A Release Control Policy Based on the Primary Bottleneck Workstation Utility of Multiple Semiconductor Manufacture Lines

Nan Xiang, Li Li
School of Electronics and Information Engineering,
Tongji University,
Shanghai, China
E-mail:1433186@tongji.edu.cn, lili@tongji.edu.cn

Abstract—Release control plays an important role in the performance of semiconductor manufacturing lines. Considering their characteristics, a release control policy based on the primary bottleneck workstation utility of multiple semiconductor manufacture lines (abbreviated as Rel_M2) is proposed. Comparing to common release control strategy CONWIP (Constant Work i n Process), Rel_M2 can make full use of the primary bottleneck equipment of each manufacture line, increase the throughput, and improve the on-time delivery rate. The simulation results demonstrate that the throughput, on-time delivery rate and on-time delivery rate of hot job are increased by 17.46%, 85.84% and 53.65%, respectively. Moreover, the mean cycle time is decreased by 32.90%.

Keywords—primary bottleneck workstation; workstation utility; multiple semiconductor manufacture lines

I. INTRODUCTION

The semiconductor industry has become one of the focuses of global industrial development over the past decades [1]. Compared with other manufacturing industries, semiconductor manufacturing has the features of numerous products, complex manufacturing processes and high bottleneck workstation utility [2].

The release control plays an important role in the semiconductor wafer fabrication facility (abbreviated as fab)[3]. Release policies are dedicated to determine when, which type and how many jobs should be released into a fab [4]. Nowadays the research focus in this area is switched from static release control strategies into the dynamic release control ones, which take the actual working condition of the production line into consideration. The dynamic release control strategies can be divided into two types. The first type is monitoring the number of WIP in the production system, for example, the strategy of Constant Work in Process (CONWIP). In CONWIP release policy, a new job is released into the fab when a job is completed in a fab. In other words, a new job can be released into the fab until the number of jobs in the fab drops down below a pre-designed constant WIP level [5]. Thus, CONWIP could adapt to the changing product varieties with strong robustness. However, it cannot control the number of WIP of each workstation, resulting in low efficiency. The second type is to monitor the workload in real time, including the strategy of SA (Starvation Avoidance). Though SA can make full use of bottleneck workstation, it is limited by the single bottleneck workstation [6].

To improve the utilization rate of bottleneck workstation of the release control system, Zhang et al.[7] proposed a minimum spanning tree with constraints based on the semiconductor packaging and testing on fine feed control method. By solving constrained minimum spanning tree, the specific feeding sequence and the specific feeding time of each product are obtained. This method can reduce the machine cost, improve equipment utilization, and shorten the production cycle.

Recently, more and more researchers focused on the release control with multiple lines. For example, Chen et al [8] proposed a Capacity Requirements Planning System (CRPS) composed of four major modules for twin fabs of wafer fabrication. Simulation results show that CRPS can balance the equipment loading between the twin fabs with shared equipment, on various days, and across various equipments at various levels of demands.

Compared to the single semiconductor manufacturing line, the multiple semiconductor lines can effectively allocate the new jobs in each line. Here, we propose a new policy for multiple lines to allocate the new jobs according to their due dates, processing characteristics, and capacity of primary bottleneck workstation. It can increase the bottleneck workstation utility and on time delivery rate of job, and achieve high throughput with short cycle time. The remainder of the paper is organized as follows. In Section 2, we propose the release control policy based on the primary bottleneck workstation utility of multiple semiconductor manufacture line (called as Rel_M2). Section 3 presents the results and analysis of the simulation results. Finally, in Section 4, we summarize our work.

II. A RELEASE CONTROL POLICY BASED ON THE PRIMARY BOTTLENECK WORKSTATION UTILITY OF MULTIPLE SEMICONDUCTOR MANUFACTURE LINE

The main idea of Rel_M2 is that in the first place we should make sure the primary bottleneck workstation in each semiconductor manufacturing line. Then we should confirm the interval level of the primary bottleneck workstation utility. At last according to the utilization of the primary bottleneck workstation of every day, we can determine which line the new job should be released.
A. Find the Primary Bottleneck Workstation

According to the statistical data of daily workstation utility implemented by simulations, we can find the primary bottleneck workstation whose utility rate is more than 80% every day.

B. Determine the Interval Level of the Primary Bottleneck Workstation Utility

We firstly conduct simulations to analyze the utility of primary bottleneck workstation for different product-mix, product categories and product quantities.

Then we choose samples with good short-term performance, such as high movements of jobs per day and high throughput per day. According to the utilization of primary bottleneck workstation every day from the good samples, we can get the interval level \([W_l, W_h]\) \((W_l: the lower level of the interval, W_h: the upper level of the interval) of the primary bottleneck workstation utility. Based on the interval level, the line to be released new job is determined.

The detailed process is show in Figure 1.

The Decision Process in the Multiple Semiconductor Manufacturing Lines

According to the interval level \([W_l, W_h]\) of primary bottleneck workstation utility and the average primary bottleneck workstation utility \(W_i\) of line \(i\) per day, we can make decision on which line the new job should be released. The decision process is illustrated by the flowchart diagram in Figure 2. In Figure 2, the meanings of the variables are shown in TABLE I:

| Step | Description |
|------|-------------|
| 1.   | Initialize \(Fab_i, Lot_{im}\) \((i = 1, m = 1, j = 0)\); |
| 2.   | If \(W_i \in [W_l, W_h]\), go to Step 3; otherwise, go to Step 4; |
| 3.   | \(j = j + 1\); |
| 4.   | If \(i < n, i = i + 1\), go to Step 2; otherwise, go to Step 5; |
| 5.   | Carry out switch \((j)\); |
|     | Case \(j > 1\): There are many lines that can process \(Lot_{im}\), so we choose the line which has the minimum number of WIP to release \(Lot_{im}\). |
|     | Case \(j = 1\): Among all lines, only one line can meet the condition to release the \(Lot_{im}\). |
|     | Case \(j = 0\): No line is available, so we wait time \(s\) to release the \(Lot_{im}\). |
|     | \(m = m + 1\) |

III. SIMULATION RESULTS

To validate and verify the proposed release control policy Rel_M2, we developed a simulation model of parallel semiconductor manufacturing lines to analyze by using the Plant Simulation Tecnomatix simulation software. There are 71 processing areas and 545 machines in the model.
A. Determine the Primary Bottleneck Workstations by Simulation

We conduct numerous simulations with different product-mix, categories and quantities. In each simulation, release strategy is set as CONWIP and the schedule rules include FIFO (First In First Out), CR (Critical Ratio), EDD (Earliest Due Date) and LS (Least Slack). The time is set as 90 days with the warm-up period of 30 days. Then the utility of the total 545 workstations is monitored every day. It can be seen that only 84 workstations are needed to complete the 9 different types of productions. After sorting, we can obtain the information of each workstation whose daily utility is above 80% and regard these stations as the primary bottleneck workstations, which are stated as follows:

BL_6FSI01, BL_6BTU13, BL_6AEI01, BL_6WET21, BL_6BTU12, BL_6OVN10, BL_6BTU24, BL_6TELD1, BL_6WET22, and BL_6TELC1.

B. Set the Interval Level $[W_l, W_h]$ of Primary Bottleneck Workstation Utility by Simulation

Step1: Select excellent samples

To select the samples with high quality, we analyze the 10 primary bottleneck workstations with daily utility is above 80% and regard these stations as the primary bottleneck workstations, which are stated as follows: BL_6FSI01, BL_6BTU13, BL_6AEI01, BL_6WET21, BL_6BTU12, BL_6OVN10, BL_6BTU24, BL_6TELD1, BL_6WET22, and BL_6TELC1.

Step2: Determine the interval level $[W_l, W_h]$

Through step1, we can get the primary bottleneck workstation utility among different kinds of simulation seniors with different WIP levels. Since the higher WIP level aggravates the performance of the production lines, causing more blocks, the utility of the primary bottleneck workstation determined by the good samples with high throughput per day is lower than that from the samples with high movement per day. Therefore we separately average the utilities of the good samples based on the throughput and movement, which are $W_l$ and $W_h$, respectively.

C. Simulation Results

In this paper, Rel_M0 refers to the release strategy of CONWIP. Rel_M1 is the method which only considers the number of WIP of each line (the new job is released into the line that has the minimum WIP). Rel_M2 represents the release control policy we propose. In simulation, we select nine products with different processing time. We make different kinds of orders with different product-mix, product categories and product quantity. Simulations with Rel_M0, Rel_M1 and Rel_M2 are carried out for comparison. These simulations have been done for 90 days with 30 days warmed-up period. The results are show in TABLE II and the comparison is show in TABLE III.

In order to describe the comparison of performance more clearly, the corresponding results are also presented in Figure 5, Figure 6, Figure 7 and Figure 8. Order $i$ ($1 \leq i \leq 6$) denote different WIP level including 3375 wafer, 5400 wafer, 6750 wafer, 8100 wafer, 9450 wafer and 10125 wafer. Rel_Compare0 and Rel_Compare1 indicate the improvement of performances compared to those in Rel_M0 and Rel_M1, respectively.

From Figure 5, one can clearly see that Rel_M0 has a higher TH at the beginning when the WIP unit is low; however, when the WIP unit exceed 6750 unit, Rel_M2 has its advantage in TH. Especially, maximum lifting range about 17.46% is obtained while the WIP level reaches 8100 unit. In terms of TH, Rel_M2 is superior to Rel_M1 in all cases. Additionally a max scope of 2.8% is achieved in the level of 8100 unit.
IV. CONCLUSION

This paper presents a release control policy based on the primary bottleneck workstation utility of multiple semiconductor manufacture line. According to the different product mix, categories, quantity, and different scheduling rules, we build different simulation scenarios and screen out samples with good short-term performance based on daily throughput, movement. Through these short term performance indicators, the interval level of primary bottleneck workstation utility \([W_l,W_h]\) is determined. We then decide whether to release a certain job on the basis that if the averaged utility of the primary bottleneck workstation falls in the interval value \([W_l,W_h]\).

The new job can choose the corresponding product line for feeding if the value falls within the interval. If more product lines are available, then select the line with minimum WIP; if there is only one product line to process the work piece, select it; otherwise, the job is waiting.

Talking about the performance of our strategy, Rel_M2 can improve the on-time delivery rate effectively. In particular, Rel_M2 can improve the performance of all the aspects including TH, CT, ODR and HODR compared with Rel_M0 and Rel_M1. Moreover, the maximum improvement can be as high as 85.84% in terms of ODR. This shows that Rel_M2 can improve the on-time delivery rate effectively.

The strategy mentioned here is based on the bottleneck analysis method by measuring the utility rate of each equipment. Sometimes it is very hard to distinguish which equipment is the bottleneck since the utility rates are very similar. Additionally, this method is limited by the stable system and we did not consider about the bottleneck shifting. All in all, our proposed release strategy can be further improved for more complicated semiconductor manufacturing.

ACKNOWLEDGMENT

This paper is supported in part by National Natural Science Foundation of China under grants No. 51475334.

REFERENCES

[1] Amit K. Gupta, Appa Iyer Sivakumar. “Simulation based multi-objective schedule optimization in semiconductor manufacturing,” In Proceedings of the 2002 Winter Simulation Conference, vol:2,pp:1862-1870,2002.

[2] G.H.Zhang, C.Liu, L.L.Yao, H.B.Shi, “A method of bottleneck detection of semiconductor assembly and test production line,” Chinese Journal of Electron Devices, vol. 38,pp.44-8,20 Feb.2015.

[3] Y.H.Guo, Z.F.Shao, X.S.Qian, “Dynamic Release Rule for Multiple Re-Entrant Wafer Fabrication,” Semiconductor Technology, vol.31, pp.645–65,2006.

[4] Chao Qi, Appa Iyer Sivakumar and Stanley B.Gershwin, “An efficient new job release control methodology,”International Journal of Production Research, vol.47, No.3,pp. 703-731,1 February 2009.

[5] Z.B.Chen, X.W.Pan, L.Li, et al. “A New Release Control Policy (WRELM) For Semiconductor Wafer Fabrication Facilities,”IEEE 11th International Conference on Networking, Sensing and Control (ICNSC),pp.64-68,2014.

[6] Y.Li, Z.B.Jiang, N.Li, C.Li, “A review on release policies in semiconductor wafer fabrication system,” Industrial Engineering and Management, vol. 16, pp.108–14,Dec:2011.
TABLE II. TABLE PERFORMANCE IN REL_M0, REL_M1, REL_M2

| Order | Rel_M0 | Rel_M1 | Rel_M2 |
|-------|--------|--------|--------|
|       | TH (lot) | CT (hour) | ODR | HLODR | TH (lot) | CT (hour) | ODR | HLODR | TH (lot) | CT (hour) | ODR | HLODR |
| Order1 | 1172 | 540 | 100% | 100% | 920 | 369 | 100% | 100% | 922 | 367 | 100% | 100% |
| Order2 | 1150 | 854 | 91.83% | 75% | 1006 | 581 | 97.62% | 89.90% | 1008 | 573 | 98.00% | 90.10% |
| Order3 | 1042 | 1048 | 61.79% | 46.15% | 1070 | 784 | 80.38% | 78.77% | 1084 | 758 | 81.81% | 79.74% |
| Order4 | 1008 | 990 | 36.22% | 56% | 1152 | 971 | 65.20% | 75.30% | 1184 | 946 | 67.31% | 78.57% |
| Order5 | 1060 | 1168 | 27.25% | 40% | 1148 | 1239 | 47.82% | 60.42% | 1148 | 1237 | 48.71% | 61.46% |
| Order6 | 1018 | 1251 | 18.09% | 37.5% | 1074 | 1356 | 27.58% | 43.29% | 1076 | 1349 | 30.80% | 44.69% |

TABLE III. TABLE COMPARISON WITH REL_M0, REL_M1

| Order | Rel_Compare0 | Rel_Compare1 |
|-------|--------------|--------------|
|       | TH (lot) | CT (hour) | ODR | HLODR | TH (lot) | CT (hour) | ODR | HLODR |
| Order1 | -21.3% | 32.04% | 6.72% | 20.13% | 0.22% | 0.54% | 0 | 0 |
| Order2 | -12.34% | 32.90% | 6.72% | 20.13% | 0.20% | 1.4% | 0.39% | 0.22% |
| Order3 | 4.03% | 27.67% | 32.40% | 72.78% | 1.3% | 3.3% | 1.8% | 1.2% |
| Order4 | 17.46% | 4.44% | 85.84% | 40.30% | 2.8% | 2.6% | 3.2% | 4.3% |
| Order5 | 8.30% | -5.91% | 78.75% | 53.65% | 0 | 0.16% | 1.9% | 1.7% |
| Order6 | 5.69% | -7.83% | 70.26% | 19.17% | 0.19% | 0.52% | 11.68% | 3.2% |