Proceedings of the 3rd Annual
IEEE Conference on Automation Science and Engineering
Scottsdale, AZ, USA, Sept 22-25, 2007

Capacity Planning for Semiconductor Wafer Fabrication with Uncertain Demand and Capacity

NA GENG and ZHIBIN JIANG, *Member, IEEE*

Abstract—**Capacity planning, the determination of optimal tools' configuration based on the current information about future demand and capacity, is very hard due to the high uncertainties in both capacity and demand. To cope with this problem, this paper proposes a scenario-based stochastic programming model to facilitate tool procurement decision-making. The uncertainty of product demand is represented with a set of scenarios with associated occurrence possibilities. And the capacity variability is described with a set of scenarios on tools' utilization ratios. All of these make the tools procurement plan robust to uncertainties to specific context. The resulting robust tools deals well with the changes in demand and capacity with the maximal expected profits. For those companies of risk-aversion, this model is preferred.**

Index Terms— **Semiconductor manufacturing, capacity planning, stochastic planning, uncertainty, optimization**

I. Introduction

PRODUCTION capacity, the tool configuration to meet the future demand, is the most important part of capital investment in semiconductor wafer fabrication (fab). The capacity invest is a billion dollars or more [1] and the cost has been on arise [2], [3]. Fordyce and Sullivan [2] regard the purchase and allocation of tools based on a demand forecast as one of the most important issues for managers of wafer fabs. Underestimation or overestimation of capacity will lead to low utilization of equipment or the loss of sales. Therefore, capacity planning, the determination of the optimal tool configuration based on the current information of demand and capacity, is very important for corporate performance. In semiconductor industries, a few reasons that make this problem especially difficult [1], [4]-[6] include rapid changes of technology and products, long lead time and high cost for capacity increment, and high uncertainties in demand and capacity.

Capacity planning is traditionally implemented by a spreadsheet [7]-[9], discrete-event simulation [10] or linear programming [2], [11]-[14] which have limited themselves to planning for a known demand forecast. Compared to these determinate methods, stochastic programming method [1], [6], [15]-[19] can provide a tool set that should be better able to handle changes in the product mix, because it accepts the uncertainty. A tool set means the type and number of the tools needed to meet the future demand. Two-stage stochastic programming model is widely used in capacity planning, which makes some decisions with uncertainty first, and then take a recourse action to compensate for the first-stage decisions when the uncertainty is disclosed. For the capacity planning, the first stage is to make the tool procurement plan here and now without the information about future possible demand and capacity. The second is a recourse action to decide the production, inventory, and stock out / outsourcing plan to compensate for the decision made in the first stage after materializing the demand and capacity. The objective is always minimizing the expected cost across all the scenarios or maximizing the expected profit. And the uncertainty of product demand is usually represented with a set of demand scenarios with associated occurrence probability. Reference [1] shows the merit of stochastic model for capacity planning in a wafer fab by comparing the deterministic and stochastic decision with uncertain demand. Swaminathan addresses the single period tool procurement for a wafer fab under uncertain demand [5], with the objective of minimizing the expected stock-out costs due to lost sales across all demand scenarios. Slack based heuristic and greedy heuristic are proposed to reduce the running time. Swaminathan extends this problem to multiple periods using scenario-based demands in [15]. Hood divides the tools into the primary and the nonprimary tool group [1], which depends on whether the tools can be designated as the preferred tools for the operations it can perform. Ahmed compares the relative merit between two-stage and multi-stage stochastic programming [18]. Multi-stage stochastic programming is used to improve the flexibility of tool purchase in later time caused by new events. In most references, the capacity planning process only considers demand uncertainty, ignoring the capacity fluctuation.

Reference [6] and [19] describe a multi-stage stochastic program with recourse which accepts the uncertainties from demand and capacity for coordinating the capacity expansion decisions among multiple fabs. The objective is to minimize the capacity expansion cost, and the expected costs of production, reconfiguration, inventory, outsource, and capacity underutilization. The demand and capacity are represented with two separate sets of scenarios with some

Manuscript received March 19, 2007. This work was supported by Research Grant from National Natural Science Foundation of China (No. 50475027), National 863 research project (No.2006AA04Z128), and a Specialized Research Fund for the Doctoral Program of Higher Education (20040248052).

N. Geng and Z. B. Jiang are with Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Min Hang District, 200240 Shanghai, China. (Corresponding Author: Phone: (+86) 21-34206065, Fax: (+86) 21-3420 6065, e-mail: zbjiang@sjtu.edu.cn.)

occurrence possibility. In these two papers, the capacity and capacity expansion plan are represented with wafer starts per week. Since the detailed tools are not included in the formulation, extra work is needed to make the tool procurement plan. Moreover, the capacity is difficult to be described in wafer starts per week. The reasons are as follows: 1) Different types of wafer may require different process time. The capacity is different if the product mix changes. 2) Because of the high fluctuant demand, the production mix is hard to forecast. So it's difficult to determine the capacity scenarios in wafer starts per week.

Also incorporating the uncertainties from demand and capacity, this paper describes a two-stage stochastic programming model for capacity planning in a single wafer fab. The demand is described with wafer forecast in one period. The uncertainty in demand is described with a set of scenarios with associated occurrence possibilities. And the capacity is represented by machine hour. In semiconductor wafer fab, the capacity changes due to the tools' uncertain downtime, unavailable operator and preventive maintenance. So we propose a set of scenarios on tools' utilization ratio to describe the uncertainty in capacity. To maximize the profit, we derive the tool deployment plan for a single wafer fab. The remainder of the paper is organized as follows. In Section 2, the problem and formulation is described as a stochastic programming model. An experimental analysis of the proposed model is described in Section 3. Finally, Section 4 summarizes the paper.

II. Problem Statement and Formulation

The capacity planning process is modeled under the following assumption.

1) A tool group is defined as a set of identical tools, and thus the capacity of a tool group with *n* tools is *n* times the capacity of a single tool.
2) Capacity is expressed in machine hour, i. e. the number of tools multiplied by the available time (in hours) and the tools' utilization ratio. The tool's utilization ratio is assumed to be the time used for processing divided by the period length, which is variable because of the uncertain downtime, unavailable operator and preventive maintenance. So the uncertainty in capacity is described with a set of scenarios for tools' utilization ratios, s_2, with the corresponding probability.
3) Demand is described with wafer requirements in this period. Taking the yield into account, the pure demand divided by the yield is what the demand we need in this formulation. And the uncertainty from demand is represented with scenarios set s_1. One demand forecast in this set may become the materialized with some occurrence probability.
4) Two sources of uncertainties - demand variation and capacity fluctuation - are modeled, which are represented with scenarios sets s_1 and s_2 separately. The scenarios for the capacity planning *s* are obtained from all the scenarios sets $s_1 * s_2$ and the occurrence possibility equals the multiplication of capacity and demand scenarios' probabilities.
5) One type of tool can process several operations. But one operation is assumed to be processed by one type of tool group. So adding up the operations the wafer needed on one tool group, we can calculate the time that unit of wafer needed to be processed by every tool groups. The process time is zero when one type of wafer does not need the processing of some tool group.
6) Due to the high tools' procuring cost, tools' utilization goal is set as 0.9. There's penalty if the tool utilization doesn't reach this target.
7) Inventory and out of stock are allowed with no beginning and ending inventory. No outsource is permitted.

The model is as following.

(1) Indexes

i: the type of tools, $i = 1,2,\ldots,I$;

j: the type of wafers, $j = 1,2,\ldots,J$;

s_1: demand scenarios , $s_1 = 1,2,\ldots,S_1$;

s_2: capacity scenarios, $s_2 = 1,2,\ldots,S_2$;

t: period of time, $t = 1,2,\ldots,T$;

(2) Decision Variables

σ_{it} : the binary variable for capacity expansion for tool i in period t;

N_{it}: the number of tool i to be procured in period t;

$x_{js_1s_2t}$: the number of wafer j to be produced for scenario s_1s_2 in period t;

$I_{js_1s_2t}$: the inventory number of wafer j for scenario s_1s_2 in period t;

$O_{js_1s_2t}$: the stockout number of wafer j for scenario s_1s_2 in period t;

$U_{is_1s_2t}$: the amount of capacity usage that is required to achieve a target utilization of tool i for s_1s_2 in period t;

(3) Parametric Data

α_t: the coefficient for procuring the tools in period t;

π_{s_1} : the occurrence possibility of demand scenario s_1;

π_{s_2} : the occurrence possibility of capacity scenario s_2;

$\pi_{s_1s_2}$: the occurrence possibility of scenario $s_1s_2, \pi_{s_1} \times \pi_{s_2}$;

m_{ij} : the processing time of tool i to process one wafer j;

ρ_{is_2t} : the utilization ratio of tool i for scenario s_2 in period t;

A_{it} : the available capacity of tool i in period t in machine*hours;

n_{i0} : the original number of tool i;

d_{js_1t}: the demand of wafer j of scenario s_1 in period t;

l_{it} : lower bound of procuring tool i in period t;

u_{it} : upper bound of procuring tool i in period t;

c_{it}: the unit cost of tool i in period t;

c_{jt}^x:the unit process cost for wafer j in period t;

c_{it}^σ: the setup cost for procuring tool i in period t;

c_{it}^U: the under-utilization cost of tool i in period t;

c_{jt}^I : the unit inventory cost for wafer j in period t;

c_{jt}^O : the unit stockout cost for wafer j in period t;

p_{jt} : the unit price for wafer j in period t;

CPM :

$$
\begin{aligned}
Min\ & \sum_t\sum_i \alpha_t c_{it}^\sigma \sigma_{it} + \sum_t\sum_i \alpha_t c_{it} N_{it} \\
& + \sum_t\sum_{s_1s_2}\sum_j \pi_{s_1s_2}(c_{jt}^x - p_{jt})x_{js_1s_2t} + \sum_t\sum_{s_1s_2}\sum_i \pi_{s_1s_2} c_i^U U_{is_1s_2t} \\
& + \sum_t\sum_{s_1s_2}\sum_j \pi_{s_1s_2} c_{jt}^I I_{js_1s_2t} + \sum_t\sum_{s_1s_2}\sum_j \pi_{s_1s_2} c_{jt}^O O_{js_1s_2t}
\end{aligned}
$$

s. t.

$$l_{it}\sigma_{it} \le N_{it} \le u_{it}\sigma_{it} \qquad \forall i,t \qquad (1)$$

$$\sum_j x_{js_1s_2t} m_{ij} \le A_{it}(n_{i0} + \sum_{\tau=1}^t N_{i\tau})\rho_{is_2t} \qquad \forall i, s_1, s_2, t \qquad (2)$$

$$\sum_j x_{js_1s_2t} m_{ij} + U_{is_1s_2t} \ge 0.9 A_{it}(n_{i0} + \sum_{\tau=1}^t N_{i\tau}) \qquad \forall i, s_1, s_2, t \qquad (3)$$

$$x_{js_1s_2t} + I_{js_1s_2t-1} + O_{js_1s_2t} - I_{js_1s_2t} = d_{js_1t} \qquad \forall j, s_1, s_2, t \qquad (4)$$

$$I_{js_1s_2t} = 0 \qquad \forall j, s_1, s_2, t = 0, T \qquad (5)$$

$$N_{it}, x_{js_1s_2t}, I_{js_1s_2t}, O_{js_1s_2t}, U_{is_1s_2t} \in \mathrm{N}^+, \sigma_{it} \in \{0,1\} \quad \forall i, j, s_1, s_2, t \qquad (6)$$

The objective of this model, denoted by *CPM*, is to maximize the expected profit. If all the tools' procurement cost is considered in the objective, stockout maybe preferred for the very high equipments' cost. So weight is proposed for this cost. And it is well known to all that the technology and products change quickly in wafer manufacturing industries. The tools are preferred to buy later. So the coefficient is related to the time. For the earlier the period, the coefficient is larger.

For each period, constraint (1) simply states the capacity expansion limitation on the tool number where capacity may be increased by adding new tools. For every scenario in every period, constraint (2) indicates that for each tool group, the production time that the wafers production need is not greater than that the tool's usable time, whereas the usable time for one tool group is the multiplication of one tool's available time, the summation of the current tools and the procured tools from beginning to this period, and the tool's utilization ratio; constraint (3) ensures for each tool group, underutilization penalty will be paid if the tools' utilization ratio does not reach the utilization ratio target of 0.9. This constraint is based on the fact that the tools couldn't be utilized 100%; constraint (4) makes sure that for each type of wafer, the demand in every period equals the production for this period plus the inventory for last period plus the stockout number in this period minus the inventory for this period; constraint (5) expresses no inventory at the beginning and ending period. Constraint (6) imposes integrality and non-negativity conditions on the procured tools and wafer numbers for production, inventory, and stockout and imposes Boolean condition for the binary variable for capacity expansion.

The model is implemented in ILOG OPL Development Studio IDE 4.2, ILOG Company's product for solving linear and integer programming. It accesses the data from the EXCEL file. The run time varies widely with different tool types, product types, and scenarios number.

III. Model Execution

An illustration example is provided to test the quality of solution. The testing example contains 18 tool groups for one section of the semiconductor fab. The fab produces 9 major types of wafers. The demand uncertainty is represented with 4 scenarios with the occurrence possibility 0.2, 0.3, 0.3 and 0.2. As the merit of capacity planning with uncertain demand has shown in [1], we just focus on the benefit of accepting capacity fluctuation. Here we assume that all the tools have the same utilization ratio which is impossible for the real world. For each period, the utilization ratio may be 0.5, 0.6, 0.7 and 0.8 with the possibility of 0.2, 0.3, 0.3 and 0.2 respectively.

To test the quality of solutions, the result provided by *CPM* is compared with those obtained by the planning only for one utilization ratio. All of these tool sets are determined with the same uncertain demand. Let N_o denote the toolset solution by the *CPM* planning. Let N_1, N_2, N_3, N_4 denote the toolset solution by the planning based on the utilization of 0.5, 0.6, 0.7 and 0.8 with occurrence possibility = 1 separately. And N_E denotes the solution based on the expected utilization ratio, 0.65 which is obtained by (0.2×0.5 + 0.3×0.6 + 0.3×0.7 + 0.2×0.8=0.65). Table I and II show the tool procurement plan (N_o, N_1, N_2, N_3, N_4, and N_E) under the same demand scenarios for the next 3 periods with α_t = 0.3, 0.2, and 0.1 and α_t =0.6, 0.4, and 0.2 separately. As shown in both of the tables, we find that there is minor difference for the tools procured for the first and second period and major difference for the third period. Because the capacity planning is made every period, the decision in first period is what we need and the demand information for the next two periods just helps the decision in the first period. So we can come to the conclusion that α_t is not very sensitive to the tool procurement decision in the first period.

Table III and IV list the profit (the value of *CPM*) of different plans (N_o, N_1, N_2, N_3, N_4, and N_E) under different capacity status (utilization ratio= 0.5, 0.6, 0.65, 0.7, 0.8, and stochastic). For one capacity status, the profit reaches maximum when the tool procured under the same capacity status. And for one of the plans, the profit increases with the improving of the tools' utilization until new bottleneck occurs. The expected value of profit under the N_o decision is not always the maximal under all the capacity statuses. Even the profits under the N_2 and N_E outweigh N_o in most of the situations. However, these two decisions perform rather poor when the utilization ratio is 0.5. And the expected profits with N_o decision outperforms all the others. So for those companies of risk-aversion, the N_o is the best decisions; while for those of risk-seeking maybe prefer the N_2 or N_E.

TABLE I
STOCHASTIC AND CONSTANT TOOLS PROCUREMENT PLAN
(α_t = 0.3, 0.2, 0.1)

Tool Type	Period 1						Period 2						Period 3					
	N_o	N_E	N_1	N_2	N_3	N_4	N_o	N_E	N_1	N_2	N_3	N_4	N_o	N_E	N_1	N_2	N_3	N_4
1	0	0	0	0	0	0	0	0	1	0	0	0	2	0	1	0	0	0
2	1	0	1	0	0	0	0	0	1	1	0	0	2	1	1	0	0	0
8	1	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
12	1	1	1	1	1	1	0	0	1	0	0	0	1	0	0	0	0	0
13	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
15	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

TABLE II
STOCHASTIC AND CONSTANT TOOLS PROCUREMENT PLAN
(α_t =0.6, 0.4, 0.2)

Tool Type	Period 1						Period 2						Period 3					
	N_o	N_E	N_1	N_2	N_3	N_4	N_o	N_E	N_1	N_2	N_3	N_4	N_o	N_E	N_1	N_2	N_3	N_4
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
2	1	0	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0
8	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
12	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
13	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
15	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

TABLE III
EXPECTED PROFIT FOR THE TOOLS PROCUREMENT PLAN
UNDER DIFFERENT CAPACITY STATUS (α_t = 0.3, 0.2, 0.1)

Capacity Status	N_1 ($ Billion)	N_o ($ Billion)	N_2 ($ Billion)	N_E ($ Billion)	N_3 ($ Billion)	N_4 ($ Billion)
0.5	34.64	33.92	29.56	27.07	24.50	23.33
stochastic	34.98	35.21	34.78	34.04	33.08	31.82
0.6	35.07	35.53	36.05	35.08	33.48	31.15
0.65	35.07	35.53	36.11	36.20	35.92	33.54
0.7	35.07	35.53	36.11	36.20	36.27	35.04
0.8	35.07	35.53	36.11	36.20	36.29	36.50
EV	34.98	35.21	34.79	34.13	33.26	31.90

TABLE IV
EXPECTED PROFIT FOR THE TOOLS PROCUREMENT PLAN
UNDER DIFFERENT CAPACITY STATUS (α_t = 0.6, 0.4, 0.2)

Capacity Status	N_1 ($ Billion)	N_o ($ Billion)	N_2 ($ Billion)	N_E ($ Billion)	N_3 ($ Billion)	N_4 ($ Billion)
0.5	33.68	30.99	29.15	26.67	24.19	23.13
stochastic	34.02	34.68	34.37	33.67	32.73	31.62
0.6	34.11	35.56	35.64	34.61	32.96	30.95
0.65	34.11	35.62	35.70	35.88	35.30	33.34
0.7	34.11	35.62	35.70	35.90	35.98	34.84
0.8	34.11	35.62	35.70	35.90	36.03	36.30
EV	34.02	34.68	34.38	33.77	32.86	31.70

IV. CONCLUSION

In this study, we propose a two-stage stochastic programming model, which incorporates the demand and capacity uncertainties. These uncertainties are represented with scenarios with associated occurrence probabilities. The resulting capacity planning decisions are robust to the variance of capacity. Though not all the profits under every capacity status perform the best, the decision made by CPM exceeds all the other decisions in the expected profit. For those companies of risk-aversion, the CPM model is preferred.

REFERENCES

[1] S.J. Hood, S. Bermon, F. Barahona, "Capacity planning under demand uncertainty for semiconductor manufacturing," *IEEE Trans. Semiconductor Manufacturing*, vol. 16, no. 2, pp. 273-280, May 2003.

[2] K. Fordyce, G. Sullivan, "A dynamically generated rapid response capacity planning model for semiconductor fabrication facilities," in *Impact of Emerging Technologies on Computer Science and Operations Research*, Dordrecht, pp.103-127, 1994.

[3] D. L. Benavides, J. R. Duley and B. E. Johnson, "As good as it gets: optimal fab design and deployment," *IEEE Trans. Semiconductor Manufacturing,* vol. 12, no. 3, pp. 281-287, Aug. 1999.

[4] R. Kotcher, F. Chance, "Capacity planning in the face of product-mix uncertainty," in *1999 IEEE Int. Symp. Semiconductor Manufacturing Conf.,* Santa Clara, CA, USA. , pp. 73-76, 1999.

[5] J.M. Swaminathan, "Tool capacity planning for semiconductor fabrication facilities under demand uncertainty," *European Journal of Operations Research*, vol. 120, pp. 545-558, 2000.

[6] S. Karabuk, "Coordinating capacity decisions for the supply chain in high-tech industry," Ph.D. dissertation, Department of Industrial Engineering, Lehigh Univ., 2001.

[7] M. Baudin, V. Mehrotra, B. Tilis, D. Yeaman and R. Hughes, "From spreadsheets to simulations: a comparison of analysis methods for IC manufacturing performance," in *Proc. of 1992 IEEE/SEMI Int. Semiconductor Manufacturing Science Symp.*, pp. 94-99.

[8] J. D. Witte, "Using static capacity modeling techniques," in *Proc. of 1996 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, pp. 31-35.

[9] T. Occhino, "Capacity planning model: the important inputs, formulas and benefits," in *Proc. of 2000 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, pp. 455-458.

[10] A. M. Spence, D. J. Welter, "Capacity planning of a photolithography work cell," in *Proc. of 1987 IEEE Int. Conf. Robotics and Automation*, pp. 702-708.

[11] S. J. Hood, "Strategic capacity planning using linear programming," *Future Fab. Intl.*, vol.9, 2000. Available: http://www.future-fab.com/documents.asp?d_ID=1189 .

[12] R.C. Leachman, T. F. Carmon, "On capacity modeling for production planning with alternative machines," *IIE Trans.*, vol.24, no.4, pp. 62-72, 1992.

[13] S. Bermon, S.Hood, "Capacity optimization planning system (CAPS)," *Interfaces*, vol. 29, no. 5, pp. 31-50, 1999.

[14] J. Yang, "An approach to determine appropriate fab development plans by taking space constraints and cost-effectiveness into consideration", in *Proc. of 9th Int. Symp. Semiconductor Manufacturing*, pp. 217-220, 2000.

[15] J. M. Swaminathan, "Tool procurement planning for wafer fabrication facilities: a scenario-based approach," *IIE Trans.*, Vol. 34, pp. 144-155, 2002.

[16] F. Barahona, S. Bermon, O. Gunluk, and S. Hood, "Robust capacity planning in semiconductor manufacturing," *Naval Research Logistics*, vol. 52, no. 5, pp. 459 – 468, 2005

[17] B. Catay, S. S. Erenguc, A. J. Vakharia, "Tool capacity planning in semiconductor manufacturing," *Computers & Operations Research*, vol. 30, no.9, pp. 1349-1366, 2003.

[18] S. Ahmed, "Semiconductor tool planning via multi-stage stochastic programming," in *Proc. of 2002 Int. Conf. Modeling and Analysis of Semiconductor Manufacturing*, Tempe, Arizona, U.S.A. pp. 153-157.

[19] S. Karabuk, S. D. Wu, "Coordinating strategic capacity planning in the semiconductor industry," *Operations Research*, vol. 51, no. 6, pp. 839-849, 2003.