Enterprise-Wide Semiconductor Manufacturing Resource Planning

Jonas Stray, John W. Fowler, *Life Member, IEEE*, W. Matthew Carlyle, and Aditya P. Rastogi

***Abstract*—We develop a model for global logistics and resource optimization in a typical semiconductor manufacturing operation. The model is designed to aid in resource allocation and strategic decisions for long term planning in the semiconductor industry. Decisions include where to make products, whether or not to open new facilities, whether or not to add new tools, and whether or not to subcontract.**

***Index Terms*—Mixed integer programming, optimization, production planning, routing, semiconductor manufacturing.**

I. Introduction

IN THIS paper, the problem considered is the one of allocating products to wafer fabrication facilities (wafer fabs) and routing the wafers with the integrated circuits for testing. The tested wafers are routed to where they can be cut into individual chips and put in a package. A package is a frame that is designed to protect the chip and provide connections for the chip to the rest of the computer. The packages are then routed to final test facilities for testing and classification. The products are classified (binned) according to performance and shipped to final inventory warehouses, or demand centers, for selling. Planning when to increase or decrease capacity at the production facilities as well as planning when and whether to build new facilities are some of the possibilities for these operations, e.g., purchasing a new tool for one of the bottleneck tool groups in a wafer fab, building a new test facility in a new region, or subcontracting to a foundry. When we speak about planning in this paper, we simply mean using forecasts in demand to create a lean supply chain, capable of satisfying contracts and meeting open market demands at minimum cost.

II. Literature Review

Optimization models for routing and production planning for multiproduct complex manufacturing found in the literature today are mostly based on linear programming (LP) [1], [6], [9], mixed-integer linear programming (MIP) [2], [6], goal programming [10], or stochastic programming [3], [11], [12]. Hung and Wang [1] focus on material planning and introduce a new way of modeling bin allocation. The authors show how to change the allocation of products problem from a transshipment problem with allocation variables to a dominant cut-set problem, ridding the formulation of the allocation variables and thereby decreasing solution times by two-thirds for the LP problems. The work of the authors helped verify the formulations of bin allocation constraints in the models in this paper.

Sebo [9] shows how LP methods for production planning can be applied in the real world. Tabucanon and Ong [2] discuss an MIP model for machine requirements planning that deals with capacity restrictions in a similar way to what is done in this paper. However, reentrant flow is not specifically dealt with and the concept of only using the bottleneck tools in modeling capacity is not used. The latter is introduced by Chang *et al.* [7] and is utilized in the capacity constraints in our models.

Padillo *et al.* [6] briefly describe a math programming-based strategic decision support system called the Manufacturing Enterprise Model (MEM). MEM defines eight long-range planning questions but does not provide detailed formulations of the models to answer these questions. Our paper is motivated by the work of Padillo *et al.* [6].

Other more descriptive methodologies employed, like discrete-event simulation [4], have also been applied with some success. Even combinations of mathematical programming models and discrete-event simulation models have been researched [5], [13]. In fact, [13] contains a linear program for production planning; however, it focuses more at a day-to-day planning level and does not provide solutions for extended horizon strategic planning. This is important because the decision maker needs to look far into the future to achieve a basis for strategic decisions. The argument for this statement is that investing large sums of money like building a new fab can never be justified unless there is enough time to pay back the investment.

In a search of papers focused on strategic planning, it appears as if quantitative models are scarce and qualitative models abound. Either the scope of the model does not encompass decisions for resource planning quantitatively, or it does not incorporate product allocation and routing quantitatively. So, there is a need for a model, especially one focused on the semiconductor manufacturing industry, which incorporates production planning and resource allocation in the same scenario. The benefit of planning production when trying to plan strategically is that all options for distributing production over the supply chain are taken into consideration. This means that production capacity is not overlooked or disregarded, which is important because simple solutions like stocking products to meet future demand can cover for fluctuations in demand. On the other hand, when it would not be financially sound to keep inventory and when it would be more profitable to build new facilities or install new tools can only be discovered if production is involved in the model. If the planning horizon is too short, decision makers tend to be too conservative in their decisions, building large inventories instead of increasing capacity.

Manuscript received April 1, 2003; revised December 28, 2005.
The authors are with the Department of Industrial Engineering, Arizona State University, Tempe AZ 85287-5906 USA (e-mail: john.fowler@asu.edu).
Digital Object Identifier 10.1109/TSM.2006.873399

III. MIP Model

The model presented below builds upon the LP formulation by Stray *et al.* [15]. The mixed-integer linear program presented herein maximizes profit when addition and removal of tools and addition and removal of facilities is allowed. Facilities are represented by binary variables indicating if they are operative or not. Addition of facilities is represented by binary variables and likewise for removal of facilities. Tools are nonnegative integer variables whose values depend on the number added, removed, and initially installed. Integer variables are included to represent addition or removal of tools. The objective function includes tool purchasing costs and facility purchasing costs. In order to be realistic when removing tools or facilities, the objective function includes an operating cost for both. It is also possible to have wafers fabricated in a foundry and sent to assembly and test.

In the LP model, the modeling of capacity relies on the fact that only the constraining tool groups are included. This is sufficient in the case where capacity does not change, but in the MIP model, capacity will change for individual tool groups. If the capacity of a tool group is increased, it might not be constraining anymore and other tool groups may become new bottlenecks. This can be handled in two ways. First, near bottleneck tool groups should be included in the model. Second, the costs for purchasing and maintaining a new bottleneck tool should include the costs for nonbottleneck tools needed to keep the original bottleneck tool as the bottleneck. For example, for each new stepper added there may be a number of metrology tools needed. This way of planning is consistent with the way most companies operate today, i.e., they plan which tool(s) will be the bottleneck(s).

All constraints in the LP model are included in the MIP model, but some modifications are needed. The number of machines in each tool group is now a variable and not a parameter in the capacity constraints. There are more constraints in the MIP than in the LP because of the integer variables themselves and the implications on the model that they have. The MIP model is presented as follows.

A. Indexing Sets

FAM	Set of product families. Letter p will be used to index product families.
PKG_p	Set of packages. There is one set PKG_p for each p in set FAM. Indexed by q.
$\mathrm{BIN}_{p,q}$	Set of bins for each product package and family. Indexed by b.
BET_b	Set of bins that can be sold as bin b. See bin allocation constraint.
L	Set of all location sets. Includes wafer fab set L_F, sort location set L_S, assembly set L_M, test set L_T, and demand center set L_D. Indexed by f and l.
MT_l	Set of all machine types in location l. Indexed by i.

B. Variables

1) Decision Variables:

$X_{p,l,t}^{S1}$	Number of lots of product to start in a facility in period t. S1 $\in$ (F, S, M, or T where F=Fab, S=Sort, M=Assembly, T=Test).
$W_{p,l,t}^{S1,S2}$	Number of lots of product to put in inventory before (B) or after (A) location l in period t. S1 $\in$ (A or B) S2 $\in$ (F, S, M, T, or D, where D=Demand center).
$Y_{p,l,d,t}^{S1,S2}$	Number of lots of product shipped between two locations in period t. $S1 \in L$ $S2 \in L$.
$Z_{p,q,b,d,t}$	Number of lots of each product sold in period t.
$\zeta_{p,q,b,d,t}$	Number of lots of each product available at demand centers in period t.
M_{ilt}^A	Number of machines added to tool group i, location l, and period t.
M_{ilt}^R	Number of machines removed from tool group i, location l, and period t.
Ω_{lt}^A	Indicator variable for adding plant l in period t.
Ω_{lt}^R	Indicator variable for removing plant l in period t.
S_{plt}	Number of wafers subcontracted of each family p to each assembly operation l in each period t.

2) Bookkeeping Variables:

M_{ilt}	Number of machines in tool group i, location l, and time period t.
Ω_{lt}	Indicator variable (0,1) for plant existence for location l in time period t.

C. Parameters

PBC_{lt}	Cost of building facility l in period t.
POC_{lt}	Cost of operating facility l in period t.
PRC_{lt}	Cost of removing facility l in period t.
MPC_{ilt}	Cost of purchasing for a single machine i in facility l in period t.
MOC_{ilt}	Cost of operating machine i in facility l through period t.
m_{il}	Number of machines initially installed in tool group i and facility l.
MAX_{il}^S	Maximum number of machines allowed in tool group i in facility l.
MAX_l^T	Maximum total number of machines allowed in all tool groups in facility l.
SC_{pt}	Cost for subcontracting one lot of wafers of family p in period t.

TPL	Length of one period in hours. All periods are of the same length.
T	Number of periods in model. t will be used to index periods.
$C_{p,l,t}$, $C_{p,q,l,t}$	Fraction of products started in period t that finish in period $(t+1)$. Complementary fraction finishes in period t.
$Q_{p,l,t}$, $Q_{f,p,q,l,t}$	Yields of the product (p,q) in location l in period t. $Q_{f,p,q,l,t}$ is indexed over wafer fab in which original wafer was manufactured.
$Q_{f,p,q,b,l,t}$	Resulting bins b of a product, depending on origin fab f, family p, package q, location l, and time-period t. Summed over q for all p, this number has to be less than or equal to one. It incorporates yield at test operations.
$G_{p,q}$	Number of chips per wafer, for family p and package q.
$T_{i,p,l}$	Total time product p takes to complete tool group i in location l.
$\alpha_{i,l}$	Maximum tool utilization for tool group i in location l.
$S_{i,l}$	Average downtime of tool i in hours in location l over period of length TPL.
$m_{i,l}$	Number of tools in a tool group i in location l.
$D_{p,q,b,l,t}$	Demand of a product p in package q and bin b at location l and period t.
$PC_{p,l,t}$	Cost of starting product p at location l in period t.
$TC_{l,d,t}$	Transportation cost from l to d in period t.
$IC_{p,l,t}$	Inventory cost for product p in location l and period t.
$PV_{p,q,b,l,t}$	Sales price for product p in package q and bin b at demand center l in period t.
$\mathrm{PEN}_{p,q,b,l}$	Penalty for not meeting demand for product p in package q and bin b in period t.
WLS_l	Number of wafers in a lot at wafer fabs and wafer sorts by location l.
CLS_l	Number of chips in a lot at assembly, test, and demand centers by location l.
LBT_l	Building time for location l.

D. Objective Function

The objective function includes revenue generated from selling the product, the costs of not meeting the demand, the production costs, and the costs for building and operating (or removing) facilities and tools. The objective function is shown in its entirety and then its terms are explained one line at a time

$$
\begin{aligned}
\text{Max} & \sum_{t,p,q,b,d\in L_D} PV_{p,q,b,t} Z_{p,q,b,d,t} \\
& - \sum_{t,p,q,b,d\in L_D} \mathrm{PEN}_{p,q,b,t}(D_{p,q,b,l,t} - Z_{p,q,b,d,t}) \\
& - \sum_{t,p,l\in L_F} \left(PC_{p,l,t} X^F_{p,l,t} + IC_{p,l,t} W^{AF}_{p,l,t}\right) \\
& - \sum_{t,p,l\in L_F,d\in L_S} TC_{l,d,t} Y^{FS}_{p,l,d,t} \\
& - \sum_{t,p,f\in L_F,l\in L_S} IC_{p,l,t} W^{BS}_{f,p,l,t} \\
& - \sum_{t,p,f\in L_F,l\in L_S} \left(PC_{p,l,t} X^S_{f,p,l,t} + IC_{p,l,t} W^{AS}_{f,p,l,t}\right) \\
& - \sum_{t,p,f\in L_F,l\in L_S,d\in L_M} TC_{l,d,t} Y^{SM}_{f,p,l,d,t} \\
& - \sum_{t,p,f\in L_F,l\in L_M} IC_{p,l,t} W^{BM}_{f,p,l,t} \\
& - \sum_{t,p,q,f\in L_F,l\in L_M} \left(PC_{p,l,t} X^M_{f,p,q,l,t} + IC_{p,l,t} W^{AM}_{f,p,q,l,t}\right) \\
& - \sum_{t,p,q,f\in L_F,l\in L_M;d\in L_T} TC_{l,d,t} Y^{\mathrm{MT}}_{f,p,q,l,d,t} \\
& - \sum_{t,p,q,f\in L_F,l\in L_T} IC_{p,l,t} W^{BT}_{f,p,q,l,t} \\
& - \sum_{t,p,q,f\in L_F,l\in L_T} PC_{p,l,t} X^T_{f,p,q,l,t} \\
& - \sum_{t,p,q,b,f\in L_F,l\in L_T} IC_{p,l,t} W^{AT}_{f,p,q,b,l,t} \\
& - \sum_{t,p,q,b,f\in L_F,l\in L_T;d\in L_D} TC_{l,d,t} Y^{TD}_{p,q,b,l,d,t} \\
& - \sum_{t,p,q,b,f\in L_F,l\in L_D} IC_{p,l,t} W^{BD}_{f,p,q,b,l,t} \\
& - \sum_{t,l\in L} \Omega^A_{lt} PBC_{lt} - \sum_{t,l\in l} \Omega_{lt} POC_{lt} - \sum_{t,l\in l} \Omega^R_{lt} PRC_{lt} \\
& - \sum_{t,l\in L,i\in \mathrm{MT}} M^A_{ilt} MPC_{ilt} - \sum_{t,l\in L,i\in \mathrm{MT}} M_{ilt} MOC_{ilt} \\
& - \sum_{p,l\in L_A,t} SC_{plt} S_{plt}.
\end{aligned}
$$

1) Revenues and Penalty for Unmet Demand:

$$
\sum_{t,p,q,b,d\in L_D} PV_{p,q,b,t} Z_{p,q,b,d,t} - \sum_{t,p,q,b,d\in L_D} \mathrm{PEN}_{p,q,b,t}(D_{p,q,b,l,t} - Z_{p,q,b,d,t}).
$$

The first term calculates generated income based on the number of chips sold and the price per chip by taking the sum over periods, families, packages, bins, and demand locations. The second term has a negative sign indicating that it represents a penalty in the model. It finds the number of chips in demand that could not—or would not—be produced and forces the model to pay a penalty for it.

2) Costs for Fab Production, Back-End Fab Inventory, Fab-to-Sort Transportation, and Front-End Sort Inventory:

$$
\begin{aligned}
&-\sum_{t,p,l\in L_F}\left(PC_{p,l,t}X^F_{p,l,t}+IC_{p,l,t}W^{AF}_{p,l,t}\right)\\
&-\sum_{t,p,l\in L_F,d\in L_S}TC_{l,d,t}Y^{FS}_{P,l,d,t}-\sum_{t,p,f\in L_F,l\in L_S}IC_{p,l,t}W^{BS}_{f,p,l,t}.
\end{aligned}
$$

The first term contains fab production costs and costs for storing wafers directly after production. The inventory levels are calculated by summing the net flow of product at the back end of the fabs over previous time periods. The second term represents the transportation costs between fab and sort facilities, and the third term accounts for inventory costs before sort facilities.

3) Costs for Sort Production, Back-End Sort Inventory, Sort-to-Assembly Transportation, and Front-End Assembly Inventory:

$$
\begin{aligned}
&-\sum_{t,p,f\in L_F,l\in L_S}\left(PC_{p,l,t}X^S_{f,p,l,t}+IC_{p,l,t}W^{AS}_{f,p,l,t}\right)\\
&\quad-\sum_{t,p,f\in L_F,l\in L_S,d\in L_M}TC_{l,d,t}Y^{SM}_{f,p,l,d,t}\\
&\quad-\sum_{t,p,f\in L_F,l\in L_M}IC_{p,l,t}W^{BM}_{f,p,l,t}.
\end{aligned}
$$

The first term is for sort production costs and back-end inventory costs, the second one is for transportation costs between sort and assembly, and the third one is for front-end assembly inventory or storage costs.

4) Costs for Assembly Production, Back-End Assembly Inventory, Assembly-to-Test Transportation, and Front-End Test Inventory:

$$
-\sum_{t,p,q,f\in L_F,l\in L_M}\left(PC_{p,l,t}X^M_{f,p,q,l,t}+IC_{p,l,t}W^{AM}_{f,p,q,l,t}\right).
$$

We see production costs and back end inventory costs for the assembly operation. Note that the summation occurs over package and family as opposed to family only for fab and sort production

$$
\begin{aligned}
&-\sum_{t,p,q,f\in L_F,l\in L_M;d\in L_T}TC_{l,d,t}Y^{\mathrm{MT}}_{f,p,q,l,d,t}\\
&\qquad\qquad-\sum_{t,p,q,f\in L_F,l\in L_T}IC_{p,l,t}W^{BT}_{f,p,q,l,t}.
\end{aligned}
$$

These two are assembly-to-test transportation costs and front-end inventory cost for the test operation.

5) Costs for Test Production, Back-End Test Inventory, Test-to-Demand Center Transportation, and Demand Center Inventory:

$$
\begin{aligned}
&-\sum_{t,p,q,f\in L_F,l\in L_T}PC_{p,l,t}X^T_{f,p,q,l,t}\\
&\quad-\sum_{t,p,q,b,f\in L_F,l\in L_T}IC_{p,l,t}W^{AT}_{f,p,q,b,l,t}\\
&\quad-\sum_{t,p,q,b,f\in L_F,l\in L_T;d\in L_D}TC_{l,d,t}Y^{TD}_{p,q,b,l,d,t}\\
&\quad-\sum_{t,p,q,b,f\in L_F,l\in L_D}IC_{p,l,t}W^{BD}_{f,p,q,b,l,t}.
\end{aligned}
$$

Notice that the first term (test production) is summed over family and package but not over bin, and the last three terms are summed over bin. This is because the number of products in each bin is not a decision variable but merely a result of the quality of the products described by the user in the matrix parameter $Q_{f,p,q,b,l,t}$. Thus, the decision variable for test production only tells how much of each package to be produced.

6) Costs for Building Facilities, Operating Facilities, and Removing Facilities:

$$
-\sum_{t,l\in L}\Omega^A_{lt}PBC_{lt}-\sum_{t,l\in L}\Omega_{lt}POC_{lt}-\sum_{t,l\in L}\Omega^R_{lt}PRC_{lt}.
$$

The costs to build a new facility and to operate it once it is built are included in the first two terms, and the costs to remove a facility are included in the last term.

7) Costs for Installing and Operating Tools:

$$
-\sum_{t,l\in L,i\in\mathrm{MT}}M^A_{ilt}MPC_{ilt}-\sum_{t,l\in L,i\in\mathrm{MT}}M_{ilt}MOC_{ilt}.
$$

The installation and operating costs are accounted for in these two terms.

8) Costs for Subcontracting:

$$
-\sum_{p,l\in L_A,t}SC_{plt}S_{plt}.
$$

Subcontracting costs in this model are on a dollar per wafer basis and do not take costs associated with setting up contracts into consideration.

E. Constraints

The LP model by Stray [15] contains network flow constraints, capacity constraints, bin allocation constraints, and demand constraints. The network flow constraints, described in two parts, enforce the material flow conservation, i.e., total inflow is equal to total outflow. The first part deals with balance of flow between outflow of products from a facility and the shipment of products to the next facility, while the second part deals with balance of flow between the inflow of materials into a facility and the amount of products started for production. The capacity constraints limit the total time available for production.

The bin allocation constraints take into account all products sold are from a given group of interclassifiable sellable chips. The demand constraint enforces the total sales to be less than or equal to the forecasted demand.

In addition to the constraints in the LP model, constraints are added that ensure practical feasibility of the mixed integer solutions. The capacity constraints of the linear programming model are altered. Facility indicator constraints and tool counting constraints are needed to find the correct status and capacity of the facilities. Production suppressing constraints and constraints that incur limits on the number of tools are also important parts of the model.

1) Production Suppressing Constraints: To avoid production in a nonexistent plant, a large number N multiplied by the existence indicator variable must be greater than or equal to the production variable X. Note that this ensures that for X to be a positive number, the indicator variable must be equal to one. The method is shown by Murty [14]

$$\Omega_{lt}N \geq X^F_{plt} \quad : \forall p \in \mathrm{FAM}, l \in L_F, t \in 1, \ldots, T.$$

A constraint like this one exists for all production variables; the only difference between them is that indexing over the originating wafer fab is done for sorts, assemblies, and tests. Currently, all facilities are allowed to be removed, hence all facilities are subject to this constraint; however, if for some reason that option is excluded for one of the facilities, the variable in question can be fixed by forcing it to always be equal to one and removing the appropriate constraints.

2) Capacity Constraints: The total time needed to produce the fraction of lots started at the current time period and the fraction of lots carried over from previous time periods should be less than the total available time on machines

$$\sum_{p \in \mathrm{FAM}} T_{ipl}\left(X^F_{pl,t-1}C_{pl,t-1} + X^F_{plt}(1 - C_{plt})\right) \leq \alpha_{il}M_{ilt}(TPL - S_{il}) \quad : \forall i \in \mathrm{MT}_l, l \in L_F, t \in 1, \ldots, T.$$

The previous constraint concerns itself with the capacity of wafer fabs. The other plant types are subjected to the same substitution and are shown at the bottom of page.

3) Facility Counting Constraints: The equations that follows guarantee that the status of the indicator variables are correct. The model sets the variable Ω^A_{lt} equal to one if a plant is to be built. Facilities that are assumed to exist from the beginning are also treated as variables, but the cost of adding them is zero

$$\Omega_{lt} = \sum_{r=1}^{t}\left(\Omega^A_{lr} - \Omega^R_{lr}\right) : \Omega_{lt}, \Omega^A_{lr}, \Omega^R_{lr} \in \{0,1\}, t \in 1, \ldots, T, l \in L.$$

4) Tool Counting Constraints: The constraint that follows counts the number of machines in each tool group in each plant that are up and running in time period t. The variable on the left-hand side also resides in the capacity constraints, indicating the number of tools available in each tool group. The model does not include all tool groups of the factories, but rather a selection of the most constraining ones (generally three to five groups). There are many procedures for finding these tool groups, one of them shown by Chang *et al.* [7]

$$M_{ilt} = \sum_{r=1}^{t}\left(M^A_{ilr} - M^R_{ilr}\right) + M_{il} : \forall i \in \mathrm{MT}_l, l \in L, t \in 1, \ldots, T.$$

5) Number of Tools Limiting Constraints: The two constraints that follow set a limit on the number of machines in each tool group and a limit on the total number of machines in a facility

$$M_{ilt} \leq \mathrm{MAX}^S_{il} \quad : \forall i \in \mathrm{MT}_l, l \in L, t \in 1 \ldots T$$
$$\sum_{i \in \mathrm{MT}_l} M_{ilt} \leq \mathrm{MAX}^T_l \quad : \forall l \in L, t \in 1 \ldots T.$$

IV. Computational Results and Validation

A. Break-Even Analyses for Strategic Decisions

1) Setting Up Scenarios: We examine the model to provide good support for real-world decision makers concerned with questions like facility building, subcontracting, and investing in tools. Real-world data has been used and the solutions have been

$$\sum_{p \in \mathrm{FAM}}\left(T_{ipl}\sum_{f \in L_F}\left(X^S_{fplt-1}C_{plt-1} + X^S_{fplt}(1 - C_{plt})\right)\right) \leq \alpha_{il}M_{ilt}(TPL - S_{il}) \; : \forall\, l \in L_S, i \in \mathrm{MT}_l, t \in 1, \ldots, T$$

$$\sum_{\substack{p \in \mathrm{FAM} \\ q \in \mathrm{PKG}_p}}\left(T_{ipl}\sum_{f \in L_F}\left(X^M_{fpqlt-1}C_{plt-1} + X^M_{fpqlt}(1 - C_{plt})\right)\right) \leq \alpha_{il}M_{ilt}(TPL - S_{il}) \; : \forall\, l \in L_M, i \in \mathrm{MT}_l, t \in 1, \ldots, T$$

$$\sum_{\substack{p \in \mathrm{FAM} \\ q \in \mathrm{PKG}_p \\ b \in \mathrm{BIN}_{pq}}}\left(T_{ipl}\sum_{f \in L_F}\left(X^T_{fpqlt-1}C_{plt-1} + X^T_{fpqlt}(1 - C_{plt})\right)\right) \leq \alpha_{il}M_{ilt}(TPL - S_{il}) \; : \forall\, l \in L_T, i \in \mathrm{MT}_l, t \in 1, \ldots, T$$

TABLE I
PARAMETER LEVELS FOR MIP MODEL

Parameter	Notation	Level	Unit
Length of a time-period	TPL	2184	Hours
Plant building costs: Fab	PBC_{lt}	1000	Million dollars
Sort	PBC_{lt}	100	
Assembly	PBC_{lt}	200	
Test	PBC_{lt}	200	
Plant operating costs: Fab	POC_{lt}	50	Million dollars per year
Sort		10	
Assembly		20	
Test		20	
Tool purchase costs:	MPC_{ilt}		
Fab: Steppers		7	Million dollars per tool
Implanters		1	Million dollars per tool
Non-bottleneck tools		35	Million dollars per stepper
Assembly tools		1	Million dollars per tool
Test: Testers		4	Million dollars per tool
Tool operating costs	MOC_{ilt}	10	% of purchase costs/year
Time lag between building decision and full capacity operability: Fab	LBT_l	18	Months
Assembly		12	
Test		12	
Tool efficiency	$\alpha_{i,l}$	70	% of total available time
Tool downtime	$S_{i,l}$	60	Hours / time-period
Total time for a product needed on a bottleneck tool:	$T_{i,p,l}$		
Fab		1.2	Hours per lot of 25 wafers
Test		0.1	Hours per chip
Total available time on a tool	$TPL - S_{i,l}$	2184 minus tool downtime	Hours
Cycle times at high load of facilities: Fab	Used to calculate	45	Days
Sort	fractions:	2	
Assembly	$C_{p,l,t}$	5	
Test	$C_{p,q,l,t}$	5	
Transportation cycle times:			Days
Fab to Sort		1	
Sort to Assembly		1	
Assembly to Test		0	
Test to Demand center		3	
Product yield of plants: Fab		96	%
Test		96	
All others		100	
Production costs (consumables costs)		1	% of operating costs of the tools
Inventory costs		6	Dollars per wafer per month
Wafer lot size		25	Wafers with 200 die each
Die lot size		1000	Die
Revenue for sold products		10K ± \$2.5K to \$100K ± \$25K	Dollars per wafer
Demand level		Varied	

reviewed by industry experts. To see how the model works, different scenarios are set up to see what break-even points there are for deciding to buy new tools, outsource wafers to foundries, and build new production facilities under different conditions. The point of doing this includes being able to examine the stability of the model's solutions, determine whether the model provides realistic answers, and whether those answers could be used in strategic planning. Looking for break-even points could possibly provide an inherent sensitivity analysis in assessing the possible future profit of the semiconductor manufacturer.

In this paper, break-even analyses will be carried out for situations where demand ranges from very low to very high. By very low, it is meant that the production facilities are running with substantial excess capacity. By very high, it is meant that the production facilities are running with too little capacity. Practically, this means setting the parameter demand level $(D_{p,q,b,l,t})$ to an even level across all periods and then solving the problem. The solution is then analyzed and the number of tools bought, wafers subcontracted, and what facilities where built are noted. The model is rerun with a different demand level and the same characteristics are noted. After running enough problem instances with different demand levels, curves are generated and examined to see what the demand level is that makes the model go from current capacity to subcontracting, from subcontracting to buying new tools, and from buying tools to buying entire wafer fabs, assemblies, and test facilities.

The foundation of the data consists of estimates for the parameters of the MIP model. The data are listed in Table I.

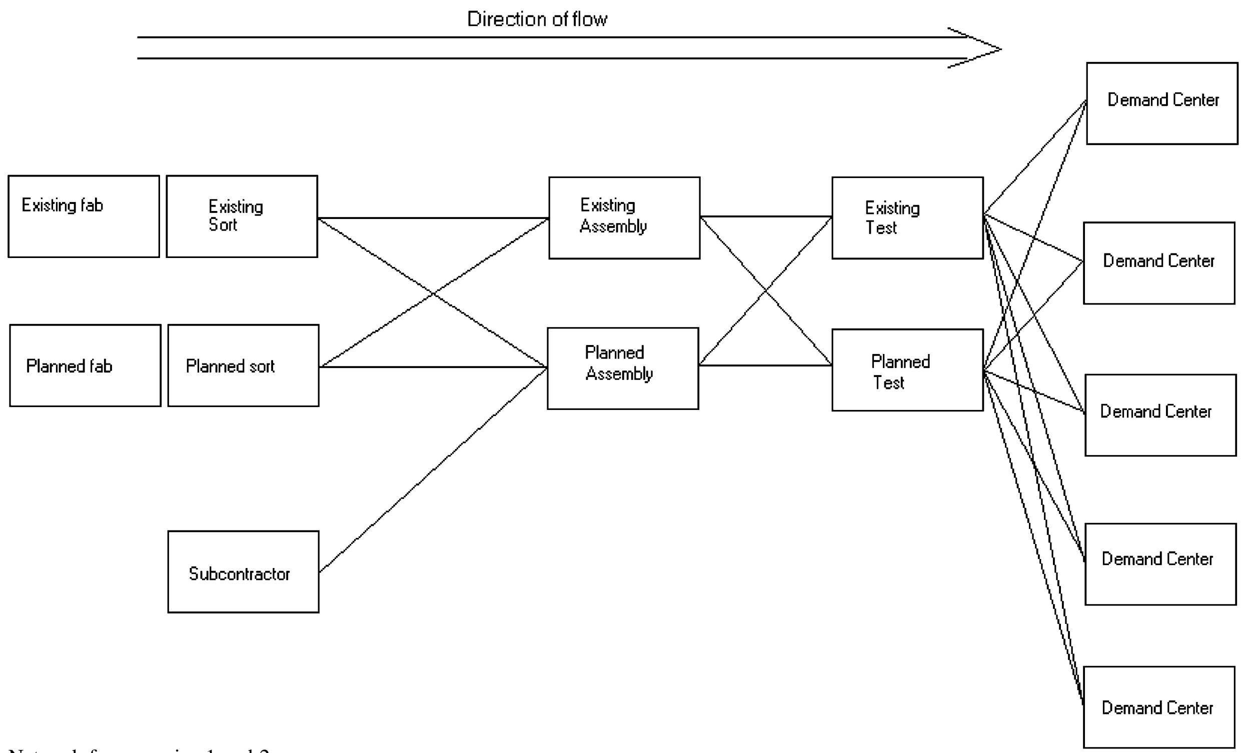


Fig. 1. Network for scenarios 1 and 2.

The network for the first two break-even analysis scenarios is shown in Fig. 1. In the first two scenarios, one fab is already up and running, together with one sort, one assembly, and one test facility and five demand centers. Capacity in this model is balanced so that each individual facility maximum capacity matches the maximum capacity of a single fab, preceding it in the manufacturing supply chain. One sort facility can thus handle the output from one fab, and one assembly can handle the output from a fab, and so on.

There is also one foundry available with unlimited capacity in the first of the two scenarios and limited in the second. The product from the foundry is ready for the assembly operation. The price of wafers from the foundry is higher than the cost of producing them in-house, as long as existing capacity is utilized. However, if a fab has to be built, the cost per unit will increase. When the costs are high enough, subcontracting becomes an interesting option. All fabs (existing and planned) have the same maximum capacity per period. The maximum capacity is defined as the capacity of a facility when the allowed maximum number of tools has been installed. Demand is varied evenly over the five demand centers. There are two product families in the model, each divided into two packages. In the binning process, two different qualities result from each package. Fabs are considered to have two bottleneck tool-groups while sort, assembly, and test have one each. In the third scenario, there are only two demand centers (instead of five).

2) Results for Break-Even Scenarios:

a) Ten-period unlimited subcontracting scenarios: Since these problems are NP-hard we will allow feasible solutions that are provably within 5% of optimum. Fig. 2 shows the optimal solutions (within 5%) for 20 different demand settings (varied along the x axis) with subcontracting of wafers unlimited. For each demand scenario, the MIP was run using AMPL [8]. The chart shows that production matches demand at all locations except at the foundry from the first demand scenario onward. The foundry starts producing wafers for the demand scenario with 320 million chips and increases its sales for each new scenario until the scenario with 600 million chips is reached. For the demand scenarios between 280 and 320 million chips, the initial test and assembly facilities run out of capacity, even though new tools have been installed, and immediately the model decides to build one more of each. Up to the demand scenario with 320 million chips, the manufacturing supply chain has been able to completely meet demand and would be able to do so up to 600 million chips if the new facilities could be used instantly. However, this is not the case, and that is why the curve for unmet demand starts to rise at the 320 and not at 600 million chips. At 600 million chips the existing two test and two assembly operations operate at full capacity, so no further subcontracting is useful.

The model chose not to build any fabs since subcontracting was available at reasonable cost to cover demand and since only one test and one assembly facility can be built. For this set of scenarios, there is not enough time (ten periods) to pay back the difference in costs of building a fab versus subcontracting no matter high the demand is. The jump in foundry production at the last step is a result of the acceptance of a suboptimal solution by the solver. From the solution file of this specific run it can be seen that some fab tools were sold in period 10 to decrease operation costs, thus partly compensating for the cost of subcontracting, enough to stay within 5% of optimality.

b) Ten-period limited subcontracting scenarios: The second set of runs was made with the same data but restricted foundry production. The restriction lies at 7 million chips (counting 200 chips per wafer), totaled over all products per period. The results are shown in Fig. 3. The only significant

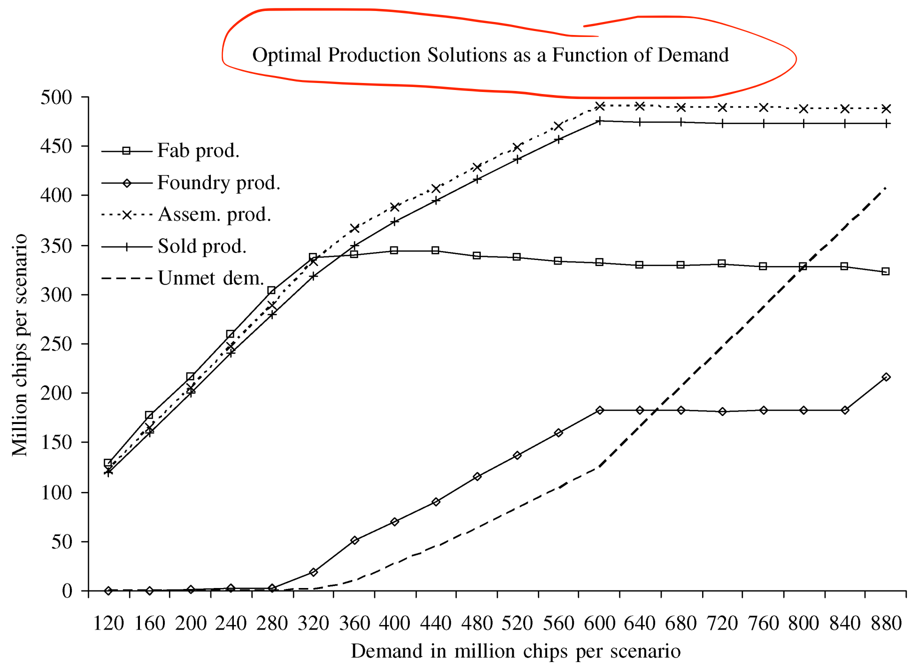


Fig 2. Optimal production solutions with unlimited subcontracting.

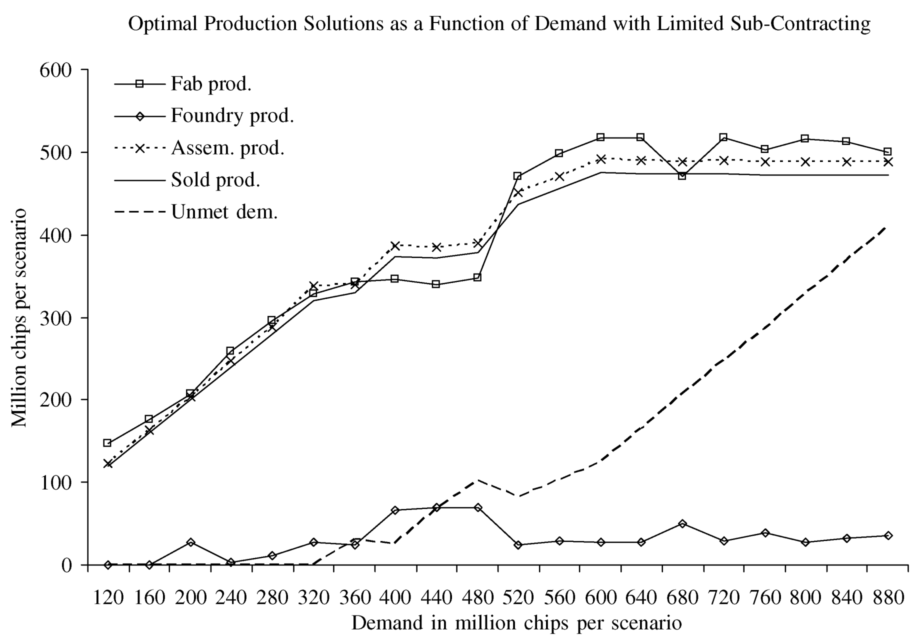


Fig. 3. Optimal production solutions with limited subcontracting.

difference between the results for this scenario and the previous is that in this one a fab is built. This rules out the possibility that building a fab cannot be justified during only ten time periods and shows that the decision of building the fab competes against the cost of subcontracting.

The fab is built at 520 million chips, 200 million more than the maximum capacity of a fab, 320 million chips. After the fab is built, foundry production goes down and stays down comfortably under the allowed maximum. The reason why foundry production is not zero is that it covers for the production capacity that is lost during the building of the wafer fab. The dip in in-house production and increase in subcontracting at the 680 million chips demand scenario can only be explained by the 5% optimality criterion, i.e., accepting the solution even though not quite optimal. The solution at 640 million chips consisted of shutting down one of the fabs in the last time period, replacing its capacity with subcontracting, saving the operation costs for the facility and its tools, and spending the money on the subcontracted wafers.

c) Sixteen–period unlimited subcontracting scenarios: The third scenario was run over 16 time periods with a network consisting of only two demand centers; otherwise, it is the same as the two first scenarios. The number of different products was reduced to one family, one package, and two bins. Costs were

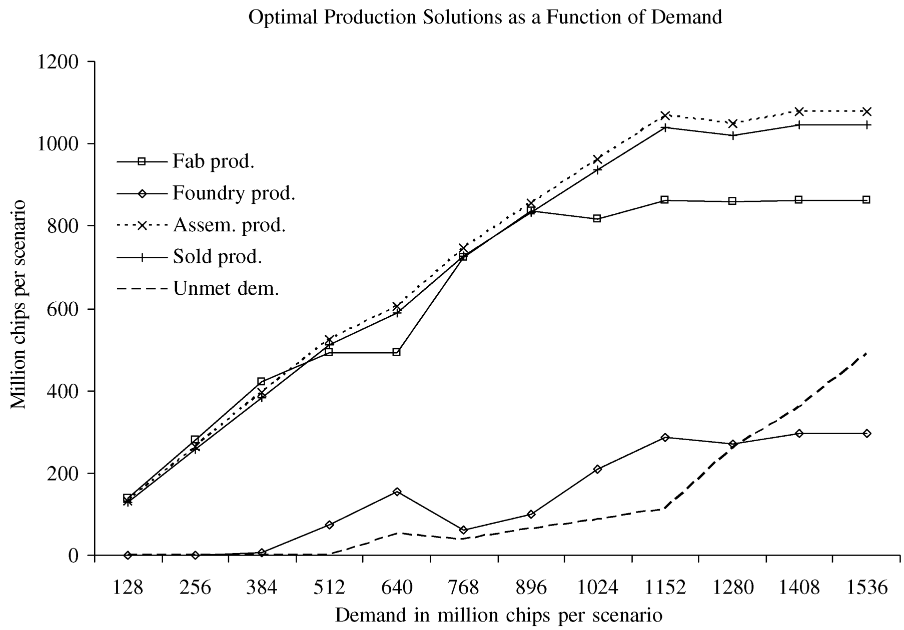


Fig. 4. Optimal production solutions for 16 periods.

kept the same but revenue per chip was reduced. The results are summarized and conclusions are drawn from the graph in Fig. 4. The option of subcontracting in this scenario was unlimited; nevertheless, the model selects the option of building a fab accompanied by a sort facility when demand reaches 768 million chips for the entire scenario. One might argue that the model does not choose to do this at the same level of demand in the ten period scenarios, but the answer to this argument lies in the number of periods available for payoffs. In the ten-period scenarios, a new fab can produce wafers for a maximum of 5 periods. From the maximum capacity of the fab (340 million chips for an entire scenario), it can be deduced that 170 million chips could be produced by this fab over these five periods. In the 16-period scenario, the new fab could produce as many as 374 million, the maximum capacity being 544 million for an entire scenario. A total 170 million chips are not enough for the fab to make up for its associated costs compared to the price of wafers from the available foundry; however, 340 millions chips are enough. This clearly shows that the longer the time horizon, the less conservatively the model acts.

V. Conclusions and Future Research

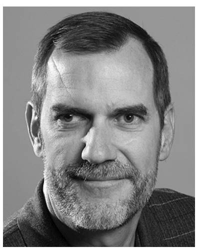
The cost of competing in the semiconductor industry continues to increase. In this paper, we present an MIP model that maximizes sales revenue minus production, transportation, and acquisition costs of a semiconductor manufacturing enterprise subject to demand and capacity constraints. Decision variables include what to produce and in what quantity, where it should be produced (including subcontracting), what equipment should be purchased (or sold), and what facilities should be built (or closed). In this paper, all values were treated deterministically. There is a need for research into how to best model situations where the demand is not known with certainty.

References

[1] Y.-F. Hung and Q.-Z. Wang, "A new formulation technique for alternative material planning—An approach for semiconductor bin allocation planning," *Computers Indust. Eng.*, vol. 23, pp. 281–297, 1997.
[2] M. T. Tabucanon and J. L. Ong, "Multiproduct, multistage machine requirements planning models," *Appl. Math. Modeling*, vol. 17, pp. 163–167, 1993.
[3] C. E. Love and M. Turner, "Note on utilizing stochastic optimal control in aggregate production planning," *Eur. J. Oper. Res.*, vol. 65, pp. 199–206, 1993.
[4] S. Shirodkar, "A modular approach for modeling and simulating semiconductor supply chains," M.S. thesis, Dept. Industrial Eng., Arizona State Univ., Tucson, 1999.
[5] Y.-F. Hung and R. C. Leachman, "A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations," *IEEE Trans. Semiconduct. Manufact.*, vol. 9, no. 2, pp. 257–269, May 1996.
[6] J. M. Padillo, R. Ingalls, and S. Brown, "A strategic support system for supply network design and management in the semiconductor industry," *Computers Indust. Eng.*, vol. 29, pp. 443–447, Sep. 1995.
[7] Y. S. Chang, M. Carlyle, J. Fowler, and S. H. Shin, "Planning for multipleflow semiconductor manufacturing systems," in *Proc. Int. Conf. Semiconductor Manufacturing Operational Modeling Simulation*, San Francisco, CA, Jan. 1999, pp. 195–200.
[8] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL a Modeling Language for Mathematical Programming*. San Francisco, CA: Scientific, 1993.
[9] T. Sebo, "A long range production/resource planning system using linear programming," in *Proc. 8th Annu. Conf. Computers Industrial Engineering*, 1986, pp. 180–183.
[10] K. E. Kendall and M. J. Schniederjans, "Multi-product production planning: A goal programming approach," *Eur. J. Oper. Res.*, vol. 20, pp. 83–91, 1985.
[11] D. Kira, M. Kusy, and I. Rakita, "A stochastic linear programming approach to hierarchical production planning," *J. Oper. Res. Soc.*, vol. 48, pp. 207–211, 1997.
[12] M. A. Bakir and M. D. Byrne, "Stochastic linear optimization of an MPMP production planning model," *Int. J. Oper. Res.*, vol. 55, pp. 87–96, 1998.
[13] M. D. Byrne and M. A. Bakir, "Production planning using a hybrid simulation-analytical approach," *Int. J. Production Economics*, vol. 59, pp. 305–311, 1999.
[14] K. G. Murty, *Operations Research: Deterministic Optimization Models*. Englewood Cliffs, NJ: Prentice–Hall, 1995.
[15] J. Stray, J. W. Fowler, C. W. Carlyle, and A. P. Rastogi, "Enterprise-wide semiconductor manufacturing and logistics planning," ASU Working Paper Series 2002-ORPS-001.

Jonas Stray received the M.Sc. degree in industrial engineering from Arizona State University, Tempe, in 2000.

He is currently a Lecturer in quality engineering and operations research at the University College of Boras (UCB), Sweden. He is also Head of the Industrial and Mechanical Engineering Department at the School of Engineering, UCB. Prior to joining UCB, he held manufacturing and quality manager positions in industrial companies. His research interests include industrial statistics for quality engineering.



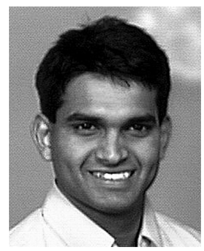
John W. Fowler (M'97) received the Ph.D. degree in industrial engineering from Texas A&M University, College Station, in 1990.

He spent the last 1.5 years of his doctoral studies as an Intern at Advanced Micro Devices. He is currently a Professor in the Operations Research and Production Systems group, Industrial Engineering Department, Arizona State University (ASU), Tempe. He is also the Center Director for the Factory Operations Research Center. Prior to joining ASU, he was a Senior Member of Technical Staff at SEMATECH. His research interests include modeling, analysis, and control of semiconductor manufacturing systems.

Dr. Fowler is the Area Editor for Manufacturing of the Society for Computer Simulation International, an Associate Editor of the IEEE TRANSACTIONS ON ELECTRONICS PACKAGING MANUFACTURING, and an Area Editor for *Planning and Scheduling of Computers* and *Industrial Engineering*. He serves on the Board of Directors of the Winter Simulation Conference and was recently named a Fellow of the Institute of Industrial Engineers.

W. Matthew Carlyle received the B.S. degree in industrial engineering from the Georgia Institute of Technology, Atlanta, and the Ph.D. degree in operations research from Stanford University, Stanford, CA.

He is currently an Associate Professor in the Operations Research Department, Naval Postgraduate School. He joined the faculty in 2002 after five years as an Assistant Professor in the Department of Industrial Engineering at Arizona State University. His research and teaching interests include network optimization, integer programming, and network interdiction. Applications of this research have included attack and defense of critical infrastructure, delaying large industrial projects and weapons programs, theater ballistic missile defense, sensor mix and deployment, naval logistics, communications network diversion, underground mining, and semiconductor manufacturing.



Aditya P. Rastogi received the M.S. degree in industrial engineering from Arizona State University, Tempe, in 2001.

He is currently working as Senior Optimization Analyst at K2B Inc., Overland Park, KS. His research interests include statistical forecasting, retail supply chain optimization, and financial analysis.

Mr. Rastogi served as Joint Editor for an online newsletter, ORMS Tomorrow, for two years. He achieved a Second Place IIE Graduate Research Award, in 2002, for the research work carried out during his M.S. thesis.