Family-Based Scheduling Rules of a Sequence-Dependent Wafer Fabrication System

Ching-Chin Chern and Yu-Lien Liu

Abstract—**Consider the dispatch problem for a wafer fabrication where production process is divided into hundreds of operations and takes a few months to complete. In the process, wafers have to go through similar operations several times repeatedly for different layers of circuit, called job re-entrances. General family-based scheduling rules are proved to perform better than the individual job scheduling rule in terms of machine utilization for multiserver and multiple job re-entrance manufacturing systems under the condition that with a positive possibility, a queue exists in front of steppers. Five special family-based scheduling rules are constructed, of which FCFS-F, SRPT-F, EDD-F, and LS-F are modified from previous well-known scheduling rules, while SDA-F is a rule-based algorithm, using threshold control and least slack principles and being developed in this research. A simulation model is built to evaluate the performances of these five family-based rules by using the information collected from a wafer fab located in Hsin-Chu, Taiwan. As a result, SDA-F is shown to perform best among all five rules, followed by LS-F and FCFS-F.**

Index Terms—**Dispatching rule, family-based scheduling, sequence-dependent manufacturing, wafer fabrication.**

I. Introduction

THE PRODUCTION of integrated circuit (IC) chips is a very complex process. The first stage of IC production is usually referred to as wafer processing or wafer fabrication, which is different from general production processes in such a way that producing one wafer needs hundreds of operation and takes a few months to complete. The basic operations for wafer fabrication are diffusion, thin film deposition, etching, photolithography, and ion implant. The bottleneck is photolithography process including operations such as coating, photomasking, and developing, of which photomasking operations performed on steppers receive most attention because steppers are very expensive tools. To apply capital effectively, most wafer fabrications schedule the work on steppers efficiently to avoid idleness. Each photomasking operation requires a different mask and mask changes are troublesome and time-consuming tasks. In addition, the factors of wafer re-entrances, equipment downtime, random yields, and so on make scheduling and dispatching in wafer fabrication particularly difficult as elaborated in Uzoy *et al.* [13], [14] as well as in Hughes and Schott [7] and Bai and Gershwin [2]. A thorough description of wafer fabrication and its specific modeling and analysis problems are seen in Johri [8].

From this point of view, to have an effective scheduling and dispatching algorithm is urgent for wafer fabrication. The discussion of general scheduling and dispatching algorithms could be found in [5], [9], and [12] though they are not directly related to wafer fabrication. Wein [15] uses a Brownian queueing network model to approximate a multiclass queueing network model with dynamic control capability which emulates wafer fabrication. However, they did not consider the effect of family-based rules on the setup time and their conclusion that input control methods are more effective than dispatch rules is obtained from a large development laboratory which is different from the regular wafer manufacturing processes. Adachi *et al.* [1] adopt a rule-based control method in a decision support expert system, called a rule-based decision support system (RBDSS), whose construction requires the involvement of experts from different areas and thus are suitable only for a few specific environments. Glassey and Petrakian [4] develop a dispatch rule based on the principle of starvation avoidance. However, their algorithm continuously adjusts the priorities of the jobs waiting before work units which results in requirements of large capacity and longer processing time of a computer unit. Lu and Kumar [10], on the other hand, evaluate buffer-based policies versus due-date-based policies through a control theory to develop the bounds of delays for a wafer fabrication plant. They use simulation in their study to compare these two types of policies. Cheng and Liao [3] develop a three-layer real-time scheduling and dispatching control framework for a semiconductor wafer fab. They adopt Lagrangian Relaxation and Network flow techniques to solve this problem, while no setup effect is considered in their model. In all of the above studies, simulation is used as a tool to either evaluate the algorithms or justify their arguments. This research is different from the above studies in ways that it includes setup factor, which is ignored in [3], [4], [10], and [15]. This research also modifies the buffer-based and due-date-based scheduling rules and re-evaluates their effects as done in [10] as well as utilizes the concept of a rule base as in [1] and develops a heuristic family-based scheduling rule.

As mentioned earlier, steppers are the bottleneck of the whole wafer fabrication and should be utilized as efficiently as possible, which implies that the reduction of mask changes is crucial for wafer fabrication. Wemmerlöv and Vakharia [16] compare the effect of setup reduction between individual job scheduling and family scheduling of a flow line manufacturing cell. They conclude the family-based scheduling performs

Manuscript received February 15, 2001; revised May 9, 2002. This work was supported by the Taiwan National Science Committee under Project NSC 88-2416-H-002-056.

C.-C. Chern is with the Department of Information Management, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: chern@chern.im.ntu.edu.tw).

Y.-L. Liu was with Texas Instruments-Acer Incorporated, Hsin-Chu Science-Based Industrial Park, Taiwan, R.O.C. She is now with Brooks-PRI Automation, Hsin-Chu Science-Based Industrial Park 300, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TSM.2002.807742

0894-6507/03$17.00 © 2003 IEEE

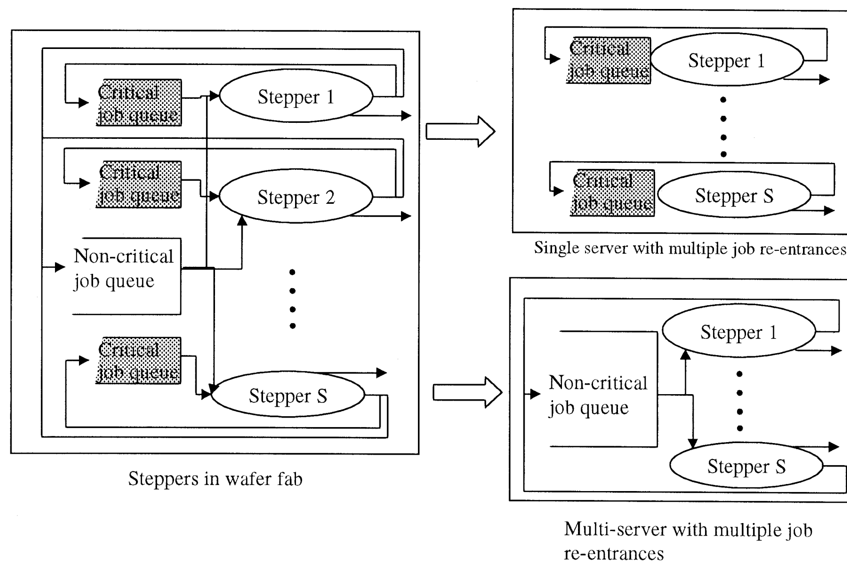


Fig. 1. Wafer fab as two manufacturing systems.

better under most scenarios. However, their result is based on a shop flow system, which is very different from wafer fabrication, in terms of makespan computations and multiserver scheduling. Missbauer [11] studies the schedule of the jobs with sequence-dependent setup times in a one-stage one-machine system and concludes a potential savings in setups if an appropriate family-based sequencing rule is used. In light of the above two studies, this work adopts and modifies the family-based scheduling rules for a multiserver and multiple job re-entrance system such as wafer fabrication by minimizing the numbers of photomask changes.

The rest of the paper is organized as follows. Section II describes the problem and characterizes a general family-based scheduling rule. Section III presents several modified family-based scheduling rules. Section IV illustrates the simulation model and displays the results of simulations. Section V summarizes the results, finally, and Appendix A demonstrates the proofs of theorems shown in Section II.

II. A General Family-Based Scheduling Rule

A typical production cycle of a wafer starts with nonphotolithography processes such as diffusion, thin film deposition, or etching. Once the wafer enters photolithography area, it is coated, photomasked, and developed. The wafer is then transferred to nonphotolithography area and stays there for some other processes such as ion implant, diffusion, and etc. The wafer returns to the photolithography area for the second-layer circuit and so on. The production cycle of a wafer stops when all the required layers of circuit are properly produced. Each layer of circuit needs to go through photomask operations once, which are processed on steppers. Two types of photomasking operations are usually performed on a stepper: critical or noncritical. Critical operations of a wafer are performed on the same steppers by requirement while noncritical operations do not have this restriction. For instance, photomasking operations 2, 3, 4, and 5 are critical operations and 6 is noncritical. That is, if operation 2 of a wafer is performed on stepper 2, operations 3, 4, and 5 of this particular wafer are performed on stepper 2 by requirement, while operation 6 could be done on any available stepper. All steppers are assumed to have the capabilities of performing either critical, noncritical, or both operations. Each layer of circuit needs a distinguishing photomask, which implies that whenever steppers process wafers for different layers of circuit from the previous ones, they have to change photomasks. For example, a stepper just finishes operation 2 for a lot of wafers. The stepper needs to change photomask if it is to process operation 4 for other lot of wafers, while it needs not to change the mask if it continues processing operation 2 for another lot of wafers.

From the above description, a wafer fab could be simplified to a manufacturing system with more than one machine as seen in Fig. 1. Like in Missbauer [11], consider two scheduling rules: FCFS, which puts no control on the job sequence except arrival time, and the family-based scheduling rule, which puts the job of the same type ahead of other jobs. Missbauer [11] has proven that the family scheduling rule performs better in saving the setup times for a single-server system. This study will expand the result to a more complicated system with multiservers and multiple job re-entrances. Define i as the job types where I is the total number of types and j as the index of visit to steppers where J_i is the total number of visits of type i. Before the theorems are stated, define the following state variables for a system with multiservers and multiple job re-entrances, or namely, some visits to steppers are critical and some are noncritical.

P_{FCFS}: Probability that a setup is necessary if we apply FCFS rule.

U_{FCFS}: Actual utilization rate if we apply FCFS.

P_{FSR}: Probability that a setup is necessary if we apply the family-based scheduling rule.

U_{FSR}: Actual utilization rate if we apply the family-based scheduling rule.

Assume the arrival of each job type follows a Poisson process with arrival rate λ_i for type i. Thus, the arrival rate of type i to steppers is $J_i\lambda_i$ and the total arrival rate is $\lambda = \sum_{i=1}^{I} J_i\lambda_i$. Assume that $\bar{J}_i$ is the total number of critical visits and J_i' is the total number of noncritical visits of type i. Thus, $J_i = \bar{J}_i + J_i'$. Further, assume that there are totally S steppers to be utilized. The following theorems are stated here first and proved later in Appendix A.

Theorem 1: $P_{\mathrm{FSR}} \leq P_{\mathrm{FCFS}}$. This theorem simply states that the family-based scheduling rule can never increase the probability that a setup is necessary for FCFS sequences.

Theorem 2: $P_{\mathrm{FSR}} < P_{\mathrm{FCFS}}$ if $\sum_{i=1}^{I} J_i > 1$, $\lambda_i > 0$ for all i, and $p_n > 0$ for at least one $n > S$ where p_n is the probability with n jobs in the system.

This theorem implies that the probabilities of necessary setups can be reduced if applying family-based scheduling rules, more than one product type exists, total number of re-entrances >1, and for a positive possibility, a queue exists in front of the steppers.

Let a_{ij} be the processing time of type i for the jth visit and r be the setup time for each group change, or in this case, each photomask change. Define δ as the mean duration of a job (including service and setup) on a stepper when a setup is necessary for each job, or $\delta = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \lambda_i/\lambda(a_{ij} + r)$. However, not each job needs a setup when applying FCFS or the family-based scheduling rule. The probabilities of setup savings from FCFS and from family-based rule are $1 - P_{\mathrm{FCFS}}$ and $1 - P_{\mathrm{FSR}}$, respectively. Therefore, the mean duration of a job on steppers for FCFS and for the family-based scheduling rule are $\delta - r(1 - P_{\mathrm{FCFS}})$ and $\delta - r(1 - P_{\mathrm{FSR}})$. Since $U_{\mathrm{FCFS}} = \lambda[\delta - r(1 - P_{\mathrm{FCFS}})]$, $U_{\mathrm{FSR}} = \lambda[\delta - r(1 - P_{\mathrm{FSR}})]$, and $P_{\mathrm{FSR}} < P_{\mathrm{FCFS}}$, the utilization rate of applying the family-based scheduling rule is higher than one of applying the FCFS rule, or equivalently, $U_{\mathrm{FCFS}} < U_{\mathrm{FSR}}$, under the conditions mentioned in Theorems 1 and 2. When examining wafer fabrication, the required conditions mentioned above are met, especially in photolithography processes. Therefore, a general family-based scheduling rule for steppers can be constructed based on the above theorems to reduce the numbers of photomask changes. Define CAS as "current available set," or "current photomask family" in wafer fabrication, which includes the job families waiting in the queues that need the same photomask and can be processed on certain steppers for critical operations or all steppers for noncritical operations.

A General Family-Based Scheduling Algorithm:

1. Schedule all jobs in front of steppers that needs the same photomask together as job families. Each stepper has its own critical job queue and there is also a joint noncritical job queue for all steppers. Place the job families into different queues according to their designated steppers and their criticality.
2. When a job, or a wafer, arrives to the queues, check whether any family the job belongs to is presented in the queues. If yes, the job is placed in a certain position of that family based on the "special family-based scheduling rules." If no, the job is placed at the end of the queues to start another job family.
3. When a stepper is available, check whether there are jobs belong to CAS. If yes, assign the job to this stepper. Otherwise, assign the jobs based on "special family-based scheduling rules."

In this algorithm, "special family-based scheduling rules" applied in Steps 2 and 3 need to be developed next.

III. Special Family-Based Scheduling Rules

In this section, five special family-based scheduling rules are proposed for wafer fabrication, of which the first four are modified from the previous studies such as [10] and [16]. The last one, SDA-F, is developed in this research and is currently adopted by a large wafer fab in Taiwan. Assume that each job has an assigned due date at its first arrival. Define slack of job k, S_k, as the due date minus current time and total processing time for all the remaining operations including operations in photolithography and nonphotolithography areas.

- Family-Based First Come First Serve (FCFS-F): FCFS-F is modified from the general FCFS rule, which schedules jobs according to their arrival times in the job family. In addition, FCFS-F processes the job families based on FCFS principle when certain steppers are available and no CAS presents.
- Family-Based Shortest Remaining Process Time (SRPT-F): SRPT-F schedules jobs in ascending order of their remaining processing times in the job family. When certain steppers are available and no CAS presents, SRPT-F processes the job families based on the shortest remaining processing time first principle, namely, the job family with the shortest remaining processing time is processed first. If more than two job families have the same shortest remaining processing time, FCFS is used to be the tie-breaking rule.
- Family-Based Earliest Due Date (EDD-F): EDD-F schedules jobs in ascending order of their due dates in the job family. When certain steppers are available and no CAS presents, EDD-F processes the job families based on the Earliest Due Date principle, namely, the job family with the earliest due date is processed first. FCFS is also the tie-breaking rule.
- Family-Based Least Slack (LS-F): LS-F schedules jobs in ascending order of their slacks S_k in the job family. When certain steppers are available and no CAS presents, LS-F processes the job families based on the Least Slack principle, namely, the job family with the least slack is processed first. FCFS is the tie-breaking rule.

The last family-based scheduling rule, Family-Based Stepper Dispatch Algorithm (SDA-F), is specially developed for scheduling jobs on steppers. SDA-F is also a rule-based algorithm which not only adopts the rule-based scheduling concept from [1] but also includes threshold control and least slack principles. The development of SDA-F is based on the following two rationales: 1) The critical photomask operations need to come back to the same steppers, and thus, should be assigned higher priorities than noncritical ones; and 2) Among critical photomask operations, the one with longer processing time should have more savings on setup than ones with shorter processing time since

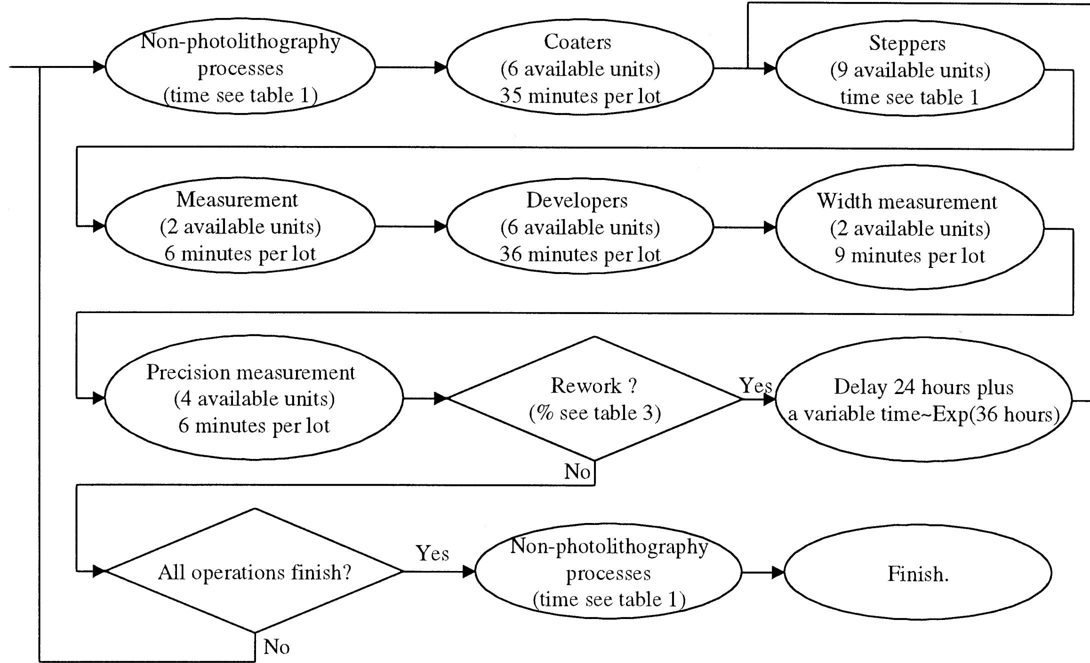


Fig. 2. Operations of the simulation model.

longer processing time results in more arrivals, which in turn, results in more jobs waiting in front of steppers. From Theorem 2, more $p_n > 0$ for $n \geq S$ results in larger differences between P_{FSR} and P_{FCFS}. As a result, critical operations are divided into two groups based on w, the threshold of photomask processing time. The processing times of critical operations of group 1 are all greater than w while those of group 2 are less than or equal to w. From the simulation experiments, the rule of thumb to choose w is that no more than 30% of total critical operations are assigned to group 1. Based on the above grouping result, the steps of SDA-F are as follows.

1) Schedules jobs in ascending order of their slacks in the job family.
2) When certain steppers are available and no CAS presents, SDA-F processes the job families based on the sequence of group 1 first, followed by group 2, and finally the noncritical jobs. In other words, the job families of group 1 are chosen to be processed first, those of group 2 second, and noncritical jobs last, based on SDA-F. Within each group, the job family with least slack is processed first.
3) Finally, FCFS is the tie-breaking rule.

IV. Simulation Model and Analysis

MAXIM, a platform specially designed for wafer fab is used to build the simulation model to evaluate the five special family-based scheduling rules. The operations, resources, and time information needed in this simulation model are shown in Fig. 2 and are collected from a wafer fabrication located in Hsin-Chu, Taiwan, R.O.C. Two types of products, L and S, are assumed in this simulation model with 60% and 40% shares of production, respectively. Each product has to go through similar processes but takes different lengths of processing times on the stepper and in the nonphotolithography area as seen in Table I. More parameters and assumptions made in this simulation model are listed as follows.

TABLE I
Times of Operations on Stepper and Nonphotolithography Areas for All Lot j (in Min)

Operation	Stepper processing time		Non-photolithography time		Critical ?
	For L	For S	For L	For S	
0			*14400*	*14400*	
1	*34*	*38*	*7200*	*7440*	
2	*31*	*36*	*7560*	*8040*	Yes
3	*32*	*42*	*10080*	*10500*	Yes
4	*33*	*36*	*10560*	*11280*	Yes
5	*32*	*45*	*3210*	*3350*	Yes
6	*35*	*37*	*5050*	*5220*	No
7	*35*	*47*	*3890*	*3600*	Yes
8	*35*	*42*	*5350*	*5820*	No
9	*34*	*40*	*4220*	*3960*	Yes
10	*34*	*36*	*14760*	*13610*	No

- The outcome of each finished photomasking operation is an independent Bernoulli trial with reworking probabilities listed in Table II and is either good or rework but not scraped.
- The time between machine failures and repair time for each stepper are both assumed to be random variables following exponential distributions with means listed in Table III. Each machine is also scheduled for regular maintenance as seen in Table IV.
- In a wafer fabrication process, special requirements for certain products with various quantities, called hot lot in a wafer fab, arise occasionally. Hot lots are assigned highest priorities, according to the engineering principle in general wafer fabs. In this simulation model, hot lot percentages are 0.07% for product L and 0.1% for product S.

TABLE II
REWORK RATE (IN %)

Operation	For L	For S
1	1.0	1.2
2	1.2	1.2
3	3.2	6.0
4	1.2	0.6
5	4.0	14.8
6	2.2	2.4
7	1.1	1.2
8	1.6	0.0
9	0.6	1.2
10	2.9	0

TABLE III
MACHINE FAILURE OF STEPPERS (IN H)

Stepper	Mean time between failures	Mean repair time
1	145	1.3
2	388	0.5
3	211	0.6
4	138	0.3
5	388	0.5
6	482	0.1
7	482	0.1
8	63	2.2
9	96	0.6

TABLE IV
REGULAR MAINTENANCE OF WORK UNITS IN PHOTOLITHOGRAPHY AREA

Work units	Type of regular maintenances	Interval (in days)	Repair Time (in minutes)
Coater	1	60	120
Developer	1	60	120
Measurer for width	1	1	10
	2	30	60
	3	180	480
Measurer for precision	1	7	20
	2	30	60
Stepper	1	30	84
	2	90	150
	3	180	240

- Four photomasks are available for each photomasking operation. Usually, the mask changing time is ranged from 6–20 min. In this model, mask changing time is assumed to be 6 min.

To validate the simulation model, 31 lots are input into the system every day with lot size equal to 25 wafers, of which 18.6 lots are product L and 12.4 lots are product S. The time between each input is a constant of 45 min. The simulation model is run for 360 days with a 90-day warmup period since after 90 days (or 3 months), the process becomes stable as seen in Fig. 3(a)–(c). The scheduling rule used for this simulation run is FCFS-F for all resources, including coaters, steppers, measurements, and developers, as used in the wafer fab located in Hsin-Chu, Taiwan, R.O.C., before SDA-F is adopted. Table V shows the information collected and summarized from the plant and from the simulation run. The average monthly throughput, the average WIP, and the mean cycle time all show that the simulation is close to the current plant situations. Therefore, this simulation model can be used to evaluate these family-based scheduling rules.

It is also interesting to see how close the approximations of P_{FCFS} and P_{FSR} computed in Appendix A are to the simulation results. Ten visits to steppers, including six critical and three noncritical visits, are assumed for each lot as seen in Table I. The average processing time for each photomask operation is 36.06 min with nine steppers available. Ignoring the maintenance and rework factors and based on the approximations in Appendix A, $P_{\text{FCFS}} \approx 93.61\%$, $P_{\text{FSR}} \approx 71.33\%$, $U_{\text{FCFS}} \approx 99.69\%$, and $U_{\text{FSR}} \approx 96.49\%$, respectively. As mentioned in Appendix A, P_{FCFS} and P_{FSR} are computed by adopting the formula from $M/M/1$ and $M/M/S$ models. Based on the properties of the functions as explained in the Appendix, they are overestimated, which means that P_{FCFS} should be greater than 93.61% and P_{FSR} should be greater than 71.33%. From the derivation in Section II, $U_{\text{FCFS}} = \lambda[\delta - r(1 - P_{\text{FCFS}})]$ and $U_{\text{FSR}} = \lambda[\delta - r(1 - P_{\text{FSR}})]$. Therefore, the overestimations of P_{FCFS} and P_{FSR} also result in the overestimations of U_{FCFS} and U_{FSR}. From the simulation, the average utilization rate of steppers after the warmup period is 93.70% without the machine down time which is almost 3% lower than the approximation result or $U_{\text{FSR}} = 96.49\%$ as expected. However, it is still clear that family-based scheduling rule performs more than 20% better in terms of setup probability. Since a 6-min mask changing time is assumed, the savings in setup reflects more than 3% in utilization. That is, the larger the setup time is, the greater the percentage of savings in utilization will be, which implies larger throughput if varying arrival rates or shorter cycle time in fixing arrival rate.

From the above analysis, family-based scheduling rules can always increase the utilization rate through the savings in setup time. More savings in setup result in more job processing capacities, which means that different job arrival rates should be applied for different rules. For comparison reasons, the average utilization rate is set at around 95% for all the simulation runs of these five rules while theoretical results in Section II and a trial-and-error method are used to find the appropriate arrival rates for different rules. It is also assumed that the special family-based rules are applied to steppers as well as all other resources. The measurements to evaluate the performance of these five family-based scheduling rules are average weekly throughput (in wafers) and mean cycle time per wafer (in hours). Fig. 4 shows the time plots of average weekly throughput rate and cycle time per wafer of these five rules. It should be noted that SDA-F is the one with larger weekly throughput and lower cycle time among all. The reason is that SDA-F saves more setup time, or photomask changing time, on steppers and thus, better utilizes the steppers in processing wafers. It is also clear that the weekly throughputs from SRPT-F and from EDD-F are less stable than the weekly throughputs from the other rules. It is caused by the principle adopted by these two rules that the jobs get higher priorities as jobs get close to the ending processes. When the system is highly utilized, SRPT-F and EDD-F tend to schedule the lots that has been processed ahead of the new entry lot. Thus, the system has a higher weekly throughput for one week followed by a lower one the next week and vice versa when applying these two rules. It is also worth mentioning that

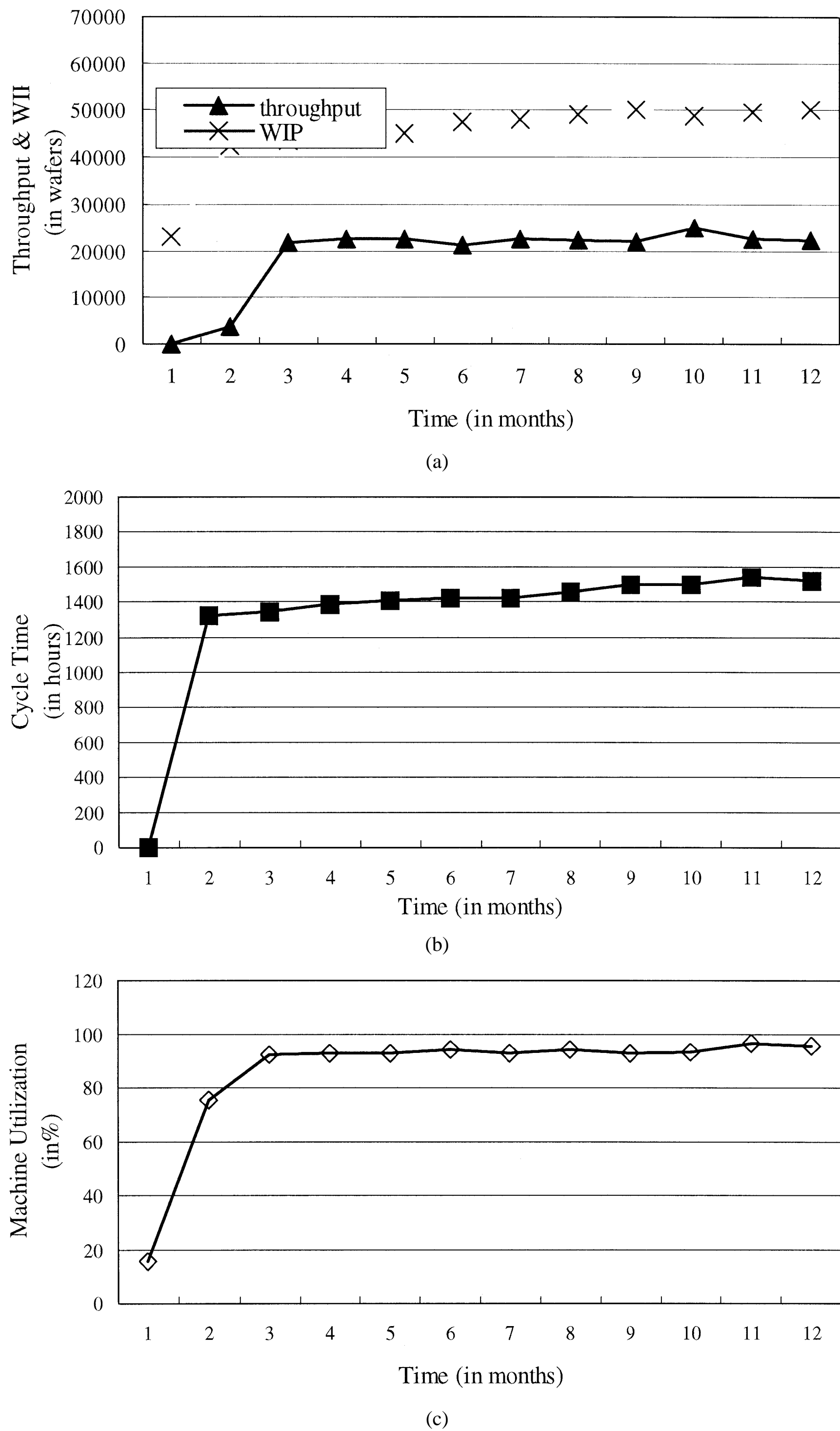


Fig. 3. (a) Time plots of throughput and WIP. (b) Time plot of cycle time. (c) Time plot of utilization.

TABLE V
DATA COLLECTED FROM THE SAMPLED PLANT AND THE SIMULATION

Sources	Average WIP	Average Monthly Throughput	Mean Cycle Time
Plant	53000	22000	68
Simulation	50000	23000	65

FCFS-F, the rule currently used in the fab located in Hsin-Chu, Taiwan, R.O.C., and LS-F performs almost the same and they are second best among these five rules, following SDA-F. To further examine these five rules statistically, ten runs of simulation are done for each rules. The means and standard deviations of the ten-run average weekly throughputs and mean cycle times among five rules for two products are plotted in Fig. 5(a)–(d). It is clear that the reduction in cycle time or increase in weekly throughput is significantly better from SDA-F to the others for both products.

After the results of this study presented to the management of this wafer fab located in Hsin-Chu, Taiwan, R.O.C., SDA-F was adopted immediately in 1998. It was estimated that over 120

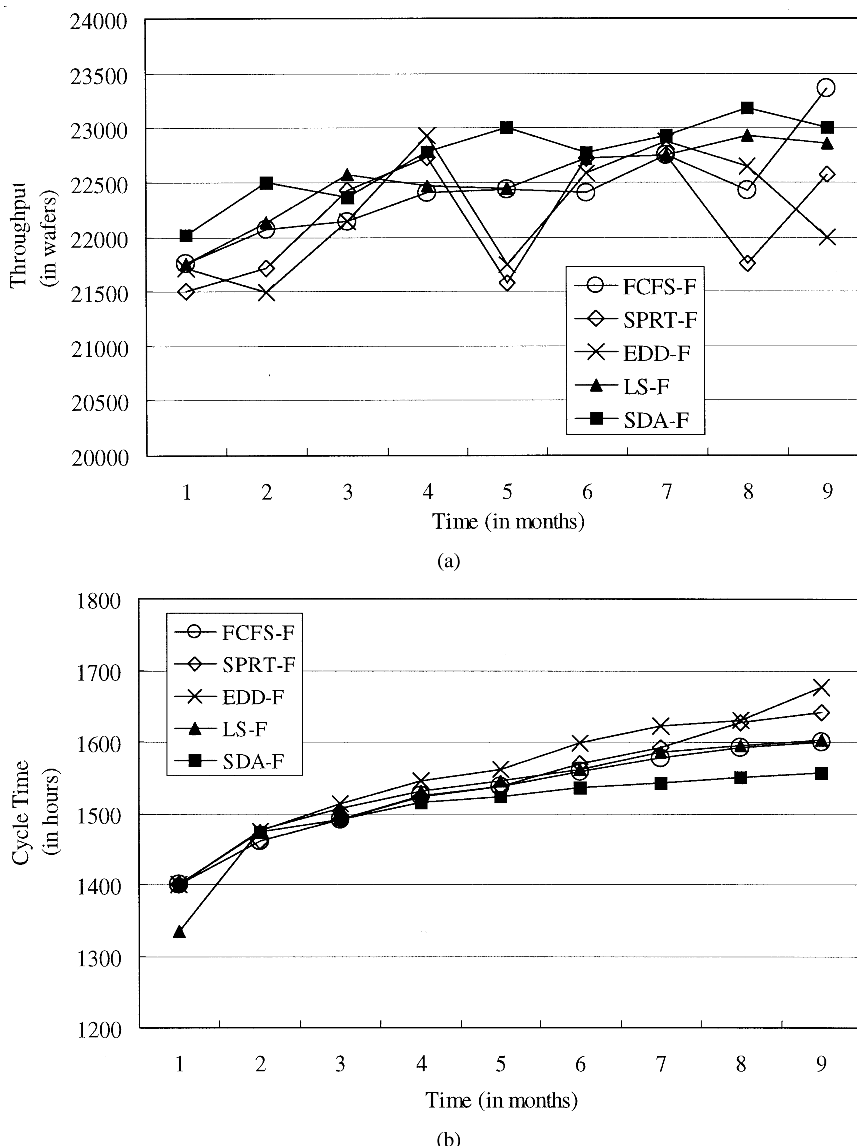


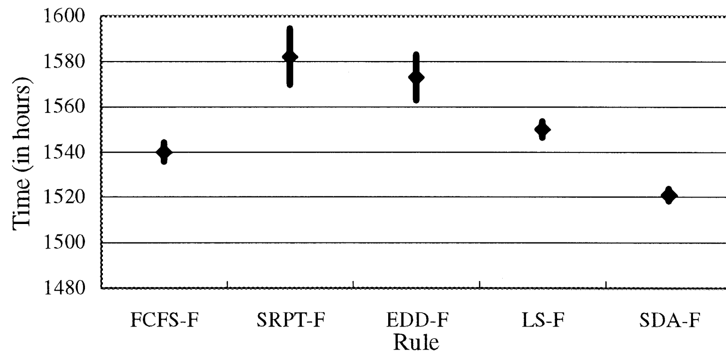
Fig. 4. (a) Time plots of weekly throughput. (b) Time plots of cycle time.

wafers or five lots more were produced per week, which is more than half of the difference between the results of previous plant situation and the simulation. This wafer fab was later merged with another IC chip manufacturer and another DRAM manufacturer in Taiwan, R.O.C., has adopted the SDA-F rule since then. There is no information reported regarding to how much it saves or earns after the plant adopts SDA-F.

V. Conclusion

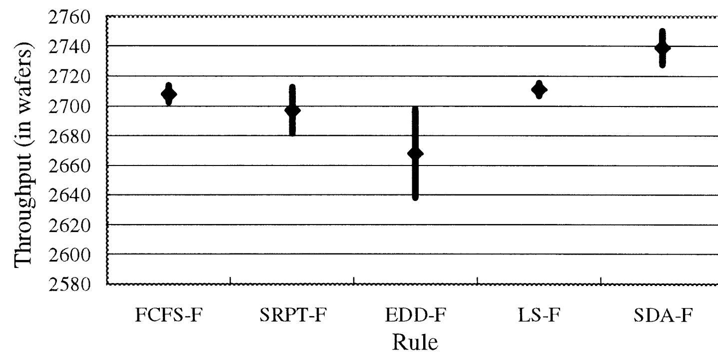
This research studies the effect of applying family-based scheduling rules to wafer fabrication and concludes that they increase the utilization of steppers if there is a positive possibility of at least a job waiting for steppers. Five family-based scheduling rules are introduced with the first four, FCFS-F, SRPT-F, EDD-F, and SLACK-F, being modified from well-known scheduling rules and the last one, SDA-F, being constructed in this study. SDA-F is a family-based scheduling rule that benefits from setup time reduction, or photomask change reduction, by including threshold control schemes and least slack principles. A simulation model is created to evaluate the performances of these five family-based scheduling rules. A wafer fab located in Hsin-Chu, Taiwan, R.O.C., is used as an example in this study. The data needed in the simulation model is collected from this wafer fab. As a result of this work, SDA-F performs significantly better than other rules when comparing the average weekly throughput and mean cycle time. Some assumptions made in this study to simplify the model might be relaxed for future research, of which is one assuming that all steppers can process both critical and no-critical operations. In many wafer fabs, only the most recent fine-line tools can handle critical layers, so that there are different pools of tools to handle the noncritical versus no-critical operations. The family-based scheduling rule can still apply to this type of system based on the proof in Appendix. However, it depends on the future research to find the suitable special family-based scheduling

Rules	Mean	Stdev
FCFS-F	1540	4.2
SRPT-F	1582	12.3
EDD-F	1573	9.9
LS-F	1550	3.5
SDA-F	1521	2.7



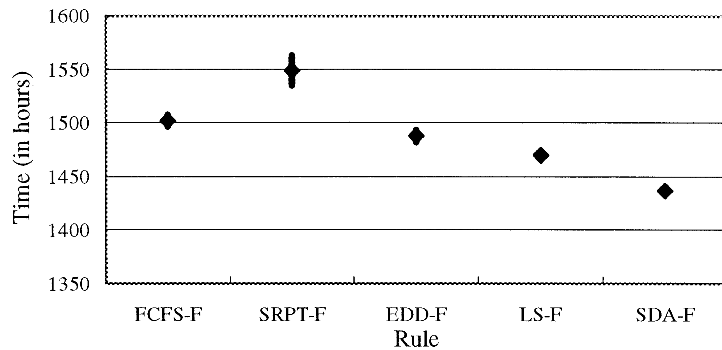
(a)

Rules	Mean	Stdev
FCFS-F	2708	5.6
SRPT-F	2697	15.6
EDD-F	2668	30.1
LS-F	2711	4.3
SDA-F	2739	11.4



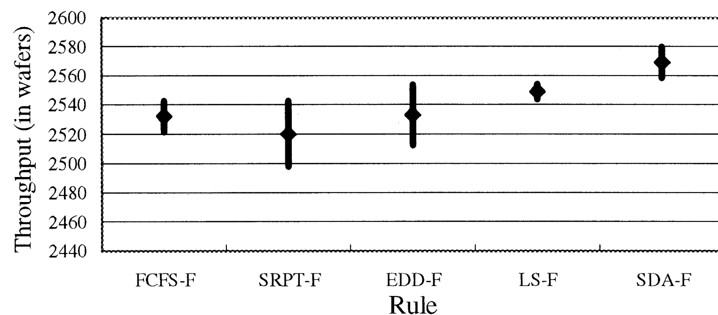
(b)

Rules	Mean	Stdev
FCFS-F	1502	5.2
SRPT-F	1549	14.1
EDD-F	1488	5.8
LS-F	1470	3.6
SDA-F	1437	2.4



(c)

Rules	Mean	Stdev
FCFS-F	2532	10.2
SRPT-F	2520	22.2
EDD-F	2533	20.4
LS-F	2549	5.3
SDA-F	2569	10.7



(d)

Fig. 5. (a) Cycle time per wafer for product L. (b) Weekly throughput for product L. (c) Cycle time per wafer for product S. (d) Weekly throughput for product S.

rule for this type of environment. Another assumption to simplify the model is for all the resources, including machines in photolithography and nonphotolithography areas, to follow the same dispatching rule. However, different working machines have different processing characteristics which may need different dispatch principles. For example, steppers process wafers one by one while furnaces for diffusion operations process wafers in batches. It is of future interest to develop a combined family-based scheduling rule for the whole wafer process.

APPENDIX

A. Proof of the Theorems

For the sake of simplicity, assume that the processing rates for all the machines are identical and the processing times of all visits for all jobs are also the same and constant. To simplify the computation, assume that the time between the jth and $(j+1)$th visits of all types follows an independent and identical exponential distribution. The first step is to evaluate a single-server system with multiple jobs re-entrances. Let $\bar{P}_{\rm FCFS}$ be the probability that a setup is necessary if applying the FCFS rule and $\bar{P}_{\rm FSR}$ be the probability that a setup is necessary if applying the family scheduling rule for a single-server system with multiple job re-entrances, or namely, all jth visits to steppers are critical operations. Assume that all J operations are critical. With these, the following lemmas are derived and proved.

Lemma 1: $\bar{P}_{\rm FSR} \le \bar{P}_{\rm FCFS}$.

Proof: From [11], the probability a setup is necessary is when the arrival to have a predecessor is processed for a different operation, or equivalently, $\bar{P}_{\rm FCFS} = \sum_{i=1}^{I} J_i\,(\lambda_i/\lambda)\,(1-\lambda_i/\lambda) = 1 - \sum_{i=1}^{I} J_i\,(\lambda_i/\lambda)^2$ and $\bar{P}_{i,j,{\rm FSR}} = p_0\,(1-\lambda_i/\lambda) + \sum_{n=1}^{\infty} p_n\,(1-\lambda_i/\lambda)^n$ where p_n is the probability with n jobs in the system. Thus

$$\begin{aligned}\bar{P}_{\rm FSR} =& \sum_{i=1}^{I} J_i \frac{\lambda_i}{\lambda} \bar{P}_{i,j,{\rm FSR}}\\ =& \sum_{i=1}^{I} J_i \frac{\lambda_i}{\lambda} \left\{ p_0 \left(1 - \frac{\lambda_i}{\lambda}\right)\right.\\ &\left. + \sum_{n=1}^{\infty} p_n \left(1 - \frac{\lambda_i}{\lambda}\right)^n \right\}. \qquad (6.1)\end{aligned}$$

From (6.1), the maximum of $\bar{P}_{\rm FSR}$ happens when $p_0 = 1$ and $p_n = 0$ for all $n \ge 1$. In other words, $\bar{P}_{\rm FSR} \le \sum_{i=1}^{I} J_i(\lambda_i/\lambda)\,(1-\lambda_i/\lambda) = 1 - \sum_{i=1}^{I} J_i\,(\lambda_i/\lambda)^2 = \bar{P}_{\rm FCFS}$. The result follows. □

Lemma 2: $\bar{P}_{\rm FSR} < \bar{P}_{\rm FCFS}$ if $\sum_{i=1}^{I} J_i > 1$, $\lambda_i > 0$ for all i, and $p_n > 0$ for at least one $n > 1$ where p_n is the probability with n jobs in the system.

Proof: Without losing generality, assume $p_1 > 0$, $p_2 > 0$, and $p_n = 0$ for all $n \ge 3$. If $\sum_{i=1}^{I} J_i > 1$ and $\lambda_i > 0$ for all i, then $(1-\lambda_i/\lambda) > (1-\lambda_i/\lambda)^n > (1-\lambda_i/\lambda)^{n+1}$ for $n > 1$. From the proof of Lemma 1, it is known that $\bar{P}_{\rm FSR} \le \sum_{i=1}^{I} J_i(\lambda_i/\lambda) \left\{p_0\,(1-\lambda_i/\lambda) + p_1\,(1-\lambda_i/\lambda) + p_2\,(1-\lambda_i/\lambda)^2\right\}$, whose right-hand side is less than $\bar{P}_{\rm FCFS} = \sum_{i=1}^{I} J_i\lambda_i/\lambda\,(1-\lambda_i/\lambda)\,\{p_0+p_1+p_2\} = 1 - \sum_{i=1}^{I} J_i\,(\lambda_i/\lambda)^2$. The result follows. □

If p_n is approximated by the probability of $M/M/1$ formula from [6], or $p_n = (1-\rho)\rho^n$, as did in Missbauer [11], (6.1) can be calculated. However, the utilization rate ρ depends on the total amount of setup time, which means ρ depends on $\bar{P}_{\rm FSR}$. Replace $p_n = (1-\rho)\rho^n$ in formula (6.1). Therefore

$$\begin{aligned}\bar{P}_{i,j,{\rm FSR}}(\rho) =& p_0\left(1-\frac{\lambda_i}{\lambda}\right) + \sum_{n=1}^{\infty} p_n \left(1-\frac{\lambda_i}{\lambda}\right)^n\\ =& \frac{\left(1-\rho+\rho\frac{\lambda_i}{\lambda}-\rho^2\frac{\lambda_i}{\lambda}\right)\left(1-\frac{\lambda_i}{\lambda}\right)}{\left[1-\rho\left(1-\frac{\lambda_i}{\lambda}\right)\right]}.\end{aligned}$$

Taking the first derivative on $\bar{P}_{i,j,{\rm FSR}}$ results in

$$\frac{\partial \bar{P}_{i,j,{\rm FSR}}(\rho)}{\partial \rho} = \rho \frac{\lambda_i}{\lambda} \frac{\left[3\rho\frac{\lambda_i}{\lambda}\left(\frac{\lambda_i}{\lambda}-1\right)-2\right]}{\left[1-\rho\left(1-\frac{\lambda_i}{\lambda}\right)\right]^2} \le 0.$$

Thus, if $\rho_1 \le \rho_2$, then $\bar{P}_{i,j,{\rm FSR}}(\rho_1) \ge \bar{P}_{i,j,{\rm FSR}}(\rho_2)$, which implies that $\bar{P}_{\rm FSR}(\rho_1) = \sum_{i=1}^{I} J_i(\lambda_i/\lambda)\bar{P}_{i,j,{\rm FSR}}(\rho_1) \ge \sum_{i=1}^{I} J_i(\lambda_i/\lambda)\bar{P}_{i,j,{\rm FSR}}(\rho_2) = \bar{P}_{\rm FSR}(\rho_2)$. To be able to compute $\bar{P}_{\rm FSR}$, ρ is assumed to be valued as the average arrival rate divided by average service rate without considering setup time, or λ/μ. Thus, ρ is underestimated, which means $\bar{P}_{\rm FSR}$ is overestimated.

Let s be the index of machine and S be the total number of machines available. Define $P'_{\rm FCFS}$ as the probability that a setup is needed if we apply the FCFS rule and $P'_{\rm FSR}$ as the probability that a setup is needed if we apply the family scheduling rule for a multiserver system with jobs re-entering the system more than one times, or namely, all jth visits to steppers are noncritical operations. With these, the following lemmas are derived and proved.

Lemma 3: $P'_{\rm FSR} \le P'_{\rm FCFS}$.

Proof: Since the $P(j$th visit of type i has a jth visit of type i as predecessors | the system is empty $) = \sum_{s=1}^{S}\binom{S}{s}\,(\lambda_i/\lambda)^s\,(1-\lambda_i/\lambda)^{S-s} = X_{i,S}$, $P($a jth visit of type i job has a jth visit of type i job as predecessors | n jobs are in the system and $n < S) = \sum_{s=1}^{S-n}\binom{S-n}{s}\,(\lambda_i/\lambda)^s\,(1-\lambda_i/\lambda)^{S-n-s} = X_{i,S-n}$, P (a jth visit of type i job has a jth visit of type i job as predecessors | $n \ge S) = \lambda_i/\lambda$, and P (an arrival job is of jth visit and type $i) = \lambda_i/\lambda$

$$\begin{aligned}P'_{\rm FCFS} =& 1 - \sum_{i=1}^{I} J_i \left(\frac{\lambda_i}{\lambda}\right)\\ &\times \left[p_0\,(1-X_{i,S})\right.\\ &+ \sum_{n=1}^{S-1} p_n\,(1-X_{i,S-n})\\ &\left.+ \sum_{n=S}^{\infty} p_n\left(1-\frac{\lambda_i}{\lambda}\right)\right]. \qquad (6.2)\end{aligned}$$

Let $P'_{i,j',{\rm FSR}}$ be the probability that a setup is necessary for a job of jth visit and type i if applying the family-based scheduling rule. From P (no jth visit of type i job in the system) $= 1 - X_{i,S}$, P (no jth visit of type i job in the system | n jobs are in the system and $n \ge S) = (1-\lambda_i/\lambda)^n$, and P (no jth visit of type i job in the system | n jobs are in the system and $n < S$) $= 1 - X_{i,S-n}$, it is known that $P'_{i,j',{\rm FSR}} = p_0\,(1-X_{i,S}) +$

$\sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n (1 - \lambda_i/\lambda)^n$ where p_n is the probability with n jobs in the system. Thus

$$P'_{\rm FSR} = \sum_{i=1}^{I} J_i \frac{\lambda_i}{\lambda} P'_{i,j',{\rm FSR}} = \sum_{i=1}^{I} J_i \frac{\lambda_i}{\lambda} \left\{ p_0 (1 - X_{i,S}) + \sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n \left(1 - \frac{\lambda_i}{\lambda}\right)^n \right\}. \quad (6.3)$$

Since $(1 - \lambda_i/\lambda)^n \leq (1 - \lambda_i/\lambda)$ when $n - S \geq 0$ and from (6.2) and (6.3), $P'_{\rm FSR} = \sum_{i=1}^{I} J_i(\lambda_i/\lambda) \{p_0 (1 - X_{i,S}) + \sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n (1 - \lambda_i/\lambda)^n\}$, which is less than or equal to $\sum_{i=1}^{I} J_i(\lambda_i/\lambda) \{p_0 (1 - X_{i,S}) + \sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n (1 - \lambda_i/\lambda)\} = P'_{\rm FCFS}$. The result follows. □

Lemma 4: $P'_{\rm FSR} < P'_{\rm FCFS}$ if $\sum_{i=1}^{I} J_i > 1$, $\lambda_i > 0$ for all i, and $p_n > 0$ for at least one $n > S$ where p_n is the probability with n jobs in the system.

Proof: If $\sum_{i=1}^{I} J_i > 1$ and $\lambda_i > 0$ for all i, then $(1 - \lambda_i/\lambda) > (1 - \lambda_i/\lambda)^n$ for all $n > 1$. From Lemma 3, $P'_{\rm FSR} \leq \sum_{i=1}^{I} J_i(\lambda_i/\lambda) \{p_0 (1 - X_{i,S}) + \sum_{n=1}^{S} p_n (1 - X_{i,S-n}) + \sum_{n=S+1}^{\infty} p_n (1 - \lambda_i/\lambda)^n\}$, which is less than $\sum_{i=1}^{I} J_i(\lambda_i/\lambda) \{p_0 (1 - X_{i,S}) + \sum_{n=1}^{S} p_n (1 - X_{i,S-n}) + \sum_{n=S+1}^{\infty} p_n (1 - \lambda_i/\lambda)\} = P'_{\rm FCFS}$ if $p_n > 0$ for at least one $n > S$. The result follows. □

If p_n is approximated by the probability of the $M/M/S$ formula from [6], or $p_0 = \left[\sum_{n=0}^{S-1} 1/n! (\lambda/\mu)^n + 1/S! (\lambda/\mu)^S (S\mu/S\mu - \lambda)\right]^{-1}$, $p_n = (\lambda^n/n!\mu^n)p_0$ for $1 \leq n \leq S$, and $(p_n = \lambda^n/S^{n-S}S!\mu^n)p_0$ for $n \geq S$ where μ is the service rate of each machine without considering setup, (6.3) can be calculated. It should be noted that p_0 increases as μ increases. However, since the computation formula is too complicated, a numeric test is conducted to see when service rate increases, how $P'_{\rm FCFS}$ and $P'_{\rm FSR}$ behave. As expected, when μ increases, $P'_{\rm FCFS}$ and $P'_{\rm FSR}$ also increas, which implies that when no setup time is assumed or service rate is overestimated (utilization is underestimated), $P'_{\rm FCFS}$ and $P'_{\rm FSR}$ are also overestimated.

According to the definitions in Section II, the total arrival rate of critical operations is $\sum_{i=1}^{I} \bar{J}_i \lambda_i$ and the total arrival rate of noncritical operations is $\sum_{i=1}^{I} J'_i \lambda_i$. Based on the above lemmas, Theorems 1 and 2 can be proved.

Proof of Theorem 1: Since $\bar{J}_i + J'_i = J_i$ and from Lemmas 1 and 3, the probabilities could be approximated as follows:

$$P_{\rm FCFS} \approx \left[\sum_{i=1}^{I} \bar{J}_i \left(\frac{\lambda_i}{\lambda}\right) \left(1 - \frac{\lambda_i}{\lambda}\right)\right] + \sum_{i=1}^{I} J'_i \left(\frac{\lambda_i}{\lambda}\right) \times \left[p_0 (1 - X_{i,S}) + \sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n \left(1 - \frac{\lambda_i}{\lambda}\right)\right] \quad (6.4)$$

while

$$P_{\rm FSR} \approx \sum_{i=1}^{I} \bar{J}_i \frac{\lambda_i}{\lambda} \bar{P}_{i,j,{\rm FSR}} + \sum_{i=1}^{I} J'_i \frac{\lambda_i}{\lambda} P'_{i,j',{\rm FSR}} = \sum_{i=1}^{I} \bar{J}_i \frac{\lambda_i}{\lambda} \left\{ p_0 \left(1 - \frac{\lambda_i}{\lambda}\right) + \sum_{n=1}^{\infty} p_n \left(1 - \frac{\lambda_i}{\lambda}\right)^n \right\} + \sum_{i=1}^{I} J'_i \frac{\lambda_i}{\lambda} \left\{ p_0 (1 - X_{i,S}) + \sum_{n=1}^{S-1} p_n (1 - X_{i,S-n}) + \sum_{n=S}^{\infty} p_n \left(1 - \frac{\lambda_i}{\lambda}\right)^n \right\}. \quad (6.5)$$

Also from Lemmas 1 and 3, the result follows. □

Proof of Theorem 2: If there is no critical operation, Lemma 4 will be the proof. If there is no noncritical operation, Lemma 2 will be the proof. Thus, assume there exist both critical and noncritical operations. From the proofs of Lemmas 2, 4, and Theorem 1, the result follows. □

To compute the approximations of (6.4) and (6.5), combine the results from (6.1), (6.2), (6.3), p_n of $M/M/1$ formula for critical operations and p_n of the $M/M/S$ formula for noncritical operations.

References

[1] T. Adachi, J. J. Talavage, and C. L. Moodie, "A rule-based control method for a multi-loop production system," *Artificial Intelligence*, vol. 4, no. 3, pp. 115–125, 1989.

[2] X. Bai and S. B. Gershwin, "A manufacturing scheduler's perspective on semiconductor fabrication," in *VLSI Memo*, 1990, pp. 89–518.

[3] S. C. Chang and D. Y. Liao, "Scheduling flexible flow shops with no setup effects," *IEEE Trans. Robot. Automat.*, vol. 10, pp. 112–122, June 1994.

[4] C. R. Glassey and R. G. Petrakian, *The Use of Bottleneck Starvation Avoidance with Queue Predictions*. Berkeley: Univ. California, 1989.

[5] S. C. Graves, "A review of production scheduling," *Oper. Res.*, vol. 29, no. 4, pp. 646–675, 1981.

[6] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. New York: Wiley, 1985.

[7] R. A. Hughes and J. D. Schott, "The future of automation for high volume wafer fabrication and ASIC manufacturing," *Proc. IEEE*, vol. 74, pp. 1775–1793, Dec. 1986.

[8] P. K. Johri, "Practical issues in scheduling and dispatching in semiconductor wafer fabrication," *J. Manuf. Syst.*, vol. 12, no. 6, pp. 474–485, 1992.

[9] C. Koulamas, "The total tardiness problem: Review and extensions," *Oper. Res.*, vol. 42, no. 6, pp. 1025–1041, 1994.

[10] S. H. Lu and P. R. Kumar, "Distributed scheduling based on due dates and buffer priorities," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 1406–1416, Dec. 1991.

[11] H. Missbauer, "Order release and sequence-dependent setup times," *Int. J. Production Economics*, vol. 49, pp. 131–143, 1997.

[12] S. S. Panwalkar and W. Iskander, "A survey of scheduling rules," *Oper. Res.*, vol. 25, no. 1, pp. 45–64, 1977.

[13] R. Uzsoy, C. Y. Lee, and L. A. Martin-Vega, "A review of production planning and scheduling models in the semiconductor industry. Part I: System characteristics, performance evaluation and production planning," *IIE Trans.*, vol. 24, no. 4, pp. 47–60, 1992.

[14] ——, "A review of production planning and scheduling models in the semiconductor industry. Part ii: Shop-floor control," *IIE Trans.*, vol. 26, no. 5, pp. 44–55, 1994.

[15] L. M. Wein, "Scheduling semiconductor wafer fabrication," *IEEE Trans. Semiconduct. Manuf.*, vol. 1, pp. 202–215, Sept. 1988.

[16] U. Wemmerlov and A. J. Vakharia, "Job and family scheduling of a flow-line manufacturing cell: A simulation study," *IIE Trans.*, vol. 23, no. 4, pp. 383–393, 1991.

Ching-Chin Chern received the bachelor degree in business administration from National Taiwan University and the M.B.A. degree from the University of Texas, Arlington. She received the Ph.D. degree, in 1995, from University of Texas, Dallas.

She is an Associate Professor in the School of Management at National Taiwan University. She was employed as a System Analyst at National Hand Tool Corp., Farmers Branch, TX, from 1987 to 1992. In 1995, she joined the faculty at National Taiwan University. During her research period as a doctoral student, she also worked as a Research Consultant for Bell North Research Lab (BNR), primarily on evaluating telecommunication systems by applying simulation models. Her teaching and research interests include applying Quantitative Methods to solve problems related to Supply Chain Management, especially in the semiconductor industry.

Yu-Lien Liu received the M.B.A. degree in information management from National Taiwan University.

She was a Factory Information Manager in Texas Instruments-Acer Inc., from 1990 to 1998. She then went to work for Compag-Hp, Taiwan, until 2000. She now works as a Solution Delivery Director for the Great China Area in Brooks-PRI Automation, Taiwan.