Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

WIP BALANCE AND DUE DATE CONTROL IN A WAFER FAB WITH LOW AND HIGH VOLUME PRODUCTS

Zhugen Zhou
Oliver Rose

Computer Science Department
University of Federal Armed Forces Munich
85577 Neubiberg, GERMANY

ABSTRACT

For a customer-oriented wafer fab, low volume products such as development lots or customer samples are often more critical than high volume products with regard to cycle time and delivery reliability because of due date commitment. In this study, a global rule combining WIP balance and due date control is developed for a wafer fab with low and high volume products. The purpose is to figure out the following two issues. Firstly, whether WIP balance of high volume products takes the cost of due date of low volume products. Secondly, how to make the trade-off between on-time delivery and WIP balance for the low volume products.

1 INTRODUCTION

In the semiconductor industry companies need to differentiate themselves in the products and services they offer nowadays. As many companies move from mass production to mass customization to satisfy the unique requirements from customers, due date becomes more and more important. A missed due date not only causes penalty but also future business loss. Inventory is also critical because it has a major influence on the overall manufacturing cost.

Wafer fabrication is considered as one of the most complex manufacturing processes because of reentrant processing flow, batch processing, sequence dependent setups, unpredictable tool failure and so on, which differentiate wafer fab from other traditional flow shops or job shops. Dispatching is one of the major techniques to help to smooth manufacturing process, lower inventory level and meet due date. Most of the current dispatching rules are related to due date. They are variants of classical rules like Critical Ratio (CR), Earliest Due Date (EDD) and Operation Due Date (ODD) (Baker and Bertrand 1981). There are also numbers of operational control policies which target the control of inventory level of work center or operation like Minimum Inventory Variability Scheduler (MIVS) (Li et al. 1996). While the first set of dispatching rules do not primarily lead to low inventory level, the later ones do not always lead to good on-time delivery performance. For this reason, some researchers address the complementary strength of WIP balance and due date control (Lee et al. 2008; Zhou and Rose 2011).

There are hundreds of wafer products in a wafer fab. Some products are referred to low volume products such as test, sample, small order and new product which have low release rate e.g. dozens of wafers are released per week, while some products are referred to high volume products like common commodity type which has a higher release rate than low volume products. Low volume products are often have a tight target due date and are more critical than high volume products with respect to cycle time and delivery reliability because of due date commitment to the customer. Low volume products are expected to go through the fab as soon as possible, at least meet the target cycle time and due date. However, there is a basic assumption that low volume products suffer from specific machine constraints like higher batch

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

time, longer setup waiting time and less qualified machine available, etc. In addition, local rules change the target function of global rules in order to make a compromise between due date and local constraints. For instance, a WIP balance target between the machines seems to reduce the weight of due date control, because a WIP balance approach would rather push a early lot to a empty machine instead of push a tardy lot to a crowded machine. Therefore, with regard to WIP balance and due date control, there are two main questions for low volume products: (1) whether due date is sacrificed by achieving WIP balance; (2) how to make trade-off if due date is desired more than WIP balance. These impose an additional challenge to the operational control in a wafer fab.

This paper is organized as follows. In Section 2, we describe the proposed WIP balance and due date control approaches in detail. In Section 3, we compare the simulation results with two cases. Section 4 is the summary.

2 GLOBAL RULE COMBINING WIP BALANCE AND DUE DATE CONTROL

2.1 WIP Balance Approach

2.1.1 Bottleneck Workload Control

According to the Theory of Constraints, the performance of the whole fab, e.g., its throughput is mainly determined by the bottleneck performance. It is necessary to determine an adequate WIP level for the bottleneck to avoid starvation and to support the whole fab to achieve its maximum throughput while running at the minimum WIP level. However, if the WIP level of the bottleneck exceeds the desired WIP level while achieving the maximum throughput of the whole fab, the cycle time is degraded. Lots will spend a significant queue time in front of the bottleneck work center, which will also cause a WIP imbalance to the line. Therefore, a minimum workload is defined for the bottleneck work center. If the actual workload of the bottleneck drops to the minimum workload, the bottleneck is fed with lots to prevent starvation. A maximum workload is also taken into account. If the actual workload of the bottleneck is higher than the maximum workload, bottleneck feeding is stopped to avoid extraordinary queue time, especially, when the bottleneck is broken down. In this study, we only consider a single dynamic bottleneck in the fab where the bottleneck is the work center with the highest utilization. The minimum and maximum workload for the bottleneck is defined as 12 hours and 24 hours respectively which are defined by the engineers.

2.1.2 Feeding Empty Non-bottleneck Work Center

Although the bottleneck is the most critical work center which determines the performance of the whole fab, feeding empty non-bottleneck work centers can also smooth the material flow, avoid capacity losses of machines, and improve product cycle times. Therefore, a minimum workload 1.5 hours is also defined for the non-bottleneck work centers. If the workload of non-bottlenecks drop to this minimum workload level, lots are scheduled to feed it to avoid starvation.

2.2 Due Date Control

2.2.1 Acceleration of Maximum Tardiness Lot

In general, WIP balance algorithms tend to push lots to work centers that are running out of WIP without taking due dates into consideration. In this case, overemphasizing WIP balance has a negative impact on on-time delivery. In fact, sometimes it would be better to push a delayed lot to a high WIP work center instead of pushing an early lot to a low WIP work center. Because of customer commitments, keeping the due date is the first priority for customer-oriented companies. Therefore, a compromise is necessary in order to meet due dates and reduce tardiness. Pushing a delayed lot despite WIP balance requirements to downstream work centers can give the delayed lot a chance to speed up, to save cycle time, and reduce tardiness, although work center capacity might be lost. The acceleration algorithm works as follows:

Step 1: In the queue of the upstream work center, if lots are delayed for the operation, we determine the lot which has the maximum tardiness 'MaxTardinessUp' for the operation.

Step 2: Then, we identify the target downstream work center where the 'MaxTardinessUp' lot will be processed. Next, we find the lot which has the maximum tardiness 'MaxTardinessDown' for operation in the queue of the target downstream work center (like in Step 1).

Step 3: If 'MaxTardinessUP' is greater than 'MaxTardinessDown', the lot which has 'MaxTardines sUp' is assigned a high priority in the upstream work center.

2.2.2 Acceleration of Lot close to Due Date

Acceleration of delayed lots can only reduce tardiness instead of improving on-time delivery performance. Thus, we also propose to speed up the lots which are close to their due dates. This provides a mechanism for those lots to catch up with their due date. If there is still 1 week left for the lot to chase after the due date and the lot's CR value is less than 1 – which means the lot is close to due date and possibly falls behind schedule – this lot will obtain a higher priority since there is a high probability that it will be delayed in the future.

2.3 Global Rule Combining WIP Balance and Due Date Control

In order to test our approaches we extended a simplified version of the global dispatching rule 'IFD' which is in use at Infineon Technologies AG Dresden, a German semiconductor manufacturer, with our ideas. As we can see from Figure 1 (a), there are 3 hierarchies of lot priority for the IFD rule. In each queue of a work center, lots are categorized into 3 classes in descending priorities according to their states.

When WIP balance approaches described in section 2.1 are incorporated into IFD rule, it becomes the one in Figure 1 (b). From priority class 2 to 4, the priority is divided into 2 sub-classes which are delayed lot and non-delayed lot. The goal is to avoid bottleneck starvation and capacity loss for the empty non-bottlenecks. Nevertheless, the first problem for the low volume products arises here. The WIP balance approaches intend to balance the workload of work center, without taking lot's due date information into account, e.g., it would prefer to feed a lot with a loose due date to a low WIP work center rather to push a lot with tight due date to a high WIP work center. This may lead to cycle time reduction with the cost of on-time delivery of products with tight target due date.

Therefore, in order to solve this problem, due date control approaches mentioned in Section 2.2 are included into the IFD rule too, which is presented in Figure 1 (c). The delayed lots which fulfill the criterion for accelerating of maximum tardiness lots belong to the second priority class. This priority class is more critical than the priority class of the bottleneck workload control method and of the feeding empty non-bottleneck method because customer commitment is more important than WIP balance in this study. Accelerating maximum tardiness lots is considered as a compromise to WIP balance. The upstream work centers would rather push the maximum tardiness lot to downstream work centers which may be highly loaded instead of pushing an early lot to downstream work centers which may be starved to maintain WIP balance. The maximum tardiness lot has to be moved to the next operation to minimize delay. Furthermore, the non-delayed lot class is also split into two sub-classes which separate lots close to their due dates from lots on schedule. According to the acceleration of lots close to due date method, lots which are close to due date are more preferential than lots on schedule. In Figure 1, if lots belong to the same priority class, the ODD rule is applied as the dispatching rule.

2.4 Simulation Model

The small wafer fab dataset MIMAC6 from Measurement and Improvement of MAnufacturing Capacities (MIMAC) is used to test our ideas. We refer the interested reader to Fowler and Robinson (1995) for details. MIMAC6 is a typical complex wafer fab model including:

- 9 products, 9 process flows, maximum 355 process steps.
- 24 wafers in a lot. 2777 lots are released per year under fab loading of 100%. All lots have the same priority of 1 when they are released in the fab.
- 104 tool groups, 228 tools. 46 single processing tool groups, 58 batching processing tool groups.
- Sequence dependent setup, rework, MTTR (mean time to repair), and MTBF (mean time between failures) of tool group.

The simulation experiments are carried out with Factory eXplorer (FX) from WWK. The proposed ideas are not provided by the FX simulation package, but FX supports customization via a set of user-supplied code and dispatch rules.

(a) IFD Rule

1. Waiting time > 48 hours
2. Delayed lot
3. Non-delayed lot

(b) IFD Rule + WIP balance

1. Waiting time > 48 hours
2. Feeding empty bottleneck
    2.1. Delayed lot
    2.2. Non-delayed lot
3. Feeding empty non-bottleneck
    3.1. Delayed lot
    3.2. Non-delayed lot
4. Lot for non-empty work center including normal bottleneck
    4.1. Delayed lot
    4.2. Non-delayed lot
5. Lot for over-loaded bottleneck
    5.1. Delayed lot
    5.2. Non-delayed lot

(c) IFD Rule + WIP balance + Due date control

1. Waiting time > 48 hours
2. Acceleration of maximum tardiness lot (only for low volume products)
3. Feeding empty bottleneck
    3.1. Delayed lot
    3.2. Non-delayed lot
        3.2.1. Close to due date
        3.2.2. On schedule
4. Feeding empty non-bottleneck
    4.1. Delayed lot
    4.2. Non-delayed lot
        4.2.1. Close to due date
        4.2.2. On schedule
5. Lot for non-empty work center including normal bottleneck.
    5.1. Delayed lot
    5.2. Non-delayed lot
        5.2.1. Close to due date
        5.2.2. On schedule
6. Lot for over-loaded bottleneck
    6.1. Delayed lot
    6.2. Non-delayed lot
        6.2.1. Close to due date
        6.2.2. On schedule

Figure 1: IFD rule combines with WIP balance and due date control approaches.

3 SIMULATION RESULTS AND PERFORMANCE ANALYSIS

The products like ‘B6HF’, ‘C4PH’ and ‘C6N3’ have a relative low release rate, while the products like ‘B5C’, ‘C5P’ and ‘C5PA’ have a relative high release rate in the original MIMAC6 model. In order to test our idea, we modified the release rate to make sure the low volume products are separated from high volume products, which is demonstrated in Table 1. In Case 1, products ‘B6HF’, ‘C4PH’ and ‘C6N3’ are considered as low volume products. They are only released 1-2 lots per week and have a tight target due date. Vice verse for the products ‘B5C’, ‘C5P’ and ‘C5PA’. The release rates in Case 1 result in a fab loading of 99.5%. In Case 2, the low volume products in Case 1 are changed to high volume products. While the high volume products in Case 1 become low volume. The release rates in Case 2 lead to a fab loading of 99.4%, which is quite close to Case 1.

Table 1: Release rate and target due date flow factor for each product in MIMAC6.

Case 1: 99.5% Fab Loading			Case 2: 99.4% Fab Loading		
Product	Release Rate (wafers per week)	Target Due Date Flow Factor	Product	Release Rate (wafers per week)	Target Due Date Flow Factor
C6N3	48	1.8	C6N3	150	2.4
B6HF	24	1.8	B6HF	165	2.4
C4PH	48	1.8	C4PH	300	2.4
C6N2	100	2.4	C6N2	100	2.2
OX2	100	2.4	OX2	100	2.2
C5F	100	2.4	C5F	100	2.2
C5P	350	2.6	C5P	48	1.3
C5PA	300	2.6	C5PA	48	1.3
B5C	150	2.6	B5C	48	1.3

Firstly, the MIMAC6 model is tested by the IFD rule with the setting in Case 1. Then the WIP balance approaches are incorporated into the IFD rule, to find out whether the target due date of low volume products are sacrificed by the WIP balance of high volume products. If it is true, the due date control approaches are integrated into the IFD rule too, to see whether the tardiness of low volume products can be minimized as much as possible without losing the cycle time achieved by WIP balance. The simulation length of MIMAC6 was carried out for 18 months. The first 6 months were considered as warm-up periods and not taken into account for statistics. The average cycle time, percent tardy lot and average tardiness of tardy lots are considered as major performance measures, and the results are presented in Table 2.

When the WIP balance approaches are incorporated into the IFD rule, from the fab viewpoint, the average cycle time of all products are improved compared with the case of only the IFD rule, whereas, from the product viewpoint, not each product’s cycle time is reduced. According the IFD rule, actually the ODD rule plays an important role. Because the low volume products have a tight target due date, the ODD rule tries to process them as soon as possible. But the WIP balance approaches reduce the weight of due date control, therefore, it has a positive effect for the high volume products. While the low volume products naturally get no benefit but lose cycle time and tardiness performances. In our case, not only the low volume products but also the normal products like ‘C5F’ are influenced. Our assumption becomes true, the due dates of low volume products are sacrificed by the WIP balance of high volume products. Look at the results that are from the combined due date control approaches, the average cycle time and tardiness performances outperform the one which includes the IFD rule and WIP balance. Because the low volume products acquire the chance to speed up when they are close to the due date or already tardy despite of WIP balance, which saves the cycle time and increases the on-time delivery obviously. In contrast, the cycle time and tardiness of other products degrade a little, because they share the cost what the low volume products benefit. Since the high volume products have enough time to spend in the fab, the cost for high volume products is reasonable and acceptable. Moreover, the cycle time and tardiness of the whole fab are superior over other 2 cases when due date control is complementary to WIP balance.

Table 2: Three performance measures of each products for Case 1.

Case 1: 99.5% Fab Loading									
	Avg. Cycle Time (day)			Percent Tardy Lot (%)			Avg. Tardiness for Tardy Lot (day)		
Product	IFD	IFD + W	IFD + W + D	IFD	IFD + W	IFD + W + D	IFD	IFD + W	IFD + W + D
C6N3	25.2	25.3	25.2	8.8	9.6	0	0.10	0.17	0
B6HF	28.8	29.1	28.7	20.4	28.8	7.1	0.15	0.22	0.06
C4PH	19.6	20.1	19.6	59.7	67.8	45.6	0.48	0.82	0.40
C6N2	28.8	28.3	28.4	0.5	0	0	0.03	0.01	0
OX2	31.2	28.4	28.6	0	0	0	0.005	0.002	0
C5F	34.7	35.0	35.1	5.2	12.0	13.3	0.06	0.26	0.28
C5P	30.4	29.4	29.5	18.2	4.6	9.4	0.14	0.06	0.10
C5PA	33	32.4	32.6	0	0	0	0	0	0
B5C	42.5	41.8	41.9	0	0	0	0	0	0
Fab	30.5	30.0	30.0	12.5	13.6	8.4	0.11	0.17	0.10
IFD + W: IFD rule combines with WIP balance approaches, IFD + W + D: IFD rule combines with WIP balance and due date control approaches.									

Next the same simulation procedures are carried out for Case 2 and the results are showed in Table 3. The low volume products have a extremely tight target due date (due date flow factor 1.3). Even though only the IFD rule is applied, the low products are the ones which have tardiness. We know that it is not possible to achieve non-tardiness, and what we desire is to reduce the tardiness as much as possible.

Table 3: Three performance measures of each products for Case 2.

Case 2: 99.4% Fab Loading									
	Avg. Cycle Time (day)			Percent Tardy Lot (%)			Avg. Tardiness for Tardy Lot (day)		
Product	IFD	IFD + W	IFD + W + D	IFD	IFD + W	IFD + W + D	IFD	IFD + W	IFD + W + D
C6N3	31.9	31.5	31.7	0	0	0	0	0	0
B6HF	36.8	36.4	36.3	0	0	0	0	0	0
C4PH	23.3	22.5	22.8	0	0	0	0	0	0
C6N2	25.6	25.7	25.4	0	0	0	0	0	0
OX2	24.5	24.0	24.2	0	0	0	0	0	0
C5F	30	29.4	29.7	0	0	0	0	0	0
C5P	15.2	15.7	15.0	63.9	72.1	50.1	0.28	0.41	0.18
C5PA	17.1	17.3	17.0	12.7	28.7	10.4	0.14	0.26	0.12
B5C	22.3	22.5	22.1	17.3	22.9	8.2	0.46	0.62	0.25
Fab	25.2	24.9	24.9	10.4	13.8	7.6	0.09	0.14	0.06
IFD + W: IFD rule combines with WIP balance approaches, IFD + W + D: IFD rule combines with WIP balance and due date control approaches.									

The performance is quite clear and similar to Case 1 when the WIP balance approaches are utilized. The cycle time and tardiness of low volume products degrade, although the cycle time of the whole fab reduces. The due date approaches prove again that they can effectively improve the cycle time and tardiness of low volume products with small cost to the cycle time of other products.

4 SUMMARY

In this study, the WIP balance and due date control approaches were proposed and incorporated into a simple global dispatching rule 'IFD' that is from Infineon AG, Dresden. The MIMAC6 model was slightly modified to become a high and low volume products environment to test our idea. The low volume products have a tight target due date, while the high volume products have a loose target due date. Firstly, the WIP balance approaches were incorporated into the IFD rule, we found out that the due dates of low volume products were sacrificed by achieving WIP balance for the high volume products. Then the due date control approaches were integrated into IFD rule too, which was considered as complementary to WIP balance approaches. The due date control approaches provided an effective mechanism to speed up low volume products to save cycle time and reduce tardiness, in the meantime, the cycle time and tardiness performances of the whole fab can still maintain a good level compared to the case of only the IFD rule and the IFD rule combined with WIP balance.

ACKNOWLEDGMENTS

This work was supported by Infineon Technologies AG Dresden, Germany. The authors would like thank Frank Lehmann and Sebastian Werner for their fruitful discussion.

REFERENCES

Baker, K. R., and J. W. M. Bertrand. 1981. "A Comparison of Due-Date Selection Rules." *AIIE Transactions* 13:123-131.

Fowler, J., and J. Robinson. 1995. "Measurement and Improvement of Manufacturing Capacities (MIMAC): Final report." Technical Report 95062861A-TR, SEMATECH, Austin, TX.

Lee, B., Y.H. Lee, T. Yang, and J. Ignisio. 2008. "A Due-date Based Production Control Policy using WIP Balance for Implementation in Semiconductor Fabrications." International Journal of Production Research 46:5515-5529.

Li, S., T. Tang, and D. W. Collins. 1996. Minimum inventory variability schedule with applications in semiconductor fabrication. IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING 9:1–5.

Zhou, Z., and O. Rose. 2010. "A Composite Rule Combining Due Date Control and WIP Balance in a Wafer Fab." In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, 2085-2092. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

ZHUGEN ZHOU is a PhD student at University of the Federal Armed Forces Munich, Germany. He is a member of the scientific staff of Prof. Dr. Oliver Rose at the Chair of Modeling and Simulation. He received his M.S. degree in Computational Engineering from Dresden University of Technology. His research interests include dispatching concepts for complex production facilities and work center modeling for wafer fab. His email address is zhugen.zhou@unibw .de.

OLIVER ROSE is the professor for Modeling and Simulation at the Department of Computer Science, University of the Federal Armed Forces Munich, Germany. He received an M.S. degree in applied mathematics and a Ph.D. degree in computer science from Würzburg University, Germany. His research focuses on the operational modeling, analysis and material flow control of complex manufacturing facilities, in particular, semiconductor factories. He is a member of IEEE, INFORMS Simulation Society, ASIM, and GI, and General Chair of WSC 2012.