Semiconductor Production Planning

Rob Leachman
University of California at Berkeley

October, 2014
Agenda

• Industry background
• Elements of production planning
• Basic planning techniques
• IMPReSS planning system at Harris Corp.
Background

• Evolution of competition in the semiconductor industry:
  – Proprietary designs
  – Price
  – Quoted lead time
  – On-time delivery ("customer service")

• Increasing need to improve delivery quotation and production scheduling
Background (cont.)

- Evolution of manufacturing network:
  - State-of-the-art production facilities now cost billions of $, located around the globe
  - 500 - 5,000 or more finished goods types
  - Cycle times to build products are now 2-3 months
  - Demand forecasts are prone to error

- Increasing need to control and direct manufacturing network
Legacy practices

- Internal manufacturing network run much like a supply chain
- **Multiple planning groups** (e.g., marketing, central planning, factory planning) exercising local control
- Limited information exchange between local systems. **Negotiations required** to achieve a plan
  - time-consuming and infrequent planning cycles
  - considerable judgment and uncertainty involved
Trends in performance

• During the 1990s, most firms in the industry improved on-time delivery percentages from 70s to 80s

• A few improved to high 90s by making fundamental changes
  • Formal, mostly automated planning system
  • Frequent and swift re-planning of entire supply chain
  • Organizational change away from multiple planning groups to single organization maintaining the data and procedures of the formal system
On-time delivery

% of line-item shipments (1H90)

- 100
- 50

ARE YOU SELLING A LOTTERY OR A SERVICE?

Production Planning

Company F1

Company F2

Shipment week minus quoted week
What makes planning hard

- Factory capacity is complex and not understood
- Yields and cycle times are variable
- Demand is uncertain
- Should we build to forecast or wait for an order?
- Binning, substitution, alternative source products
- Organizational barriers: Decentralization of planning, lack of information or even incorrect information
Two types of planning cycles

• Incremental planning cycle: add new demands to production plan without adjusting the rest of the plan

• Full regeneration or batch planning cycle: re-plan everything

• Companies with limited capacity and a variety of products perform batch planning cycles in order to allocate capacity efficiently

• Incremental planning cycles are used to plan build-to-order products without waiting for next batch cycle

• Both kinds of planning cycles can be done in same company
Basic Planning Concepts: The Production Planning Cycle

• Elements of the production planning cycle:
  • Quantify and prioritize demands
  • Requirements planning
    – How much product do we still need to start?
  • Capacitated loading
    – Smooth the net requirements into a feasible plan for factories
  • Computation of availability
    – What is the planned, uncommitted supply line?
Quantify and prioritize demands

• Generate “unconstrained” market forecasts
• Track forecasts vs. actual customer requests for quotes or vs. actual customer orders (forecast error)
• Sort demands into priority classes. Example:
  – previous customer commitments
  – replenishment to target inventory levels
  – sales forecasts discounted by historical forecast errors
  – rest of (i.e., risky portion of) sales forecasts
• Decide boundary between build-to-order or build-to-plan for each product
Requirements Planning

• **Standard methodology: MRP logic**
  – Start with time-phased gross demands
  – “Net” out the inventory and WIP
  – "Explode" net demands onto predecessor products
    » Account for yield loss and shift back by a cycle time

• **Many software packages available**

• **MRP works fine if no binning or product substitution or alternative factories are involved**

• But it's only for Requirements Planning, not Capacitated Loading!
Capacitated Loading

• Smooth the net requirements ("starts requests") into a capacity-feasible schedule
  – No standard logic for discrete parts industry; linear programming optimization is standard logic in process industry

• Typical method in many companies:
  – Compute in a spread sheet the approximate loads from starts requests on machines or on some artificial capacity limiters
  – Compare loads to rough estimate of capacity
  – Negotiate adjustments to starts requests
  – Re-compute loads and re-negotiate as necessary
Computation of Availability

- Availability is computed based on comparing cumulative supply to cumulative orders on hand

\[ S_t = \text{cumulative supply at time } t \text{ (including on-hand inventory and planned output)} \]

\[ O_t = \text{cumulative orders on hand at time } t \]

\[ A_t = \text{cumulative available-to-promise at time } t \]

\[ A_t = \text{Min } \{S_\tau - O_\tau \mid \tau \geq t\} \]
Computation of Delivery Quote

- Delivery quote is computed based on comparing cumulative availability to cumulative delivery request

\[ R_t = \text{cumulative delivery request at time } t \]

\[ A_t = \text{cumulative available-to-promise at time } t \]

\[ Q_t = \text{cumulative quoted delivery at time } t \]

\[ Q_t = \text{Min} \{ A_t, R_t \} \]

If quote is accepted, we update \[ A_t \leftarrow \text{Min} \{ A_\tau - Q_\tau \mid \tau \geq t \} \]
## Quotation Example

<table>
<thead>
<tr>
<th>Time period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cum supply</td>
<td>220</td>
<td>320</td>
<td>440</td>
<td>560</td>
<td>680</td>
<td>800</td>
</tr>
<tr>
<td>Cum orders</td>
<td>100</td>
<td>220</td>
<td>350</td>
<td>455</td>
<td>605</td>
<td>635</td>
</tr>
<tr>
<td>Difference</td>
<td>120</td>
<td>100</td>
<td>90</td>
<td>105</td>
<td>75</td>
<td>165</td>
</tr>
<tr>
<td>Cum availability $A_t$</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>165</td>
</tr>
<tr>
<td>New order request (cum)</td>
<td>30</td>
<td>60</td>
<td>90</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery quote (cum)</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>60</td>
<td>75</td>
<td>120</td>
</tr>
<tr>
<td>Revised $A_t$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
</tbody>
</table>
Application of Mathematical Optimization to Production Planning

• A linear optimization problem is called a linear programming (LP) problem

• Heavy process industry (petrochemical refineries, paper mills, steel and aluminum plants have used LP to do planning since the 1950s

• LP is mostly unknown in discrete parts industry

• A Tutorial on Linear Programming (LP)

• Connection between LP and MRP
### A Simple Example

<table>
<thead>
<tr>
<th>Product</th>
<th>Market demand per week</th>
<th>Sales price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product</th>
<th>Processing dept.</th>
<th>Processing time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>3</td>
</tr>
</tbody>
</table>
L. P. Tutorial (Cont.)

A Simple Example (cont.)

<table>
<thead>
<tr>
<th>Processing Dept.</th>
<th>Capacity (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
</tr>
</tbody>
</table>

Problem: How much should we make of each product?
Formulating the L. P.

1. Define Variables
   - $x = \text{amount of product 1 produced}$
   - $y = \text{amount of product 2 produced}$

2. Define Objective Function
   - Maximize the profit $P = 10x + 5y$

3. Define Constraints on the variables
   - Capacity of Dept. A: $5x + 2y \leq 30$
   - Sales limits: $x \leq 5, y \leq 10$
   - Capacity of Dept. B: $3x + 3y \leq 30$
   - Non-negativity: $x \geq 0, y \geq 0$

4. Put in matrix form:

   $\begin{array}{ccc}
   10 & 5 & P \\
   5 & 2 & 30 \\
   3 & 3 & 30
   \end{array}$

   1 Column per variable + 1 Row per constraint + 1
Geometric Interpretation of L.P.

Make $P$ as big as possible while staying inside the constraint set.
Geometric Interpretation (cont.)

Note that the *optimal solution* for the variables is always a corner.

In this case, the optimal corner is defined by the intersection of the lines

\[ 5x + 2y = 30 \]
\[ 3x + 3y = 30 \]

or \[ x = 3.33 \text{ units per week and } y = 6.67 \text{ units per week.} \]

The *feasible region* for the variables defined by the constraints is called the *Simplex.*
Solving Linear Programs

• A corner of the Simplex is defined by $n$ equations in $n$ variables, where $n$ is the number of constraints (rows)

The trick is to pick the **optimal** corner.

• **The Simplex Algorithm**(1947) is a means to successively evaluate corners, always moving to a better one
  • Solution time of the Simplex Algorithm is ~ $(3 - 6)(\text{number of rows})$

• **Interior Point Methods** (1980s) cut across the Simplex, trying to get to the best corner more quickly

• L. P. software packages today can solve problems with 100,000 rows in a few hours
Time-Phased Linear Program

• **Data and variables are expanded by time period** (e.g., sales demand by period, capacity by period, amount produced of each product by period.

• Time-phased problems model the possibilities of producing early (inventory) or producing late (backorders).

• To model inventory and backorders, we need to define variables for same and costs for same, and we need to link them in constraints to the production variables.

• We will illustrate the case of inventory variables. Assume the same data as before (applicable to all periods), plus inventory holding costs of one dollar for each product in each period.
Time-Phased L.P. (cont.)

1. Define variables for each time period:
   \[ x_t = \text{amount of Product 1 produced in period } t \]
   \[ y_t = \text{amount of Product 2 produced in period } t \]
   \[ I_{1t} = \text{amount of Product 1 left in inventory at end of period } t \]
   \[ I_{2t} = \text{amount of Product 2 left in inventory at end of period } t \]

2. Define objective function to include all variables in all periods:
   \[ P = \sum_{t} \left( 10x_t + 5y_t - I_{1t} - I_{2t} \right) \]

3. Define constraints on the variables for each time period:
   - Capacity of Dept. A: \( 5x_t + 2y_t \leq 30 \)  \hspace{1cm} \text{Non-negativity: } x_t \geq 0, \ y_t \geq 0
   - Capacity of Dept. B: \( 3x_t + 3y_t \leq 30 \)  \hspace{1cm} \text{Non-negativity: } I_{1t} \geq 0, \ I_{2t} \geq 0
   - Sales limits:
     \[ \sum_{\tau=1}^{t} x_{\tau} \leq 5t, \sum_{\tau=1}^{t} y_{\tau} \leq 10t \]
Time-Phased L.P. (cont.)

3. (cont.) Define constraints on the inventory variables for each time period:

\[
I_{1t-1} + x_t - d_{1t} - I_{1t} = 0, \text{ where } d_{1t} \text{ is the demand for product 1 in period } t
\]

\[
I_{2t-1} + y_t - d_{2t} - I_{2t} = 0, \text{ where } d_{2t} \text{ is the demand for product 2 in period } t
\]

These constraints make sure all demands are met on time while maintaining inventory balance.
Connection Between MRP and LP

- Is the MRP problem an L. P.?

Of course!

- MRP problem constraints
  - Meet all demands on time
  - Maintain inventory balance

- MRP objectives
  - Minimize total production
  - Produce as late as possible

- Equivalent objective: Minimize discounted production “cost”

- Although LP can be used, MRP calculus is quicker than solving an L. P.
IMPReSS
A System for Automated Production Planning and Delivery Quotation

Harris Corporation - Semiconductor Sector and University of California at Berkeley
Introduction to Harris

- Harris Corporation - $3.5 Billion electronics company based in Melbourne, FL
- Semiconductor Sector - $670 Million annual sales, based in Palm Bay, FL
  - Wafer fabrication plants ("Front End" plants) in Florida, Ohio and Pennsylvania
  - Device packaging and test plants ("Back End" plants) in Malaysia, Florida and Pennsylvania
Introduction (cont.)

- Sector historically focused on military and aerospace markets
- Acquisition of General Electric Solid State products and factories in late 1988
  - Sector tripled in size
  - Greater focus on commercial businesses
- After acquisition, 6 major product lines:
  - Discrete Power
  - Signal Processing
  - Data Acquisition
  - Intelligent Power
  - Digital
  - Military & Space
Introduction (cont.)

• After GESS acquisition, on-time delivery became a crisis issue
  – Many delinquent orders and inferior delivery performance
  – Estimated $100 million in lost sales in calendar 1989
  – Sector reported a loss of $75 million in fiscal 1990-91
Introduction (cont.)

• In 1990, project launched to develop and install automated production planning and delivery quotation system

• “IMPReSS” (Integrated Manufacturing Production Requirements Scheduling System)
IMPReSS Information Flow

**Demand Forecast System**
- Prioritized demands and build rules
- Factory capabilities and status
- Factory Plans (start and out schedules)

**Quotation & Order Entry System**
- Order Board
- Product Availability
- Queries & Orders
- Quotes

**Planning Engine (BPS)**
- Customer
- Material availability
- Material requirements
- Product structure and sourcing rules

**Raw Materials System**
- Bill of Materials System
- Product Availability

**Factory Floor Systems**
- Day-to-day factory operations

**Production Planning**
- Resource utilization and capacity management
The Planning Engine

- The **Berkeley Planning System (BPS)** serves as the IMPReSS Planning Engine.
- BPS combines linear programming and MRP techniques.
- BPS developed in research at the University of California at Berkeley, 1985-present.
- Sponsorship by Harris beginning in 1987.
Planning Engine Scope

Prioritized Demand Inputs:
- Order Board
- Safety Stock Replenishments
- Sales Forecasts

World-wide finished goods inventory
Factory WIP-out projections and static inventory
Product structure
Factory capability models

Compute net demands for Back End and Front End production

Compute factory production plans

“Requirements Planning”

Compute availability for quotation
Availability

“Capacitated Loading”
Technical Challenges

• Need for **standardized data structure**
• Product structures with **binning and substitution**
• Capacity analysis of semiconductor process flows using **dynamic production functions**
• **Incorporate marketing controls** on plan generation
• Cope with **immense problem scale**
Standardized Representation of the Product and Process Structure Within BPS
Requirements planning of binning structures: the simplest case

Simple Example Data:

<table>
<thead>
<tr>
<th>Bin Split</th>
<th>FG Demands</th>
<th>t=1</th>
<th>t=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin 1/Type 1</td>
<td>0.20</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Bin 2/Type 2</td>
<td>0.80</td>
<td>80</td>
<td>110</td>
</tr>
</tbody>
</table>
Try MRP logic with a "driver bin"

Try Bin 1 as the driver bin:

\[
\begin{align*}
& t=1 \\
& 25/0.2 = 125 \\
& 10/0.2 = 50 \\
& \text{(Net Shortage of 50 Bin 2's!)}
\end{align*}
\]

Try Bin 2 as the driver bin:

\[
\begin{align*}
& t=1 \\
& 80/0.8 = 100 \\
& 110/0.8 = 138 \\
& \text{(Shortage of 5 Bin 1's!)}
\end{align*}
\]

(12.6 Bin 1's left over)
Formulate as a linear programming problem

Maximize total discounted cash flow considering production costs and demand revenues

Subject to constraints for demand satisfaction and bin inventory balance
Most general product structure

- Alternative Assembly/Test Flows
- Bin Inventories
- Demands for Finished Goods
Failure of Usual Capacity Analysis

Product 1 - machine M1 in week 1, machine M2 in week 2

Product 2 - machine M2 in week 1, machine M1 in week 2

Capacity of machine M2 - 2000 units per week

Production plans:

<table>
<thead>
<tr>
<th>Plan</th>
<th>Product</th>
<th>Week 1</th>
<th>Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>P1</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>#2</td>
<td>P1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>#3</td>
<td>P1</td>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0</td>
<td>2000</td>
</tr>
</tbody>
</table>

NOT FEASIBLE!
Capacity Analysis Using Dynamic Production Functions

Issue: Process routes that visit key resources repeatedly ("Re-entrant flows")

Diagram:
- Fab → Probe
- Mask Align → ... → Mask Align → ... → Mask Align
Assume *rate-based* schedule of production:

- **Cumulative Starts**
- **Actual Starts Curve**
- **Target Starts Curve (Basis of Capacity Model)**

Production Planning
Dynamic Capacity Analysis (cont.)

Need to express loads on key resources in terms of period-by-period starts rates

Example: Load on P&E 240 Positive Aligner from starts of process route P411

Upload extract from factory floor databases:

<table>
<thead>
<tr>
<th>Process Step ID</th>
<th>Cum TPT (Weeks)</th>
<th>Cum Yield (%)</th>
<th>UPH (Units Per Hour)</th>
<th>Derived Load Per Start (Machine Hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.368</td>
<td>97.98</td>
<td>56</td>
<td>0.0175</td>
</tr>
<tr>
<td>9</td>
<td>1.330</td>
<td>95.10</td>
<td>45</td>
<td>0.0211</td>
</tr>
<tr>
<td>12</td>
<td>1.744</td>
<td>92.76</td>
<td>36</td>
<td>0.0258</td>
</tr>
<tr>
<td>16</td>
<td>2.290</td>
<td>88.95</td>
<td>39</td>
<td>0.0228</td>
</tr>
</tbody>
</table>
Let $x(t)$ denote the rate of starts of process route P411 at time $t$. Then the load on the P&E 240 machines at time $t$ is, theoretically,

$$0.0175 \ x( t - 0.368 ) + 0.0211 \ x( t - 1.330 ) + 0.0258 \ x( t - 1.744 ) + 0.0228 \ x( t - 2.290 )$$

<table>
<thead>
<tr>
<th>Process Step ID</th>
<th>Cum TPT (Weeks)</th>
<th>Cum Yield (%)</th>
<th>UPH (Units Per Hour)</th>
<th>Derived Load Per Start (Machine Hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.368</td>
<td>97.98</td>
<td>56</td>
<td>0.0175</td>
</tr>
<tr>
<td>9</td>
<td>1.330</td>
<td>95.10</td>
<td>45</td>
<td>0.0211</td>
</tr>
<tr>
<td>12</td>
<td>1.744</td>
<td>92.76</td>
<td>36</td>
<td>0.0258</td>
</tr>
<tr>
<td>16</td>
<td>2.290</td>
<td>88.95</td>
<td>39</td>
<td>0.0228</td>
</tr>
</tbody>
</table>
Consider the load from performing step 12 in week 3:

Assume starts are made uniformly within each week. Then the load from step 12 in week 3 is

\[0.0258 \{ 0.744 \text{ (starts in week 1)} + 0.256 \text{ (starts in week 2)} \}\]
Let $x_t$ denote the starts of process P411 in week $t$.

Then the load on the P&E 240s in week $t$ is

$$0.0175 \left[ 0.632 x_t + 0.368 x_{t-1} \right] + 0.0211 \left[ 0.670 x_{t-1} + 0.330 x_{t-2} \right]$$

$$+ 0.0258 \left[ 0.256 x_{t-1} + 0.744 x_{t-2} \right] + 0.0228 \left[ 0.710 x_{t-2} + 0.290 x_{t-3} \right]$$

or

$$0.01106 x_t + 0.02718 x_{t-1} + 0.04235 x_{t-2} + 0.0066 x_{t-3}$$
Dynamic Capacity Analysis (cont.)

Represent Capacity of P&E 240 Machines:

Upload extract from factory floor capacity database:
  • Quantity in service: 7
  • Max utilization (of total working time): 0.66
  • Hours worked per week: 168

Machine hours per week available to run product:
  \[7 \times (168) \times (0.66) = 554.4\]
Dynamic Capacity Analysis (cont.)

The constraint on production schedules is then

\[ 0.01106 x_t + 0.02718 x_{t-1} + 0.04235 x_{t-2} + 0.0066 x_{t-3} \]
\[ + \{\text{expressions for loads from other process routes}\} \leq 554.4 \]

BPS constructs similar constraints for all other key resources automatically from the factory floor data.

- Accounts for factory working calendar
- Allows time-dependent yields, cycle times, UPH's, equipment assignments, equipment quantities, down time factors, etc.
Accuracy of BPS Capacity Model

• BPS schedules have been fed into detailed simulations of actual wafer fabs

• Agreement with detailed simulations within 1%, in terms of product cycle times and equipment utilizations
Need for Marketing Controls

- All demands can't be filled as soon as requested
- All demands are not equally important, so they must be prioritized
  - Key accounts, very late customer orders, "lines down"
  - Other customer orders
  - Safety stock replenishments
  - Sales forecasts
- How far through the product structure one should build-to-forecast needs to be different for different products
Modeling Marketing Priorities

Demands are categorized by *priority classes* of three types (orders, safety stock rebuilds, forecasts).

Multiple classes of each type are allowed.
Modeling Marketing Priorities (cont.)

Priority classes used at Harris:

• **Class 1**: Orders sorted by Harris promise date
• **Class 2**: Orders sorted by customer request date
• **Class 3**: Safety stock rebuilds
• **Class 4**: Sales forecasts
BPS Approach Using Priority Classes

• Demand classes are loaded one by one in a series of L.P. calculations. (Make Class 1 as on-time as possible before considering Class 2 demands, etc.)

• In the L.P. calculation for each demand class, we maximize total discounted cash flow
  – subject to available capacity and subject to maintaining on-time delivery in the higher-priority classes.
Priority Classes (cont.)

• To protect customer service, demands corresponding to orders are placed in higher priority classes. Suggested classes:

Class 1: Orders sorted by delivery promise date
Class 2: Orders sorted by customer request date
Class 3: Safety stock replenishments
Class 4: Reliable portion of forecasts (e.g., subtract one sigma of forecast error from forecast)
Class 5: Remaining (risky) portion of forecasts
BPS Algorithm for Incremental Loading of Prioritized Demands

Solve Linear Programs to load each class (in order)

- L. P. model:
  \[ D^r_t = \text{Cumulative demand at time } t \text{ (Cum over time and Cum over classes } 1, 2, ..., r) \]
  \[ BO^r_t = \text{Back orders at time } t \text{ for demand input } D^r_t \]
  \[ I^r_t = \text{Inventory at time } t \text{ for demand input } D^r_t \]
  \[ X^r_t = \text{Cumulative production output in period ending at time } t \text{ for demand input } D^r_t \]

Constraints in L. P. for Demand Class \( r \):
\[
X^r_t + BO^r_t - I^r_t = D^r_t
\]
\[
BO^r_t \leq \overline{BO^r_t} = BO^{r-1} + D_t - D^{r-1}_t
\]
where \( BO^{r-1}_t \) is solution to previous L. P.
Incremental Loading (cont.)

\[ BO_r^r \leq \overline{BO}_t^r = BO_{r-1}^r + D_r^r - D_{r-1}^r \]
Incremental Loading (cont.)

Actually like solving only 1 LP:

- Stop after each class to adjust bounds, add demands, and perhaps change objective function
- Solution to previous class is feasible for next class (after adjusting backorder variables)
- In practice, time to optimize 5 classes is about 2X the time to optimize the first class
Build-to-Forecast Controls

- **Products** are declared as either “build-to-order” (BTO) or “build-to-plan” (BTP).

- Upper bounds are placed on period 1 production variables of BTO products in formulations for forecast classes.

- Production in period 1 is orders only; production planned for future periods includes response to forecasts, so that availability will be populated.
BPS Objective Function

Cash Flow Components:

- **Production costs** = avoidable costs of factory starts (direct materials, subcontract rates, interplant shipment costs)

- **Backorder costs** for output late to customer commitments

- **Revenue (ASP) from output**, awarded at time of shipment to demand

BPS discounts the costs and revenues
Many objectives in one

- We maximize on-time delivery.
- We protect original delivery promise dates, but if we can pull in delivery towards the customer request date, we do so.
- We don’t schedule factory starts before we have to ("demand pull").
- We capture market potential as soon as capacity permits.
Objective function (cont.)

- We maximize bottleneck equipment utilization.
- Given alternative factories with different yields or different cycle times, we load the factory with the shorter cycle time or the higher yield (provided there is capacity to do so).
- We use up in-house capacity before subcontracting.
Issue: Persistence in planning

• A concern in planning is “persistence” from one plan to the next. What aspects of plan need to persist:

• Maintain on-time delivery for on-hand orders (including “booked” orders as well as contractual guarantees)

• Do not re-schedule factory WIP or in-transit WIP
  – Do not overload factory with excessive new starts

• Company politics: maintaining “fair” allocation of capacity to various marketing product lines
Impact of re-scheduling WIP

- Some companies get caught in a “vicious circle”:

  - Increasing Sales Forecast Error
  - Since Manufacturing compensated for error, Sales dept. feels no need to improve forecasting
  - Manufacturing reprioritizes WIP
  - Cycle time gets longer
Benefit of freezing WIP schedule

• Replace with “virtuous circle”:

Since Sales dept. sees inventory, it feels need to improve forecasting

Cycle time gets shorter

Manufacturing only adjusts starts as allowed by bottlenecks

Increasing Sales Forecast Accuracy
What elements do not need to persist in next plan?

- Future factory starts and corresponding outs
- Future inventory allocations and interplant shipments
- Uncommitted product availability (modulo the company politics issue)
BPS persistence strategy

• Variables of production plan are the starts of each product in each factory
  – Projected WIP-outs of each factory are an input to planning, not an output
  – Capacity consumed by WIP flush is subtracted from RHS to insure WIP-out will be on-time

• Demands for each product are divided into priority classes which are incrementally loaded in separate optimization calculations to obtain overall plan
Coping with Problem Scale

- Harris planning problem involved **8,000 finished goods**, **1.5 year time horizon**, **20+ production lines**
- Formulated as a **single linear program**, the Harris planning problem (with only one demand class) would have had about **a half million variables and a half a million constraints**.
- Decomposition scheme is necessary.
IMPReSS Operation

• Official plan generated every weekend, with additional mid-week planning cycles.

• Five IBM Model 560 RS6000 work stations are used to run the IMPReSS Planning Engine (BPS).

• IBM’s OSL software is used to solve the linear programs formulated by BPS.
  – According to IBM, Harris is solving the largest LP’s using OSL on a regular basis of all their users.
Implementation History

• Small-scale BPS application first, jointly sponsored by Sector R&D and manufacturing (1987-90)
  – Demonstrate concept on 2 wafer fabs
  – Design interfaces to factory floor systems
  – Get user feedback, revise model and system
  – Get some success
Implementation History (cont.)

• IMPReSS project (1990-92)
  – Standardized databases at each factory and at HQ, with interfaces to factory floor systems and to marketing and materials systems
  – BOM and capacity data development
  – Install demand forecasting system
  – Enhance quotation system
  – Develop BPS for company-wide application
Implementation Issues

• Management acceptance of OR/MS approach
  – Previous small-scale success helped
  – Technical background of management helped
  – Company crisis helped

• At project kick-off, executive leadership helped a lot
Implementation Issues (cont.)

• Conversion to standard manufacturing data model
  – Various factories defined different boundary points to process flows and different product structures
  – **Conflicts** with long-held intuitions, conventions and factory floor systems
  – Tremendous one-time data conversion task
Implementation Issues (cont.)

- Discipline of Formal Data Management
  - If data not in right table in right system, then not in plan!
  - Many “sanity checks” of the data were programmed to identify missing or inconsistent data.
  - “Culture change” to formally maintain all data in exact format in proper place.
Implementation Issues (cont.)

• Management frustration
  – One year after project start, system still could not be tested on large scale
  – 1.5 years after start, data quality < 50%
  – 2.0 years after start, data quality < 70%
  – Large-scale testing and debugging finally completed during 1992
Implementation Issues (cont.)

• Conflicts with organizational goals and incentives
  – Traditional policy of building inventory to meet budget targets
  – Change to demand driven, constraint driven manufacturing paradigm (TOC campaign helped)
IMPReSS Project costs

- **One-time: $3.8 million**
  - $0.7M software
  - $1.5M hardware
  - $1.4M consulting
  - $0.2M project travel

- **Annual: $600K** (5 new head count + software maintenance)
On-Time Delivery Performance

Before IMPReSS - 75%

After IMPReSS - 94-95%

% of Line Items Delivered On Time

- BPS Installed
- Regular Operation Begins
- Official Use Begins
- 90% Product Data Quality

1990 1Q92 2Q92 3Q92 4Q92 1Q93 2Q93 3Q93 4Q93 1Q94 2Q94 3Q94 4Q94

70 80 90 100
Delinquent Orders

Number of Delinquent Order Line Items

Delinquency and on-time delivery metrics improved dramatically, yet inventories as a % of sales remained flat and lead times were reduced.
Customer Survey Results

Customer Advocacy Rates

% who will continue to buy from Harris

% who will continue to buy AND recommend Harris without reservation

Production Planning
Annual semiconductor sales increased 28% in the two years after IMPReSS implementation ($530 million to $680 million). Orders increased even more.
After heavy losses in 1990-91 and 1991-92, the Sector experienced increasing net income in 1993 and 1994, a trend that is projected to continue this year and next.
Other Benefits

- **Penetration of new markets** (e.g., telecom in Japan)
- **Lead times and cycle times were significantly reduced.**
- Increased data maintenance and accuracy permitted **accounting improvements, leading to improved pricing decisions.**
- **Improved annual capital spending decisions.**
  - Savings in first-year equipment purchases alone exceeded cost of project.
A “Cultural” Transformation

- IMPReSS provided an integrated, globally optimized plan, replacing local optimization efforts.
- After IMPReSS, there was a global, common understanding of demands and constraints.
- Better “pipeline management”
  - Other semiconductor companies were amazed at the level of coordination and communication between Harris’ Front End and Back End plants.
Harris IMPReSS’ed the World!
Subsequent trends in practice

- Most semiconductor companies worked to integrate and automate their supply chain management
- Typical strategy: integrate ERP system with planning engine
- There are now 5 major vendors of planning engines to the semiconductor industry
  - some are optimization-based, others use rule-based logic
  - all claim to incorporate features pioneered in BPS
What Happened at Harris

• Instead of dying, the Semiconductor Sector survived and thrived

• In 1999 it was spun off as a new company named Intersil
  – The IPO was the largest in semiconductor industry history and raised more than $1 Billion

• The IMPReSS planning system ran the company until 2004, when it was replaced by one of the commercial software offerings