Economic Analysis and Optimization of Tool Portfolio in Semiconductor Manufacturing

Yon-Chun Chou and Chuan-Shun Wu

***Abstract*—The tool portfolio of a plant refers to the makeup, in quantity and type, of processing machines in the plant. It is determined by taking into consideration the future trends of process and machine technologies and the forecasts of product evolution and product demands. Portfolio planning is also a multicriteria decision-making task involving tradeoffs among investment cost, throughput, cycle time, and risk. Tool portfolio planning is a complex task that has strong bearing on manufacturing efficiency. In the first part of this paper, a multicriteria economic decision model is presented for optimal configuration of the portfolio and to determine the optimal factory loading. The second and third parts of the paper contain applications of the model. If plants are closely located or have a twin-plant design, portfolio planning at multiple plants can be integrated to enhance the overall effectiveness of portfolios. In the second part, a novel methodology for arbitrating capacity backup between plants is described. Because the economic model is constructed upon a valuation of both cycle time and throughput, it is a suitable method for the evaluation of cycle time reduction projects. The application procedure is outlined in the third part.**

***Index Terms*—Configuration design, cycle time reduction, resource portfolio planning.**

I. Introduction

The tool portfolio of a plant refers to the makeup, in quantity and type, of processing machines in the plant. Tool portfolio planning in the semiconductor industry is an important task that has strong bearing on manufacturing efficiency. Besides the implication of high investment cost of semiconductor processing tools, there are several other reasons. First, the tool portfolio is determined by taking into consideration the trends of process and machine technologies and the forecasts of products and product demands. Due to rapid changes in the technology environment, there is a high uncertainty of mismatch between actual demands and the right types of capacity. Because the market environment is volatile, there is a high risk of under- or over-capacity. The tool portfolio must be evaluated and fine-tuned on a continuous basis. Also, a wafer fabrication plant is a complex manufacturing system that contains several hundred machine tools and the manufacture of a product requires several hundred processing steps. Because the tools and storage systems in a plant are tightly interconnected, by automated material transport systems and by computer control, a wafer plant exhibits complex queuing network behaviors. Not only is its throughput complex to estimate, but there are other requirements such as cycle time and work-in-process (WIP) level which should be simultaneously taken into consideration in performance evaluation.

There are three important issues in portfolio planning, namely performance evaluation, configuration design, and risk analysis. A tool portfolio can be represented as an ordered set of tool quantities, $TP(=n_1, n_2, \ldots, n_k, \ldots)$, where k is the index for tool groups. The number of tool groups is in the order of one hundred. Tool groups are not always distinct. A tool group can be used as alternative tools for some other tools after setup operation. The solution space is very large for configuration design, especially when the tool portfolios of multiple time periods under multiple demand scenarios are to be determined [1], [13].

Portfolio planning requires a capacity model of the plant as a basic tool for performance analysis. Static capacity modeling, queuing capacity modeling, and simulation are three common modeling techniques of capacity analysis. Static capacity models are the most popular because of its ease of use. A static model may take into account all or part of the capacity factors: tool availability, tool efficiency, process yield, lead-time offset, and changeover time loss [3]. The availability, efficiency of serial tools, and process yield can be easily included in the calculation logic. The other factors are more involved and have been the focuses of several recent studies. The loading policy of batch machines has a strong effect on machine utilization and it has been studied extensively. Depending on whether job arrival information is utilized or not, loading policies can be classified as either static [5] or dynamic [6], [7], [14]. The efficiency of batch tools can be estimated using regression analysis of re-entry times and work release quantity or by formulas [3], [7]. The changeover loss can be estimated from historical data of the occurrence times of tool idleness [10]. The lead-time offset was shown to be a significant factor in computing workload [4]. Static models are very useful; the major shortcoming is the lack of flow time information.

Queuing models are used in the analysis of the steady state. They provide the flow time, utilization, and WIP performance information. Programs based on queuing capacity models require a very short time to run and can provide results with modest accuracy. It was reported that the accuracy of queuing models as compared with actual factory operation data was within 5% to 10% for throughput and within 10% to 30% for flow time [2]. Therefore, queuing capacity models are useful tools for capacity planning. Simulation is usually used to analyze operation dynamics at very detailed levels. In principle,

Manuscript received May 8, 2002; revised July 15, 2002. This work was supported in part by the Semiconductor Manufacturing Corp. under Grant 201-NJ-879 and the National Science Council, Republic of China, under Grants 89-2212-E-002-094 and 89-2213-E-002-117.

The authors are with Industrial Engineering, National Taiwan University, Taipei, Taiwan.

Digital Object Identifier 10.1109/TSM.2002.804885

0894-6507/02$17.00 © 2002 IEEE

simulation models can provide very accurate capacity information. However, their construction and maintenance require very significant effort, and their computer run time is usually long. They are not suitable for early stages of portfolio planning, in which many alternative portfolios need evaluating. Instead, they should be used to verify and fine-tune the final portfolio.

The importance of portfolio configuration under demand uncertainty has been well recognized in the literature. Because capacity requirements are dependent on the product mixes, Kotcher describes a heuristic procedure for identifying tool sets that are prone to become bottleneck tools as the product mixes are changed [8]. The loading of each tool set is first computed based on a fixed product mix. Tool sets with high utilization are selected for further sensitivity analysis under varying product mixes. These tools are called capacity-constraining tools. They are prioritized along with their triggering product mixes. Swaminathan addresses the tool procurement decision over a planning horizon of multiple time periods [13]. The uncertainty of product demand is modeled by a set of demand scenarios where each scenario reflects a possible set of demands for different products and each scenario is associated with a probability of occurrence. A large Integer Program model is described that minimizes the expected stock-out cost over all demand scenarios. The paper also presents two heuristics to generate upper bounds and two relaxation methods to generate lower bounds. A stochastic programming approach to capacity planning under demand uncertainty has also been described [1]. The uncertainty in demand is modeled by a set of demand scenarios. The tool purchase, wafer starts, and work assignment decisions are formulated as a very large mixed-integer program of 2500 integer variables, 230 000 continuous variables, and 140 000 constraints.

Portfolio planning is a multicriteria decision task involving tradeoffs among investment cost, throughput, cycle time, and risk. The decision is complicated. Not only are there multiple portfolios that will satisfy a specified set of production goals, but also that each portfolio can be operated in a multitude of load scenarios, yielding various combinations of multiple performance measures. There is little literature on this tradeoff analysis except by Ozawa [12]. Furthermore, most works reported in the literature deal with how to determine the tool portfolio using one method of workload allocation or another. However, in capacity planning, the planner would also be interested in knowing the risk that any capacity plan entails. There is also very little treatment of this risk issue in the literature.

The focus of this paper is on the multicriteria issue of portfolio planning. An optimization and economic analysis model and its applications are presented. The remainder of the paper is organized as follows. In Section II, a procedure to generate a multitude of feasible portfolios is described. Those portfolios constitute a solution space. In Section III, an economic decision model is presented for optimal configuration of the portfolio and to determine the optimal operation loading. In Section IV, a novel methodology for capacity sharing between plants is described. Finally, the procedure to apply the economic model to cycle time reduction projects is described in Section V, and conclusions are found in Section VI.

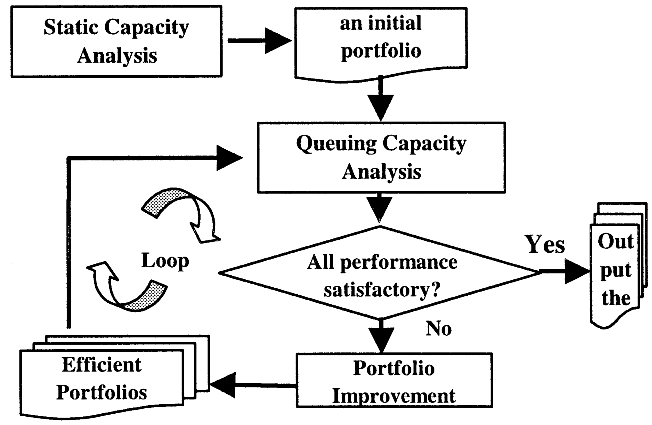

Fig. 1. Generating a solution space of efficient portfolios.

II. Generation of the Solution Space

In configuring a portfolio, marginal analyses of multiple performance measures can be used to incrementally adjust tool quantities [3], [12]. Suppose there is a set $\boldsymbol{F} = \{1, \ldots, \mathrm{F}\}$ of products to be produced and a set $\boldsymbol{K} = \{1, \ldots, \mathrm{K}\}$ of distinct tool groups. The raw input data for portfolio planning are the process routing for each product $f \in \boldsymbol{F}$, tool information for each tool group $k \in \boldsymbol{K}$, and product demand (D_f). The tool requirements (R_k) are computed based on processing time, machine availability, process yield, and other factors, but for brevity it can be expressed as

$$\boldsymbol{R_k} = \sum_{\boldsymbol{f}} \boldsymbol{D_f} \cdot \boldsymbol{W_{f,k}}$$

where $W_{f,k}$ is the standard processing time of product f on tool k. Fig. 1 is a flow diagram for a two-stage procedure we used to generate the solution space of portfolios. A static capacity model is first applied to generate an initial solution. Because the logic of static capacity models involves calculating the average workload, they provide the first order analysis of capacity requirements. In comparison, both first-order and second-order analyses are addressed in queuing analysis. (Queuing analysis requires second-order statistics such as the variance of job arrivals and job processing rates.) The resultant portfolio of static capacity analysis is a lower bound portfolio. In the second stage, the initial portfolio is evaluated using a queuing model to estimate its performance in throughput, flow time, and utilization. Based on the estimated performance, the portfolio is then modified by increasing the machine quantity of the bottleneck tool group. This improvement process continues for a number of iterations to collect a set of efficient portfolios. There is more than one strategy to improve any given portfolio. This will be discussed next.

The bottleneck of a plant is a very complex phenomenon. A common point of view is to take the stations with high utilization as the bottlenecks. However, as evident in many studies, there are other indicators or concepts of bottleneck that are rooted in the synergistic effects of machine availability, utilization, and maintainability [9], [11]. Because the available time of a tool can be divided between regular utilization (ρ), utilization due

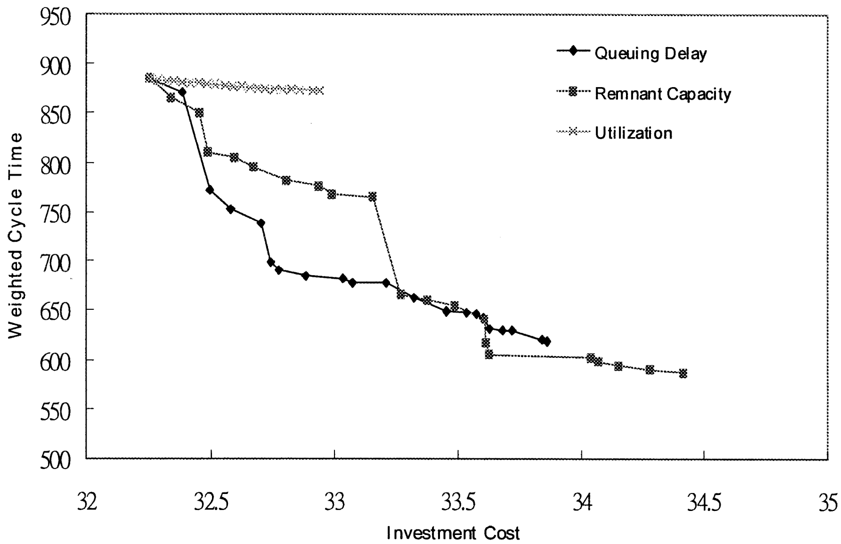

Fig. 2. Effectiveness of different bottleneck indicators.

to incapacitation (ρ^{inc}), and idleness, in this study three bottleneck indicators have been used to guide the search process of portfolio generation, namely utilization (regular plus incapacitation), queuing delay, and remnant capacity (rc), where the remnant capacity of a tool group g is defined as

$$rc_g = \begin{cases} \dfrac{1-\rho_g^{inc}}{\rho_g} - 1, & \text{for nonbatch tools} \\ \dfrac{(1-\rho_g^{inc})\cdot \text{MaxBatchSize}}{\rho_g \cdot \text{MeanBatchSize}} - 1, & \text{for batch tools.} \end{cases}$$

The remnant capacity indicator has a value greater than or equal to 1. It is a concept related to safety capacity buffer and a measure for the likelihood of being a bottleneck. (The lower the value, the more likely.) To compare their effectiveness, the three indicators have been used in the portfolio planning procedure (the second stage). Each indicator is used as a strategy to improve the current portfolio. Each strategy leads to a different set of portfolios. It should be noted that there are many other possible strategies. For example, given a current portfolio (TP_0), one can use the following cost-based formula to rank all tool groups (TP_i):

$$\text{Ratio} = \frac{\text{CycleTime}\,(TP_i) - \text{CycleTime}\,(TP_0)}{\text{Cost}\,(TP_i) - \text{Cost}\,(TP_0)} = \frac{\text{CycleTime}\,(TP_i) - \text{CycleTime}\,(TP_0)}{\text{Unit Cost of Tool}\,k}.$$

The results of this empirical study are partly shown in Fig. 2. Using each indicator, a series of portfolios were generated. In the figure, each point represents a portfolio for the same nominal output level. Points to the lower left can be considered as more cost effective. Three empirical conclusions can be drawn. First, because the criteria are different, it is not absolute that the series merge to the same portfolio. There is substantial "variation" between the response surfaces of the solution space. Second, the queuing delay and remnant capacity indicators are more computation-effective than the utilization indicator. Third, the queuing delay indicator is more effective than the remnant capacity indicator.

Although each portfolio in Fig. 2 is characterized by a cycle time and an investment cost, the performance level achievable by a portfolio is not a single value but a range of values. The cycle time as indicated in the figure is conditioned on a throughput that is equal to the nominal capacity. As the operation loading is varied, the cycle time will change accordingly. The cycle time under different throughput rates can be plotted in the first quadrant of Euclidean space with two dimensions: throughput and cycle time. Fig. 3 shows 20 plotted curves for those portfolios of Fig. 2 that are generated using the queuing delay indicator. Each curve represents the operation options of a portfolio. These curves will be called option curves and the space that all option curves lie in, i.e., the two-dimensional (2-D) Euclidean space depicted in Fig. 3 will be called the option space.

The above option curves are actually composed of discrete points. For convenience, regression analysis is applied to construct approximate option curve (OC) functions. It can be observed that option curve functions are consistent with the queuing delay phenomenon of general queuing networks. As the throughput is increased, utilization and cycle time will increase. We have found that the function form $(m - n\cdot x)^{-1}$ is a good fit for the OC(x). This can be explained by the fact that a wafer plant is a large queuing network and in queuing analysis the formulas (such as that of M/M/1) that relate queuing delay, service time, and utilization have this general form: queuing delay = (average service time)/(1-utilization).

III. A Decision Model for Portfolio Optimization

Portfolio planning is a multicriteria decision-making task involving tradeoffs among investment cost, throughput, and cycle time. In this section, we will present a methodology for determining the optimal portfolio. While the value of throughput is relatively easy to quantify, the benefit of cycle time is somewhat subjective and is dependent on the business situation. Utility function is a method used in economics to express the change in perceived value that is assigned to goods as its consumption quantity increases. Goods can be distinguished as regular

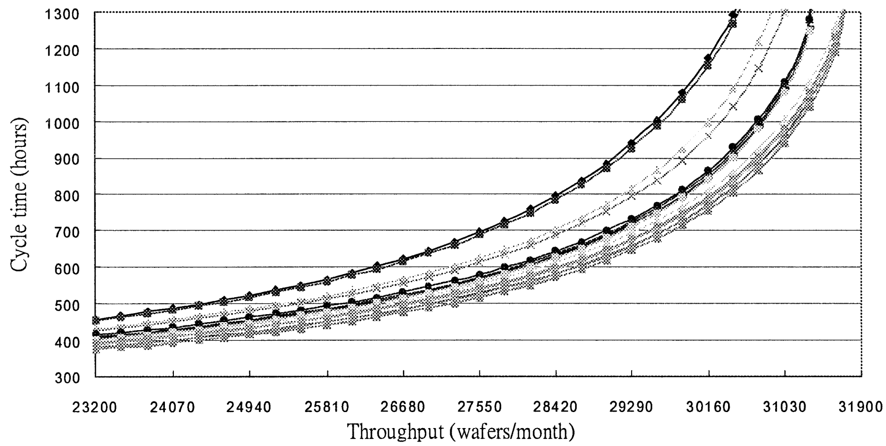

Fig. 3. Option curves in the option space.

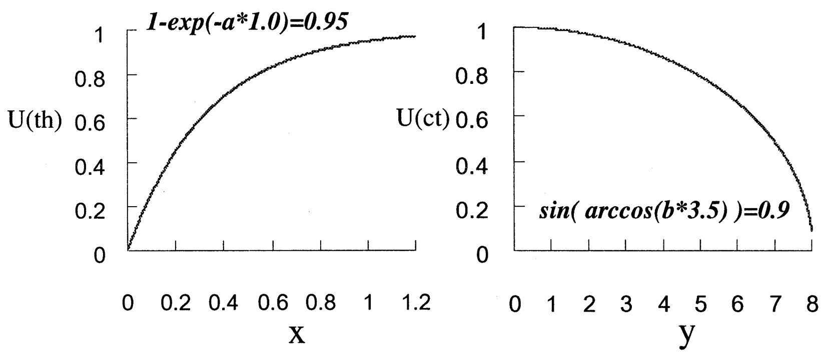

Fig. 4. Utility functions of throughput and cycle time.

goods and nonregular goods. For a rational decision maker, the total utility will increase as the consumption quantity of regular goods increases. Regular goods also have the property that their marginal utility decreases with the consumption quantity. For portfolio planning, throughput can be regarded as regular goods. In contrast, cycle time is not a regular goods because the total utility decreases as the cycle time increases.

We use the following functional forms (shown in Fig. 4) to model the utility of throughput (th) and cycle time (th):

$$U(th) = f(x) = 1 - e^{-ax} \quad \text{where } x = \frac{th}{C_0}$$

$$U(ct) = g(y) = \sin(\arccos(b \cdot y)) \quad \text{where } y = \frac{cycle\,time}{RPT}$$

where throughput and cycle time are normalized with respect to the nominal capacity (C_0) and the sum of raw processing times (RPT). The parameters a and b affect the curvature of the functions. Due to the physical limit to the throughput for any portfolio, there is an upper bound to x. Two questions have been designed to aid the assignment of parameters a and b: 1) What is the utility of a throughput that equals 100% of the nominal capacity? 2) What is the utility of a cycle time that equals 3.5 times that of the raw processing time? If the answers are 0.95 and 0.9, respectively, then the corresponding value of a and b will be 2.9957 and 0.1245 (see Fig. 4).

The total utility function $h(x, y)$ is defined as the weighted sum of functions $f(x)$ and $g(y)$ using an assigned weight w. A

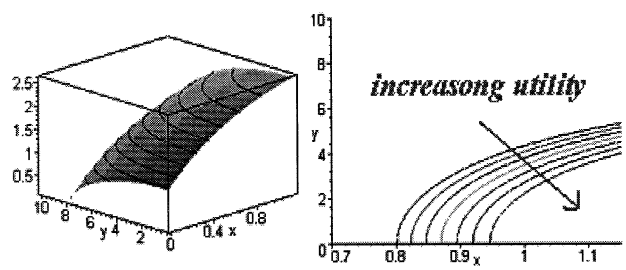

Fig. 5. The total utility function and indifference curves.

method to determine the value of the weight w will be discussed later

$$h(x, y) = w \cdot f(x) + g(y).$$

Using this representation, each option curve can be represented as a hyper surface in the three-dimensional (3-D) space of $U(th)$, $U(ct)$, and $h(th, ct)$. Fig. 5 graphs the total utility in the z axis of a 3-D plot for the case of $w = 2$. Each horizontal cross section of the response surface represents an indifference curve between throughput and cycle time (also shown in the right panel). That is, the total utility of all points on an indifference curve (IDC) is the same.

The optimal portfolio and its optimal operation loading can be obtained by evaluating the total utility of all points in the option space, using data of Figs. 3 and 5, as follows. Each option curve is a hyper surface in the 3-D space of Fig. 5. The intersection between the hyper surface and the response surface of the total utility function is a hyper curve. The optimal operation loading is then the highest point of the hyper curve. This point can be conveniently visualized in the x–y space as shown in Fig. 6 (for three portfolios). The option curve is convex and the indifference curves are concave. Therefore, for each portfolio, there exists a tangential indifference curve that yields the maximum total utility. The tangent point is the optimal loading.

Alternatively, the optimal loading for each tool portfolio can be solved mathematically by using the LaGrange multiplier method as follows. The total utility function is regarded as the objective function to be maximized and the option curve as a constraint relating throughput and cycle time. For (4), the objective function $h(x, y)$ and the hyper surface for the constraint OCS(x, y) are of the forms

$$h(x, y) = w \cdot (1 - e^{-a \cdot x}) + \sin(\arccos(b \cdot y))$$
$$OCS(x, y) = \frac{1}{m - n \cdot x} - y.$$

The LaGrange function is $L(x, y, \ell) = h(x, y) - \ell \cdot OCS(x, y)$, where ℓ is the multiplier. Take the partial derivative of the L function with respect to x, y and l, and set the derivatives to zero

$$L(x, y, \ell) = w \cdot (1 - e^{-a \cdot x} + \sin(\arccos b \cdot y)) - \ell \cdot \left(\frac{1}{m - n \cdot x} - y\right)$$
$$\frac{\partial}{\partial x} L(x, y, \ell) = w \cdot a \cdot e^{-a \cdot x} - \frac{\ell \cdot n}{(m - n \cdot x)^2} = 0 \quad (1)$$
$$\frac{\partial}{\partial y} L(x, y, \ell) = -\frac{b^2 \cdot y}{\sqrt{1 - b^2 \cdot y^2}} + \ell = 0 \quad (2)$$
$$\frac{\partial}{\partial \ell} L(x, y, \ell) = -\frac{1}{m - n \cdot x} + y = 0. \quad (3)$$

Because the gradient of an option curve is a monotonically increasing function of the throughput and that of an indifference curve is a decreasing function, it can be easily derived that there is only one optimal solution for each portfolio.

Once the optimal loading is determined for each portfolio, the maximum utility that can be achieved with each portfolio is also determined. The capital efficiency of a portfolio is defined as the ratio of its maximum utility to its investment cost. The portfolio with the highest investment efficiency is considered optimal.

We now return to the issue of setting the weight w. Using the above procedure, an optimal loading level is determined using a required input of weight w between two utility functions. The weight is a subjective judgment of the relative utility between cycle time and throughput. As shown in Fig. 6, if a weight of 2 is used, the resultant optimal point will be ($x = 0.940$, $y = 3.05$). The derivative of the OC function at that point is equal to 14.606. Since the derivative can be interpreted as the relative utility between cycle time and throughput, a logical impasse now surfaces. That w equals to 2 means that the utility of throughput is twice as important as that of cycle time. That the derivative equals 14.606 means that the utility of throughput is 9.28 ($= C_0 \cdot RPT^{-1} \cdot 14.606^{-1} = 29\,000 \cdot 214^{-1} \cdot 14.606^{-1}$) times more important than that of cycle time. In the following we will show the existence of an equilibrium weight that is inherent to the option curve of each portfolio.

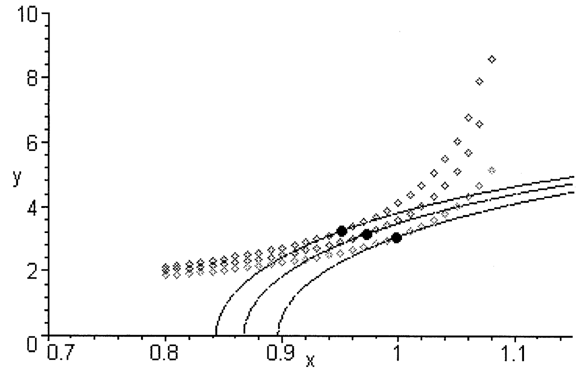

Fig. 6. The optimal operation loading.

We present two numerical examples; one starts with a small value of w and another with a large value of w, to layout a framework of analysis. Each example involves a number of iterations to compute the optimal operation loading. The procedure is as follows.

1) Iteration $i = 1$. Give an arbitrary initial weight $w = w_1$.
2) Compute the optimal loading O_i using w_i as input. Calculate the derivative of the OC at O_i. Let the derivative be D_i.
3) Set $w_{i+1} = (C_0/RPT) \cdot (1/D_i)$.
4) If $|w_i - w_{i-1}| < \varepsilon$, stop. Otherwise, set $i = i + 1$ and go to Step 2).

The results for an initial weight of 1.0 and 8.0 are summarized in Table I. In both examples, the weight w_i converges to a value of approximately 5.29. This convergence value is called the equilibrium weight. Alternatively, the equilibrium weight can be derived by adding two more constraints to the LaGrange multiplier formulation

$$w_1 = \frac{C_0}{RPT} \cdot \left(\frac{d}{dx} OC(x)\right)^{-1} = \frac{C_0}{RPT} \cdot \frac{(m - n \cdot x)^2}{n} \quad (4)$$
$$w = w_1. \quad (5)$$

An OC function is constructed out of different loading levels and corresponding flow times. Therefore, it implicitly represents tradeoffs between marginal throughput and flow time for a fully loaded plant, given the specified utility functions. In practical application of this methodology, there could be other constraints on the value of x. For example, the constraint $x < 1.15$ will impose a restriction such that the throughput is not greater than 1.15 times that of nominal capacity. In practical application, the weight w should also be specified. In a business environment that favors throughput to cycle time, one could use a weight greater than the equilibrium weight. In contrast, if one favors cycle time to throughput, a smaller weight could be used.

IV. Capacity Sharing and Backup

In this section, we describe an application of the above decision model. Some modern plants have a twin-plant design. Two clean rooms are built side by side or stacked up one on top of another to share common utility facilities. In still some other occasions, plants are close to each other. The bottleneck tools of two plants may not be the same at all times. The proximity of plants allows capacity sharing to take place. If

TABLE I
CONVERGENCE OF THE WEIGHT REVISION

	W1	W2	W3	W4	W5	W6	...	W13	W14	W15	W16	W17
Case1	1.0	13.79	3.16	7.10	4.50	5.80	...	5.284	5.298	5.293	5.291	5.292
Case2	8.0	4.21	6.03	4.92	5.51	5.17	...	5.296	5.292	5.292	5.292	5.294

tool capacity is shared between plants, the overall performance will be improved. In this section, a novel methodology for capacity sharing between plants is described.

The solid curve in Fig. 7 is the option curve of a plant. If additional capacity of the bottleneck tools is obtained from a partner plant, the option curve will shift downward to the right, i.e., the dotted curve. Suppose the current operation loading is at point O. With the borrowed capacity, either the throughput could be increased from ω_1 to ω_2 (point A), or the cycle time could be reduced from τ_1 or τ_2 (point C), or any other points on the dotted curve will be achievable.

Table II shows the results obtained by applying the above analysis to a set of industry data. It is shown that if 20 h of capacity of the bottleneck tool group is borrowed from the partner plant, the throughput would be increased by 52 wafers per week, or the cycle time could be reduced by 10 h. This methodology provides an objective arbitration for capacity sharing between plants.

The economic benefit of increased throughput can be computed from the revenue that it brings in and the inventory cost of WIP. From the queuing theory, it is known that a reduction in cycle time would affect the level of throughput and WIP. Therefore, we used an economic model to correlate the economic benefit of cycle time to that of throughput and WIP as follows. Let (ω_i, τ_i) and (ω_j, τ_j) be two points in the option space and $\omega_j > \omega_i$. The value of cycle time reduction can be computed using the following formulas with average asking price (ASP) of processed wafers, material cost (MC), production cost (PC), and a rate of return (r).

Revenue from throughput:

$$R = \omega \cdot (ASP - PC - MC) \cdot 12 \text{ (month/year)}$$

WIP inventory cost:

$$WIC = WIP \cdot \left(MC + \frac{PC}{2}\right) \cdot r$$

Value of cycle time reduction

$$= \frac{(R_j - WIC_j) - (R_i - WIC_i)}{\tau_i - \tau_j}.$$

Let $r = 0.30$, $ASP = 1800$, $PC = 1000$, and $MC = 71.88$, the marginal value of cycle time reduction equals to approximately US \$40 000/h. It should be noted that this figure is derived from the condition of optimal loading. This information could be used as a reference in evaluating projects of cycle time reduction.

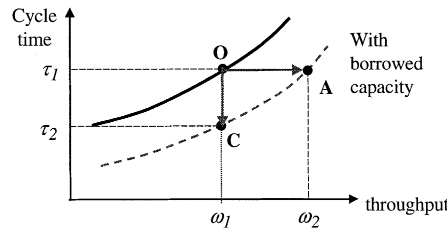

Fig. 7. The effect of capacity sharing.

TABLE II
EFFECT OF CAPACITY SHARING

	Point O	Point A	Point C
Throughput	29,000	29,052	29,000
Cycle time	883.4	883.4	873.9
WIP level	35,583	35,647	35,197

V. APPLICATION TO CYCLE TIME REDUCTION

Besides throughput, cycle time is an important metric of wafer plant performance. Given a tool portfolio, it is desirable to have both high throughput and low cycle time. Unfortunately, these performance measures tend to conflict with each other. Because the operation of a wafer plant is complex and the process and technology portfolios might undergo significant changes, the state of its performance might gradually shift toward one extreme or another over time. Therefore, periodically the need to reign in the cycle time will arise [9], [11] and projects will need initiating. Like all other projects, cycle time reduction projects will need evaluating for their economic benefits. The value of cycle time is difficult to assess because it is really dependent on the level of plant loading. When a plant is underloaded (compared to its designed capacity), the cycle time would be short. The payback of reducing cycle time is minimal. When a plant is fully loaded, the cycle time is high and tends to spike up with disrupting events such as machine failures and process instability. When throughput is near the designed capacity, cycle time is an excellent metric of overall plant efficiency; the utility of reducing cycle time then becomes significant.

The decision model presented in this paper is a suitable method for the evaluation of cycle time reduction projects. The procedure is as follows.

1) Construct an option curve of the plant. This can be achieved by using a simulation program that simulates the operation of the plant. The option curve would be as shown in Fig. 8.

2) Analyze the room of improvement of cycle time reduction. The current state of the plant operation should be determined.

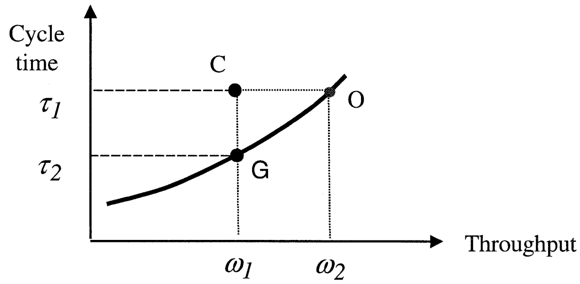

Fig. 8. The room of improvement of cycle time reduction.

Suppose point C represents the current operation performance (ω_1, τ_1). Points O and G are two corresponding points on the option curve. Point G is an ideal goal of cycle time reduction. The distance between points C and G then is an estimate for the potential room of improvement.

3) Analyze the economics of cycle time reduction.

VI. Discussion and Conclusion

This paper describes a methodology for tool portfolio planning, focusing on portfolio optimization and economic analysis with multiple performance criteria. The proposed procedure can be summarized as follows.

1) Assign a value to the parameters a and b of the utility functions $f(x)$ and $g(y)$, respectively.
2) Generate a number of efficient portfolios.
3) Find a fitting function for each portfolio and determine the parameters m and n of the option curves.
4) For each portfolio, solve for the equilibrium weight w and the optimal loading level.
5) Determine the optimal portfolio using the capital efficiency criteria.

Following this procedure, the optimal portfolio and its optimal loading level can be determined. This procedure has been run through a set of industry data. Furthermore, this methodology can be applied to objectively arbitrate capacity sharing between plants and to evaluate the economic value of marginal reduction of cycle time.

Acknowledgment

The authors would like to thank Dr. J. K. Wang and Dr. E. Liou of TSMC and Dr. Y.-C. Shih of Macronix Co. for their comments.

References

[1] F. Barahona, S. Bermo, O. Gunluk, and S. Hood, “Robust capacity planning in semiconductor manufacturing,” IBM Res. Division, Res. Rep. RC22196, 2001.
[2] H. Chen, J. M. Harrison, A. Mandelbaum, A. van Ackere, and L. M. Wein, “Empirical evaluation of a queuing network model for semiconductor wafer fabrication,” *Ops. Res.*, vol. 36, no. 2, pp. 202–215, 1998.
[3] Y.-C. Chou and R.-C. You, “A resource portfolio planning methodology for semiconductor wafer manufacturing,” *Int. J. Advanced Manuf. Technol.*, vol. 18, pp. 12–19, 2001.
[4] Y.-C. Chou, I.-H. Hong, C.-Y. Kuo, and L.-C. Lu, “Product mix planning in semiconductor manufacturing,” in *Proc. Int. Symp. Semiconductor Manuf.*, Santa Clara, CA, USA, Oct. 1999, pp. 19–22.
[5] R. K. Deb and R. F. Serfozo, “Optimal control of batch service queues,” *Advances Applied Probability*, vol. 5, pp. 340–361, 1973.
[6] J. W. Fowler, D. T. Phillips, and G. L. Hogg, “Real-time control of manufacturing bulk service semiconductor manufacturing process,” *IEEE Trans. Semiconductor Manuf.*, vol. 5, no. 2, pp. 158–163, 1992.
[7] C. R. Glassey and W. W. Weng, “Dynamic batching heuristic for simultaneous processing,” *IEEE Trans. Semiconductor Manuf.*, vol. 4, pp. 77–81, June 1991.
[8] R. Kotcher, “Capacity planning in the face of product-mix uncertainty,” in *Proc. Int. Symp. Semiconductor Manuf.*, Santa Clara, CA, Oct. 1999, pp. 73–76.
[9] A. Maiorana, “Improving cycle time through managing variability in DRAM production line,” in *Proc. Int. Symp. Semiconductor Manuf.*, San Francisco, CA, Oct. 1997, pp. A29–A32.
[10] J. Neudorff, “Static capacity analysis using Microsoft Visual Basic,” in *Proc. Int. Conf. Semiconductor Manufacturing Operational Modeling and Simulation*, San Francisco, CA, 1999, pp. 207–212.
[11] M. Nishimura, “Strategic inventory control for stable production,” in *Proc. Int. Symp. Semiconductor Manufacturing*, San Francisco, CA, Oct. 1997, pp. A1–A4.
[12] K. Ozawa, “Optimal tool planning using the X-factor theory,” in *Proc. Int. Symp. Semiconductor Manuf.*, Santa Clara, CA, Oct. 1999, pp. 49–52.
[13] J. M. Swaminathan, “Tool capacity planning for semiconductor fabrication facilities under demand uncertainty,” *Eur. J. Ops. Res.*, vol. 120, pp. 545–558, 2000.
[14] W. W. Weng and R. C. Leachman, “An improved methodology for real-time production decision at batch-process work stations,” *IEEE Trans. Semiconductor Manuf.*, vol. 6, pp. 219–225, Sept. 1993.

Yon-Chun Chou received the Ph.D. degree from Purdue University, West Lafayette, IN, in 1988.

He is currently a Professor and the Director of the Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan. He was an Assistant Professor and an Associate Professor at the University of Massachusetts, Amherst. His research interests include the areas of production and logistics systems design, production scheduling, and information systems. His recent research work includes electronics and semiconductor manufacturing and supply chain management.

Chuan-Shun Wu received the M.S. degree in industrial engineering from National Taiwan University in 2001.

He currently works at the Industrial Technology Research Institute at HsinChu, Taiwan.