

行政院國家科學委員會專題研究計畫 期中進度報告

計畫編號: NSC92-2213-E-009-044- 4 92 08 01 93 07 31

執行單位: 國立交通大學工業工程與管理學系

計畫主持人: 鍾淑馨

報告類型: 精簡報告

計畫參與人員: 高清貴、陳益參、范國基

。<br>在前書 : 本計畫可公開查

報告附件: 出席國際會議研究心得報告及發表論文

93 5 31

# $(2/3)$

# **Production Schedule Planning and Synthetic Performance Evaluation of Wafer Fabs Under an Non-steady State Environments (2/3)**

NSC 92-2213-E-009-044 92 08 01 93 07 31

# ABSTRACT

In semiconductor industry, dynamic changes in demand force companies changing the product mix that makes the production planning being challenging. This project aims at the environment that product mix changes weekly and presents a production scheduling system to plan the wafer lot release and throughput. The proposed system is designed on the make-to-stock basis and comprises two modules. Preliminary analysis module analyzes throughput and cycle time distributions for different product mixes so as to derive the cycle time for bidding for each product type. The analysis results are used as the inputs to the production planning. In production scheduling module, with the considerations of the achievement of demand forecast, production smoothing, and commitment of due dates of the released job orders, the job release schedule and completion time table are prepared. Simulation model of a semiconductor fab is used as the base case to demonstrate the effectiveness and efficiency of the proposed system.

*Key Words* - Cycle time, product mix, lot release scheduling, wafer fabrication.

# **1. Introduction**

As the competition become much fiercer, semiconductor companies must quickly respond to customers' fluctuating demand in order to survive. Dynamic changes in demand force companies to make changes on the product mix. Product mix changes complicate the already-complex system. In a semiconductor fab, machines are shared by a huge number of different products, resulting in a heavy load sharing of precious resources and consequently a long queue may be present. Product mix level has considerable impact on production throughput, cycle time, and the capability of meeting due dates. Production throughput, cycle time, machine utilization, and work in process (WIP) inventory, are highly interrelated. (Chou 2000, Dümmler 2000) Under different product mixes, the overall manufacturing performance of the system also will be different. Facing the environment with volatile demand, production planning for make-to-stock wafer fabrication is an even complicated problem compared to other manufacturing industries.

The planning tool for master production scheduling can be categorized into simulation, mathematical programming, and the combination of the fore-mentioned. By using simulation tool, a system can be constructed easily to match with the production activity control (Liu *et al.* 1995). However, building and/or running a simulation model is time-consuming. Linear programming can quickly derive the best production plan based on capacity constraints, but there are too many assumptions in describing the real phenomenon (Hackman and Leachman 1989, Chu 1995). For this reason, some researches combined the simulation and mathematic algorithm for better describing the environment than by separately adopting one of the two methods (Burman *et al.* 1986, Thompson and Davis 1990, Hung and Leachman 1996).

The purpose of this project is to present a production scheduling system that deals with the product mix changes periodically under the environment with volatile demand. The proposed system is designed on the make-to-stock basis. Cycle time for bidding are first determined for each product type with the data collected from the environments of different product mixes. Wafer release schedule and completion time table are then prepared. Such planning results can be valuable for preparing available-to-promise information and for decision-making in demand management system.

# **2. System Environment**

For a make-to-stock environment, such as DRAM chip manufacturing, the primary goals of production planning are the maximization of throughput and resource utilization while maintaining production smoothly. The wafer stepper machine is the most expensive

and key operation in wafer fabrication factories. So, to guarantee the utilization of the stepper machine in sufficient and effective way is the basic idea for designing a production scheduling system. In this project, a planned utilization rate is kept in a certain range for the wafer stepper machine, which is treated as the bottleneck (BN) resource. In addition, to be consistent with the processing batch size of thermal oxidation process so as to raise the utilization rate of equipment, the batch size of wafer release is six lots where one lot consists of 25 pieces of wafers.

Practically, in order to achieve a certain degree of stability in the production system, the first week of the schedule is frozen, which means all but the most critical changes in this period cannot be made to the production schedule. Therefore, in the frozen period, if there is any shortage resulted from demand fluctuation or forecast error, the shortage will be treated as back order and be promised for a later time period.

The planning horizon for master production schedule is 12 weeks including a one-week frozen period. The planning period is one week and the product mix can only be changed by weekly. Wafer lots are released under a CONWIP release policy. For the wafer lots already being processed in the plant, first-in-first-out rule is used. The framework for the proposed system is shown in Figure 1 and the notations used in this project are defined in Table 1.



Figure 1. The framework of the proposed system

The proposed system comprises two modules. In preliminary analysis module, the distribution of cycle times is analyzed. In order to grasp the fluctuation of demand, several representative product mixes are input to the simulation model to collect cycle times. All the collected cycle times are used to calculate the mean and the standard deviation for each product type. The percentile cycle time for each product type is then determined as the bidding cycle time. Those results will be the inputs to the production scheduling module.

In production scheduling module, the release schedule is prepared with the considerations of demand fulfillment, production smoothing, and protecting due date commitment of the released job orders. The release schedule and corresponding job order completion time will be planned in consequence.

$\dot{i}$	: index of product type;
l	index of layer;
$\mathcal{f}$	index of product family;
h	index of planning horizon;
$\boldsymbol{t}$	. index of planning week;
$\boldsymbol{k}$	: index of workstation:
$k_f$	: the workstation with the maximum difference in average theoretical process time between product family $f$ and any other family;
fpw	The first planning week;
$bat_k$	batch size setting for workstation $k$ ;
$cap_k$	planned capacity of workstation $k$ ;
$CT_{i,r}^{\text{BBCT}}$	cycle time of product type $i$ estimated by BBCT algorithm at simulation run $r$ ;
$CT_i^{\alpha}$	$\alpha$ percentile cycle time for product type <i>i</i> ;
$AD_{i,t}$	actual demand of product type $i$ at time period $t$ ;
$DF_{i,t}$	demand forecast for product type $i$ at time period $t$ ;
$DN_{i,t}$	net demand for product type $i$ at time period $t$ ;
$EI_{i,t}$	ending inventory for product type $i$ at planning period $t$ ;
$Q_f^{\max,k}$	maximum affordable quantity for product family $f$ processed on workstation $k$ ;
$Q_f^{\max}$	the production quantity limit for product family $f$ in the system;

TABLE 1. MATHEMATICAL NOTATIONS.



#### **3. Preliminary Analysis Module**

The main work of this module is to calculate the mean cycle time and cycle time spread under different product mixes. In this module, quantity limit for each product family is first determined in order to prevent bottleneck shifting and cycle time variance increasing.

Several representative product mixes in the plant are then selected on the basis of the seasonal demand of each product type For each product mix, cycle times of the job orders belong to each distinct product type are collected by running a simulation model, which describes the system environment and operation behavior. Cycle time analysis is performed for each set of collected cycle times. After that, to have a bidding cycle time for due date setting, the  $\alpha$  percentile cycle time for each product type is determined as the input for planning lot release schedule. The  $\alpha$  percentile cycle time is defined as the longest cycle time of  $\alpha$ % out of the wafer lots completed during a specific time period.

#### *3.1 Setting quantity limit for product family*

Product types of the same product family often have similar manufacturing process. Releasing a great quantity of products belonging to the same product family at one time period may cause some specific workstations being overly demanded. This will result in bottleneck shifting and cycle time variation in shop floor. Therefore, the quantity limit for each product family *f* needs to be established before product mix setting.

The quantity limit setting is based on the capacity that can offer to produce the maximum quantities of the corresponding product family in the system. Two kinds of capacity constraints must be considered, namely the system constraint and the family constraint. The system constraint comes from the capacity of bottleneck while the family constraint comes from the capacity of the workstation  $k_f$ , which has the maximum difference in average theoretical process time between product family  $f$  and any other families

$$
d(k, f) = \max\{ \left| PT_f^{avg, k} - PT_{f'}^{avg, k} \right| \forall f' \notin f \}
$$
\n
$$
(3-1)
$$

$$
k_f = \max_{k} \{d(k, f)\}\tag{3-2}
$$

$$
Q_f^{\max,k_f} = (cap_{k_f} \times U_{k_f} \times bat_{k_f}) / PT_f^{avg,k_f}
$$
\n(3-3)

$$
Q_f^{\max, \rm BN} = (cap_{\rm BN} \times U_{\rm BN} \times bat_{\rm BN}) / PT_f^{\text{avg}, \rm BN}
$$
\n(3-4)

$$
Q_f^{\max} = \min\left(Q_f^{\max,k_f}, Q_f^{\max,BN}\right) \tag{3-5}
$$

#### *3.2 Simulation model description*

Cycle times are collected by running a simulation model, which is built with eM-Plant (Tecnomatix 2000). Actual production data of a wafer plant in Taiwan is used as input data, including machine, product, and product routes information. In this wafer fab, there are 83 workstations and five product types. Product A and B belong to the family of SDRAM, and C, D, and E belong to the family of DDR. Each process contains process steps in a range of 276 to 338 operations. For the information of workstations, the distribution of MTBF, MTTR, MTBPM, and MTTPM for each workstation is known.

For each simulation run r, the product mix is predetermined and fixed. Based on the achievement of planned bottleneck capacity utilization, the weekly throughput target, *Or* , is calculated by dividing the bottleneck capacity by the theoretical process time on bottleneck according to the given product mix, as shown in Equation (3-6).

$$
O_r = (cap_{BN} \times U_{BN} \times bat_{BN}) / \sum_{i} (PT_i^{BN} \times \pi_{i,r})
$$
\n(3-6)

The WIP level for each product type  $i$ ,  $L_i$ , is estimated by applying Little's law (Little 1961),  $L_i = \lambda_i \times CT_i$ . The arrival rate,  $\lambda_i$ , and the corresponding cycle time for the specific product type, *CT<sub>i</sub>* require to be estimated first. Since the CONWIP policy is adopted in the system, wafer lots are released into the plant only when the equal quantity of wafers are finished and transferred out. Therefore, the arrival rate for each product type is equivalent to the corresponding throughput rate.

The block-based cycle time estimation algorithm (BBCT), developed by Chung and Huang (1999), has a remarkable performance in cycle time estimation under the throughput target is given. BBCT algorithm estimates the cycle time by cutting the production process into several blocks based on the characteristics of material flow. In each block, the load factor waiting time and batch factor flow time are estimated. Finally, the cycle time of a product is calculated by summing up all the block cycle times related to the process of the corresponding product type. To utilize the advantages of quick response and satisfactory accuracy, BBCT algorithm is applied to estimate the cycle time for each product type in the preliminary analysis.

#### *3.3 The estimation of* α *percentile cycle time*

The simulation horizon is set to be 24 weeks, in which the first 12 weeks are the warm-up period. In order to eliminate simulation errors, 15 replications with different random seeds are run.

For each product type under each product mix, the histogram of the cycle time frequency distribution is plotted to figure out the pattern of the data. The related parameters, such as average, variance, can be determined.

When estimating cycle time for bidding, we concern not only the effects of product mix composition but also the meeting of the planned on-time delivery rate under the environment that product mix changes periodically. In order to tackle the interference in periodical product mix changes, the percentile cycle time,  $CT_i^{\alpha}$ , is derived in order to ensure the achievement of planned on-time delivery rate, %. All the information derived is saved in the database and is the input to production scheduling plan.

# **4. Production Scheduling Module**

The objective of production scheduling module is to establish wafer lot release schedule and the completion timetable under the environment that product mix changes period by period. On the premise of meeting demand forecast, the initial release schedule is prepared by backward planning. In order to ensure the feasibility of the release schedule, cycle time examination of released job orders and workload examination are performed and then to fine-tune the schedule by forward planning.

# *4.1 Setting the initial release schedule by backward planning*

The initial release amount is determined by net demands offset cycle time for planning. The net demand of each product type in one specific time period is calculated by deducing the wafer lot scheduled receipt and the amount of beginning inventory from demand forecast. The scheduled receipt is the job orders that have been released earlier and planned to be finished in the corresponding time period according to the completion timetable.

To achieve the planned on-time delivery rate, 95%, the 95-percentile cycle time is used as the cycle time for bidding. Since the job released to the plant at period one will be finished after the length of cycle time, the first planning week, that can have the demand of product type *i* being matched through our wafer release plan equals the cycle time for bidding away from current date. Thus, the first planning week,  $fpw$ , equals  $\max_i \{Np_i\}$ , and we derive the net weekly demand in the planning horizon starting from the first planning week.

The time unit for cycle time collection in simulation model is in hours while the planning period is in weeks, therefore, the number of time periods covered for producing product type *i* needs to be calculated first. The procedures for deriving the initial release schedule are as follows.

Step 1. Calculate the number of time periods needed,  $N_{p_i}$ , according to the length of 95-percentile cycle time,  $CT_i^{95\%}$ , for each product type *i*.

$$
N p_i = \left\lceil \frac{C T_i^{95\%}}{24 \times 7} \right\rceil, \text{ for each product type } i. \tag{4-1}
$$

- Step 2. Set the initial planning week as  $t = fpw$ .
- Step 3. Calculate net demand,  $DN_{i,t}$ , and expected ending inventory,  $EI_{i,t}$ , for each product type  $i$  in the planning period  $t$ .

$$
DN_{i,t} = \max[(DF_{i,t} - EI_{i,t-1} - SI_{i,t}), 0], \text{ for each product type } i. \tag{4-2}
$$

$$
EI_{i,t} = \max[(EI_{i,t-1} + SI_{i,t} - AD_{i,t}), 0], \text{ for each product type } i.
$$
 (4-3)

$$
t = t + 1. \tag{4-4}
$$

Step 4. Repeat steps 3 until  $t > fpw + h$ .

Step 5. Calculate the planned release amount for product type *i* at planning period  $t - Np_i$ ,  $PR_{i,t-Np_i}$ , equals net demands,  $DN_{i,t}$ , offset by numbers of time period,  $Np_i$ .

$$
PR_{i,t-Np_i} = DN_{i,t}
$$
, for each product type *i* and for time period  $t = fpw$  to  $fpw + h$ 

$$
(4-5)
$$

Step 6. In each planning period, compare the planned release amount of each product family to the quantity limit for the corresponding product family,  $Q_f^{\max}$ . If the planned release amount is larger than  $Q_f^{\text{max}}$ , we need to fine-tune the planned released quantity in the successive planning period.

#### *4.2 Planning the wafer release date*

The lot release timing is related to the system WIP level under the CONWIP release mechanism. In the environment of product mix changes, the weekly planned throughput could also vary because of the need to maintain the bottleneck utilization. When using Little's Law (Little 1961) to estimate the suitable WIP level for each product  $i$ , the system WIP level may consequently change with the time period because the arrival rate for product  $i$  is assumed to be equivalent to the corresponding throughput rate.

For each time period *t*, Little's Law (Little 1961),  $L_{i,t} = \lambda_{i,t} \times CT_i^{\alpha}$ , is applied to determine the suitable WIP level for each product *i*, where  $\lambda_i$  is the average throughput rate and  $CT_i^{\alpha}$  is the corresponding percentile cycle time for the specific product type. The estimated system WIP level,  $L_t$ , is calculated by summing up all the estimated values of  $L_{i,t}$ .

To keep production smoothly, the system WIP level requires to be properly distributed to all layers of each product type such that the fabrication of wafers will not be concentrated on some specific layers. When the predetermined system WIP level changes periodically, if we release wafer lots only according to the time when the WIP level is lower than the determined value, the first layer of each product type could increase or decrease drastically in the beginning of each time period. Material flow is disturbed in consequence. For solving such a problem, two boundaries are set in the release policy: with the consideration of system WIP level and the WIP level of first layer. Under the revised CONWIP policy, a fixed number of wafer lots, with the product type being determined by the largest accumulated unreleased quantity, will be released into the plant only when the system WIP level is lower than the planned system WIP level,  $L_t$ , and the WIP level of first layer is lower than the planned WIP level of first layer,  $L_t^1$ .

$$
L_{i,i}^{l} = L_{i,t} \times \frac{LPT_{i,l}}{PT_{i}}, \text{ for each layer } l \text{ and each product type } i.
$$
 (4-6)  
where  $\sum_{l} LPT_{i,l} = PT_{i}.$ 

# *4.2.1 The planning of release timing and sequence*

Once the system WIP level is lower than  $L_t$  and the WIP level of first layer is lower

than the  $L<sup>1</sup>$ , under the revised CONWIP policy, six lots of the product type which has the largest accumulated unreleased quantity is released into the plant. The calculation of "accumulated unreleased quantity" is based on the planned daily release amount. When the product type is assigned to releasing, six lots, the release batch size, is deducted from the corresponding unreleased quantity. On the other hand, if there are remaining quantities not released to the plant, the unreleased quantities will be accumulated to the next day.

After the release sequence is determined, the release schedule can then be derived.

### *4.3 Cycle time examination*

As product mix level affects the length of cycle time, to protect the commitment due date of earlier released job orders from the disturbance of new product mix, the cycle times of those job orders are examined to ensure that their new completion date will not be later than the date recorded in the completion table.

For each earlier released job orders, the layers need to be processed in the following period with newly plan are first identified. We then estimate the updated layer cycle time based on the corresponding product mix. So as to estimate the completion time of the job order by summing up the remaining layer cycle times. If the completion time is later than the original record, we need to fine-tune the planned released quantity in the successive planning period.

#### *4.4 Bottleneck workload examination*

In wafer fabrication, wafer lots released to the plant at one time spot will induce workloads in several time periods. To accomplish the planned bottleneck utilization and to have a leveling bottleneck workload, bottleneck workload is examined for each planning period. The circuit layer segmentation concept is adopted to estimate the bottleneck workload. Restated, the last operation of each layer is processed on the bottleneck. Based on the information of release date and layer cycle time, the timing and bottleneck capacity required by each job orders can be easily estimated. Finally, the daily capacity required can be estimated by summing up the required bottleneck capacity in the corresponding date. If the bottleneck workload is not satisfied, we also need to fine-tune the planned released quantity from the successive planning period.

### **5. Conclusions**

Quick response to customers' fluctuating demand is one of the critical issues for market competence. This research has presented a planning system, which comprises two modules, for the make-to-stock environment with volatile demand. With the preliminary analysis, the quantity limit for product family is set to provide a fast check of the feasibility of the product

mix. In addition, mean cycle time and cycle time distribution have analyzed for all the possible different product mix choices so as to decide the cycle time for bidding.

The production scheduling module is to plan a release schedule and completion time table with the considerations of demand varying, production smoothing and due date protection for released job orders. The output of this module includes the suitable WIP level, release amount for each product type, release timetable, and completion time table. Such information is very helpful in demand management and production control. The example cases showed that the proposed planning system can effectively solve the product mix setting problem for the make-to-stock wafer fabrication under demand fluctuating environment. Future research can stress on the product mix optimization for each planning period under demand fluctuating make-to-stock environment.

# **REFERENCES**

- [1] Burman, D. Y., Gurrola-Gal, F., Nozari, J. A., Sathaye, S. and Sitarik, J. P., 1986, Performance analysis techniques for IC manufacturing lines. AT&T Technical Journal, 46-56.
- [2] Chou, Y. C., and Hong, I. H., 2000, A methodology for product mix planning in semiconductor foundry manufacturing. IEEE Transactions on Semiconductor Manufacturing, 13(3), 278-285.
- [3] Chu, S. C. K., 1995, A mathematical programming approach towards optimized master production scheduling. International Journal of Production Economics, 38, 269-270.
- [4] Chung, S. H., and Huang, H. W., 1999, The block-based cycle time estimation algorithm for wafer fabrication factories. International Journal of Industrial Engineering, 6(4), 307-316.
- [5] Dümmler, M. A., 2000, Analysis of the instationary behavior of a wafer fab during product mix changes. Proceedings of the 2000 Winter Simulation Conference, 1436-1442.
- [6] Hackman, S. T., and Leachman, R. C., 1989, A general framework for modeling production. Management Science, 35.
- [7] Hood, S. J., Bermon, S., and Barahona, F., 2003, Capacity Planning Under Demand Uncertainty for Semiconductor Manufacturing. IEEE Transaction on Semiconductor Manufacturing, 16(2), 273-280.
- [8] Hung, Y. F., and Leachman, R. C., 1996, A production planning methodology for semiconductor manufacturing based on interactive simulation and linear programming calculations. IEEE Transaction on Semiconductor Manufacturing, 9, 257-269.
- [9] Little, J. D. C., 1961, A proof for the queueing formula *L= W*, Operations Research, 9, 383-387.
- [10] Liu, C., Thongmee, S., and Hepburn, P., 1995, A methodology for improving on-time delivery and load levelling starts. Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 95-100.
- [11] Tecnomatix Technologies Ltd., 2000, eM-Plant Objects Manual, Tecnomatix Software Company, Germany.
- [12] Thompson, S. D., and Davis, W. J., 1990, An integrated approach for modeling uncertainty in aggregate production planning. IEEE Transactions on Systems, Man, and Cybernetics, 20, 1000-1012.

 $(2/3)$ 

 $\sqrt{ }$ NSC 92-2213-E-009-044 92 08 01 93 07 31

 $($  ; in the set of the

 $\Box$ 





 $3.$ 

**2.**本項研發成果若尚未申請專利,請勿揭露可申請專利之主要內容。