Dynamic Scheduling Rule Selection for Fab Operations*

Bo-Wei Hsieh[1], Shi-Chung Chang[1], Chun-Hung Chen[2]

[1]Dept. of Electrical Engineering, National Taiwan University, Taiwan, R.O.C.

[2]Dept. of Systems Engineering, University of Pennsylvania, Philadelphia, USA.

Abstract

In this paper, we exploit the speed of the OO-based simulation tool designed by Hsieh et al. to investigate dynamic selection of scheduling rules when significant amount of WIPs are held due to engineering causes and when major machine failures occur. Four prominent dispatching rules, FSVCT, FIFO, LDF, and OSA, combined with the workload regulation release policy constitute 256 rule options over a 4-week horizon, where the dispatching rule may change weekly. The rule options are evaluated and ranked by the OO-based simulation tool under the performance index of per layer mean cycle time and throughput rate. Results clearly indicate that dispatching rule should be changed dynamically to handle these unusual events.

1. Introduction

Major fab scheduling problems include how wafers should be released into a fab and how they should be dispatched among machines for processing. A popular approach by pracitioners for scheduling the production in a fab is to select from the many empirical scheduling rules available for IC fabs. To quickly select a good enough scheduling rule from a rule library, Hsieh et al. developed a fast simulation tool (Figure 1) based on the ordinal optimization (OO) method [HCC99]. Our previous studies indicate that rule selections vary with initial states, performance indices and time horizons. The "state-dependent" feature of rule selection implies that scheduling rule should be selected dynamically according to current system state and operation objective.

In this paper, we exploit the speed of the OO-based simulation tool to investigate dynamic selection of scheduling rules when significant amount of WIPs are held due to engineering causes and when major machine failures occur. A 10-product, 60-stage and 12-tool-group fab model is extended from that of Lu et al. [LRK94] and is adopted for our simulation study. Four prominent dispatching rules, FSVCT,

FIFO, LDF, and OSA, combined with the workload regulation release policy constitute 256 rule options over a 4-week horizon, where dispatching rule changes weekly. The 256 options of dispatching rules changing weekly are evaluated and ranked by the OO-based simulation tool under the performance index of per layer mean cycle time (LMCT) and throughput rate. Simulation results clearly indicate that dynamic selection of dispatching rules is needed to achieve the operation objective of minimizing LMCT when the unusual events occur.

The remaining paper is organized as follows: In Section 2, we present the motivation of dynamic selection of scheduling rules. Details about the simulation model and scheduling rules are described in Section 3. Dynamic selections of dispatching rules when significant amount of WIPs are held due to engineering causes and when major machine failures occur are given in Sections 4 and 5 respectively. Section 6 concludes this paper.
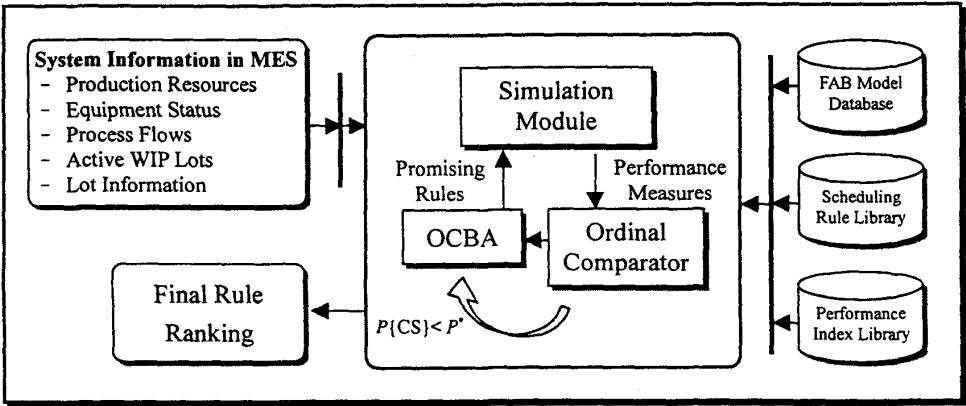

Figure 1 An Ordinal Optimization-Based Simulation Tool

2. Dynamic Selection of Scheduling Rules

For a fab, especially a foundry fab, the operation objectives change dynamically as well as the machine/inventory states. In view of the dynamic nature of short-term scheduling in wafer fabs, dynamic selection of a scheduling rule is intuitively desirable for competitive fab operations. Experimental studies of static rule selection by Hsieh et al. indicate that rule selections vary with factors of initial state, performance index and time horizon [HCC99]. The initial-state and operation objective dependent feature implies that scheduling rule is supposed to be selected dynamically according to system state feedback and operation objective in order to handle unexpected circumstances. In this paper, we exploit the speed of the

OO-based simulation tool to investigate dynamic selection of scheduling rules when significant amount of WIPs are held due to engineering causes and when major machine failures occur.

Both engineering hold of WIPs and machine failure may cause bottleneck capacity loss. An engineering hold event means that a certain amount of WIPs are held from further production due to engineering problem and will be allowed to continue after the problem is cleared. Engineering hold events lead to enormous variance in WIP distribution of a fab. When this occurs, the available WIPs at each stage are reduced which leads to shortage of WIP for processing. But when the held WIPs are released back to the production line, WIPs are suddenly increased and result in longer waiting time of WIPs. Unscheduled machine failures may also cause significant impact on a production line. When major machine failures occur, WIPs cumulate at the stages of the failed machines and machines of later stages are starved which may cause capacity loss. The operation objective might be changed to reduce the impact of these two abnormal events. Fab practitioners may concern about achieving throughput target rather than reducing mean or variance of cycle times. Due to the enormous variation of fab states and operation objectives, dynamic selection of dispatching rule is intuitively beneficial to fab performance under these circumstances.

3. Simulation Model Description

In this paper, a 10-product model (named FAB1) extended from a single-product model designed by Lu et al. [LRK94] is adopted. There are three types of processing technologies in model FAB1: $T1$, $T2$ and $T3$. Among the ten product types, four product types use technology $T1$, three product types use $T2$, and the other three use $T3$. Processing times of different product types using the same technology are different. The model involves 12 failure-prune processing stations, each having one or more identical but independent machines. Among the processing stations, Station 8 is modeled as a batch-processing machine group whose maximum capacity is 6 lots/batch. Processing times, times between failures and times to repair are exponentially distributed. The numbers of operation stages of $T1$, $T2$ and $T3$ are 60, 41 and 30 respectively. With release rates of 0.3 lots/hour for $T1$, 0.2 lots/hour for $T2$ and 0.12 lots/hour for $T3$, the bottleneck machine is Station 6 whose percentage utilization is 95.3%. Detailed model parameters of FAB1 are given in Table 1.

Table 1 Plant Data of FAB1

Station	No. of	No. of	No. of	No. of	MPT	MTBF	MTTR	% Utilization

	Machines	Visits (T1)	Visits (T2)	Visits (T3)				
1	4	14	10	8	0.500	150	5	92.73%
2	3	12	9	7	0.375	200	9	82.31%
3	10	7	5	4	2.500	200	5	91.94%
4	1	1	0	1	1.800	200	1	76.10%
5	1	2	1	1	0.900	200	1	83.30%
6	2	3	2	2	1.200	200	6	95.31%
7	1	1	1	0	1.800	200	1	90.50%
8	4	8	6	4	0.800	150	5	84.83%
9	1	3	0	0	1.000	200	5	92.44%
10	9	5	4	1	3.000	130	5	84.37%
11	2	3	2	1	1.200	200	5	87.64%
12	2	1	1	1	2.500	200	5	79.94%

In our experimental study, two performance indices are considered: per layer mean cycle time (LMCT) and total throughput rate, which are among the most frequently used fab performance indices. There are four prominent dispatching rules and a representative wafer release policy considered in this study as listed in Table 2. Workload regulation release policy proposed by Wein [Wei88], FSVCT dispatching rules proposed by Lu et al. [LRK94], and the OSA rule proposed by Li et al. [LTC96] are known to be good for reducing mean and variance of cycle time. We designed the LDF rule for controlling production smoothness and for tracking production targets.

Table 2 Scheduling Rules

Rule	Symbol	Description
Release policy	WR(C)	Workload regulation release for one bottleneck system. When the expected work in fab for bottleneck machine drops to C hours, then release a new lot.
Dispatching rules	FSVCT	Choose the lot with smallest $(a_n + C_p - \zeta_i)$, where a_n is the release time of lot n.
	LDF	Choose a stage with the largest deviation of completed moves from the desired moves and then choose a lot from the stage by FSVCT rule.
	OSA	Choose a stage according to the following priorities: Priority I: stage i such that $N_i(t) > \overline{N}_i$ and $N_{i+1}(t) < \overline{N}_{i+1}$; Priority II: stage i such that $N_i(t) < \overline{N}_i$ and $N_{i+1}(t) < \overline{N}_{i+1}$; Priority III: stage i such that $N_i(t) > \overline{N}_i$ and $N_{i+1}(t) > \overline{N}_{i+1}$; Priority IV: stage i such that $N_i(t) < \overline{N}_i$ and $N_{i+1}(t) > \overline{N}_{i+1}$, where $N_i(t)$ is the WIP at time t at stage i, $\overline{N}_i$ is the average WIP at stage i. Choose a lot with the same priority using FSVCT.
	FIFO	Select the lot which arrived in the queue at the earliest time.

4. Rule Selection under Engineering Holds

Engineering hold of WIPs causes temporary shortage of WIPs on certain processing stations. Shortage of WIPs may lead to capacity loss which is critical to fab performance. In this set of experiment, we study dynamic selection of dispatching rules under engineering hold circumstances. Suppose that the fab is initially operated under scheduling rule of WR-FSVCT. It is clearly indicated that if the operation objective changes when engineering hold event occurs, dispatching rules will change from our previous experimental studies of static rule selection. If the operation objective does not change, does dispatching policy need to change? To avoid unnecessary capacity loss, dispatching rule of LDF may be needed under engineering hold. LDF gives higher priority to WIPs of the same product type as the held WIPs during the period of engineering hold. In so doing, the capacity loss of machine stations that process the same product types as the held WIPs may be reduced.

Consider half of WIPs of technology *T1* are held completely held over the whole line for one week in FAB1 model which has been operated under the scheduling rule of WR-FSVCT for one year. Since that Station 9 can be used to process products of technology *T1* only, the engineering holds lead to 12% capacity loss of Station 9 whose utilization rate is about 90%. These held wafers due to engineering problem are supposed to be released back after one week. Mean cycle time value is supposed to decrease during engineering hold period since total workload and utilization rate of bottleneck station is decreased. Throughput rate is also supposed to decrease by a certain amount. When the held wafers are released back, the utilization rate of Station 9 is raised to about 95% for the rest of the week which results in the raise of mean cycle time and throughput rate.

We first conduct simulations to investigate what the best dispatching rule combination is over a four-week horizon where engineering hold occurs and the holding period is supposed to be one week. The initial state of the experiment is obtained by running the FAB1 simulation under the scheduling rule of WR-FSVCT over a one-year time period with half of WIP of *T1* set to be held and will be released back after seven days. Dispatching rules may be changed weekly. Four dispatching rules FSVCT, FIFO, LDF, and OSA are considered. There are therefore 256 (4x4x4x4) options. The release policy is the workload regulation policy for all these options. The workload levels of all scheduling rule combinations are all set at 106 hours, which is the long-term average workload to achieve the required throughput rate of 0.62 lots/hr. The 256 options of dispatching rules changing weekly are evaluated and ranked by our OO-based simulation tool under the

performance index of LMCT and throughput rate.

Simulation results listed in Tables 3 and 4 clearly indicate that (1) throughput of 0.62 lots/hr can be achieved by many options (2) to minimize LMCT dispatching rule should be changed from WR-FSVCT to WR-LDF. The throughput of 0.62 lots/hr can be achieved by many options since that the capacity loss of Station 9 is only 12%. The capacity loss at the first week can be recovered by higher utilization rate of Station 9 for the rest three weeks. The best rule listed in Table 4, A-C-D-D, achieves the maximum throughput rate where rule A, rule C, and rule D represent WR-FSVCT, WR-LDF, and WR-OSA respectively. The result shows that if the operation objective is to maximize throughput rate under engineering hold, rule changes from using WR-FSVCT throughout to A-C-D-D over a four-week horizon.

As shown in Table 3, the best selection for minimizing LMCT is to apply WR-LDF throughout the four-week horizon under engineering hold. WR-FSVCT is no longer the best in LMCT under the engineering hold situation. Although not the best in throughput rate performance, WR-LDF obtains throughput rate of 0.618 lots/hr, which is very close to that of rule option A-C-D-D. It is observed from the top-9 rules in Table 3 that WR-LDF should be applied for the first two or three weeks to reduce LMCT when engineering hold occurs. It is not unexpected since that WR-LDF gives higher priority to WIP of $T1$ technology during the holding period of other WIP of $T1$. In so doing, WIPs of $T1$ move faster to Station 9, which can reduce the capacity loss of Station 9 and cycle time of WIPs of $T1$. The LMCT performance of C-C-C-C, 12.388 hour, as compared with that of A-A-A-A, 13.690 hour, is shorter than 10%, which is quite significant in practice. This result clearly indicates the necessity of rule change under engineering hold.

Table 3 Dynamic Rule Selection under Engineering Holds (ranked by LMCT)

Rank	Rule [+]	LMCT	%	Throughput
1	C-C-C-C	12.388	-	0.618
2	C-C-C-B	12.395	0.06%	0.616
3	C-C-C-D	12.492	0.84%	0.617
4	B-C-C-B	12.644	2.07%	0.602
5	B-C-C-D	12.653	2.14%	0.607
6	B-C-C-C	12.705	2.56%	0.600
7	A-C-C-C	12.733	2.78%	0.608
8	A-C-C-B	12.740	2.84%	0.607
9	D-C-C-B	12.759	2.99%	0.601
10	C-D-C-C	12.781	3.17%	0.608
131	D-D-D-D	13.628	10.01%	0.610
145	A-A-A-A	13.690	10.51%	0.612

[+] Rule-A, Rule-B, Rule-C, and Rule-D represent WR-FSVCT, WR-FIFO, WR-LDF, and WR-OSA respectively.

Table 4 Dynamic Rule Selection under Engineering Holds (ranked by throughout)

Rank	Rule	Throughput	%	LMCT
1	A-C-D-D	0.621	-	13.514
2	A-A-B-A	0.621	0.00%	13.970
3	A-A-D-A	0.620	0.16%	13.995
4	A-C-C-A	0.619	0.32%	13.279
5	A-B-C-B	0.619	0.32%	14.043
6	A-B-B-D	0.619	0.32%	14.170
7	C-C-C-C	0.618	0.48%	12.388
8	D-D-C-C	0.618	0.48%	13.106
9	D-D-C-A	0.618	0.48%	13.605
10	D-B-A-D	0.618	0.48%	13.642
46	A-A-A-A	0.612	1.45%	13.690

5. Rule Selection after Unusual Machine Failure

Unusual machine failure may also cause capacity loss of bottleneck machine as engineering hold does. Unlike indirect capacity loss caused by shortage of WIPs under engineering hold, machine failure directly causes capacity loss over the down time period. In this set of experiments, we study dynamic selection of dispatching rules under an unusual machine failure. Again, consider the machine of Station 9 has been down and will be repaired after five days in FAB1 model, which has been operated under the scheduling rule of WR-FSVCT. As the MTTR of Station 9 machine is 5 hours, a more than five days of down time is obviously unusual. Station 9 then becomes a new bottleneck after repaired within the four weeks. A five-day down time will cause 18% of capacity loss of Station 9 and significant impact on other processing stations. Unlike the capacity loss due to engineering hold in Section 5.1, capacity loss of Station 9 cannot be recovered in this four-week horizon. When this unusual machine failure occurs, WIPs cumulate at the stages processed by Station 9 and machines of later stages are starved. It is undoubted that throughput rate will decrease by a certain amount after unusual machine failures. For such an unusual event, the most important of all might be maintaining the throughput, which is set as 0.62 lots/hour in this experiment, and reduce the impact of capacity loss. A natural question is whether the scheduling rule of WR-FSVCT stays the best under such an unusual fab state. An experiment similar to that of Section 4 is conducted. The release policy, dispatching rule combinations, performance indices and time horizon are the same as those of Section 4 except that the initial state is obtained by running the FAB1 simulation with a machine of Station 9 down for five days.

Simulation results listed in Tables 5 and 6 indicate that (1) throughput rate is

decreased by more than 10% under unusual machine failure for all rules and (2) to minimize LMCT dispatching rule should be changed from WR-FSVCT to WR-LDF. It is under our expectation that the throughput rate decreases due to unrecoverable capacity loss of Station 9. If the operation objective is to maximize throughput rate, rule option A-B-A-A whose throughput rate performance is 0.555 lots/hr is the best as shown in Table 6, where rule A and rule B represent WR-FSVCT and WR-FIFO respectively. But the differences of throughput rate of top-ranking rules in Table 6 are not significant. If there is no machine failure occurred, it is supposed that using WR-FSVCT throughout the four weeks achieves the best LMCT performance and reaches the required throughput rate. However, scheduling rule of WR-FSVCT is no longer the best under the machine failure situation. The best selection for LMCT is to apply WR-LDF throughout the four weeks as shown in Table 5. The difference in LMCT performance as compared with using WR-FSMCT throughout range is about 7%, which is very significant in practice.

Table 5 Dynamic Rule Selection under Unusual Machine Failure (ranked by LMCT)

Rank	Rule [+]	LMCT	%	Throughput
1	C-C-C-C	14.527	-	0.553
2	C-C-C-D	14.702	1.20%	0.549
3	C-C-C-B	14.795	1.84%	0.546
4	B-C-C-C	14.833	2.11%	0.538
5	D-C-C-B	14.845	2.19%	0.543
6	D-C-C-C	14.850	2.22%	0.537
7	C-C-A-D	14.912	2.65%	0.546
8	B-C-C-D	14.919	2.70%	0.543
9	C-A-C-C	14.940	2.84%	0.543
10	A-C-C-C	14.966	3.02%	0.545
69	A-A-A-A	15.547	7.02%	0.545

Table 6 Dynamic Rule Selection under Unusual Machine Failure (ranked by throughput)

Rank	Rule	Throughput	%	LMCT
1	A-B-A-A	0.555	-	15.889
2	D-B-D-D	0.555	0.00%	16.419
3	C-C-C-C	0.553	0.36%	14.527
4	C-B-D-D	0.553	0.36%	15.946
5	D-A-A-A	0.552	0.54%	15.163
6	D-A-D-A	0.552	0.54%	15.865
7	D-A-A-B	0.551	0.72%	15.243
8	C-D-A-B	0.551	0.72%	15.843
9	B-D-A-A	0.551	0.72%	15.936
10	B-D-D-A	0.551	0.72%	16.092
59	A-A-A-A	0.545	1.80%	15.547

[+] Rule-A, Rule-B, Rule-C, and Rule-D represent WR-FSVCT, WR-FIFO, WR-LDF, and WR-OSA respectively.

6. Conclusions

In this paper, we investigated dynamic selection of scheduling rules under engineering holds and unusual machine failure, which may cause capacity loss to bottleneck station. We exploit the speed of the OO-based simulation tool to select a good scheduling rule over a 4-week horizon, where rule changes weekly. Experimental results clearly indicate that dispatching rule should be changed dynamically to handle these unusual events.

References

[HCC99] B. W. Hsieh, C. H. Chen, S. C. Chang, "Fast Fab Scheduling Rule Selection by Ordinal Comparison-Based Simulation," *Proceedings of International Symposium of Semiconductor Manufacturing* 1999, Santa Clara, Oct. 1999, pp.53~56.

[LRK94] S. H. Lu, D. Ramaswamy, P. R. Kumar, "Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants, " *IEEE Transactions on Semiconductor Manufacturing*, vol. 7, no. 3, Aug. 1994.

[LTC96] S. Li, T. Tang, and D. W. Collins, "Minimum Inventory Variability Schedule with Applications in Semiconductor Fabrication, " *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, No. 1, February 1996.

[Wei88] L. M. Wein, "Scheduling Semiconductor Wafer Fabrication, " *IEEE Transactions on Semiconductor Manufacturing*, Vol. 1, pp. 115-1C0, August 1988.