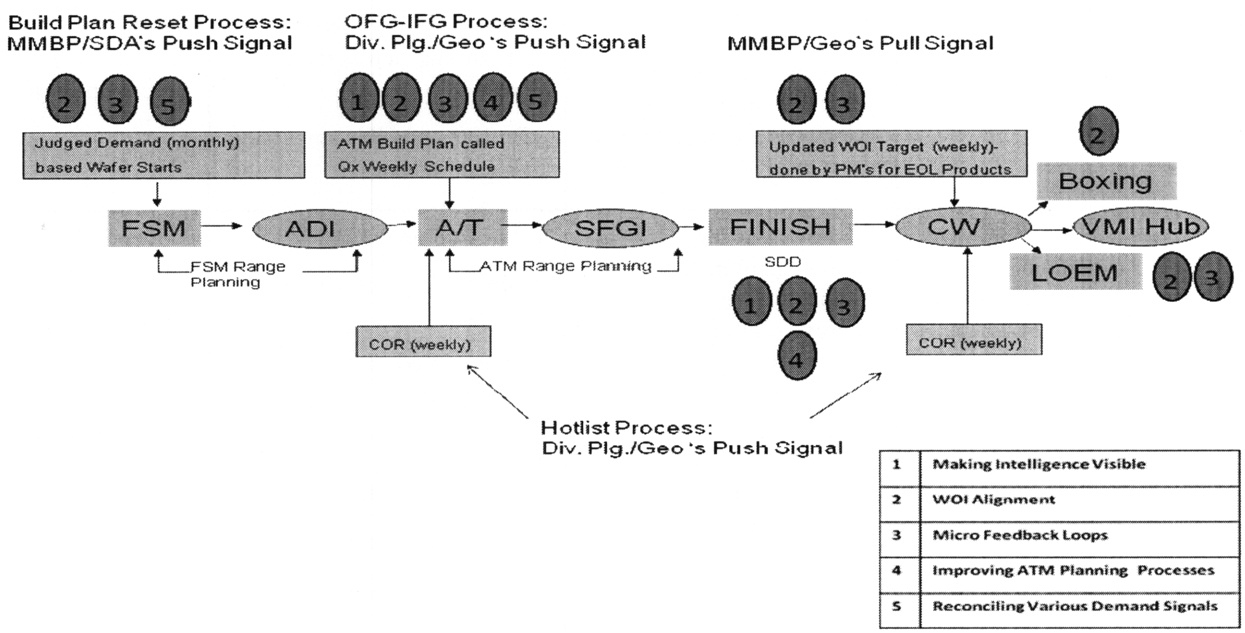



**Figure 3-6. Responsiveness Improvement Possibilities-Interception Points along Chain**




It can be seen from the above that option 5 reconciling various signals is a significant effort on an ongoing basis though it is highly impactful while option 2 on the WOI alignment is relatively easier though still significantly useful. The other options are all impactful and though they require some work to setup are easier than the demand signal work.

### 3.2.3 Decision Making Tree:

Given the above options it is worth putting a framework together to structure the options and understand how they all fit together. The decision tree shown below provides a simple way to do so dividing up the problem into three main parts –

1) Identifying and making changes in production plans to improve responsiveness

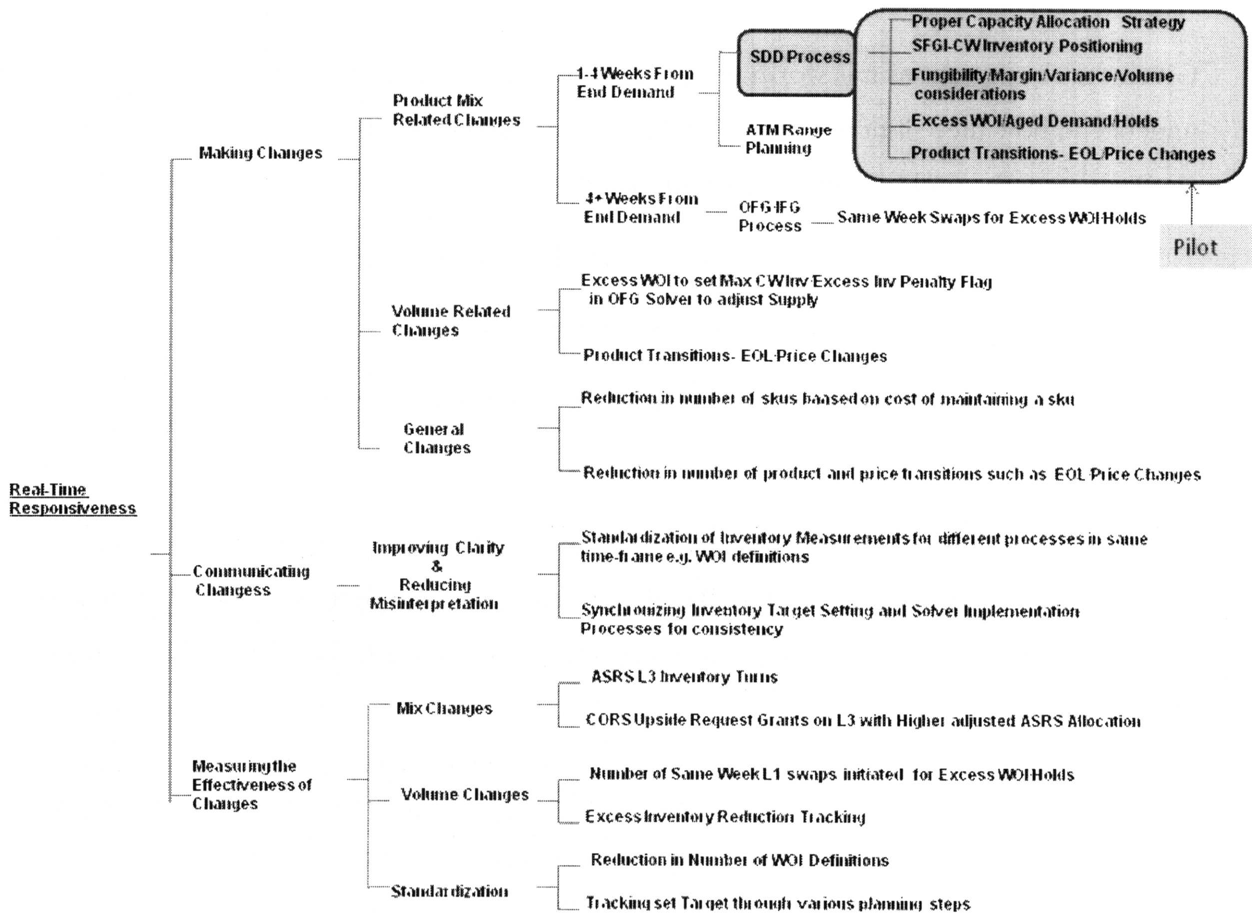
2) Communication these plan changes clearly and consistently across the supply chain

3) Measuring the effectiveness of these changes through some key performance indicators to close the learning loop

A last but related part is the identification of a largely self-contained pilot effort to improve responsiveness based on the above outlined options. In this case SDD has been chosen since it is both self-contained and has some special features which magnify the improvements made here.

The figure below shows a decision tree framework for thinking about the various options to improve responsiveness broadly classifying them under the umbrella of making changes, communicating changes and measuring changes. Each of these are separately described below-

**Figure 3-7. Framework for Improving Real-Time Responsiveness :**




## 1) Making Changes:

The changes that can be made to improve responsiveness can be broadly classified as either general changes or specific changes i.e. ones designed to correct for changes related to volume or product mix. Of the 5 broad options listed before, all except the one on improving communication fall under the category of making changes efficiently. While the 5 options are broad themes to keep in mind while thinking of specific options, the responsiveness framework above lists the specific changes and whether they affect volume or mix decisions or are general in nature. It is easier to think of Mix decisions as either ones correcting for delivery in the next 4 weeks (mainly including SDD and ATM Range Planning described in detail in the next chapter) or ones outside that period (and again here we focus on ATM planning i.e. OFG-IFG).

The SDD related changes were used as the pilot and are described in detail in chapter 4. The idea for the L1 level demand swaps between product families will also be described in more detail there. The volume related changes include a suggestion to have a micro-feedback loop between an excess inventory flag when raised at CW warehouse to the OFG-IFG planning process and solvers used there. The solvers have an option to penalize builds on certain sku's and the excess inventory flag at CW should be linked to the OFG_IFG solver as an excess inventory penalty flag. Currently excess inventory flagged at CW is used to reduce wafer starts in JD. The link to OFG-IFG is not formalized and relies on human communication to planners etc. The benefit of creating a feedback loop to solvers is the formalization of the process.

Another area is changes that happen during price or product transitions. The key here is to create visibility for ATM planners planning OFG-IFG , SDD etc. as well as making decisions on CW and VMI hub stock levels into the transitions about to take place. Most product transitions such as EOL and NPI are on the product plan (roadmap). But changes due to off roadmap transitions, or ones that are trending to be faster or slower than earlier projected, as well all price transitions, are not clearly visible to planners in the databases or solvers they use such as Qx weekly databases, OFG-IFG solvers, SDD interface, etc. If directional guidance and visibility can be given on these transitions without revealing sensitive price or margin information etc. then this could be of great help in leveling the informational disparity and improving the quality of decision making. Finally two broad and general improvement suggestions are to reduce either the number of sku's or the number of product/price transitions if the cost of managing them is more than the incremental profit contribution achieved by having them.

**2) <u>Communicating changes – Standardizing to communicate well</u>**

Making changes to improve responsiveness are not very effective if they can't be communicated across a global supply chain clearly and without misinterpretation. The key way MMBP and other planning groups try to communicate demand changes across the Intel supply chain are through inventory targets specified as x WOI (weeks of inventory).There are two main types of issues that have been identified by this project.

**Issue 1-**

The first issue is that inventory targets are set with a static view of the world e.g. MMBP might set a goal such as 2 WOI need to be kept for a particular family or sku. Where 1 WOI maybe calculated based on a projection of the next quarters demand projection/13 weeks. But the next quarter is a fixed set of 13 weeks when targets are set. But in the solvers which division planners use the next 13 weeks are defined on a rolling basis so that when week 12 is being looked at the solver is looking at the 13 weeks from week 12. This means that the solvers are doing something totally different than how targets are set. But it gets even more challenging. For a division planner trying to implement this goal in a dynamic (looking ahead on a rolling basis) OFG-IFG solver, they first need to have week to week demand requirement filled out and the die or Test-out's coming in. Then they need to iteratively populate the ATM builds to look at the resulting difference between outs and ins which is the inventory level. If this in any week doesn't meet the inventory target, then they need to adjust the optimization control flags in the solver which influences things like priority etc. iteratively to get them the desired inventory answer by changing builds. In the process of doing this they have now prevented the solver from optimizing builds.

One way around this is to have standardized business process guidelines on how solver flags are set, for inventory target setters like MMBP to use the same demand signal as the Planners trying to execute on their target with solvers and finally for the targets to be set by MMBP in the same way or with the same solver interface as used by planners. Currently there is no standardized business process for how solver flags are leading to wide variation, both planners and MMBP use different sets of demand signals, and MMBP sets its targets using excel spreadsheets with different formulae between different vertical groups, while planners use difficult to manipulate optimization solvers. The result is that after a lot of iterations by planners either the MMBP targets are met at the expense of optimizing production planning or production is optimized while MMBP targets are not met. Constraining the optimizer to get the required sub-optimum answer in some cases requires up to twice as much time for planning. Similarly based on the flag setting the resulting WOI (weeks of inventory) can vary by up to a range of two (2x).

**Issue 2-**

If WOI is the main term used for communicating all important demand and inventory projections and changes then it has to be totally unambiguous what it means i.e. what demand signal and in what period is it referring to, over what period is the average taken to get to an average week requirement projection and then how is the actual projection e.g. x WOI calculated. Unfortunately however this is not standardized across the chain. In fact it gets very confusing as there are several definitions used across the chain, and they all refer to not only different signals but also different time periods for averaging etc. The definitions are shown listed below and have been referred to the inventory team at Intel which has recognized this as a major issue and is working on standardizing it.

**Figure 3-8. WOI Definition By Production Planning Process**

| | Weeks Of Inventory (WOI) Definition | Process |
|---|---|---|
| 1 | $[{}^{1}\sum^{13}$ (Next Qtr Demand )]/13 | MMBP Corporate Target |
| 2 | $[{}^{1}\sum^{13}$ (Next 13 weeks Demand)]/13 on weekly rolling basis | FAB Div. Planning Wafer Starts FAB and ATM Build Plan |
| 3 | $[{}^{1}\sum^{4}$ (Next 4 weeks ATM Plan)]/4 ,weekly rolling | ATM BOM Request Solver |
| 4 | $[{}^{1}\sum^{4}$ (Next 4 weeks ATM BOM Plan )]/4 on weekly rolling basis | ATM BOM Request Solver WOI |
| 5 | $[{}^{1}\sum^{4}$ (Next 4 weeks ATM BP)]/4,weekly rolling | ATM Capacity Plan Solver |
| 6 | {$[{}^{1}\sum^{14}$ or ${}^{1}\sum^{18}$ (D1 or D2)]/(14 or 18)} x (14 or 18) | ATM Finish SDD Algorithm |
| 7 | $[{}^{1}\sum^{4}$ (Next 4 weeks Orders)]/4 | Excess Inv. Measurement |

One of the key challenges faced in communicating tactical demand updates for precision adjustments is that inventory measurement definitions are defined differently at different locations based on the time frame and process. Understanding various processes and measurements to standardize these prevents misinterpretation. As seen in the table above a major challenge is the use of different Demand signals (Judged Demand, Backlog, Commits, OFG Adjusted Demand, IFG, Division Adjusted Demand, etc.) Another issue is the use of different types of product groupings between Supply and Demand Processes with many to many mapping at each level (e.g. Wafer types, Stepping's, BOM, etc. for Supply and Sku's, Marketing Family, MM's, etc. for Demand)

### 3) Measuring the effectiveness of changes – closing feedback loops –

There are two kinds of system philosophies involving demand forecasting-

1) Systems which acknowledge there are limits to how accurate a forecast will be and have quick feedback loops (as regards information on demand changes etc. that are important for planning etc.) to course correct for demand changes improving accuracy by increasing responsiveness and

2) Systems based on the belief that the best way to improve responsiveness is to primarily focus on improving the accuracy of the forecast.

The primary focus of the improvement suggestions here has been the effort to improve feedback loops for two reasons –

1) There is a significant amount of work being done in the area of improving forecast accuracy, while not enough has been done on the area of improving feedback loops

2) In a model of distributed decision making the best way to improve the quality of decisions is to improve information visibility (Lee, October 04) through a system of interlocking feedback loops.

A key area of improvement here is the feedback regarding how well the system is doing i.e. some metrics to measure the quality of forecasting, build algorithms, decision making etc. As shown in the responsiveness framework, these can again be sub-divided into mix, volume and standardization related metrics. To assess mix improvement achieved through the SDD Pilot improvements, the inventory turns in SFGI and the increased numbers of upside change order requests after a change is made are key metrics to look at. To assess volume improvements reductions in excess inventory and the number of demand swaps from one sku to another (explained in next chapter) pro-actively implemented can be used. For work standardization efforts the reduction in the number of WOI definitions in use across the chain as well as the fluctuations in demand or inventory quantity being planned (caused purely as an artifact of the planning software or process or interpretation of signal being used) can be used.

**4) Identifying a pilot**

The best pilot location in the supply chain is in the SDD SFGI Delayed differentiation location. The main reasons for this are as follows-

1) The entire operation is self contained within a single group in terms of planning and execution.

2) SDD currently is the main push-pull interface as far as production planning goes. The OFG-IFG process based ATM plan is the push based production plan under which the SFGI goods are made and stored in the ASRS (Automatic Storage Retrieval System) space. From here only the volume and mix requested in the immediate few weeks by customers, as evident through CW (Component or Finished Goods warehouse) depletion, are finally finished for shipping and delivery.

3) The SDD business process guidelines are unclear resulting in inconsistent usage across products. So this location offers the most promise of improvement in responsiveness.

4) This location being the last process before customer shipments through CW or VMI etc., the forecast accuracy is highest here.

Given these reasons the next chapter describes in detail the actual pilot process and its results.

## 3.3 Harnessing Business Intelligence to improve Responsiveness

### 3.3.1 Types of Business Intelligence: Both New and Already Available

There are many different types of business intelligence which following the general approaches outlined above can be used to improve responsiveness and make precision real-time adjustment. They are discussed below under the categories in which they fall

**1) Intelligence visibility:**

- Since so much of the supply chain volatility is caused due to the price/ product transitions that happen frequently, flagging them across planning databases for visibility across the chain while addressing the confidentiality issues is a key lever for responsiveness improvement.

- Having visibility into the orders on record to judge their quality and likelihood of materializing may also be useful for prioritization purposes. Examples of this are the amount of customer backlog which is on hold due to reasons such as customers failing certain checks, to the portion of demand forecast that is projected upside rather than actual demand.
- Making planners who accept change order requests aware of the ability to prioritize orders, even inside the ATM frozen period through SDD and range planning, is very useful.

**2) <u>WOI Alignment:</u>**

- Standardizing WOI definitions will help communication across the chain
- Setting WOI targets on a rolling basis weekly just like in optimization solvers helps align target setting with the solvers used to execute on these targets.

**3) <u>Micro-feedback Loops</u>:**

- Often an inertial time lag is seen in information flow in planning processes, where even after receiving new information (e.g. An excess inventory signal at CW) nothing happens (e.g. WIP is processed the same way at ATM's) till the next planning cycle comes through. Efforts to have more frequent planning cycles are worthwhile. Given a certain cycle, faster transfers of this information to all the intermediate upstream interception points even before the main planning decision point can make use of this information to do things differently. This can improve agility, coordination and the rate of dissemination of information or learning substantially at a fraction of the effort of an additional planning cycle e.g. using a trend to reset WSPW takes several months to take effect, using it to affect OFG-IFG takes up to a month, while using it to affect ATM range planning takes less than a week and for SDD takes less than a day.
- The loop wherein feedback from price transitions affecting demand trends to price setters needs to be sped up for more nimble action on discounts etc. Similarly Product Transition trends are very disruptive when EOL products linger on longer or NPI products fail to swap more demand off older products sooner. Providing clearer feedback on this to both planners and pricing/branding groups helps respond sooner on changing forecasts and builds.

- Other signals that could be exploited for projecting sooner than anticipated EOL are when customer order commitments fall below projected demand, or when there is available inventory for upside requests but it isn't being used up.
- Tying in the Excess Inventory report outs at CW to set flags such as Excess Inventory Penalty, maximum inventory for underweighting certain builds in OFG, or increasing set aside quantities through ADI Minimum Flags can also be effective. Similar are the SDD and range planning statistics to set OFG flags.

**4) Improving ATM Planning Processes**

- Having a Sku Level acceptance list for deciding whether to allow ATM's to pull-in next week's production plan to this week for lot size or capacity utilization can be helpful.
- Using weekly linearity or breakout of demand in forecasts such as JD, Geo Demand etc. may prevent the misuse of assuming an equal breakout across weeks of monthly forecasts.
- Providing confidence bands or variability on forecasts which is currently done via WOI target adjustments to the up or down side would help planners downstream
- Adding a global solve step for the individual solves done by planners helps to really put extra inventory builds where needed for proper optimization

**5) Reconciling various demand signals**

- One of the biggest problems facing ATM planner's down-stream is what to do when the multiple signals that they have access to (such as commits, backlog, JD, Geo Demand, etc.) diverge vastly. They typically will act conservatively choosing the largest amount for every sku. This would be greatly improved by either having a signal (a planning andon) bringing different forecasters and planners together when things diverge by a certain percentage, or by having a pre OFG-IFG step for reconciliation of top 5 or top 10 biggest divergences.
- Similar reconciliation in cases of overbooking for the 0-4 week tactical horizon or when sales forces still have large amounts of demand on supply reservation without having an assigned and ready customer will be useful.

### 3.3.2 Delivering Intelligence

**1) Aggregating Intelligence:**

Once the intelligence identified above is obtained, the format in which it is delivered is important. Whether it is an information packet or visible right through various planning databases such as Qx weekly in the case of ATM planning, for example, is a key choice. Another issue is to prevent unsophisticated users of this information from misinterpreting the information. A product identified for EOL may not be transitioning as fast as another product similarly identified. How guidance or extra commentary is provided when providing broad access to information or intelligence is therefore a key issue. In addition, aggregating the intelligence can take multiple forms depending on what type of actions one wants the user to take. Conceivably a combination of all the intelligence can lead to an adjusted signal or, through an algorithm or business process guideline, lead to the making of certain decisions. A range can also be a useful way of communicating intelligence where different types of information either increase or decrease confidence bounds on a given range or change the range around the forecast itself.

**2) Integrating Intelligence with Production Planning Processes:**

In each of the above described cases the planning professional can make or alter certain planning decisions as and when new intelligence is available. This requires the close alignment of intelligence in a way that is actionable to the regularly scheduled production planning process. As such to integrate intelligence delivery seamlessly with planning processes the following considerations should be looked at –

- Delivery location – SDD, ATM Range Planning, OFG-IFG, ATM Planning
- Type of intelligence from list described in section 3.3
- A choice of strategic product categories – NPI/EOL products, High Volume parts, High Price or most Price volatile parts etc.
- Type of Decision being influenced

## 3.4 Case Study: Using Intelligence to set Responsiveness Improvement Strategy:

The previous sections have discussed broadly the types of strategies useful for improving responsiveness, as well as the types of business intelligence available for use with these strategies. This section discusses a case study trying to use a specific type of business intelligence for drawing insights into actions that would help improve responsiveness.

The numbers used here are normalized to disguise them. The analysis used a sample month and a certain statistically significant set of sku's in a product category. The signals used were both "push" based obtained from production planning and "pull" based obtained from an effort to replenish inventory depletion caused by customer ordering. Four main insights for the direction responsiveness improvement efforts should take result:

### 1) Push vs. Pull:

A Push based system is a made to forecast system while a Pull based system is a made to order system. Demand signals can be forecast based or Push based using this terminology. Or they can be demand signals generated based on customer ordering and in that sense be Pull based. The reason why the Pull signal still has an error is that the customer demand typically changes up to the very end due to the zero penalty cancellation policy that Intel has. The 2 charts below show both for 2 weeks and for 4 weeks, error distributions related to 2 signals denoting cumulative demand: a "push" based signal called commits denoted by CMT and a "pull" based signal called backlog denoted by BAB for a certain sample number of sku's over a sample month. The X axis denotes a normalized measure of absolute percentage error i.e. Magnitude of Error and the Y axis denotes the number of data points in each of the Error buckets. So in chart1 for example there are about 60 sku's for which the CMT signal had an error of about 500 on an logarithmic error scale, while there were only 30 sku's with error in that range for the BAB signal.

The absolute percentage error is calculated by comparing the demand signals with "true" demand as measured by actual sales. The charts show the frequency distribution of data points or sku's falling in each error bucket. Regardless of whether it is 2 or 4 weeks, the "push" signal is worse than the "pull" signal when it comes to forecast accuracy, suggesting that to improve "responsiveness", the "pull" signal should be used to the extent possible for production planning.

**Figure 3-9. Push vs. Pull Signal Comparison at 4 weeks**

Note: All numbers are representative but disguised

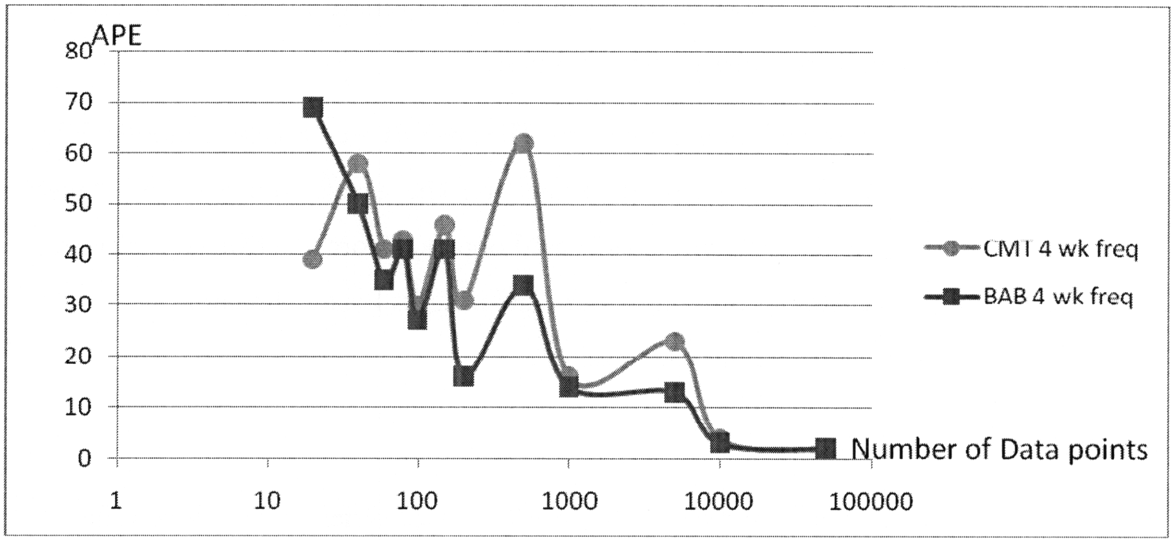
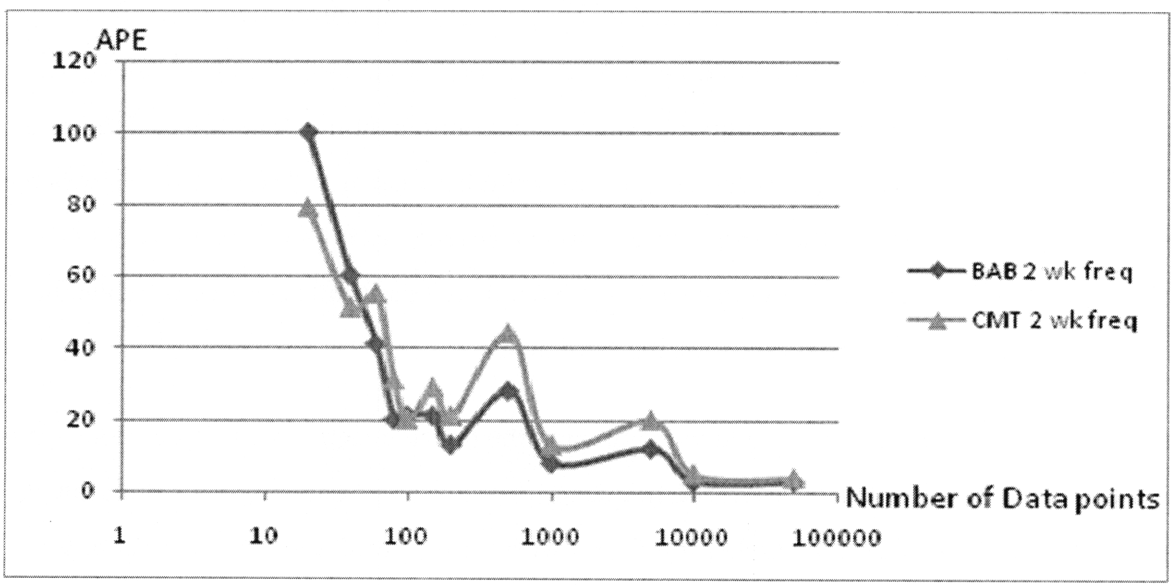



**Figure 3-10. Push vs. Pull Signal Comparison at 2 weeks**

Note: All numbers are representative but disguised



**2) Cumulative 2 week vs. Cumulative 4 week Demand Signal:**

Note: All numbers are representative but disguised

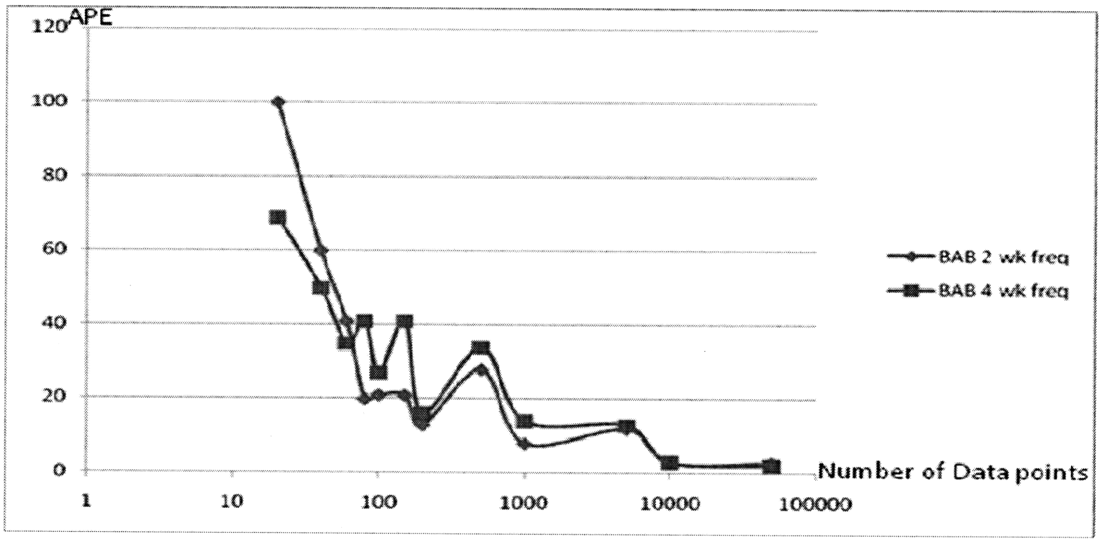



**Figure 3-11. Cumulative 2 week vs. 4 week signal Comparison for BAB**

The above chart shows that forecast error for a given "pull" based signal, called "BAB" in this case, the cumulative 2 week forecast errors are typically smaller than the cumulative 4 week errors. Currently 4 or 8 week forecasts are done in OFG-IFG for ATM planning. So while in the case of FABs this might not be possible given the 3-4 month long lead time, if ATM Production throughput time and OFG-IFG production planning time can be improved below a certain critical point, then using 2 week customer pull as the demand signal instead of a 4 or longer week "push" signal would improve accuracy and customer responsiveness. Note that the averaging of forecast errors over a longer period is more than dwarfed by the significant inaccuracy due to a lack of visibility in forecasting.

### 3) Level of Forecasting Granularity for Production Planning:

Note: All numbers are representative but disguised

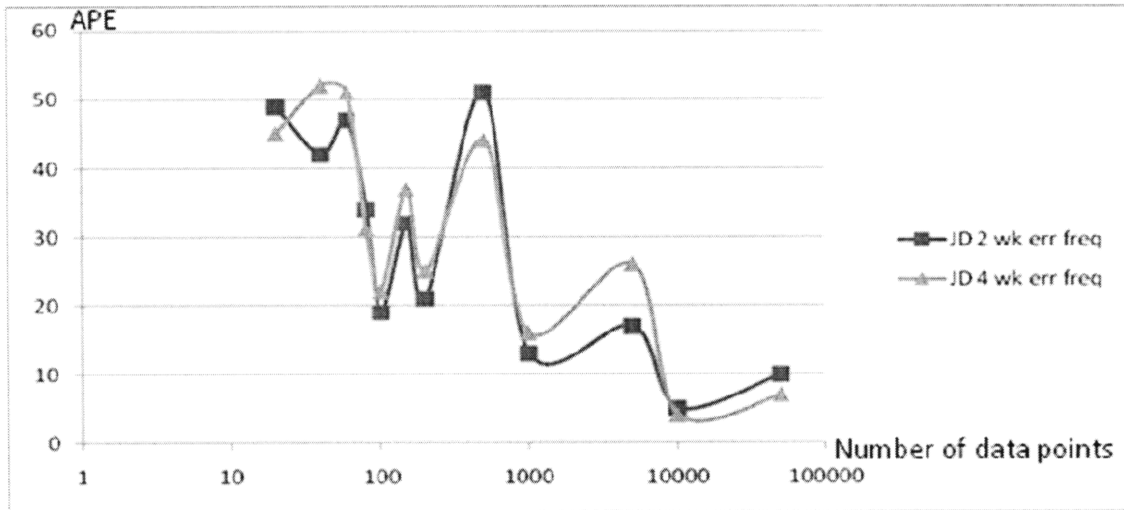



**Figure 3-12. Cumulative 2 week vs. 4 week signal Comparison for JD**

The above chart shows that forecast error for a given "push" based signal, called "JD" in this case, the cumulative 2 week forecast errors are not (statistically significantly) necessarily smaller than the cumulative 4 week errors. This is interesting given the result in the previous bullet. The reason for this is that the JD signal is predicted on a monthly basis and then using a linear "4-4-5" weekly schedule per quarter it is broken into a linear weekly forecast(even though it is known that the demand is not evenly distributed). Due to this artifact the accuracy advantages of forecasting and planning 2 weeks ahead vs. 4 weeks ahead (see description for point 2 on previous page) is lost. This demonstrates the importance of forecasting at a granularity that is consistent with the production plan being developed. A monthly JD forecast may be suitable for the monthly wafer starts plan but is not suitable for use in the ATM production plan which is used with a weekly granularity. In addition there are several well known phenomena causing non linear demand distribution across the weeks such as efforts to meet end of the month or end of quarter sales quotas, typical price changes in 4$^{th}$ week of Quarter, product and price transitions on products at end of quarter etc. In all these cases accuracy and responsiveness decreases when trying to artificially force fit a monthly forecast into a weekly plan.

### 4) Judged vs. Unjudged Signals in the Tactical Horizon (Near Term Horizon):

Note: All numbers are representative but disguised

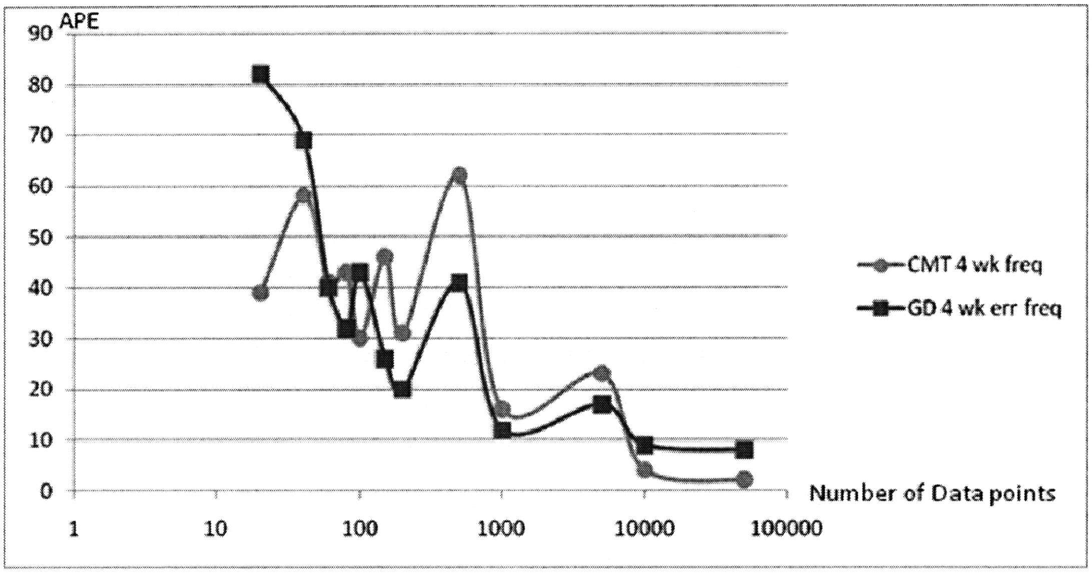



**Figure 3-13. Judges vs. Unjudged signal Comparison in Near Term Tactical Horizon**

The above chart shows 2 signals an "Unjudged" or relatively unadjusted signal from the customer through sales called "GD" and a "Judged" or adjusted version of that signal using business intelligence about supply constraints and past order patterns. The interesting thing to note is that the Judged signal is arguably worse or at best no better than the Unjudged signal despite significant effort at trying to improve on the customer signal. This suggests that while forecasting a few months out forecasting judgment maybe necessary to correct for optimism and bias, over the next few weeks, the accuracy and visibility is high enough that efforts at judging might only succeed in introducing more error.

# 4. SDD

## 4.1 SDD Process Overview

SDD or SFGI Delayed Differentiation is Intel's Postponement Strategy for the Final Finishing Operation. Final product characteristics like Speed and Power rating can be fused (blowing a fuse) to one of many values by down-binning the part. Current production relies on Push based ATM starts to SFGI and End Customer Demand based Pulls from SFGI. SDD is the critical Push-Pull interface process in the Supply Chain located at SFGI. The section below explains the problems that give rise to the need for SDD as well as what the SDD process entails. The sections after that describe the pilot that was carried out at SDD to improve responsiveness and the results of that pilot.

### 4.1.1 Why SDD

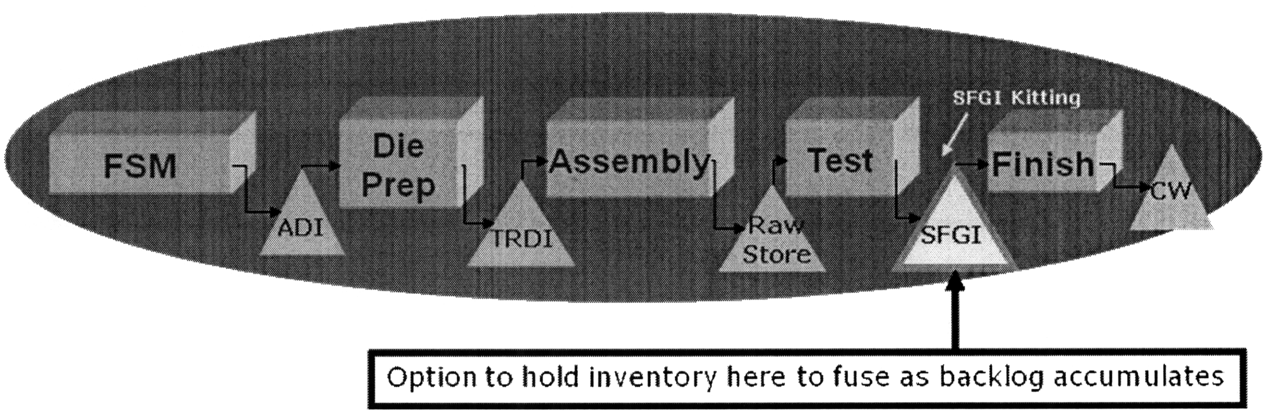
The ATM Factory currently builds to a push based Build Plan based on the OFG/IFG process. In the absence of any postponement strategy, the final MM or L1 level sku is completed through the Finish process based on a 4-8 week old forecasted demand. As previously discussed the forecast accuracy significantly improves as the final delivery date from CW approaches. This is because there is a no penalty order cancellation policy and the customers change the composition and volume of orders substantially as they have more visibility into what the end customer desires. Inherently higher uncertainty over a longer duration causes higher demand variance but there are also other reasons. Since there are more steps and players involved over a longer duration versus a shorter duration, the bullwhip effect is inherently higher over a longer duration.

The customer order on backlog fluctuates daily till the very last day resulting in significant mismatches between what was made based on OFG-IFG versus what is actually needed leading either to shortages or excess inventory held at CW at the sku level. This leads to both lost upside opportunities and higher costs from obsolescence and holding of unneeded inventory. A large portion of the unmet opportunity is actually within the same product family and differs only from what was made in the final Finish step. As such postponing the final Finish or Fusing decision to the very last minute and basing it on a day to day Pull Signal based on customer backlog makes a lot of sense to get the right mix and volume. This is what SDD or SFGI delayed differentiation does. It allows the ATM to build needed mix and volume of products based on latest customer backlog.

### 4.1.2 SDD Business Process

The SDD algorithm bases the Finish or fusing decision on real time CW supply and demand, also called customer Backlog. The decision on what and how much to make is based on an SDD algorithm run each shift. It remixes the current week builds within fuseable families to where need is as constrained by the product family levels being built in the build plan (called Qx weekly schedule). Normally production change decisions in the current week and next week are off-limits to planners to avoid ATM manufacturing from being whipsawed. This 2 week period is called the Frozen Horizon. But with SDD the planners can actually submit same week demand swaps (at L1 level) based on the mismatch between current customer backlog and the OFG-IFG based Build Plan for fungible sku's (i.e. sku's within same product family that only differ in the final finish step). The diagram below shows the relative locations of the SFGI inventory that is stored in a space called ASRS (Automatic Storage Retrieval System). The SFGI is at the end of the ATM and feeds the final Finish step before the finished goods are sent to the CW warehouse.

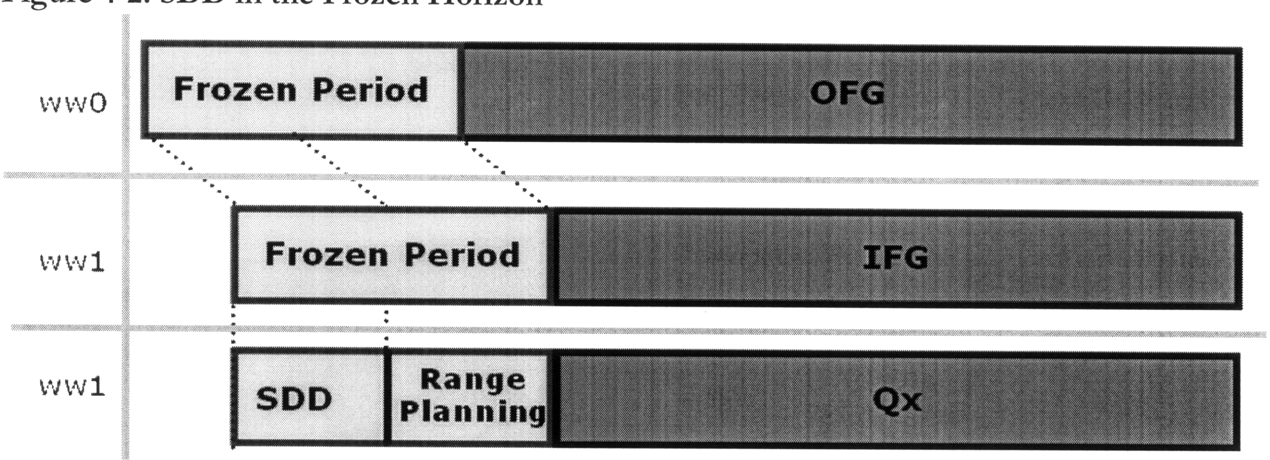
**Figure 4-1. SDD in the Supply Chain**



### 4.1.3 Where SDD Fits in the Frozen Horizon

The figure below shows how SDD relates to OFG-IFG as well as to the other process in the Frozen Horizon viz. ATM Range Planning. The current week and the next week are frozen from a ATM build plan change perspective. Within this frozen horizon SDD dictates current week production and any build plan changes have to be limited to Finish or Fusing related changes called L1 swaps within the same L3 level Product family. The ATM production for next week can be changed via a process called ATM Range Planning. While we won't go into the details of this process, the Range planning effort allows factories to make more or less (at the L3 or product family level) than the OFG-IFG plan based on a calculation that determines whether there is more or less demand now as compared to when the plan was made. Beyond the frozen horizon the production plan of record is still the one based on previous OFG-IFG cycles. So in essence SDD and Range Planning are the only two POR (plan of record) processes that can be used to make production plan changes within the frozen horizon and their relation to each other and OFG-IFG is shown below..

**Figure 4-2. SDD in the Frozen Horizon**

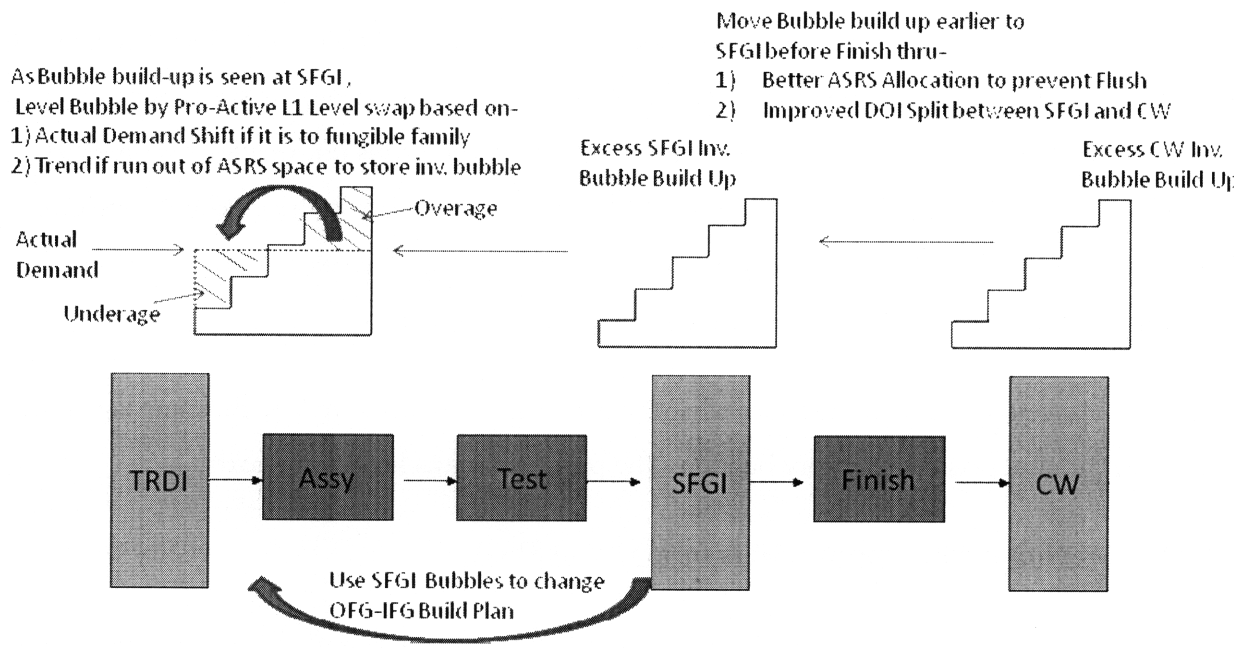



## 4.2 SDD Pilot Overview

### 4.2.1 Pilot Concept

The figure below shows the concept of the SDD Pilot. A bubble refers to an inventory buildup at either CW or SFGI. A swap refers to a demand swap which is primarily a shift in requested production i.e. a shift in demand from one sku to another. Currently the response of ATM Planning processes to change in demand is very slow. As described before SDD is a critical Push-Pull point in the chain where the forecast-based OFG-IFG push demand meets the customer driven pull as measured by same week customer backlog as well as CW and VMI depletion. The pilot tries to outline what can be done at the SDD point to improve responsiveness while remaining without broadly changing the main OFG-IFG process. In the last chapter on recommendations, the ideal improvement step is outlined which would require bigger structural changes in OFG-IFG.

**Figure 4-3. SDD Pilot Overview**




The Responsiveness Pilot is based on 2 main concepts –

1) Facilitate better bubbling up of customer Pull vs. IFG Push mismatches through earlier/clearer inventory bubbles- Appropriate ASRS(SFGI ) allocation is used to address this

2) Facilitate balancing out these bubbles through pro-active Demand Swaps- Make bubbles visible and encourage Division Planning to dynamically rebalance through swaps to address this

When SDD capacity isn't allocated right it results in a high rate of Flush or Push based (on 1-2 month old OFG-IFG based Demand forecast) Starts into CW. This results in 2 main issues:

1) Fewer Inventory Bubbles or buildup of unneeded Sku's in SFGI (ASRS space)-

SFGI Inventory Bubble build up is the signal to Division Planners of unexpected demand shifts. An unclear Pull signal prevents demand swap action from unwanted sku's to wanted sku's. This reduces possibility of upside grants of customer change requests and helps add to excess inventory

2)High Flushes resulting in Risk Finish decisions and builds into CW off months old Forecast Plan

This adds to unneeded inventory. It's a bigger issue in case of transitions like EOL, Price transitions etc. where inventory may not be able to be sold. Currently risk decisions are not prioritized for Profit margin increasing the cost of the risk taken.

The figure above shows the 3 main inventory locations influencing SDD i.e. TRDI(or ADI), SFGI and CW. Currently an excess inventory flag is raised when an inventory bubble starts building up at CW above a certain range around 2-6 weeks for a sku. However as this information is conveyed to the OFG-IFG process and there is still another 4-6 weeks of inventory in the pipeline from TRDI to CW, this is too late and the inventory bubble keeps building up. The execution of the decision to stop building CW inventory in other words lags by about 4-6 weeks.

The figure above conceptually describes what the pilot aspires to do to reduce this time lag in the prevention of unnecessary inventory buildup. The problem is attacked in two parts –

**PART 1: Moving the inventory bubble from CW to SFGI-**

The idea here is to move the bubble buildup to an earlier point where something can be done about it. At CW it is too late to change the inventory configuration since it has already been finished into a final configuration. If this not needed nothing can be done about it. But had the bubble been building up at SFGI it provides ample opportunity to configure or Finish only what is actually needed and store the remaining parts at SFGI in a space called ASRS space (automatic storage and retrieval system).

How can the bubble buildup at CW be moved to SFGI. To understand this we need to understand the reason why it builds up at CW instead of SFGI. Two main reasons were identified for this.

i) The available ASRS space is limited and improperly allocated across product families(sku's)-

As a result once parts have been started in ATM, even if an excess inventory flag goes up at CW, once the limited ASRS space fills up for SFGI, it is finished and ends up increasing the CW inventory bubble. While the overall ASRS space in ATM's is very limited and increase in this space is a key recommendation in chapter 6, due to capital considerations, we assume this is a longer term solution to be worked on. So given the ASRS space constraints, the next best thing that can be done is to ensure that ASRS space allocation between product families is properly done. A new capacity allocation algorithm helped reduce the allocation mismatches of this key ASRS resource and is described later on .

ii) The business process and decision making determining the SFGI-CW split is not optimum-

Even if there is enough ASRS space allocated to a product family and it can be kept at SFGI when unneeded at CW, if the SFGI-CW split for inventory positioning is wrongly set, the CW bubble will grow e.g. 0 at SFGI and 7 DOI (days of inventory) at CW, the SDD process will end up finishing and putting 7 DOI at CW adding to the inventory bubble at CW. To address this the pilot proposes a way to properly allocate the SFGI-CW Split

iii) Dynamically changing ASRS allocation to allow large bubbles to build-

As part of the effort to allocate ASRS space efficiently to allow large bubbles to build up at the SFGI side instead of having to a flush of parts due to a lack of space, we recommend dynamically changing ASRS allocations on daily or weekly basis. This wasn't tested in the pilot as it will require further development work. This is a worthwhile pending further IT development..

iv) Using SFGI Bubble Feedback to change demand plan in OFG-IFG-

Currently the excess inventory flag once raised is used to alter demand decisions at OFG-IFG. However buildup trends in SFGI are not looked at since the space is so small that most of the bubbles buildup in CW. However once the ASRS allocation issue is fixed, SDD needs to send feedback to OFG-IFG about SFGI bubbles to affect demand swaps

**PART 2: Pro-actively leveling the inventory bubble at SFGI-**

Once SDD capacity is allocated correctly based on discussions in part 1 above and flushes are reduced to a minimum, then much clearer and earlier inventory bubbles should start primarily being detected at SFGI. If no pro-active demand swaps are done, or they are done only on receiving upside change order requests( as happens now), then, the Pull information in SFGI inventory bubbles, which is an early warning system to change OFG-IFG builds, is being ignored for another cycle or two. Hence a Pro-active demand swap policy is needed to prevent unneeded inventory by leveling the production plan. There are two ways suggested for this in the pilot –

i) Pro-actively swapping demand from one sku to another within same fungible family-

A large portion of demand variations happen from one sku to another for the same fungible item. A fungible item or L3 level is defined as the semi-finished goods form of a sku sitting at SFGI, where it can be converted into several different sku's (called L1 level items) by SDD. In this case when inventory bubble starts building up on one sku in SFGI, demand trends and other information on other sku's within the same fungible family are used to make a L1 demand swap decision from one sku to another for the same L3 item. While in some cases there will actually be more demand on another sku than planned for requiring this swap, often there will be adequate production planned for other sku's, requiring a risk decision for pro-active demand swapping based on trends in demand rather than actual demand.

ii) Using Pro-active L1 Swap Feedback to OFG-IFG to swap demand-

Along with using the SFGI bubbles to change OFG-IFG demand plan, trends in pro-active L1 swaps must be used to as an additional data point to modify demand decisions at OFG-IFG.

### 4.2.2 Description of Key Algorithms

#### 1) ASRS Allocation Algorithm-

Currently ASRS Allocation is based roughly on Projected Product Volume. ASRS Allocation. The new proposed algorithm is based on Relative Ranking using Price, Volume, Variance & Fungibility. Items that have higher value are valued more by this algorithm as evidenced by a higher score since:

i) The higher the price the higher the value of a item.

ii) The greater the volume sales of a sku the greater the profit contribution of a SKU.

iii) Highly fungible skus are more valuable as they can utilize SDD to a greater degree because they can be fused to form many different SKU's.

iv) Items with higher variance have higher value as finishing them peremptorily increases the unsold inventory risk.

The higher the score the higher the SFGI allocation percentage or ASRS space. The more this space, the more of these products are finished only when needed.

1) Product Volume- The Pilot used average weekly volume over last 2 weeks of Billings with a proposal to eventually migrate to last 2 weeks of Customer Commits when this can be setup.

2) Product Price- Since Price is significantly higher than cost and cost data is difficult to get at the sku level, the algorithm uses Price as a proxy for Profit Margin. This is currently set quarterly and is based primarily on listed (CAP) Price.

3) Demand Variance- The pilot currently looks at demand variance over last 4 weeks of customer Billings or sales. We then rank top, middle and bottom third of product family variances giving variance score of 1,2 or 3.

4) Product Fungibility- Use (L3) Higher Level Fungible Item to (L1) Sku level Fungibility to get mean Product Fungibility. We rank top, middle and bottom third, giving a score of 1,2 or 3

Final Allocation Score for each sku = (Volume x Price x Variance Score x Fungibility Score). The Relative Allocation Rank = (Volume x Price x Variance Score x Fungibility Score)/(∑ Allocation scores for all products). Relative Ranks are calculated for all products per site. If ASRS capacity per ATM site is C above, Relative Allocation per Product in a site = C x Relative Product rank. A theoretically developed allocation model is described in the appendix at the end. The key difference in this model is that the ASRS allocation in the theoretical model is inversely proportional to the variance. We recommend eventually moving to this formula. During the project due to the large numbers of SKU's with a rapidly changing mix, several practical issues related to IT and data integrity in getting reliable variance numbers, the heuristic allocation approach described in the previous section was used. Simplifications used to generate the variance rankings is described next along with the Fungibility rankings. But eventually once the variance calculations can be quickly and reliably calculated, implementation and validation on the suggested theoretical formula might be carried out.

While the volume and Prices are straightforward, the Variance score and the Fungibility score are calculated as shown below –

Variance Score- A typical Product Family contains many sku's (L1 level ).The variance score for the family is calculated as described above. A typical score calculation is shown below. For all the sku's in a family, demand variance for the last 4 weeks is calculated. For simplification it also assumes zero correlation among sku's in product family so that Variance of ∑ sku's in product = ∑ Variance of sku's. Once the scores for each family are there we just rank them in three equal size groups for low, medium and high variance with scores of 1,2 or 3 respectively

**Figure 4-4. Variance Score Table**

| Skus | Week 1 | Week 2 | Week 3 | Week 4 | Variance |
|---|---|---|---|---|---|
| 1 | 6.576 | 7.748 | 7.817 | 8.384 | 0.57610625 |
| 2 | 11.657 | 12.576 | 15.047 | 11.45 | 2.724303 |
| 3 | 0.776 | 0.878 | 0.818 | 0.479 | 0.03150825 |
| 4 | 12.474 | 20.805 | 18.612 | 15.528 | 13.21454625 |
| 5 | 11.314 | 11.842 | 11.319 | 13.726 | 1.30943225 |
| 6 | 0.911 | 0.135 | 0.376 | -0.012 | 0.164213667 |
| 7 | 15.736 | 21.084 | 35.259 | 19.821 | 72.274458 |
| 8 | 11.241 | 20.671 | 16.173 | 12.559 | 17.84034933 |
| 9 | 1.304 | 0.85 | 0.585 | 0.828 | 0.089950917 |
| 10 | 20.844 | 35.597 | 25.312 | 24.404 | 40.18138092 |
| 11 | 0.292 | 0.475 | 0.818 | 1.31 | 0.200285583 |
| 12 | 6.147 | 11.743 | 8.964 | 12.03 | 7.58883 |
| 13 | -0.009 | 0.342 | 0.009 | 0.164 | 0.026671 |
| 14 | 0.065 | 0.11 | 0.318 | 0.147 | 0.012219333 |
| 15 | 0.326 | 1.726 | 0.507 | 1.874 | 0.647134917 |
| 16 | 1.689 | 2.8 | 2.696 | 1.968 | 0.296602917 |
| 17 | 13.243 | 12.143 | 10.007 | 5.897 | 10.509377 |
| 18 | 0.175 | 0.365 | 0.05 | 0.217 | 0.016875583 |
| Assuming skus are uncorrelated variance of sum (Product) = sum of variances of skus | | | | Product Variance | 167.7042452 |

Fungibility Score- The table below shows a calculation of a Fungibility rank for a Product family. A product family can have several L3 level Fungible items. This means each fungible item can be finished into many different sku's by SDD. These are listed vertically below from 1 to 40. Each L3 level can be converted into several sku's or L1 level items . These L1 levels are shown horizontally below from 1 to 19. A "1" in the matrix means that for that particular horizontal L3 level item it can be converted to that vertical L1 level sku. For example in the first row that L3 Level item at SFGI can be converted to 11 different sku's in CW. Since a product family as shown below has many different L3 items (40 in this case), each of which has a different amount of Fungibility or flexibility to be converted into different sku's (e.g. the first L3 item here can be converted to 11skus but the next one can only be converted into 10 sku's) , we need to find a product family level Fungibility score. We do this summing up the total Fungibility(167 in this case)and calculating an average L3 level Fungibility to normalize for different number ofL3 items in different product families(167/40

=4.175). Once we get this score for all product families we just rank them in three equal size groups for low, medium and high Fungibility with scores of 1,2 or 3 respectively.

**Figure 4-5. Fungibility Score Table**

| MM-(L1)-> L3 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | L3-->L1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 11 |
| 2 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 10 |
| 3 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 9 |
| 4 | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | | | | | | | | | 6 |
| 5 | | | | | | 1 | 1 | | | | 1 | | | | | | | | | 3 |
| 6 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| 7 | | | | | | | 1 | | | | | 1 | 1 | | | | | | | 3 |
| 8 | | | | | | | 1 | | | | | 1 | 1 | | | | | | | 3 |
| 9 | | | | | | | 1 | | | | | 1 | 1 | | | | | | | 3 |
| 10 | | | | | | | 1 | | | | | 1 | 1 | | | | | | | 3 |
| 11 | | | | | | | 1 | | | | | 1 | | | | | | | | 2 |
| 12 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| 13 | | | | | | | | | | | | | | 1 | 1 | | | | | 2 |
| 14 | | | | | | | | | | | | | | | | 1 | 1 | | | 2 |
| 15 | | | | | | | | | | | | | | | | | 1 | | | 1 |
| 16 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 9 |
| 17 | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | | | | | | | | | 6 |
| 18 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| 19 | | | | | | | | | | | | | | | | | | 1 | | 1 |
| 20 | | | | | | | | | | | | | | | | | | 1 | | 1 |
| 21 | | | | | | | | | | | | | | | | | | 1 | | 1 |
| 22 | | | | | | | | | | | | | | | | | | 1 | | 1 |
| 23 | | | | | | | | | | | | | | | | | 1 | | 1 | 2 |
| 24 | | | | | | | | | | | | | | | | 1 | 1 | | | 2 |
| 25 | | | | | | | | | | | | | | | | | 1 | | | 1 |
| 26 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 9 |
| 27 | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | | | | | | | | | 6 |
| 28 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| 29 | | | | | | | | | | | | | | 1 | 1 | | | | | 2 |
| 30 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 1 | | | 13 |
| 31 | | | | | | | | | | | | | | | | 1 | 1 | | | 2 |
| 32 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 9 |
| 33 | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | | | | | | | | | 6 |
| 34 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| 35 | | | | | | | | | | | | | | | | | 1 | | 1 | 2 |
| 36 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 1 | | | 13 |
| 37 | | | | | | | | | | | | | | | | 1 | 1 | | | 2 |
| 38 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | 9 |
| 39 | | | | | 1 | 1 | 1 | 1 | | 1 | 1 | | | | | | | | | 6 |
| 40 | | | | | | | 1 | | | | | | | | | | | | | 1 |
| L1<--L3 | 4 | 3 | 9 | 9 | 14 | 15 | 26 | 14 | 9 | 14 | 15 | 5 | 4 | 2 | 2 | 6 | 10 | 4 | 2 | 167 |

## 2) CW-SFGI Split

Using an LP formulation, a solver was used to recommend an optimum SFGI-CW Split to align with ASRS Allocation. Currently SFGI-CW Split settings is set based on either a 3 or 7 DOI (current business process preferences for the algorithm) at SFGI. The Proposed SFGI-CW determination is based on optimizing the total available SFGI space using the same relative allocation rank calculated for ASRS allocation. The staging decision is based on philosophy that if you can relatively value/rank inventory, you can stage it appropriately. The more valuable a product, then in the absence of concrete demand, the more it should be conserved i.e. less risk should be taken with it in terms of finishing it and moving it from SFGI to CW. Valuable in our context means it can derive more value from SDD i.e. more profitable, higher demand volume, more fungible and more variable.

Solver constraints can be setup as follows :

1) ASRS relative allocation can be used as a relative rank value metric such that it adds up to 100

2)The value of 1 DOI for each Product is also used to convert DOI results from Solver to a unit value with total units constrained by ASRS Capacity

3) The number of DOI (in SFGI) is constrained to be less than 7 for now

4) Product Family Inventory values are obtained by multiplying relative value metric & solver DOI optimizing to get maximum total inventory value

LP Formulation-

Objective Function: Maximize total inventory value = sum of product family inventory values

Decision Variables: DOI

Constraints-

1) 0 < DOI <7

2) DOI = integer variable

3) sum of units in SFGI ≤ Maximum SFGI Capacity (fixed per ATM).[units=DOI x value(1 DOI)].

Solver results for an example optimization solution for a given ATM site is shown below:

SFGI-CW SPLIT SOLVER

| ATM Site 1 | ASRS | Value | DOI | 1 DOI | units | Inv. Value =DOI x Val. |
|---|---|---|---|---|---|---|
| Product Family 1 | 24% | 24 | 7 | 10,000 | 70000 | 168 |
| Product Family 2 | 2% | 2 | 1 | 15000 | 15000 | 2 |
| Product Family 3 | 31% | 31 | 7 | 12000 | 84000 | 217 |
| Product Family 4 | 1% | 1 | 1 | 13000 | 13000 | 1 |
| Product Family 5 | 0% | 0 | 0 | 14000 | 0 | 0 |
| Product Family 6 | 10% | 10 | 7 | 15000 | 105000 | 70 |
| Product Family 7 | 16% | 16 | 7 | 18000 | 126000 | 112 |
| Product Family 8 | 2% | 2 | 0 | 19000 | 0 | 0 |
| Product Family 9 | 10% | 10 | 7 | 20000 | 140000 | 70 |
| Product Family 10 | 3% | 3 | 7 | 13000 | 91000 | 21 |
| TOTAL | 99% | 99 | | | 644000 | 661 |
| | | Max DOI= | 7 | Max Capacity= | 645000 | |

**Figure 4-6. SFGI-CW Inventory Split Solver**

### 3) Pro-active L1 swap algorithm

Current L1 Swaps are typically triggered only when customers request a change (COR) or a negative Left to Book or Available to Commit signal (LTB/ATC) necessitates it.
This typically does not take into account –
1) An Excess Inventory Report stating we should not be building certain L1's
2) A Qx Bubble with potential to hit ASRS Capacity Limit triggering a Flush in the near future

One could potentially look at ASRS Capacity Limit as well as SFGI DOI target and do –
i) Use Excess Inventory Report: Prevent L1 build on L1's listed in Excess Inventory report in addition to stopping L1 build in IFG while triggering same week swap request to another L1
ii) Potential L1 Flushes due to Qx (SFGI) bubble: SFGI Delayed Differentiation will hold inventory in SFGI instead of closing to CW, leading to larger deltas between what was planned and what was executed or finished to CW. These are treated as misses in the ATM build schedule. Delta (Miss) is rolled into the current weeks ATM Build (Qx) Schedule using the formula
New schedule = Old schedule + Delta from prior week. If SDD keeps avoiding the Finish of a product because there is no demand on it and if the OFG has a build target for it every week , then unneeded inventory piles up in SFGI leading to a bubble of inventory. This can be addressed with a weekly remix by triggering automatic L1 Swap Request, signaling the need to change IFG Request
In both cases we can use either relative ranking within L3 family for decision or use recent SDD builds as indicator. The advantages of this approach would be –

1) Immediate and automatic swap triggering to reduce excess inventory instead of waiting for next IFG cycle or CORS request
2) Reduces and smooth's out Qx bubbles based on relative ranking i.e. optimizes risk decision to take risk on least "valuable product" instead of randomly taking a risk.
3) Dynamically optimizes SDD settings to reduce flushing so on average more products get SDD

**Proposed Algorithm for Demand Swap**

**Deciding the Swap Status for a sku**

Swap Status = $_1\sum^4$ (BL – Qx schedule – Next week Pull-in) where, Pull-in = Positive if Next wk builds were pulled-in to current week,

= Negative if lagging behind last week plan.

BL = Customer Order Backlog.

Qx Schedule = ATM Build Schedule as planned by OFG-IFG.

If,

Swap Status = Positive, then Build as per Schedule. Additionally Swaps to this L1 might be needed,

= Negative, then Swap from this L1.

**Methodology for Ordering Sku's**

Rank Swap status for all L1's in each ATM site.

Swaps should be made from Negative L1's (& Excess WOI L1's).

Swaps should be made to Positive L1's.

Keeping L1 mappings to multiple L3's in mind, highlight potential swaps.

A possible LP formulation for this problem is given below. The variables used are as follows –

1) The L3 level semi-finished inventory item at SFGI at time step t is $I_{jt}$ and after optimization at time step (t+1) is $I_{j(t+1)}$

2) The L1 level finished inventory item at CW at time step t is $I_{it}$ and after optimization at time step (t+1) is $I_{i(t+1)}$

3) Demand in units from CW at time of optimization is $D_{it}$ and the fractional portion of demand actually met, again in units, is $E_{it}$

4) For a given L3 level item j there are many different L1 level items I that can be made out of it. The actual quantity of a particular L1 level item I that we have decided to make from a particular L3 level item j is denoted by the quantity $I_{jit}$

5) For a particular L1 level item or sku I, $P_i$ denotes the price or profit on it

6) Ci and Cj are weighting factors or "costs" assigned to the finished L1 level item I in CW or semi-finished item j in SFGI respectively. These can be adjusted based on a combined rank factoring variance and Fungibility such that given a choice of many L3's from which an L1 can be made, high variance, high Fungibility, items sit in SFGI and low Fungibility, low variance items get finished and put in CW.

7) $X_i$ is a penalty factor that we can design to weigh certain sku shortages more heavier than other shortages. So in a situation where not all the L1 level demand can be met from the L3 's sitting in SFGI, finishing of certain sku's will be preferred over others. This is to account for the fact that during times of product transition, i.e. NPI strategic build-ahead's and EOL or other price transitions, we may want to avoid shortage of certain sku's.

Objective Function Maximize: $\Sigma P_i D_{it} - {}_i\Sigma C_i I_{i(t+1)} - {}_j\Sigma C_j I_{j(t+1)} - {}_i\Sigma X_i (E_{it} - D_{it})$

Decision Variables: $D_{it}$, $I_{jit}$

Constraints:

i) $I_{i(t+1)} = I_{it} + {}_j\Sigma I_{jit} - D_{it}$

ii) $I_{j(t+1)} = I_{jt} - {}_i\Sigma I_{jit}$

iii) $D_{it} \leq E_{it}$ ( for all i)

iv) ${}_i\Sigma I_{jit} \leq I_{jt}$

**IFG Tie-In**

If an L3 has all Positive L1's then request increased allocation for next IFG cycle

If an L3 has all Negative L1's then request reduced allocation for next IFG cycle

### 4.3 Pilot Results

#### 4.3.1 Pilot ATM Results on ASRS allocation

#### i) ATM Re-Allocation Results-

The tables below show the results of using the allocation algorithm and differences between the current percentage allocation and the new allocation of scarce ASRS capacity across product families. The first column gives a percentage allocation solely based on revenue (price x volume). The next 2 columns give the calculated variance and Fungibility ranking. Multiplying these 3 columns gives a overall score next. Normalizing this gives the Allocation ranking in the next column. However this is unconstrained by Finish tool capacity constraints for individual product families. A 1 or 0 in the next column indicates which products have or don't have Finish constraints. The next column then gives a normalized but Final Finish constrained relative allocation for products. The next column gives the current allocation before pilot for comparison. The next two columns give the volumes these allocations translate into. The various site results and the conclusions are given for each site below

### a) ATM Site 1:

### Figure 4-7. Results for ATM Site 1

| ATM Site 1 | ASRS Relative Revenue Percentages | Variance Rank | Fungibility Rank | Overall Score | Unconstrained Allocation Percentages | Constraint | Constrained Allocation Percentages | Current Allocation Percentage | New Allocated Constrained Volumes | Currently Allocated Volumes |
|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 24% | 2 | 4.68 | 1.44 | 21.05% | 1 | 22.89% | 0% | 93863 | 0 |
| Product Family 2 | 2% | 2 | 0.1 | 0.04 | 0.58% | 0 | 0.00% | NEW | 0 | |
| Product Family 3 | 31% | 3 | 4.66 | 2.79 | 40.79% | 1 | 44.36% | 13% | 181860 | 53300 |
| Product Family 4 | 1% | 1 | 2.63 | 0.02 | 0.29% | 1 | 0.32% | 3% | 1304 | 12300 |
| Product Family 5 | 0% | 1 | 0.1 | 0 | 0.00% | 1 | 0.00% | 0% | 0 | 0 |
| Product Family 6 | 10% | 3 | 3.83 | 0.6 | 8.77% | 1 | 9.54% | 60% | 39110 | 246000 |
| Product Family 7 | 16% | 3 | 6.2 | 1.44 | 21.05% | 1 | 22.89% | 23% | 93863 | 94300 |
| Product Family 8 | 2% | 2 | 3 | 0.08 | 1.17% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family 9 | 10% | 2 | 2.2 | 0.4 | 5.85% | 0 | 0.00% | NEW | 0 | |
| Product Family 10 | 3% | 1 | 0.1 | 0.03 | 0.44% | 0 | 0.00% | NEW | 0 | |
| TOTAL | 97% | | | 6.84 | 100.00% | | 100.00% | 99% | 410000 | 405900 |

<u>Note:</u>

- Algorithm suggests increased allocation of 93k for Product Family 1.
    - 110k new capacity has been added to Family 1 reducing gap to suggested allocation.
- Algorithm suggests significant reduction in Product Family 6 allocation of 246K to 39k.
    - Significant excess capacity and under-utilization confirmed and being reduced currently.
- Algorithm suggests higher allocation to Product Family 7 instead of Product family 6.
    - This rebalance is currently being looked at.
- Algorithm suggests 17% of Total affected by Finish Constraints. Family 9 should be looked at.
    - This is currently is being looked at.

## b) ATM Site 2

| ATM Site 2 | ASRS Relative Revenue Percentages | Variance Rank | Fungibility Rank | Overall Score | Unconstrained Allocation Percentages | Constraint | Constrained Allocation Percentages | Current Allocation Percentage | New Allocated Constrained Volumes | Currently Allocated Volumes |
|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00% | 2% | 0 | 10200 |
| Product Family 2 | 18% | 3 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family 3 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family 4 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family 5 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family 6 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family 7 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family 8 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family9 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family10 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family11 | 0% | 1 | 0 | 0 | 0.00% | 0 | 0.00%NEW | | 0 | |
| Product Family12 | 4% | 1 | 2 | 0 | 0.00% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family13 | 6% | 1 | 2 | 0 | 0% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family14 | 33% | 2 | 3 | 1.98 | 56% | 1 | 55.93% | 64% | 285254 | 326400 |
| Product Family15 | 0% | 1 | 0 | 0 | 0% | 0 | 0.00% | 0% | 0 | 0 |
| Product Family16 | 0% | 1 | 0 | 0 | 0% | 0 | 0.00%NEW | | 0 | |
| Product Family17 | 39% | 2 | 2 | 1.56 | 44% | 1 | 44.07% | 33% | 224746 | 168300 |
| TOTAL | 100% | | | 3.54 | 100% | | 100% | 99% | 510000 | 504900 |

**Figure 4-8. Results for ATM Site 2.**

<u>Note:</u>

- Algorithm suggests Site 2 Finish constrained allocation is approximately in line with optimum.
    - Corrections from 64 to 56% for Family 14 and 33 to 44% for Family 17 being studied.
- Algorithm points out issues with Family 2 architecture and suggests Design for SDD approach.
- Algorithm suggests 10% of allocation affected by Finish constraints for Families 12 and 13.

## c) ATM SITE 3

| ATM Site 3 | ASRS Relative Revenue Percentages | Variance Rank | Fungibility Rank | Overall Score | Unconstrained Allocation Percentages | Constraint | Constrained Allocation Percentages | Current Allocation Percentage | New Allocated Constrained Volumes | Currently Allocated Volumes |
|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 56% | 3 | 1 | 1.68 | 47% | 1 | 47.19%NEW | | 345084 | |
| Product Family 2 | 19% | 1 | 1 | 0.19 | 5% | 1 | 5.34%NEW | | 39027 | |
| Product Family 3 | 17% | 3 | 3 | 1.53 | 43% | 1 | 42.98%NEW | | 314273 | |
| Product Family 4 | 8% | 2 | 1 | 0.16 | 4% | 1 | 4.49%NEW | | 32865 | |
| | 100% | | | 3.56 | 100% | | 100% | 100% | 731250 | 731250 |

**Figure 4-9. Results for ATM site 3**

Note: Algorithm suggests Optimum allocation for newer implementation in Site 3.

### d) ATM Site 4

| ATM Site 4 | ASRS Relative Revenue Percentages | Variance Rank | Fungibility Rank | Overall Score | Unconstrained Allocation Percentages | Constraint | Constrained Allocation Percentages | Current Allocation Percentage | New Allocated Constrained Volumes | Currently Allocated Volumes |
|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 39% | 2 | 2 | 1.56 | 34% | 1 | 33.55% | 0% | 226452 | 0 |
| Product Family 2 | 41% | 3 | 2 | 2.46 | 53% | 1 | 52.90% | 0% | 357097 | 0 |
| Product Family 3 | 21% | 1 | 3 | 0.63 | 14% | 1 | 13.55% | 100% | 91452 | 675000 |
| TOTAL | 101% | | | 4.65 | 100% | | 100% | 100% | 675000 | 675000 |

**Figure 4-10. Results for ATM site 4.**

Note:

- Algorithm suggests financial rationale for increased allocation to Product Families 1 and 2.
  - Need for this increased allocation was agreed on; awaiting resolution of ATM logistics.

### ii) Using the Algorithm for Installing New ASRS capacity-

The algorithm was used to guide the decision on where to install newly available ASRS capacity. To do that a normalized revenue affected by SDD for each additional 32k of capacity (the lowest granularity for installing capacity) was calculated and is shown below. What the analysis suggested was that per unit capacity the best site to capture additional revenue for postponement was Site 2 or site 1. This confirmed existing intuition about which site would have been most beneficial.

**Figure 4-11. New ASRS Capacity Payoffs by Site**

| ATM Sites | Additional Revenue/32k of additional ASRS capacity |
|---|---|
| Site 1 | 8.2M |
| Site 2 | 10.6M |
| Site 3 | 6.6M |
| Site 4 | 5.0M |

## iii) Using the Algorithm to Make Different Tradeoffs while Stocking:

The two tables below show a disguised hypothetical example of one amongst many possible tradeoffs that were looked at with the algorithm. In the tradeoff shown below, once both the New DOI policy and the allocation percentage are decided, we get-

i) A 2.2x increase in the Relative allocation Rank Value of goods affected by the postponement strategy which factors in Price, Volume, Fungibility and Variance.

ii) Since the above increase has too many factors, if we just look at the increase in product revenue affected by postponement strategy it comes to about 1.3x.

iii) Similarly if the aim is to have the most fungible mix in the SFGI the allocation can be checked either as change in fungible volume at SFGI (1.28x here) or in fungible revenue (0.98x here).

iv) Similar to the metrics for Fungibility the change can be evaluated on the basis of variance weighted volume(1.47x increase) or variance weighted revenue at SFGI (1.39x here).

The above type of analysis can then be used to evaluate and tweak the decisions being made.

### Figure 4-12. SDD Algorithm ROI

ASRS ROI

| ATM Site 1 | Ranking | Old DOI | New DOI | 1 DOI | Old Units | New units | Old Inv. Value | New Inv. Value | Price ($) | Old Rev($) | New Rev($) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 24 | 7 | 3 | 10000 | 30000 | 70000 | 72 | 168 | 642 | 19260000 | 44940000 |
| Product Family 2 | 2 | 1 | 3 | 15000 | 45000 | 15000 | 6 | 2 | 1936 | 87120000 | 29040000 |
| Product Family 3 | 31 | 7 | 3 | 12000 | 36000 | 84000 | 93 | 217 | 118 | 4248000 | 9912000 |
| Product Family 4 | 1 | 1 | 3 | 13000 | 39000 | 13000 | 3 | 1 | 527 | 20553000 | 6851000 |
| Product Family 5 | 0 | 0 | 3 | 14000 | 42000 | 0 | 0 | 0 | 159 | 6678000 | 0 |
| Product Family 6 | 10 | 7 | 3 | 15000 | 45000 | 105000 | 30 | 70 | 45 | 2025000 | 4725000 |
| Product Family 7 | 16 | 7 | 3 | 18000 | 54000 | 126000 | 48 | 112 | 114 | 6156000 | 14364000 |
| Product Family 8 | 2 | 0 | 3 | 19000 | 57000 | 0 | 6 | 0 | 409 | 23313000 | 0 |
| Product Family 9 | 10 | 7 | 3 | 20000 | 60000 | 140000 | 30 | 70 | 1419 | 85140000 | 198660000 |
| Product Family 10 | 3 | 7 | 3 | 13000 | 39000 | 91000 | 9 | 21 | 415 | 16185000 | 37765000 |
| | | | | | | | | | | | |
| TOTAL | 100 | | | | 644000 | 644000 | 297 | 661 | | 270678000 | 346257000 |
| | | | | | | | | 2.2x | | | 1.3x |

ASRS ROI

| ATM Site 1 | Old Units | New units | Price ($) | Fungibility score | Old Fungible volume | New Fungible volume | Old Fungible revenue | New Fungible Revenue | Variance score | Old Variance weighted Vol | New variance weighted Vol | Old Variance weighted Rev | New variance weighted Rev |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Product Family 1 | 30000 | 70000 | 642 | 1 | 30000 | 70000 | 19260000 | 44940000 | 2 | 60000 | 140000 | 38520000 | 89880000 |
| Product Family 2 | 45000 | 15000 | 1936 | 2 | 90000 | 30000 | 174240000 | 58080000 | 1 | 45000 | 15000 | 87120000 | 29040000 |
| Product Family 3 | 36000 | 84000 | 118 | 3 | 108000 | 252000 | 12744000 | 29736000 | 2 | 72000 | 168000 | 8496000 | 19824000 |
| Product Family 4 | 39000 | 13000 | 527 | 3 | 117000 | 39000 | 61659000 | 20553000 | 3 | 117000 | 39000 | 61659000 | 20553000 |
| Product Family 5 | 42000 | 0 | 159 | 1 | 42000 | 0 | 6678000 | 0 | 1 | 42000 | 0 | 6678000 | 0 |
| Product Family 6 | 45000 | 105000 | 45 | 2 | 90000 | 210000 | 4050000 | 9450000 | 3 | 135000 | 315000 | 6075000 | 14175000 |
| Product Family 7 | 54000 | 126000 | 114 | 1 | 54000 | 126000 | 6156000 | 14364000 | 1 | 54000 | 126000 | 6156000 | 14364000 |
| Product Family 8 | 57000 | 0 | 409 | 2 | 114000 | 0 | 46626000 | 0 | 2 | 114000 | 0 | 46626000 | 0 |
| Product Family 9 | 60000 | 140000 | 1419 | 1 | 60000 | 140000 | 85140000 | 198660000 | 2 | 120000 | 280000 | 170280000 | 397320000 |
| Product Family 10 | 39000 | 91000 | 415 | 1 | 39000 | 91000 | 16185000 | 37765000 | 1 | 39000 | 91000 | 16185000 | 37765000 |
| TOTAL | 644000 | 644000 | | | 744000 | 958000 | 432738000 | 413548000 | | 798000 | 1174000 | 447795000 | 622921000 |
| | | | | | | 1.28x | | 0.98x | | | 1.47x | | 1.39x |

In the actual pilot efforts were made to increase the relative allocation rank within the system constraints. This resulted in a forecast for increase in product revenue affected by SDD for 2 of the sites and a decrease for one of the sites while a fourth site had a new allocation. Overall the revenue affected by postponement strategy went up by 10% without any capital outlay for increase in capacity. If only the two sites that had an increase were implemented and the third site had been kept as is, the increase would have been much larger at 32%. In fact the largest increase was about 68%.

### 4.3.2 ROI calculation for L1 swap

The L1 swap idea was discussed and will be considered for a future project once higher priority ongoing projects are implemented by the IT team. As such, the rough calculations below are based on certain conservative assumptions to get a sense for the yearly returns on implementing a L1 swap scheme. Note that the actual implementation is a simple IT project that would not cost much. The bigger issue is getting the commitment from hundreds of planners to spend a few minutes daily looking at L1 swap suggestions and doing regular house-keeping as regards inventory.

**<u>A) General Pre-Calculation:</u>**

% of ATM revenue on SDD = 30%

Total CPU revenue through ATMs = $30 B

Total SDD revenue = 30B x 30% = $10B

% of Flushes due to space = 10% = $ 1B

% of Flushes that could be avoided with proper ASRS allocation and SFGI-CW split setting and which result in Qx bubbles = 5% = $50M

% of Qx Bubbles eligible for L1 reallocation given L1 Fungibility within L3 level = 5% = $ 2.5M

**B) ROI Components for L1 Swap:**

ROI = (1) Savings from Avoiding Inventory Reserve Charges + (2) Savings through Inventory Shelf Life Reduction + (3) Profit from meeting perishable unmet demand

ROI Component (1): Prevention of Avoidable Inventory Reserve Charges

% of reallocated L1's that would otherwise have required an inventory reserve charge = 10%

Inventory Reserve charge saved = 10% x 2.5M = 250k$

ROI Component (2): Savings through Inventory Shelf Life Reduction

Average shelf life for remaining 90% L1's had they not been swapped from unneeded L1's = 6 weeks (since SDD looks out min of 4 weeks)

Average shelf life of swapped L1's = 2 weeks (typical)

Average shelf life saved = 4 weeks on 90% of 2.5M= 4 weeks of cost of capital (~24%) of 2.25M ~ (1/12)x(24)~2% of 2250k ~45k$+5K holding cost= 50K$

ROI Component (3): Profit from meeting perishable unmet demand

Average time that unmet demand is met earlier due to L1 swap = 2 weeks (time gap between 2 OFG-IFG cycles assuming 2x/month)

Unmet demand met = same as Amount of reallocated L1 swaps = 2.5M $

% of unmet demand that is perishable in 2 weeks = 10% = $250K

**Total Yearly ROI for L1 swap** = 250K +50K +250K = $ 550K

# 5. Goals, Incentives and Organizational Behavior

## 5.1 Human Factors in Supply Chain Management:

The goal of this thesis was in improving Intel's supply chain responsiveness to changing customer needs especially in the area of fulfilling demand changes for products. Managing a global supply chain is as much if not more about managing human behavior as it is about managing processes and supply-demand of products. So managing the communication between individuals in involved planning organizations, identifying the processes for clearly dealing with areas of shared responsibility such as with inventory, and aligning incentives with desired behavior for coordination is a key function of managing the global supply chain.

The supply chain performs as its participants do. They in turn behave as they are rewarded or punished. What is rewarded, which mistakes receive minor reprimands and which are severely punished affect people's behavior and so the behavior of the chain. For example doing hot demand swaps based on early trends may be right most of the time, but if there are no rewards for being right most of the time and punishment for being wrong a few times then people behave conservatively. The system of incentives needs to be looked at very carefully, since it is a statement of what the organization implicitly values. No amount of analysis will be useful if a perverse set of incentives exist.

The business dynamics facing external participants such as consumers cause similar behavioral change. While customers are currently asked to estimate their demand a quarter ahead the reality is that they don't know any more than Intel what the end market will do. Given the zero penalty cancellation policy they make projections for supply assurance with frequent updates almost till the last minute as demand firms up. However given the long lead time and the rigidity of the demand projection and production planning process responding to this is difficult. While the preceding chapters have explored technical solutions to these issues, the goal of this chapter is to explore strategic, political and cultural issues related to factoring incentives, organizational design and human behavior in crafting the right solution.

## 5.2 STRUCTURAL ISSUES:

### Organizational Structure and Strategic Challenges

The strategy of the Microprocessor Marketing and Business Planning- Supply Demand organization is to coordinate the operations to meet current year business goals e.g. market share objectives, profit and revenue targets etc. Though the MMBP organization has responsibility for meeting quarterly financial commitments to Wall Street through operational levers, it must rely on other organizations for doing this which creates difficulties. Geographic Sales Organizations own meeting sales targets. To do so factories have been asked to respond to requests from various sales organizations. This means that Geographic Sales Offices have control over inventory produced and the mix of products affected. This despite the fact that MMBP is held responsible for excess inventory hits to the bottom line. Similarly MMBP spends most of its time with customers and Geographies to create a composite view of demand. However it must rely on the Divisional Planning organization to implement its policies. The planning organization however uses different operational definitions for purposes of communicating marketing demands specified at marketing product family levels to factory level languages based on processing nuances. This creates significant time lag in translating any plan change from demand language to supply language quickly.

### Role related Challenges:

The various roles involved in implementing operational directives influence the current project. Product Managers set objectives at marketing levels and try to balance the risk of over producing with the risk of being unable to meet customer demand at an Intel world-wide level. They may not realize, however, what is misinterpreted along the supply chain as it tries to respond to the targets they set. The Divisional Planners who try to translate from marketing level demand language to factory level supply language sometimes may not have access to profit margins per product. Lastly Geographic Sales is heavily focused on wringing out every last dollar of sales in its geography without having visibility into how their actions to secure supply for potential sales affect other geographies worldwide or inventory risks at the Intel level. Work is needed to interface with and coordinate across all 3 roles. While some coordinating systems are in place through emails/weekly meetings, etc. making this more agile is what the current project is all about.

**Suggested Structural Changes:**

**Structural and Co-ordination Related Changes**

1) Since so much of the decision making is disaggregated, a "wisdom of the crowds" approach to alignment is advocated. The focus throughout the project has been on utilizing collective judgment by making available intelligence in MMBP easily accessible and visible across organizations.
2) Standardizing measurements and interpretations will aid in speaking the same language. To this end, collaborating across functions such as with the Supply Transformation Team's inventory metric standardization effort will be important.
3) Educating individual stakeholders on how other organizations function and use their inputs will help in responding to tactical updates quickly

**Role Related Changes:**

1) MMBP Demand-Supply Analysts as also Product Managers are focused on getting the more strategic wafer starts number right at a marketing family level and also at a monthly and Quarterly granularity. There needs to be a new role to coordinate the tactical supply demand response across the supply chain. A new role is suggested for this.

## 5.3 CULTURAL ISSUES:

**Symbolism:**

Any communication coming from a corporate organization symbolizes control to various other organizations. Also as a result of this, the demand report published by MMBP called "Judged Demand" connotes command and control to some in the sense of judging customer inputs. There is a lot of aggregated world-wide intelligence in this. When different organizations use other demand signals instead of MMBP inputs, they lose some of this intelligence regarding things such as Price changes, Branding changes etc. This project, by purporting to affect change by making intelligence visible to the myriad decision makers, tries to use this distributed experience base. Care has to be taken to assure people that the goal is to aid them in their decision making instead of trying to make their decision for them with an algorithm. New initiatives have to be introduced to other organizations as such. In that sense just the fact that MMBP is reaching out to understand what these other organizations are doing will be greatly appreciated. Trying to make existing processes better rather than advocating a completely new process is always preferable. In this sense these other groups are supportive of like-minded MMBP efforts.

## Biases

Given the rich engineering history of Intel, there is a strong reliance on quantitative approaches. While this is normally good, it does manifest itself sometimes in deleterious ways, as with a lack of ease in using qualitative intelligence. Sometimes this leads to an over-emphasis on quantitative approaches at the expense of non-quantifiable softer issues. These can only be attacked by increased efforts to gather human intelligence. An example maybe an under-exploitation of frequent meetings at the planning and operational level with VMI Hub customers to understand customer ordering psychology and behavior and the use of such intelligence in planning strategies. There is also a danger of being underprepared for rarer events which maybe more common than expected due to increased uncertainty in times of structural changes in the industry. These can't be predicted with trend analysis but might be hinted at by gathered human intelligence. Supply chain responsiveness might be hampered by the preference for choosing the certainty of a fixed plan to the ambiguity of a flexible plan. From an organizational perspective there is also the danger of a bias towards searching for solutions in improving quantitative metrics and levers. An example of this might be a preference for changing pricing to stimulate demand while under-utilizing marketing.

There is something to learn from warnings given to numbers-driven professionals (Walsh 2008). Keynes once said it is better to be vaguely right instead of precisely wrong while warning economists against "specious precision" at the expense of "concealed factors" not capable of quantification. He specifically warned against dismissing qualitative elements as unworthy of analysis, while remarking that when statistics don't make sense he preferred sense to statistics. Hayek, another titan in the field of economics, warned classical economists in his Nobel Prize lecture, against laboring under a "scientistic" attitude, disregarding those factors that cannot be confirmed by quantitative evidence while happily proceeding on the fiction that the factors they measure are the only relevant ones. An attribution to Einstein provides the final thought on this topic who suggested that not everything that counts can be counted and not everything that can be counted counts.

**Cultural Levers:**

Reaching out to other organizations helps address some of the symbolism of corporate planning being in its own ivory tower pushing decisions down the chain. In addition, since the "judged" terminology is a sensitive term, any updates should be packaged as "available intelligence". To address the concerns over centralized control points, changes have to be made widely visible to various decision-making organizations in the form they currently use to ease adoption. Another cultural lever is the use of taskforces. Being part of a cross-functional taskforce and using it to achieve long term success on goals aligned with its charter can be a key tool in securing buy-in from various organizations provided a voice by such efforts.

## 5.4 POLITICAL ISSUES:

**Stakeholder Analysis:**

MMBP is highly supportive of product prioritization efforts but has a key concern on the extent to which price information is shared. This has been addressed for this project by sharing a ranking based on price rather than price itself. As also sharing the published price rather than price with all the discounts included is preferable for confidentiality reasons. Similarly Division Planning and SDD have been supportive of initiatives where they don't have any current ones themselves such as ASRS space allocation and are open to an algorithm based suggestion on other initiatives. Their chief concern is making it easy to use and maintain without sacrificing accuracy. To take this into account, the project uses a number of approaches such as simplified product family groupings within the algorithms.

**Alignment:**

The various groups in the Intel supply chain are compatible. However the focus of some groups which are more mathematically oriented is to develop complex highly accurate and optimized algorithms. The end users typically want it to be simple and just good enough to serve the purpose while requiring the least amount of work. The end users can simply bypass complex algorithms or worse yet work at odds with it. Hence this is the group that ends up influencing the implementation the most. All groups are aware of this and are amenable to this focus on simplicity.

In the case of having to share sensitive price data to base certain ranking decisions, limiting the groups having this access and making a relative ranking based on price more widely available are much more realistic and also more acceptable to MMBP.

While there is no conflict between groups, the bigger danger is organizational inertia and making sure somebody adopts and keeps responsiveness initiatives going after a project. To address this, tactical problems that serve the larger strategic goal are chosen in this project and addressed first to develop momentum. Also simplicity and ease of use is emphasized at every point. Finally work is directed based on continuous real-time, end-user feedback since these are the users that will need to continue this initiative after a particular project is completed. Most change agents are experienced enough to know that initiating end user acceptance is the biggest step. Once this is done, continuous improvement can always be carried out to improve algorithms. As a result the focus of this project has been on simplicity as a key imperative to achieve strategic goals.

**Incentives:**

A last key political issue is incentives. Geographies have a big incentive for every additional dollar of sales they can achieve in their geography, but they have less of the risk of excess inventory. They also are penalized more heavily for missing projected customer demand than for producing unnecessary inventory, while they attempt to adapt to changing customer demand. However there is recognition of these incongruence's in incentives and there is a key task force working on aligning these incentives. In the meantime, the focus of the internship was to make intelligence visible to make excess inventory visible for proper decision making.

# 6. Conclusions and Recommendations

In looking at the supply chain and looking at ways of improving the real –time responsiveness or ability to make precision adjustments to production, the thesis provides both general recommendations related to Intel and specific ones related to organizations such as MMBP and Division Planning. MMBP recommendations were primarily related to tracking processes in pricing and supply-demand and suggestions on the product supply manager role. In the case of Division Planning they were related to the interactions of SDD with VMI hubs as well as various production planning processes. Lastly there were suggestions for both specific tools and projects that would help in making the supply chain more agile.

### 6.1 General Recommendations

1. **Simplification and Housekeeping**

Over the years, in order to price-differentiate, there has been a proliferation of sku's. While this can help capture value from discerning customers or be price competitive, there is also a supply chain cost associated with supporting each new sku. Similarly there has been a substantial increase in the number of price and product transitions. Quantifying this cost and simplifying the sku offering is key. Tracking and reducing unneeded Price/Product Transitions which also add supply chain complexity goes hand in hand with this as well.

Killing off unneeded sku's and product-price transitions on a regular basis is a housekeeping task worth institutionalizing. Having a review board (set up by product groups or vertical business segments and facilitated by MMBP) rather than one-off taskforces is recommended to both approve and delete sku's.

Another area for simplification is the way upside requests are granted. Instead of a weekly meeting, having a up to date list of products available for upside request grants, will help with expediting these kinds of customer change request reviews.

## 2. Visibility

Price based guidance is the most important piece of information MMBP could share with the Supply Chain. If confidential information is a concern, then a relative ranking for sku's based on price or profit margin without the underlying price information would be equally useful if shared with Division Planning and the factory network. Large differences in demand signals such as JD, commits and backlog are another area for improved visibility. If MMBP can share the explanation for this with Division Planning and ATM's then it's easier for them to be consistent with MMBP strategy.

The area of Product and Price Transitions i.e. both on and off roadmap product EOL, NPI and price/branding moves is a key area for improvement. Currently price moves, inventory strategies and supply chain knowledge on planned or unplanned roadmap changes is very limited. This can result in examples for instance where a decrease in price to flush inventory is seen by planners downstream as increased demand who build more inventory based on the surge in demand. Visibility is the best antidote for this, not just to division planners but also between pricing and supply-demand teams

Providing the entire supply chain visibility on trends is also desirable. Similar to central banks communicating a forward looking bias when they make interest rate changes, MMBP and Division Planning should focus not just on communicating the current demand and inventory requirement, but also whether they expect significant changes(steep declines or increases) or not to ATM's and FSM's. The outlook bias can be at a product family level with tightening or expanding bias etc. An example might be in solvers where additional capacity is allocated towards target inventory on products without much differentiation. Or during the initial phases of an economic slowdown, when numbers do not reflect anticipated activity reduction, additional precaution might be advised thru tighter control with precise determination of inventory.

## 3. Organization

While MMBP currently has a Supply lead position, they are typically busy with the work of managing the monthly Reset process for Fab Wafer starts typically. Hence creating a new Position in charge of tracking and improving Tactical performance i.e. daily signals, interaction and handoffs with Geographies and Division Planning, would help immensely.

Additionally, clarifying the roles and responsibilities for decision making on demand and inventory positions across organizations would streamline decision making and eliminate redundancy. Decision making should be devolved to the level where the most intelligence and hands-on experience resides about the operation in question. In addition to JD for wafer starts, only the things that need coordination should be addressed at the MMBP level, and an effort should be made to minimize this, allowing MMBP to focus more on product-pricing strategy and process improvement. Currently the MMBP is deeply involved in tactical operations and has very little time to focus on the dynamics of coordination, strategy and process improvement through observation.

Strategic planning under the aegis of MMBP should be looking at the system dynamics issues such as with human behavior, organization and incentive design and feedback loops to improve tactical agility. Moving to a model where it gives more tasks to Geographies and Division Planning partners, while supervising them for consistency, process improvements and alignment with strategic imperatives will be more productive and less duplicative.

External intelligence gathering is also an area where the above change in organizational focus would help. Current intelligence gathering at Intel focuses much more on analytical trend analysis from historical data as compared to speaking with the ordering customer real-time. Having more frequent and institutionalized interaction with customers to understand ordering behavior is highly recommended, especially given VMI Hubs and the devolution of near term forecasts to Geo's. Understanding the dynamics of each customer in terms of corporate vs. consumer, variety of product lines, channel, region etc. and the implications for forecast accuracy is key.

### 4. Incentives, Measurements and Accountability

Currently due to the nature of how inventory positions eventually get set, Geographies and Division Planning contribute much more to inventory decisions than MMBP. Since they set the de facto tactical inventory policy, they should more clearly have metrics and measurements for their inventory responsibilities including in bonus targets. On the other hand there should also be a penalty for missing business by being over conservative, so that incentives are balanced on both sides. Similarly Over-booking and late cancellations, percent unallocated reservations etc. should factor in employee bonus goals for Geographies and Division Planning. To avoid the risk of driving them to under-commit and then ask for upsides bonuses for sales versus penalties for cancellations must be calibrated carefully. Current measurements tilt too heavily on the side of supply assurance. This should be balanced with measurements on inventory performance. Delineating accountability in areas of shared objectives such as inventory management is an important area where things can fall through because of unclear accountability, hand-offs and incentives.

Currently customers do not have an incentive to improve their forecast accuracy if excess inventory helps garner discounts. Some sort of rebate system to share the savings resulting from accurate customer forecasts that also factors in the difference in demand volatility between corporate and retail customer bases, different geographic footprints and so forth must be devised. While it might be true that customers do not know demand over time any better than Intel does, it is also true that there is not any incentive to spend more time and commit more resource to improving it since Intel guarantees them the inventory and service level asked for.VMI hubs were an attempt to get improved forecast accuracy from customers with Intel paying the bill for forward positioned inventory with the tacit agreement that customers would have to improve their forecast accuracy. But to date not much has arguably improved. One way to jump start this might be a set of metrics for both sides of the chain committed to by senior management on both sides, with weekly review meetings to gauge progress and understand large deviations from forecasts or service levels.

**5. <u>Clearly Defining Quality and Success Metrics for Planning in a stochastic environment-</u>**

The profit maximization goal owned by multiple planning organizations has to be disaggregated to set clear objectives for each part of the chain- MMBP, Division Planning etc. A fine balance has to be achieved between fostering risk taking and the cost of a stock out. Currently given the capital intensive nature of the chip business, a very high cost is put on a stock out and almost none on excess inventory. This is reflected in incentives and penalties. The danger is this focus keeps changing as the pendulum swings from concern over customer service during booms to those over excess inventory in slowdowns. Counter cyclical increases in marketing and drops in pricing have to be carried out to dampen these effects during times of significant flux. Similarly contract manufacturer agreements might be enabled e.g. with older chipsets to offload peak demand.

## 6. Knowledge Capture and Continuous Improvement

The average tenure of a planning analyst is very short and by the time they become experts at their job they tend to move. Knowledge capture in such cases, particularly the softer knowledge acquired by analysts in the course of doing their job becomes important. Benchmarking consultants might be a worthwhile exercise for this. Similarly the addition of a field in the planning/pricing databases asking analysts to provide reasons for over-ruling suggested decisions by optimizer models(e.g. in OFG-IFG why they chose a certain demand signal over another.) will help both the model learn unarticulated qualitative knowledge trends which can be captured via data-mining the beliefs of analysts and planners. Making unsaid belief systems evident in actions but not explicitly stated, searchable and explainable captures what the essence of an art is, increasingly converting it into a science while helping ramp new hires in areas with short tenures by spreading best practices. Working on improving the process for analysis continuously needs to be as high a priority as improving the analysis. Something akin to the Army After Action Review (AAR) method can be used for debriefing to capture learning after each planning cycle, each major transition, completion of product lifecycle, quarterly or yearly etc. (Garvin, March 08) This requires thinking about

- What did we set out to do and what actually happened?
- Why did it happen and what do we do next time?
- What activities do we sustain or keep and which do we discontinue and what do we improve?

## 7. Extracting the maximum value out of VMI

Another learning area is in the area of customer learning at VMI hubs. Certain authors have suggested the following factors as key to extracting the maximum value for providing higher service levels with VMI-

- Developing a clear dual supplier /partner relationship accountability scorecard to facilitate information sharing and improving forecasting accuracy clarifying two questions- who is capturing the appropriate information on a routine, timely basis, and who is ultimately responsible for improvement?
- A single point of responsibility on each side for VMI learning per customer.
- Forming a cross-functional team on both sides responsible for performance management.

- Performing consistent periodic reviews.

In fact a co-sponsored cross-company team or LFM internship for example between Intel and Dell looking at the leaning out of inventory while improving service level at VMI hubs might be helpful.

### 8. <u>Improving Forecasting</u>

Intel is continuously working on improving forecasting accuracy. This section suggests some ideas for this:

- To avoid the danger of group think in forecasting, Intel uses third party and internal reports, both top down(from macro environment to micro issues) and bottoms up(building up from sku by sku basis), forecasting for size and mix as well as reconciliation between geographies. Similarly measures should be taken for other forecasting biases to prevent peer pressure and anchoring. Additionally it might be worthwhile to ask Geo's to predict not just their market, but adjacent markets that they may have insight into, to cross-verify relative assumptions on market share gains, etc. However any information asymmetries across groups, as regards product-price transition strategies, need to be leveled.
- Adding continuously updated product lifecycle forecasts might help sanity check whether inception-to-date projections are consistent with the lifetime projected revenue of a product as well as historical lifetime revenue. The non-linear complexities associated with diffusion curves can be used to compare s-curves and inflection points to against part performance.
- A type of scenario analysis that could be added both for forecasting accuracy as well as for responsiveness is an business-cycle inflection-point metric, benchmarking past trends during periods of sharp changes in customer or market behavior e.g. 2001 tech correction, 2008 recession crash etc. Associated metrics could be used for risk management.
- Since forecasts tend to extrapolate the immediate present into the immediate future, a sensitivity analysis should be carried out for strategic assumptions such as how fast a product transition will happen, how a price cut will stimulate demand, demand distribution across price tiers, market trends such as from laptop to netbook or discrete to embedded processors, etc. Special attention should be paid to severe risks with small probabilities e.g. financial meltdown, earthquake in Taiwan, etc.

## 6.2 Specific Recommendations

### A. Recommendations on Tracking and Transitions

### 1. Pricing

MMBP Pricing Team should track regularly the number of price transitions (average per sku and total) a quarter, as well as metrics such as the mean time between transitions for an average sku. Trends in this can be correlated with supply chain costs. Each price move causes a bull-whip effect on the supply chain planning systems. If there is an upward trend in number of transitions per sku or total number of transitions per quarter adjusted for number of sku's, then this might signify increased market fragmentation or added competitive pressures. Either way this only increases the pressure for a more responsive supply chain as well as tighter integration between pricing and the entire supply chain planning team. Currently the pricing team coordinates with MMBP supply-demand teams, but, given the frequent tactical pressure increasing price moves put on planning, division planning, ATM's and SDD also need to be more tightly integrated with pricing teams. This could be a separate role in Pricing or MMBP Supply. Eventually a better IT system and more visibility across the chain on price changes and their anticipated demand effects on skus is needed, so that it can be tied tactically to processes like SDD, ATM pull-ins, ATM Range Planning etc.

### 2. Supply-Demand

The MMBP Supply Demand team should track Supply commitments like allocations in DSS (Decision Support Systems) as tactical demand signals. This is because that is how the rest of the supply chain views as their true demand signal for planning purposes. Product transitions such as number of instances where a product crosses PSLV type segment boundaries, the number of instances where a product drops to bottom rung of price ladder, and, especially, the number of instances of products that disappear from a roadmap per quarter should be tracked. I.T. capabilities to track end to end inventory position daily or demand and inventory build signals through all the production planning processes should be built up. This will help make OFG-IFG and post SDD results visible to product planners and managers.

### 3. Managing Transitions

Given the especially disruptive nature of transitions it is worth thinking what else can be done about them. One possibility is to apply the Lean principle of Heijunka – leveling fluctuations in demand by staging price and product transitions on a family basis using a strategy similar to Intel's famed tick-tock strategy.

Inventory targets need to follow the product life cycle going up for NPI and automatically ramping down as EOL approaches. Service levels for different products should be based not just on lifecycle – NPI/EOL but also metrics such as high volume, high profit contribution, high margin, highly variable, strategic build ahead, etc. Service levels at all inventory locations (CW, VMI etc.) should thus be time phased, financially differentiated and lifecycle dependent. Also certain business processes should be differently set based on these categorizations etc. shorter frozen horizon for EOL's etc.

## B. Recommendations for Product Supply Managers

### Inventory target Setting

Product Supply Managers should look to standardizing inventory target setting methodology across verticals. Simplicity should be given precedence over complex customization for each vertical. Focus should be on compliance across the supply chain over theoretical accuracy. To this end targets need to be set as simply as possible while anticipating the environment in which solvers try to implement their targets (using a rolling basis for target setting) especially in the areas of minimum and target inventory numbers. Using current BOH (Beginning On Hand), discontinuing breaking targets down into ADI and CW targets and focusing on Wafer starts targets i.e. only the ADI target bear special consideration. In addition, SDD should be factored in target setting since it plays an increasingly large role in decoupling the ATM production plan (OFG-IFG) from actual CW Finishes.

### 4. Miscellaneous

PM's should look to tracking metric such as percentage requests answered within a day using more frequent supply commits, approval meetings, and expedited list of sku's available, for quick upside request grants, etc. More proactive approaches such as work with Division Planning through SDD and ATM Range Planning to convey and make tactical demand updates should also be used more regularly.

### C. Recommendations on approaching SDD-VMI Interactions

Intel has more recently decided to forward stage inventory in VMI (Vendor Managed Inventory) hubs for its biggest OEM customers. The VMI strategy of forward staging of significant quantities of next month's projected demand has significant implications for the delayed customization processes such as SDD. The key impact is that SDD now has to react to monthly demand projections rather than true customer demand pulls from CW. Thus the strategy reduces the benefit of SDD which increases inventory target overruns. Ideally this should be implemented as a replenishment strategy with Intel replenishing the amount taken from VMI regularly with some intelligence around product transitions like EOL and NPI. However there are some roadblocks to doing this effectively while still guaranteeing a very high level of service level. The reasons include the weekly frequency of the SAP SNP solve for VMI replenishment, the transportation times from CW to VMI hubs, and the lack of tie-ins or micro-feedback loops to OFG-IFG, SDD and ATM Range Planning. This results in increased reliance on slow, fortnightly planning processes. This still could be carried out partially for sku's with lower profit contribution, variance, and higher Fungibility if the concept of financially differentiated sku service levels is implemented. Certain types of EOL and Non High Volume (or sporadic demand distribution) runners might also be suitable for this.

**Staging Inventory:**

For Non-VMI customers, it is preferable to keep more in SFGI and less in CW e.g. 18 DOI in SFGI & 3 DOI in CW. This increases the Risk-pooling effect. This should be done in addition to significantly increasing SFGI capacity and reducing Finish capacity constraints. For VMI customers, assuming build ahead of significant amounts of next month's projected demand, CW inventory must be reduced substantially, keeping enough only to service the remaining CW Direct & Disti-Box or Tray customers.

## 2. SDD Algorithm Settings:

The SDD settings that determine SFGI-CW inventory splits must add a VMI component. Additionally during Flush operations, SDD currently looks ahead to the next month of orders to decide what to make. With VMI, this component should be included in the consideration effectively increasing the look-ahead period to more than a month

## 3. VMI:

Given that the idea behind SDD is to delay customization to as late a point in the supply chain as possible, with the new VMI strategy installing Finish capability in VMI hubs, possibly in addition to that at SFGI, must be investigated. With significant forward staging of inventory in VMI's, VMI hubs are effectively the new CW's. As discussed before, if we continue to postpone through SDD only at SFGI and then introduce VMI, several new issues are created. We have to build ahead several weeks out dilutes SDD's impact, which now will be Pull based only for a small portion of volume. Also asking SDD to rely on monthly projections make it less accurate versus reacting to weekly demand pulls from customers. Finally the additional SNP solve sitting between VMI hubs and SDD introduces additional judgment and time lag as it is done once a week instead of multiple times a day. All these reasons point to Finish in VMI as a viable idea to be considered.

# D. Recommendations for Production Planning

## 1. Converting OFG-IFG Planning for ATM's to Pull from Push:

There is a real opportunity to improve ATM Responsiveness, in addition to improving forecasts on the ATM side, by moving to a pull signal based on replenishment versus a forecast based push signal as shown in figure below. This will require –

i) Reducing the ATM frozen horizon from 2 week to 1week,

ii) Reducing OFG planning time by half by efforts to streamline the production plan even if it is at the expense of decreased tool utilization, and removing the need for IFG with only 1 demand signal plus a very simple inventory signal, and

iii) Replacing the Geo demand Push signal with a Pull signal based on inventory replenishment for CW +VMI, using better information systems to pull information rapidly versus having to piece it together.

**Figure 6-1. Current Push-Based OFG-IFG Process**

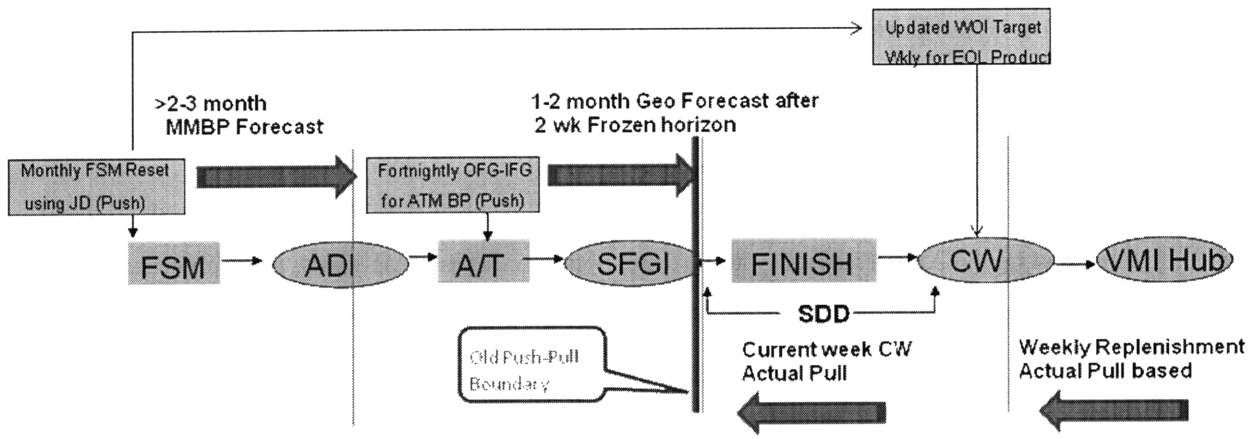
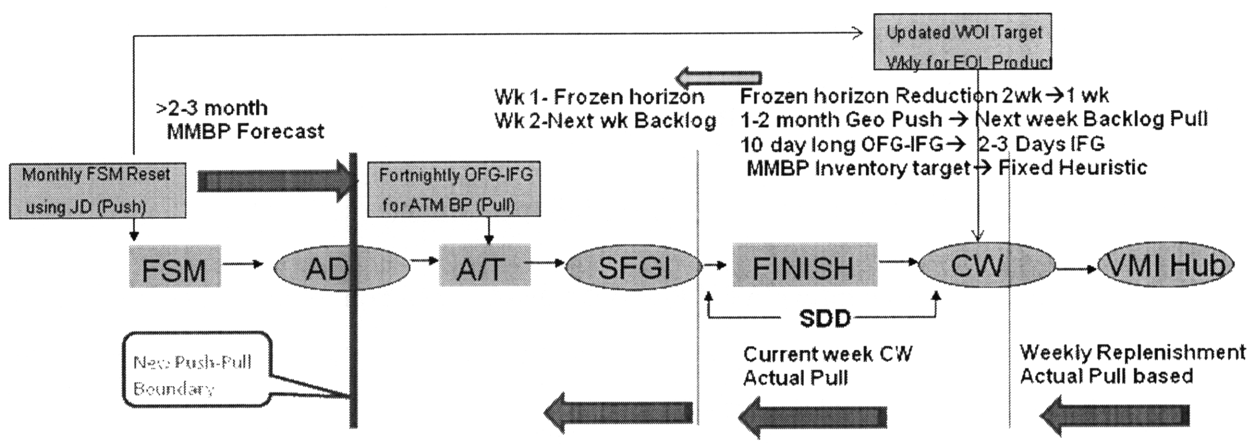



**Figure 6-2. Alternate Pull-Based OFG-IFG Process**




**2. Inventory Positioning Between CW and VMI:**

The profile below shows the current inventory positioning on a relative basis first and the suggested positioning later. The reason the CW inventory is moved to SFGI is to Finish only what is necessary. An advantage of this is the risk pooling on unfinished parts rather than finished parts. However a disadvantage is that the long lead time between CW and VMI makes it less responsive.

An alternative approach to improve responsiveness, while moving to postponement as far as possible into the chain, is to move SFGI and Finish to the VMI hub shipping semi-finished goods to VMI hubs. The disadvantage is the risk pooling is now on finished goods at CW. This is showed in the next set of figures.

**Figure 6-3. Current and Suggested Inventory Profile**

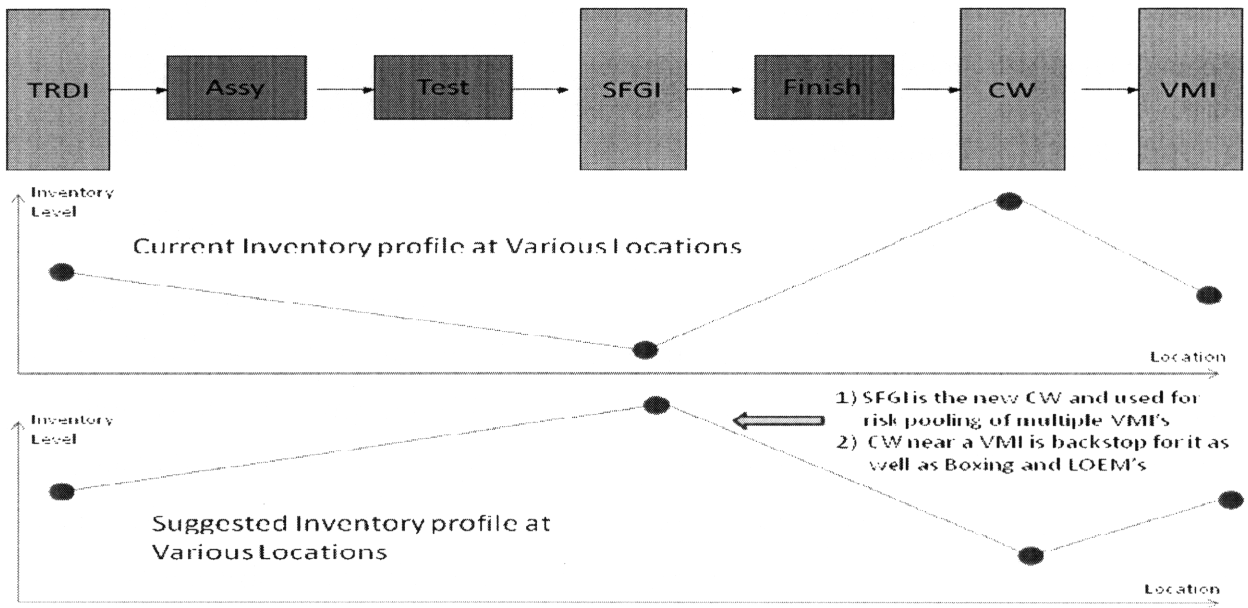
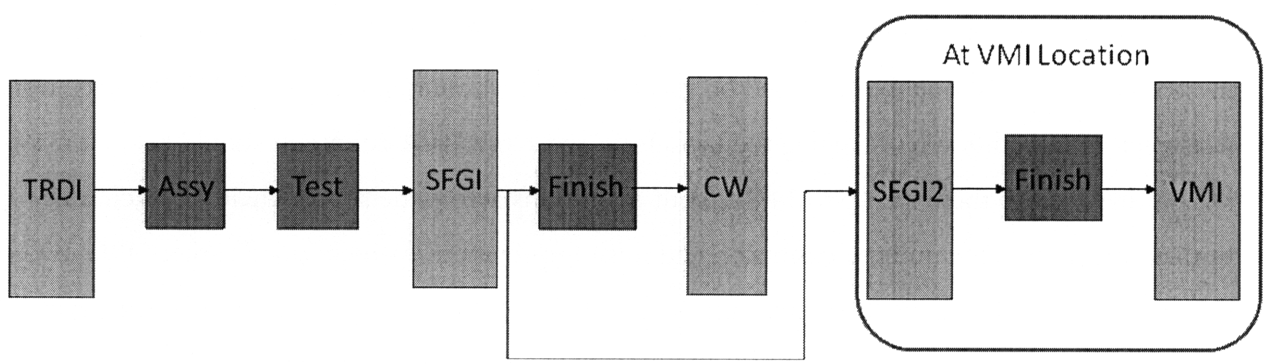



**Figure 6-4. Alternative SFGI Arrangement to Delay SDD Postponement to VMI**




### 3. ATM Processes in the Frozen Period:

The two key ATM processes for build decisions inside the frozen horizon are SDD and ATM Range Planning. To make them more useful for demand updates they need to be made more effective. For SDD, the limitations on ASRS (SFGI Storage Space) space are the key constraint and need to be removed by expanding available ASRS space. Dynamic allocation of SDD SFGI capacity must also be considered to improve ASRS utilization ratios. Scarce ASRS space should be allocated factoring in profit margins, and not purely on volume as discussed earlier.ATM Range Planning should be expanded to include affecting mean build signals and recoveries instead of just providing build ranges. In addition to the above two processes, a new process for dynamically remixing builds should be considered (L1 swaps) based on Excess Inventory, Aged Demand in Qx Bubble, SDD Flush Probability.

**4. <u>OFG-IFG Solver Optimization :</u>**

Currently the OFG-IFG production planning process is run individually by various planners for other products. This means that products sharing common BOM (Bill of Materials viz. die and packages) are not optimized globally if minimum planner requirements are met. Globally optimized runs factoring in Margins should be run post these individual fragmented solves. Additionally, the IFG process should be run without having to consume total requested capacity in OFG since a lot can change in a week. . In addition, the process of adjusting demand inputs and flags to meet the required demand as closely as possible would benefit from a more defined business process to prevent sub-optimization.

**5. <u>Solver Mechanics:</u>**

The inventory targets used in solvers have a huge impact. Currently they are set in direct proportion to volume but they should be set based on revenue (price x volume), variance and Fungibility, not just volume. In addition to changing the basis for targets, the WOI definition should be standardized, since there are up to 14 different definitions across the supply chain for each time horizon which causes immense confusion and misinterpretation. As mentioned earlier OFG-IFG solvers should have a final global optimizer step to balance inventory build above needed target on relative ranking basis (incorporating Price, Volume, Fungibility and Variance)

**6. <u>Micro Feedback loops & Trend Consistency Checks:</u>**

Early warning systems should be built based on short run feedback loops. Current feedback loops are long and introduce significant delays in learning. The excess inventory flag is a good example of this. When the flag is raised, it affects WSPW on a macro scale with its effect on JD, but it needs to be done on micro scale with ATM OFG-IFG based build plans and Pro-active L1 demand swapping being affected by CW direct pulls.

So feedback loops should be well nested and interlocked with other planning processes. A comparison of trends must also be done e.g. increased wafer starts on a particular product while current OFG-IFG plan or swaps show a decreasing trend may signify an issue such as a demand pickup of NPI products away from EOL products or product cannibalization.

### 6.3 Future Work

Future Work for continuing to improve the supply chain responsiveness lies in executing in two key areas- new tools and new projects.

1. **New Tools:**

a) A standalone tool to enable quick conversion of different product levels (Sku, MM, L3, Stepping, L1 etc.) is probably one of the key tools required for improving communication across the chain. This will help in quick conversion and communication of tactical demand updates down the chain. Logic exists inside solvers, other division planning tools, e.g. TDF and in supply team spreadsheets. The tool must be propagated to Division, ATMs, the SDD team, Geographies etc. to enabling company-wide communication.

b) A tool showing current inventory position across supply chain from target setting to implementation through build plan, OFG and IFG, SDD, to the resulting inventory positions in CW would help product supply managers manage the supply chain.

c) A Division and MMBP tool showing products at the L1 level, flagged for Excess Inventory, potential SDD flushes & bubbles in the ATM build schedule (Qx bubbles), while suggesting proactive L1 swaps for rebalancing builds, would help immensely.

d) A tool that would help maintain an up-to-date list of products, with enough supply available for expedited approval of last minute customer upside requests in the change order request (CORs) database, would help improve customer responsiveness.

e) A key issue for strategists at MMBP is that the implications of their strategy for product prioritization don't easily get transmitted down the supply chain to the factory network. A relative product prioritization tool based on MMBP strategy and pricing would help downstream decision makers, e.g. ATM, SDD etc. make the proper decisions. A web based information distribution system like the above could also work for single view of discounts for example where different business units can offer different types of discounts on the same sku and where having one view of the actual selling price of a sku on an enterprise wide basis may be difficult.

f) A software patch or process for treating placeholder values in the SAP planning system, e.g. allocation as a safety stock for field sales to prevent adding inventory on this volume, should be developed.

## 2. Projects:

There are many other projects that might help with responsiveness. Some of these might be small enough for internships while others might require the IT department to work on them.

a) Pricing changes are currently done for strategic and competitive response, but not as much to manage inventory. It should be used to shape demand in line with supply e.g. when excess inventory flags go up or SDD Qx bubbles start building up, causing increased L1 swapping activity. Vendors such as Dell which are famous for demand shaping should be bench marked for best practices.

b) Currently SDD isn't modeled in Inventory Target Setting, but it should be, since it is a key push-pull interface in the supply chain that decouples planned production from customer pulls.

c) The decision on what gets stocked in VMI Hubs is done primarily on volume. But just like the decision for SDD capacity allocation, this decision should be made on a combination of Revenue (Volume x Price), Variance, Fungibility etc.

d) Product transitions are notoriously difficult for planning purposes. Making EOL/NPI visible to the ATM Factory Network would help with production planning.

e) While there has been an effort to quantify the cost of maintaining a new sku, there hasn't been a similar study done for studying the impact of price or product transition reduction. This would help tremendously, as there definitely is a big cost to every additional transition in terms of supply chain inefficiency.

f) The impact of positioning substantially more inventory in semi finished goods instead of finished goods should be evaluated and implemented if feasible.

g) The possibility of keeping the bulk of consistent volume items on an automatic system instead of planning for them every month should be evaluated for cost savings and speed during the Reset and OFG-IFG process.

h) Currently blanket service levels are set for all sku's. Having different service levels for different customers or sku's based on their strategic value or profit margins etc. should be studied (Raiij 05).

i) Since VMI hubs might reduce the ROI of SDD, this should be modeled along with the ROI of installing SDD at VMI hubs

j) The impact of architectures such as Atom (which aren't customizable to a large extent at SDD) and their growing sales should be studied, especially as it relates to Inventory buildups.

# 7. BIBLIOGRAPHY

[1] Drucker P. & Maciariello J.(2004). The Daily Drucker. New York: Harper Business

[2] Surowiecki J.(2005). The Wisdom of Crowds. New York: Anchor Books

[3] Howe J.(2008).Crowdsourcing: Why the Power of The Crowd is Driving the Future of Business. New York: Crown Business.

[4] Walker W.(2005). Supply Chain Architecture: A blueprint for networking the flow of material, information and cash. Boca Raton: CRC Press.

[5] Beckman S. & Rosenfield D.(2007). Operations Strategy: Competing In The 21st Century. Boston:Irwin/McGraw-Hill

[6] Senge P.(2006). The fifth discipline: the art and practice of the learning organization. New York: Doubleday/Currency.

[7] Fine C.(1998)Clockspeed: winning industry control in the age of temporary advantage. Reading: Perseus Books.

[8] Taleb N.(2007). The black swan: the impact of the highly improbable. New York: Random House

[9] Taleb N.(2004). Fooled by randomness: The hidden role of chance in life and in the markets. New York: Thomson/Texere.

[10] Walsh J.(2008). Keynes and the Market: how the world's greatest economist overturned conventional wisdom and made a fortune on the stock market. Hoboken: John Wiley and Sons.

[11] Hsu M.(2006). Improving customer service level through centralized supply flexibility. MIT.

[12] Levesque J.(2004). Analysis of variability in semiconductor supply chain. MIT.

[13] Margeson W.(2003). A Forecasting and Inventory model for short life-cycle products with seasonal demand patterns. MIT.

[14] Page A.(2001). Forecasting mix-sensitive semiconductor fabrication tool set requirements under demand uncertainty. MIT.

[15] Rassey L.(2003). Enterprise Strategy: Leveraging the dynamics and behaviors in supply chain for operational excellence. MIT.

[16] Sonnet M.(2005). Cost of stock outs in the microprocessor business and its impact in determining the optimum service level. MIT.

[17] Chow J.(2004) Analysis of new approaches to improve the customer responsiveness of Intel's microprocessor supply chain. MIT.

[18] Hopman J. (August 2005). Managing Uncertainty in Planning and Forecasting. Intel Technology Journal. Volume 9, Issue 3.

[19] Bean J., Devpura A., O'Brien M. & Shirodkar S.(August 2005). Optimizing Supply Chain Planning. Intel Technology Journal. Volume 9, Issue 3.

[20] Kraiselburd S. & Watson N.(April 2007). Alignment in Cross-Functional and Cross-Firm Supply Chain Planning. HBS Working Paper.

[21] Garvin E. & Gino F.(March 08). Is yours a learning organization? Harvard Business Review.

[22] Slobodow B.et al. (January 2008). When supplier partnerships aren't. MIT Sloan Management Review

[23] Maynard M. Taking the wheel as Toyota skids. New York Times Article on February 15 2009.

[24] Summers Larry. Bloomberg Interview on US Financial Crisis on February 13 2009.

[25] Shiller R. Open Yale Lecture Series on Financial Markets.

[26] Weinstein J., Reyner A. & Kempf K.(June 2008) Internship Problem Statement.

[27] Pai N., Schwartz J. & Kempf K.(2008) Control Diagram for Demand Forecasting Process.

[28] Reyner A.(2008). Class on Demand Planning Processes.

[29] CPLG Training Document on OFG-IFG

[30] Kempf K.(2008).The Illusion of Control: Complexity and the Enterprise.

[31]O'Brien M.(2008).Presentation on SDD Overview.

[32] Crane B.(February 2009). MIT Presentation on Improving the Profitability of Hewlett Packard's Supply Chain.

[33] Bragg S. Presentation on SDD.

[34] Lee H.L. (October 04). The Triple–A Supply Chain. Harvard Business Review.

[35] Forrester J.W. (1961). Industrial Dynamics. Cambridge:MIT Press.

[36] Raiij E. (2005). The strategic value of customer profitability analysis. Marketing Intelligence and Business Planning, Volume 23, Number 4

[37] Fisher M. (January 1981). The lagrangian relaxation method for solving integer programming problems. Management Science, Volume 27, Number 1.

## Appendix - Note on theoretically setting service levels to optimize aggregate service

The Semi-Finished Goods Packaged Chip Inventory (SFGI) sits at the input of Finish process and are said to be at the L3 level. Each product at an L3 level can be mapped to several sku's or L1 level product (Intel Terminology) or MM level (SAP terminology). This is said to be a highly fungible L3. Note that each L1 can also be obtained from many L3's. In this sense this is a many to many mapping. The final mapping or Finish process is done via fusing (blowing a fuse) and down binning, either from a speed or power perspective.

In a typical case the L3 level volumes sitting in SFGI warehouse, are based on an approximately 2 month old forecast based production plans. The L1 level demand and shortages are based on end customer demand from the finished goods warehouse post-Finish called CW (component Warehouse). In this sense the SDD (SFGI delayed differentiation) algorithm that controls the Finish decision making process represents the Push-Pull interface for the Production process. The issue at hand is to figure out the best way to make the decision for rebalancing the L1 level finish requests based on Push-Pull mismatches. This can be done by trying to maximize Revenues (variable costs i.e. COGS and holding costs being much smaller than price due to high profit margins means Revenues dictate profits.)

Objective: To distribute L3 level fungible product family across L1's based on certain criteria. In other words we are looking to establish weights for a weighted average capacity allocation approach subject to a fixed inventory capacity budget.

Strategy: This can be done using an approach focusing on maximizing Revenues subject to a fixed inventory capacity budget using the techniques of non-linear optimization and Lagrangean relaxation (Fisher, January 81). Let:

$K_i$ = constant for item i representing number of standard deviations of inventory to hold

$S_i$ = shortage on item i

$\sigma_i$ = standard dev of item i during lead time

$V_i$ = volume of item i

$P_i$ = Price of item i

$R_i$ = Relative Ranking of item i

B= Fixed Budget

**(A) Theoretical Derivation of Weights:**

The Total Revenue is $\sum V_i P_i F(K_i)$

Where $F(K_i)$ is Distribution of demand for item i centered around 0

Such that $\sum K_i \sigma_i = B$ (a constant). The latter follows from a normal distribution.

So, using a Lagrangean approach (see Fisher, January 81) let's looks at aggregate service with the budget variable in the Lagrangean.

$H(\lambda)$ = [Function to be maximized] + $\lambda$ [Zero value of constraint]

$$H(\lambda) = \sum V_i P_i F(K_i) + \lambda (B - \sum K_i \sigma_i)$$

$\lambda$= Lagrange Multiplier Constant.

So to get optimality condition is for each item i

$\Rightarrow$ Max $\sum V_i P_i F(K_i) + \lambda(\sum K_i \sigma_i)$

$\Rightarrow$ Max $V_i P_i F(K_i) + \lambda (K_i \sigma_i)$ for each item i w.r.t. $K_i$

$$\Rightarrow \sum \frac{d}{dK_i} V_i P_i F(K_i) + \lambda \frac{d}{dK_i} \sum K_i \sigma_i = 0$$

assuming F is concave, we get

$$\Rightarrow V_i P_i F'(K_i) + \lambda \sigma_i = 0$$

$$\boxed{F'(K_i) = -(\lambda \sigma_i)/(V_i P_i)}$$

$$\boxed{K_i = F'^{-1} [-(\lambda \sigma_i)/(V_i P_i)]}$$

Note: 1) $\lambda$ is typically negative making operand positive.

2) High Volume or Low sigma implies a lower density and hence a larger value for $K_i$

**(B)Heuristic Method to Derive Weights:**

1) Get Rank by $[(V_i P_i)(\text{Fungibility})]/\sigma_i = \text{Rank}_i$

2) Normalize to get Relative Ranking by calculating $(Rank_i)$/Average Rank = $R_i$ .

3) Calculate average number of sigma's = K = Total Safety Stock divided by sum of sigma's e.g.

$K\sigma_1 + K\sigma_2 + \ldots =$ Total Safety Stock, which is given.

$K(\sum\sigma_i)$ = Total Safety Stock

K = (Total Safety Stock)/ $(\sum\sigma_i)$

4) Determine $K_i$ the safety factor for each sku by approximating the non linear function $F'^{-1}$ (derived in previous section A theoretically) for simplification, with a linear function f .

From previous section we got theoretical value as $K_i = F'^{-1}[-(\lambda\sigma_i)/(V_i P_i)]$

We now approximate $K_i = F'^{-1}(x)$ with $K_i = a+f(x)=a+f(R_i)=a+f(R_i K)$

5) Normalize $K_i$ again to ensure we meet constraint B by

$K_i = (\sum K_i\sigma_i)/B$ to normalize $K_i$ to make sure Budget is within limits set by B.

**Difference between Theoretical optimum and Heuristic Implementation-**

In the heuristic ASRS allocation algorithm described in section 4.2.2, we had Allocation $\propto$ (Profit)(Volume)(Variance)(Fungibility). In the theoretical optimum calculated in this appendix , we get Allocation $\propto$ [(Profit)(Volume)]/[Variance] . If Fungibility considerations are added it can be shown that Allocation $\propto$ [(Profit)(Volume)(Fungibility)]/[Variance]

Thus eventually we recommend going to a system where we move variance to the denominator. During the project due to the large numbers of SKU's whose composition is always changing, several practical issues related to IT systems and data integrity in getting reliable variance numbers a different simpler heuristic was used. Simplifications were used for variance rankings and the simpler heuristic approach was used. Eventually once the variance calculations can be quickly and reliably be turned around, implementation and validation on the suggested theoretical formula might be carried out.