Int J Adv Manuf Technol (2001) 18:12–19

The International Journal of
Advanced
Manufacturing
Technology

A Resource Portfolio Planning Methodology for Semiconductor Wafer Manufacturing

Y.-C. Chou and R.-C. You

Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan

Resource portfolio planning is a frequent task in semiconductor wafer fabrication plants, as process, machine and product technologies evolve rapidly and the plants go through capacity expansion. As wafer fabrication plants are complex integrated factories with conspicuous queuing effects, portfolio planning must take into consideration machine use, factory throughput, and total flow-time simultaneously. This paper describes a resource portfolio planning methodology for wafer fabrication foundry plants. An improved static capacity model is first presented. A portfolio planning procedure based on static capacity estimation and queuing analysis is next described. This procedure enables the solution space of resource portfolios to be explored effectively and has demonstrated a capability superior to the current planning method in an industry case study. A software implementation of the procedure is also used to clarify planning dilemmas. It is shown that empirical formulae can be used to estimate the efficiency of batch machines. It is also used to show three types of portfolio adjustment action: flow-time reduction, cost reduction and effectiveness improvement.

Keywords: Batching efficiency; Portfolio planning; Semi conductor manufacturing; Static capacity modelling; Queuing capacity modelling

1. Introduction

In the semiconductor industry, a wafer plant is run in the make-to-order production mode, and there are usually a large number of customer orders in the plant at any time. The important measures of operation performance include machine use, factory throughput, and total flow-time. A full-scale wafer plant contains more than 100 types of machine, and the quantity of machines may be as high as 300 to 400. The manufacture of a product requires 300 or more processing steps. A machine group (of identical or similar machines) may be visited more than once as successive circuit layers are added. This is known as the re-entrant property of process routeing. Because of the long routeing cycle, the total flow-time for a batch of products usually exceeds one month. Modern wafer fabrication plants are highly automated factories, with advanced material handling systems and factory-wide computer control. There are conspicuous queuing phenomena. Depending on the mix of products, the bottleneck machines shift with time and work-in-process lots can be regarded as moving through a network of queues.

A resource portfolio refers to the makeup, in quantity and type, of the set of processing machines in a plant. In the semiconductor foundry industry, there are rapid changes of product mix, process technologies and machine technologies; multiple generations of technologies coexist in a plant as it goes through capacity expansion. Therefore, wafer plants are faced with two related decisions: the product mix and resource portfolio [1]. Given a resource portfolio, the best product mix should be determined to maximise the profit or to achieve other corporate goals. Conversely, given the expected product mix in the future, the best resource portfolio should be planned for. Product mix planning and resource portfolio planning are frequent and continuous tasks in wafer plants. They are essential tasks that support business strategy planning to exploit market opportunities and to reduce the risk of machine obsolescence.

Capacity estimation is a fundamental issue in resource portfolio planning. Because wafer plants manifest the behaviour of complex queuing networks, the performance measures are complexly interrelated. Discrete event dynamic simulation, queuing models [2] and static models [3–6] are three common methods for capacity analysis, of which static models are usually used owing to their relatively quick response time and ease of use. However, static models suffer from two major drawbacks, namely, inaccuracy of estimation and lack of queuing delay information.

There have been many research studies on the configuration design of flexible manufacturing systems [7–10]. Queuing network models and mathematical programming techniques were used in configuration design and optimisation. In [11], a qualitative reasoning model was used to provide guidance for configuring integrated manufacturing systems. However, the special characteristics of wafer fabrication plants, including

Correspondence and offprint requests to: Y.-C. Chou, Institute of Industrial Engineering, National Taiwan University, 1 Roosevelt Road, Sec. 4, Taipei 106, Taiwan. E-mail: ychou@ccms.ntu.edu.tw

batch tools, tool back-up and tool dedication, have not been addressed.

The focus of this paper is to describe a resource portfolio planning methodology that has been developed for semiconductor foundry manufacturing plants. It makes use of a static model and a capacity model for capacity analysis. The static capacity model is an improvement over those that have appeared in the literature. The queuing capacity model is adapted from [2] and is designed for the task of capacity planning. A planning procedure has been developed to explore the solution space of resource portfolios. The remainder of this paper is organised as follows. In Section 2, the workload computation logic for the static capacity model is described. Our proposed method for estimating batching efficiency and its capability are presented in Section 3. A portfolio planning procedure based on the trade-off between flow-time, cost and throughput is presented in Section 4. The issues of tool back-up and dedication are analysed in Section 5. Finally, conclusions can be found in Section 6.

2. The Logic of Workload Computation

Several studies have focused on improving the accuracy of static models. Table 1 summarises the scope of static capacity models described in the literature, with the factors of inaccuracy listed down the table. (In the semiconductor industry, the word tool refers to processing machines. In this paper, the word tool and machine will be used interchangeably.) The first four factors are relatively straightforward to include in a capacity model, but the next three factors require some mathematical treatment. In calculating tool usage time, the total time is divided between operational time and non-operational time, and operational time is further distinguished between uptime and downtime. The tool availability is defined as the ratio of tool uptime to tool operation time. The tool efficiency is the statistical mean of the ratio of actual throughput to maximum throughput. The process yield is the ratio of good product output to the amount of input materials. The yield adversely affects workload in the form of reworks and scrap. Semiconductor manufacturing is characterised by a long process flow-time. Jobs released to the plant in one time period may introduce workload in other time periods, even when a large time bucket, such as one month, is used in planning. To calculate the workloads accurately, the expected arrival time of the workloads due in a job for individual tools should be calculated using lead-time offset and process routeing information. Normally, the tool availability and efficiency, the process yield, and flow-time data are collected on the shop floor [5], or they can be assumed as premises in capacity planning.

Table 1. Scope of static capacity models.

Factors	Witte [3]	Hsih et al. [4]	Wu et al. [5]	Neudorff [6]	Chou and You
Tool availability	✓	?	?	?	✓
Tool efficiency	✓	?	?	?	✓
Yield	✓	?	?	?	✓
Lead time offset	?	✓	✓	?	✓
Batching efficiency				✓	✓
Tool dedication		✓	✓		✓
Tool backup		Rudimental		Rudimental	✓

?, uncertain but presumed.

Incorporating the first four factors of Table 1 will produce a basic static capacity model [3]. To further enhance the precision of capacity estimation, the effect of nominal operation policies on the efficiency of using machine resources can be incorporated. Semiconductor processing tools can be broadly classified as serial or batch tools. Serial tools can be regarded as regular tools. A set-up changeover is usually required between two runs of different products. Batch tools, such as furnaces, are machines that have a large capacity to accommodate multiple lots of products. Batch tools may or may not be loaded to capacity before they are run. In the former case, the use is more efficient; however, work-in-process lots will spend more time waiting to be batched. In the latter case, the average waiting time for each lot is less, but a fraction of the machine capacity will be lost. The loading policy of batch tools has been well studied to control queuing delay time [12–14]. For capacity planning, however, the information that is needed is not queuing delay time, but the average batch size of loading.

Batching efficiency is expressed as the statistical mean of the ratio of actual loading size to machine capacity. In [6], batching efficiency is estimated through regression analysis of the "visits to starts" ratio and the observed efficiency, while the total set-up time is estimated by analysing the occurrence durations of tool idleness based on historical data.

Tool dedication and back-up are the second source of inaccuracy. In the process routeing of a product, each process step is usually assigned to a machine group (a machine group is made up of one or more machines of the same capability). Owing to special processing requirements, individual machines may be dedicated to the manufacture of particular products or to processing steps. Conversely, alternative machines may be specified for a process step [15]. When a primary tool has a high use, its alternative tools may be used as back-up tools to offload its workload. Both tool dedication and back-up impose limitations on resource applications, thus, they affect the factory throughput. In [5], process constraints on dedication and back-up are maintained *a priori* in a database. To forecast equipment loading, workloads are assigned manually or shifted while observing constraints on process capability, priority, and tool availability [4]. Tool back-up planning has also been dealt with in a simple static manner [6]. The practice of tool dedication and back-up complicate capacity estimation. Both issues are involved with combinatorial optimisation [4,6] and, thus, are beyond the capability of static capacity modelling. There have been no satisfactory treatments in the semiconductor manufacturing literature.

The major input to portfolio planning includes product demands, process routeings and yields. The product demands are expressed as demand batches, D_{it}, each characterised by the quantity D, product type i, and due time t. To represent the routeing information, the process steps of a product are indicated by the subscript j and the required tools by the subscript k. Because of the re-entry property, the tool subscript

is also represented, in a function form, as *k(i,j)* to indicate that *k* depends on *i* and *j,* and that there is a many-to-one relationship between steps and tools. The due time for D_{it} is the time period *t*, but the workloads for some tool groups may fall in time periods other than *t*. Let *l(i,j)* be the leadtime offset of the workload for the *j*th step of product *i*, starting from time *t*. For capacity planning, the yield of each step is also required and is specified as *sy(i,j)*.

Let the unit workloads be the amounts of workloads generated by a demand batch for the tool groups. For each demand batch, many unit workloads can be generated. This step can be symbolically represented as:

$$D_{it} \longrightarrow\longrightarrow w_{i,\{j\},k(i,j),t\text{-}l(i,j)}$$

where *w* represents a unit workload, the double arrow implies that one or more unit workloads are generated, the {*j*} represents the set of all process steps of product *i*, and the term *t*-*l*(*i*,*j*) indicates that the occurrence time for the unit workload is the due time *t* offset backward by the lead time *l*(*i*,*j*). Each unit workload is identifiable by a product-step pair (*i*,*j*). Let $P_{i,j,k}$ be the processing time and let *J(i)* be the last step of product *i*. Adjusting for yield allowances (*ya*), the unit workload of each product-step pair of (*i*,*j*) at time *t* is calculated as

$$w_{i,j(i),k(i,j),t(i,j)} = \frac{D_{i,t}\, p_{i,j,k}}{ya_{i,j,t}} \forall i$$

where

$$ya_{i,j} = sy_{i,j} * ya_{i,j+1} \ \forall j$$

$$ya_{i,J(i)} = sy_{i,j}$$

Here, the yield allowance for each step is computed backward from the last step to the first step of the process flow. The yield allowance for the last step is set to be equal to its step yield. The yield allowances for all other steps are iteratively accumulated backward from the last step. The total workload for tool *k*, $W_{k,t}$, and tool requirement, $q_{k,t}$, can be computed as

$$W_{k,t} = \sum_{i} \sum_{j(i)} w_{i,j,k,t} \ \forall k,t \tag{1}$$

$$q_{k,t} = \frac{W_{k,t}}{(availability)_k\ (operation\text{–}time)_t (efficiency)_k} \forall k,t$$

Tool quantity is finally determined by rounding up $q_{k,t}$ and by considering the time trajectory of tool requirements.

3. Batching Efficiency

As mentioned in Section 1, to reduce the flow-time, batch tools may not be loaded to full capacity. The loading policy of batch tools has been thoroughly studied and reported in the literature [12–14]. With only local information (without arrival forecasts), the general conclusion is that the greedy loading rule is close to optimal [12]. Let λ be the arrival rate of jobs, *E*[*S*] be the expected value of the service time, and let c_g be the quantity of tools in tool group *g*. Define the traffic intensity for a tool group as the sum of all workloads and downtimes over its capacity in a period. A lower bound formula for the average batch size ($\bar{b}$) was derived for the special case of independent job streams, constant arrival rates and no downtimes [12]:

$$\bar{b} \geq \left(1 - e^{-\lambda E[S]}\right) \frac{\lambda E[S]}{c_g}$$

where the second term on the righthand side represents the traffic intensity. When the traffic intensity is high, the first term approaches zero. This formula contains just one parameter, namely c_g. To enhance the estimation accuracy of the batching efficiency, the effect of tool capacity (B_g^{max}), tool quantity and tool downtime (ρ_g^{inc}, as a fraction) on batching efficiency has been analysed in this study. The objective is to identify formulae that can be used to predict the batching efficiency. The Monte Carlo simulation was used to generate data for eight scenarios of various levels of tool capacity, tool quantity and tool downtime. The eight scenarios are defined in Table 2.

The resultant characteristic curves that relate the average batch size (the vertical axis) with traffic intensity (the horizontal axis) are shown in Fig. 1. On the righthand side of each figure, where the traffic intensity is high, the average batch size is close to the tool capacity. On the lefthand side, even though the traffic intensity is low, the average batch size cannot be lower than one. Therefore, the characteristic curves would past through the point (1, B_g^{max}) and the point (ρ_g^{inc},1).

To construct the prediction formulae, piecewise lines (lines *A* and *B*) are fitted to the simulated data to obtain the following approximation formulae (Fig. 2):

$$\bar{b}_g = \begin{cases} 1 & \text{when } 0 < \rho_g \leq D \\ 1 + \left(\dfrac{B_g^{max} - 1}{1 - D - \rho_g^{inc}}\right)(\rho_g - D) & \text{when } D < \rho_g < 1 - \rho_g^{inc} \end{cases}$$

$$D = [\ -0.0674 \ln (B_g^{max}) + 0.2447\ ] + \frac{C_g}{20\ (B_g^{max})^2} \quad (B_g^{max} \geq 2)$$

where ρ_g is the traffic intensity less ρ_g^{inc}. The formulae also show that tool capacity and quantity do have an effect on the location of the excursion point *D*.

To validate the accuracy of the above formulae, a separate fabrication plant simulator that captures the interaction between multiple job streams has been run on product, process routeing, and machine data provided by a semiconductor wafer plant of the United Microelectronic Corporation. The data includes 4 representative memory and logic products of multiple technology generations, 116 tool groups, and 34 back-up relations. Of the 116 tool groups, 41 are batch tools. The simulator uses a greedy loading policy. Figure 3 shows a comparison of the

Table 2. Scenarios of simulation.

Scenario	1	2	3	4	5	6	7	8
Tool capacity	2	10	2	10	2	10	2	10
Tool quantity	1	1	10	10	1	1	10	10
Downtime fraction	0.1	0.1	0.1	0.1	0.3	0.3	0.3	0.3

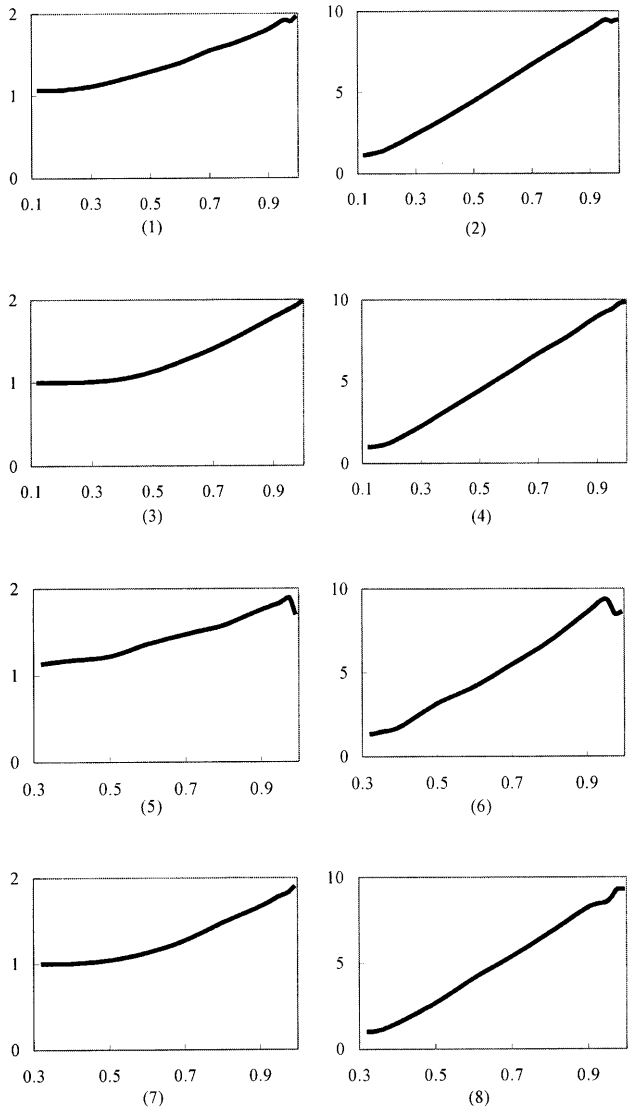

Fig. 1. The relationship between the average batch size and the traffic intensity.

resultant data of the average batch size generated by the simulator and the formulae. The average difference is approximately 4.93%. Based on the above analysis, the tool requirement (Eq. (1)) for batch tools can be calculated as:

$$q_{k,t} = \frac{W_{k,t}}{(operation\text{–}time)_t(availability)_k(efficiency)_k} \frac{B_g^{max}}{B_g} \forall k,t$$

The current industry practice is to use the maximum batch size or historical data of the average batch size in capacity analysis. This study shows that formulae, which are based on the traffic intensity, breakdown time and the maximum batch size, can be used to predict the average batch size with good accuracy.

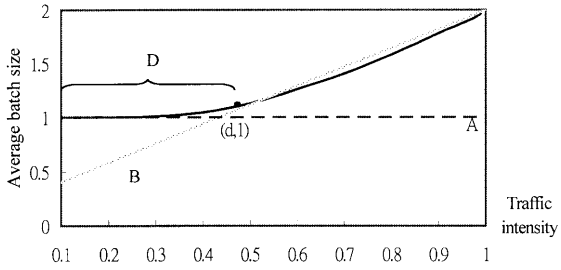

Fig. 2. Characteristic curves for the mean batch size.

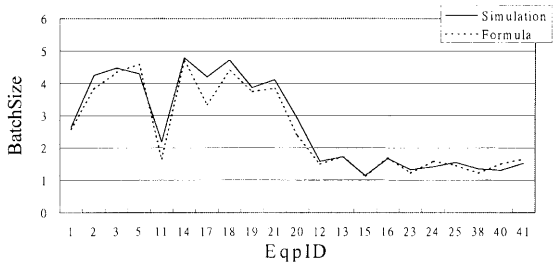

Fig. 3. The performance of the proposed batch size formula.

4. The Planning Procedure

A resource portfolio can be represented as an ordered list of tool quantities, such as $s = (n_1, n_2, \ldots, n_N)$, where n_i is the number of tools in tool group i. The performance measures of interest to capacity planning are throughput, use, and flow-time. Since static capacity models provide limited information about throughput and do not provide flow-time information, an open queuing network capacity model (Appendix A) was used to evaluate resource portfolios in this study. This queuing model is adapted from [2], and uses the following premises:

1. No scrap and rework.
2. Two classes of customer: work-in-process and machine breakdown.
3. Two types of tool: batch and non-batch tools.

The motivation for this adaptation is twofold. First, many of the scrap and rework probabilities in [2] are either not available during capacity planning or are crude estimates for future products. Secondly, the yield rates of products already in production are fairly high in well-run factories. By setting the step yields *sy(i,j)* to zero (Section 2), a static capacity model can generate a minimal portfolio. Any portfolio that has fewer tools in any tool group will not be able to meet the output requirements. However, this minimal portfolio may not be sufficient to meet the flow-time performance requirements.

Our portfolio planning procedure makes use of the static capacity model of Section 2 and the queuing capacity model (Appendix A). Both models have been implemented as software modules. The static module is first used to generate an initial resource portfolio. The portfolio is then evaluated using the queuing module and then adjusted based on the flow time data.

Figure 4 illustrates the basic concept of portfolio adjustment. Each portfolio can be characterised by three attributes: through-

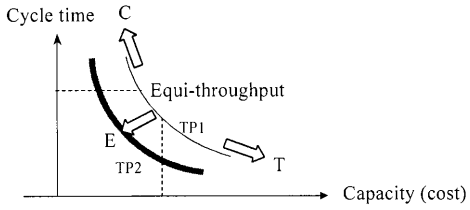

Fig. 4. The strategies of portfolio adjustment.

put, weighted flow-time and investment cost. The throughput of a portfolio must be greater than the product demands. The weighted flow-time is computed from the flow-times and product demands of all products. For a given portfolio, the quantity of certain tool groups can be increased (or decreased) to reduce (or increase) the flow-time while meeting the same level of throughput requirement. This phenomenon is shown by equi-throughput curves in the figure. Two curves, such as TP_1 and TP_2, represent the relative effectiveness of resource portfolios. In Fig. 4, the curve TP_2 has a higher throughput than curve TP_1. For the same investment cost (the vertical dotted line), TP_2 has a more balanced portfolio, resulting in a higher throughput and lower flow-time. In contrast, to achieve the same flow-time (the horizontal dotted line), TP_1 will require a higher investment. Figure 4 reveals three types of portfolio adjustment action for improving flow-time, investment cost, and portfolio effectiveness as indicated by the arrows *T*, *C*, and *E*, respectively.

Portfolio adjustment is an iterative process based on marginal analysis. For each tool group of a current portfolio, the tool quantity is increased or decreased by one, to generate neighbouring portfolios. All neighbouring portfolios are evaluated using the queuing capacity module. The ratio of flow-time decrement (or increment) over cost increment (or decrement) is computed. Two separate lists are maintained, one for flow-time reduction and another for cost reduction. The two lists can be sorted to rank the order of action types *T* and *C*, respectively. Type *E* actions are composed of type *T* and type *C* actions, as shown in Fig. 5. Portfolios *b* and *c* are obtained from a current portfolio (point *a*) by adding and subtracting a tool, respectively. If the combined effect of two actions (in the two lists) results in a reduction of both cost and cycle time, they constitute a type *E* action (point *d*).

The above procedure has been applied in an industry case study. Figure 6 summarises the results of the portfolio planning process. Two initial tool portfolios are shown at the top of the figure: one provided by our industry sponsor (denoted by the number 1), and another generated by the above static model (denoted by the number 2). Portfolio 1 was deemed infeasible

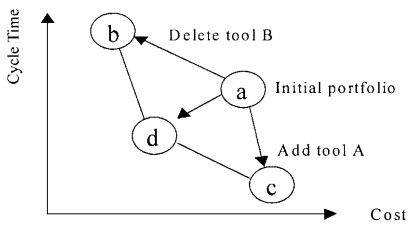

Fig. 5. Type *E* adjustment action.

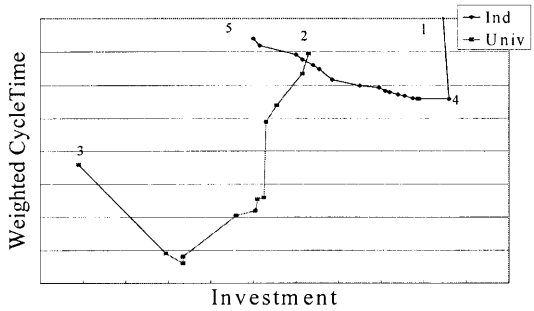

Fig. 6. Tool portfolio adjustment trajectory.

after an analysis showed that the use for some batch tools is greater than 1.0 based on the batching efficiency formula in Section 3. Therefore, the tool quantity for those tools is increased and portfolio 4 is obtained. Portfolios along the curve connecting points 4 and 5 represent trade-offs between flow-time and cost. Starting with portfolio 2, a sequence of portfolios is obtained (from points 2 to 3) by applying the same adjustment procedure. Figure 6 demonstrates that the solution space of resource portfolios can be searched more effectively using the developed static and queuing capacity models. The planning procedure can be used to generate alternative portfolios of different investment costs and flow-time performances.

Figure 7 illustrates the *T* action in greater detail for a second data set of product demands. The resultant configuration and performance data can be found in Appendix B. (The investment and flow-time data are business sensitive information, therefore, they have been normalised. Similarly, machine groups are identified by numbers.) Let the marginal cost of flow-time be defined as the ratio of the flow-time reduction to the investment cost increment that is associated with adding a machine to the portfolio. The marginal cost is computed for each machine group and the top five machine groups are listed in the second column of Appendix B. In each iteration, the top one is selected. Starting with a given initial portfolio (at the upper-left of the trade-off curve), a sequence of portfolios is then generated. This figure shows that 4% more investment (on critical machine groups) could reduce the flow-time by 27%.

2. Tool Pooling and Dedication

The objective of tool pooling is to reassign workloads such that the required tool quantities are reduced without significant

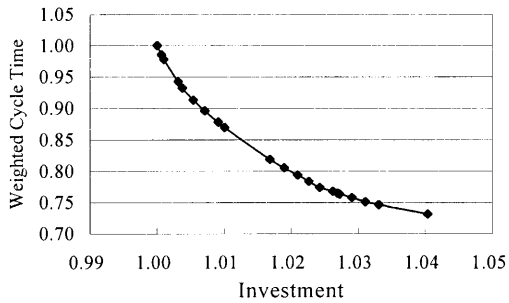

Fig. 7. The *T* action.

Table 3. Effect of tool pooling.

Tool type	Workload (h)	Required tool quantity	Reassigned workload	Improved tool quantity
A1	550	1	610	1
A2	700	2	640	1

impact on factory performance and productivity. The opportunity for tool pooling exists when a tool type can be backed-up by other types. Table 3 is a numerical example for illustrating the effect of tool pooling in reducing tool investment cost. Suppose tool type A1 can be backed-up by type A2 and there are 650 work hours in a month. Before pooling, a total of three (1+2) tools will be required. By shifting 60 hours of workload from tool type B to A, only two tools will be required. This task is also called tool pooling, because both tool types will be considered as a pool of tools in capacity planning.

Tool back-up and dedication has been implemented in a mixed integer linear program. The workloads for each tool are first calculated by time period. These workloads are then reassigned according to constraints imposed by dedication decisions and back-up relationships. Tool dedication is implemented by specifying tools in the process routeing. The objective function of the mixed integer linear program is to minimise the total tool investment. The major constraints are listed below:

BG_k tool groups that can back up tool type k
$BE_{m,k}$ back-up efficiency of tool type m w.r.t. tool type k
$Y_{m,k,t}$ workload shifted to tool m from tool k at time t
$Q_{k,t}$ tool quantity required for tool type k at time t
$a_{m,t}$ tool availability

$$\sum_k Y_{m,k,t} \le Q_{m,t}\, a_{m,t} \qquad \forall m,t$$

$$\sum_{m \in BG_k} Y_{m,k,t}\, BE_{m,k} = W_{k,t} \qquad \forall k,t$$

The first inequality is the capacity constraint. For ease of coding, a tool type is considered a back-up tool for itself. Thus, workloads may be "shifted" from a tool to itself. Tool pooling is suitable for either single- or multi-period planning. Figure 8 shows the effect of tool pooling. The curve at the top represents two sequences of portfolios. The curve at the bottom represents the resultant portfolios after tool pooling is applied. The relative position of the two sets of portfolios indicates that tool pooling is an E type action. It should be cautioned, however, that pooling tools together would change the dynamics of material flow and set-up frequency. The "improved" portfolio should be subjected to further analysis regarding cycle time, use and work-in-process inventory.

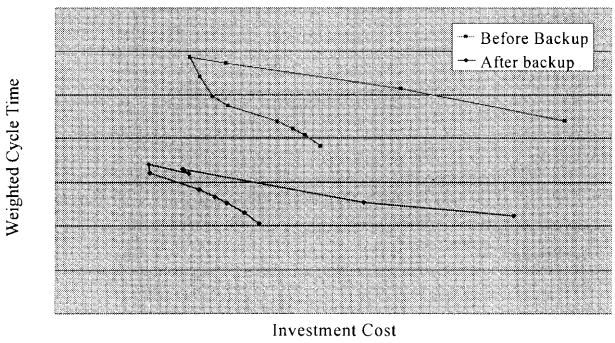

Fig. 8. The effect of tool back-up adjustment.

6. Conclusions

A resource portfolio planning methodology for semiconductor wafer fabrication plants is presented in this paper. The methodology has three major components: an improved static capacity model, a queuing capacity model and a portfolio adjustment procedure. It is shown that batching efficiency of batch tools can be predicted accurately by using formulae. The methodology enables capacity planners to explore effectively the solution space of portfolios and to evaluate better the trade-off between flow-time, investment cost and factory throughput. This methodology has been implemented in a software decision system, and has demonstrated its capability in generating superior solution in a benchmarking case study.

Acknowledgements

This study has been partially funded by United Integrated Circuits Corporation of United Microelectronic Corporation and National Science Council of ROC (NSC 88-2212-E002 and 89-2212-003).

References

1. Yon-Chun Chou, I-hsuan Hong, C. -Y. Kuo and L. -C. Lu, "Product mix planning in semiconductor manufacturing", International Symposium on Semiconductor Manufacturing, Santa Clara, California, USA, pp. 11–14, October 1999.
2. D. P. Connors, G. E. Feigin and D. D. Yao, "A queuing network model for semiconductor manufacturing", IEEE Transactions on Semiconductor Manufacturing, 9(3), pp. 412–427, 1996.
3. J. D. Witte, "Using static capacity modeling techniques in semiconductor manufacturing", IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, pp. 31–35, 1996.
4. H. W. Hsih, H. C. Wu et al. "Equipment loading dynamic forecasting system", The Seventh International Symposium of Semiconductor Manufacturing, Tokyo, Japan, pp. 83–86, October 1998.
5. W. F. Wu, J. L. Yang and J. T. Liao, "Static capacity checking system with cycle time considered", The Seventh International Symposium of Semiconductor Manufacturing, Tokyo, Japan, pp. 307–310, October 1998.
6. J. Neudorff, "Static capacity analysis using Microsoft Visual Basic", International Conference on Semiconductor Manufacturing Operational Modeling and Simulation, San Francisco, USA, pp. 207–212, January 1999.
7. Andrew Kusiak, Flexible Manufacturing Systems: Methods and Studies, North-Holland, 1986.
8. K. E. Stecke and R. Suri, (ed.), Proceedings, Third ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, Elsevier, 1989.
9. Rajan Suri and Richard R. Hildebrant, "Modeling flexible manufacturing systems using mean-value analysis," Journal of Manufacturing Systems, 3(1), pp. 27–38, 1983.

10. Yves Dallery and Yannick Frein, "An efficient method to determine the optimal configuration of a flexible manufacturing system," Proceedings, Second ORSA/TIMS Conference on Flexible Manufacturing Systems: Operations Research Models and Applications, pp. 269–282, August 1986.
11. Yon-Chun Chou, "Configuration design of complex, integrated manufacturing systems", International Journal of Advanced Manufacturing Technology, 15, pp. 907–913, 1999.
12. C. R. Glassey, F. Markgraf and H. Fromm, "Real time scheduling of batch operations", Optimization in Industry, John Wiley, pp. 113–137, 1993.
13. D. Fowler, J. Philips and G. Hogg, "Real-time control of multiproduct bulk service semiconductor manufacturing processes", IEEE Transactions on Semiconductor Manufacturing, 5, pp. 294–297, 1992.
14. C. Roger Glassey and W. Willie Weng, "Dynamic batching heuristic for simultaneous processing", IEEE Transactions on Semiconductor Manufacturing, 4, pp. 77–82, 1991.
15. Pravin K. Johri, "Practical issues in scheduling and dispatching in semiconductor wafer fabrication", Journal of Manufacturing Systems, 12, pp. 474–485, 1993.

Appendix A. A Queuing Capacity Model for Portfolio Planning

An open network capacity model based on [2] was used to evaluate the flow-time in this study. That model has been modified to fit the task of capacity planning. This modified model has the following premises:

1. No scrap and rework.
2. Two classes of customer: work-in-process and machine breakdown.
3. Two types of tool: batch and non-batch tools.

There is a set $F = \{1, \ldots, F\}$ of product families and a set $G = \quad 1, \ldots, G\}$ of distinct tool groups. The raw input data for capacity planning are the process flow for each product $f \in F$, tool information for each tool group $g \in G$, and product demand (D_f). The process steps (k), processing time $(s_{f,k,g})$ and tool group for the manufacture of a product are specified in its process flow. Other tool information includes cost, MTTR, MTBF and quantity c_g.

Let the traffic rate for tool group g be Γ_g. Let S_g and S_g^B be the random variables denoting the processing time and machine downtime of tool group g. Let the arrival rate of an incapacitated event at group g be λ_g^B. (For lack of tool breakdown variability data, machine incapacitation is assumed to be Poisson distributed.) Let the squared coefficient of variation of service time be v_g^s and that of the interarrival time be v_g^a.

The output of the queuing module includes tool use ρ_g, machine downtime proportion ρ_g^{inc}, average batch size $\bar{b}_g$, and queuing delay D_g. For non-batch tools,

$$\rho_g = \frac{\Gamma_g E[S_g]}{c_g}$$

$$\rho_g^{inc} = \frac{\lambda_g^B E[S_g^B]}{c_g} = \frac{c_g}{\text{MTTR} + \text{MTBF}} \frac{E[S_g^B]}{c_g} = \frac{\text{MTTR}}{\text{MTTR} + \text{MTBF}}$$

$$E[D_g] = \frac{(\rho_g^{inc} + \rho_g)^{(c_g - 1)}}{c_g^2} \times \frac{\varphi_b \lambda_g^B E[(S_g^B)^2] + \varphi_g(\Gamma_g E[S_g^2])}{2(1 - \rho_g^{inc})(1 - \rho_g^{inc} - \rho_g)}$$

where φ_g and φ_b are two correction factors [2]. Assuming Poisson breakdown, φ_b is equal to 1 and $\varphi_g = (v_g^s + v_g^a)/(v_g^s + 1)$.

To compute the performance measures for batch tools, the processing time is adjusted to account for incapacitation. The adjusted service time is denoted as $Z_g = S_g + B_g$, where

$$B_g = \begin{cases} S_g^B & \text{w.p.} \quad \lambda_g^B/\Gamma_g \\ 0 & \text{w.p.} \quad (1 - \lambda_g^B/\Gamma_g) \end{cases}$$

Let $\phi_g = \Gamma_g E[Z_g]$. For batch tools with a maximum batch size β_g^{max}, the total use $\tilde{\rho}_g$ (including serving incapacitation events) is

$$\tilde{\rho}_g = \pi_{0,0} \sum_{i=1}^{c_g - 1} \frac{i\phi_g^i}{c_g\, i!} + \pi_{0,0} \frac{\phi_g^{c_g} x}{c_g!\,(x - 1)}$$

where π_{mn} denote the probability that m servers are busy and there are n jobs in the queue and x is the solution lying in the interval $(1, c_g\beta_g^{max}/\phi_g)$ of the following polynomial equation.

$$\frac{\phi_g}{c_g} x^{(\beta_g^{max} + 1)} - \left(1 + \frac{\phi_g}{c_g}\right) x^{\beta_g^{max}} + 1 = 0.$$

The tool use, downtime proportion, queuing delay, and average batch size $(\bar{b}_g)$ for batch tools are:

$$\rho_g = \frac{\tilde{\rho}_g}{1 + E[B_g]/E[S_g]}$$

$$\rho_g^{inc} = \tilde{\rho}_g - \rho_g$$

$$E[D_g] \approx \frac{\pi_{c_g,0x}}{\Gamma_g(x - 1)^2} \frac{(v_g^a + v_g^z)}{2}$$

$$\bar{b}_g = \frac{\Gamma_g \cdot E[S_g]}{\rho_g\, c_g}$$

Appendix B. Portfolio and Performance

Data for Fig. 7.

Table 4.

Iteration	Top 5 candidates	Tool selected	Investment cost	Cycle time
0	Initial portfolio		1.000	1.00
1	48, 82, 42, 49, 26	48	1.001	0.99
2	82, 42, 49, 26, 4	82	1.001	0.98
3	42, 49, 26, 4, 77	42	1.003	0.94
4	49, 26, 4, 77, 83	49	1.004	0.93
5	26, 4, 77, 83, 85	26	1.005	0.91
6	4, 77, 83, 85, 44	4	1.007	0.90
7	77, 83, 85, 44, 75	77	1.009	0.88
8	83, 85, 44, 75, 47	83	1.010	0.87
9	85, 44, 75, 47, 79	85	1.017	0.82
10	44, 75, 47, 79, 21	44	1.019	0.81
11	75, 47, 79, 21, 48	75	1.021	0.79
12	47, 79, 21, 48, 82	47	1.023	0.78
13	79, 21, 48, 82, 71	79	1.024	0.77
14	21, 48, 82, 71, 74	21	1.026	0.77
15	48, 82, 71, 74, 18	48	1.027	0.77
16	82, 71, 74, 18, 16	82	1.027	0.76
17	71, 74, 18, 16, 58	71	1.029	0.76
18	74, 18, 16, 58, 42	74	1.031	0.75
19	18, 16, 58, 42, 14	18	1.033	0.75
20	16, 58, 42, 14, 13	16	1.040	0.73