Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

ACCELERATING PRODUCTS UNDER DUE-DATE ORIENTED DISPATCHING RULES IN SEMICONDUCTOR MANUFACTURING

Oliver Rose

Institute of Computer Science
University of Würzburg
Würzburg, 97074, GERMANY

ABSTRACT

In semiconductor manufacturing facilities, there is often the need to speed up certain product types. This is usually done by either assigning higher priorities or by reducing due dates. In this paper, we study the effects of accelerating one product type by a tighter due date on the on-time delivery performance of the other products manufactured. It turns out that the results depend on the considered factory, its load, and the accelerated product. As a consequence, it will be hard for production planners to find simple rules of thumb for the effects of accelerating products. In general, detailed simulation experiments will be required.

1 INTRODUCTION

In semiconductor industry, a variety of production control techniques are applied in order to increase throughput, to decrease cycle times, and to achieve on-time delivery of the products (Fowler and Robinson 1995, Wein 1988). Some manufacturers use scheduling approaches but still the majority of the fabs are run under the regime of dispatch rules. For an overview of dispatch rules typically applied in semiconductor industry see Atherton and Atherton (1995). Looking for simple dispatch rules for on-time delivery (OTD) control, the effectiveness of the Critical Ratio (CR) dispatch rules in wafer fabs is often discussed. In Rose (2003), we showed that CR is outperformed by the Operation Due Date (ODD) rule for tight due dates. In semiconductor manufacturing, the due dates are usually determined by means of Flow Factors (FF) which are defined as the target cycle times divided by the raw processing times (RPT). For instance, a FF of 2 says that a lot spends half of its cycle time in processing state and the other half in non-processing states like waiting. Then the due date is the release date plus RPT times target FF. In the above study, the FF values for all products were equal. In a real factory this is usually not the case because some products are more important than others. Generally, make-to-stock products will have larger FF targets (typically larger than 2.5) than make-to-order ones (1.5 to 2.5). Of course, the minimum FF that can be achieved depends on the fabrication facility (fab) and its product mix. Production planners may set even lower FF targets but this will result in very long cycle times and huge inventory levels. In some cases the fab can even become unstable (Rose 2002).

The promising results of the ODD study with equal FF values motivated us to investigate the scenario of having different FF targets. We devoted our interest to a special case of this problem. One product type has to be accelerated, i.e. should have a low FF, while all other product types have the same but higher FF values. Then, the question arises about the functional relationship between the FF reduction of the accelerated product and the FF increase for the remaining products while keeping a given level of OTD performance. For production planners it would be very helpful if there would be an easy to understand heuristic for the solution of this problem.

The paper is organized as follows. In the next section, there is an overview of the ODD rule, the fab models, and the simulation experiments. Then, the simulation results are discussed. The paper closes with concluding remarks.

2 SIMULATION EXPERIMENTS

In this section, we introduce the ODD dispatch rule, the fab models, and experimental setup for our simulation runs.

2.1 The Operation Due Date Dispatch Rule

This rule is not mentioned in the classic dispatching rule overviews for semiconductor manufacturing like (Atherton and Atherton 1995). We heard about ODD the first time from simulation and planning practitioners from Infineon Technologies AG, a German semiconductor manufacturer.

Unlike CR which only takes into account the due date of the product, i.e., the due date of the final processing operation, ODD also considers due dates for all intermediate steps.

The ODD of operation i is defined as the Release Time + RPT(i) * FF where RPT(i) denotes the RPT for a sequence of processing steps or operations from operation 1 to operation i (including operation i). For the final operation of a lot the ODD is equal to the classical due date as used in CR.

In Rose (2002), we saw that tight due dates lead to considerable problems when CR dispatching is applied. Tight in this case means target average cycle times that are lower than the corresponding cycle times using FIFO dispatching. The CR fabs with tight due dates showed the following behavior. During warm-up more and more lots enter the fab and the average cycle time increases because of the increase of waiting times at the tools. In front of highly loaded machines the lots experience more waiting times. In addition, tools break down and the CR value of the lots shrinks again. Because of the fact that the due date for the lots is less than the average FIFO cycle time, a considerable number of lots tend to become late. The CR rule assigns higher priorities to these lots to speed them up. Then, fresh lots have to wait. Because of the large amount of lots with higher priority they are using up their slack time to the due date while still being in the first part of their route. The waiting times at the first steps grow up to a certain balance threshold. From that moment on the fab is stable but operating at a very high inventory and cycle time level. This issue arises because problems to keep the due date that arise at late operations are propagated backwards in the operations sequence over time. Because all lots are only focused on their final due date the CR dispatching does not provide a mechanism to speed up lots before they are very close to their due dates.

The ODD rule does not have this problem by design. Because there are strict due dates for each operation the lots are already kept at the right pace to meet their due date from the early operations on. Slack times for fresh lots are shorter than in the CR case and thus they do not have to let old lots pass before they are processed. Therefore, it is not possible for ODD dispatching that problems at operations at the end of the processing sequence propagate back to the operations at the beginning.

2.2 Experimental Environment

As test models we used the MIMAC (Measurement and Improvement of MAnufacturing Capacities) test bed datasets 1 and 6. These date sets were chosen based on the experiences from prior studies (Rose 2001, Rose 2002, Rose 2003). Table 1 shows the basic properties of the model fabs.

Table 1: Considered MIMAC Datasets

Fab	Tool Groups	Tools	Products	max. Steps
1	83	265	2	245
6	104	228	9	355

For further details on the datasets and their download: see <http://www.eas.asu.edu/~masmlab>.

The simulation runs were carried out with Factory Explorer 2.8 from WWK. We simulated 7 years of fab operation. The first two years were considered as warm-up phase and not taken into account for the statistics. We checked the length of the initial transient both by the cycle time over lot exit time charts and the Schruben test. If there was an indication of initial bias problems the warm-up phase was increased appropriately. The measurement interval was 5 years in all cases.

The only performance measure we were interested in was OTD percentage for the accelerated and remaining product types. For our experiments we set a target value of 95%, i.e. more than 95% of the lots have to meet the due date. For each fab we selected product types to be accelerated.

- Fab 1: products 0 and 1,
- Fab 6: products 9, 3, and 8.

We considered factory loads of 91% and 98% in order make the results comparable to prior studies.

First, we conducted several pilot studies. After selecting the fab model, the accelerated product type, and the fab load, a series of simulations was run where the FF value of the accelerated product was increased from 1.2 to 2.8 in steps of 0.2. For each of these FF values we determined the corresponding FF for the remaining products by gradually increasing this value until either 95% OTD percentage for all lots or a FF of 5.0 was reached. The 5.0 limit was necessary to keep the study tractable. Even with this limit we had to run several hundred simulation experiments.

Based on the pilot studies, we selected a few interesting cases for this paper, where we decreased the FF step size from 0.2 to 0.1.

3 SIMULATION RESULTS

Figure 1 shows a typical result for fab 6 at 98% load where product 3 was accelerated. On the x axis there is the FF of the accelerated product and on the y axis the minimum

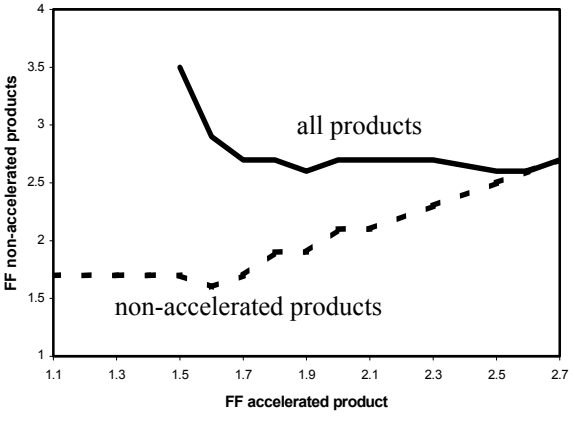

Figure 1: FF Curves for Product 3, Fab 6 @ 98%

flow factor for the remaining products in order to achieve 95% OTD percentage. The dashed curve depicts the values where the non-accelerated products reach their OTD target, the solid line represents 95% OTD for all products, i.e. accelerated and non-accelerated ones.

While the non-accelerated products meet their OTD target already with rather low FF values it requires a considerable increase in FF to achieve 95% OTD percentage for both non-accelerated and accelerated products.

The graph can be split into three parts or FF ranges.

- Low FF (1.1 to 1.4): the FF of the accelerated product is too low to be compensated by increasing the FF values for the remaining products. For practical planning purposes this FF range should not be considered.
- Medium FF (1.5 to 2.5): the FF of the accelerated product can be achieved and it is lower than the FF of the remaining products. This is the most interesting FF range where we will focus on in the remainder of this paper.
- High FF (2.6 and above): all products meet the FF target. There is no reason to try to accelerate a product. Thus this FF range is of little interest.

The border between medium and high is exactly the FF where a fab with equal FF values for all product types reaches 95% OTD percentage (Menth 2003, Rose 2003). In addition, it is the FF value where the dashed and solid curves meet.

So far, we found no approach with small or no simulation effort to determine the border between low and medium FF values, i.e., the minimum FF which can be used for the accelerated products where it is still possible to find a FF value for the remaining products that is less than 5.0.

For the remainder of this section we will focus on the functional relationship between reducing the FF for the accelerated product and increasing the FF values for the remaining products. Therefore we transform the above graph in the following way. Reference FF is the minimum FF value where in fab with equal FF targets all products achieve 95% OTD percentage or better, i.e. the border FF between medium and high FF as defined above.

Figure 2 shows the solid curve after transformation. The reference FF (with no reduction) is 2.6. Then the graph reads as follows. If we want to decrease the FF of product 3 by, say, 1.0 (resulting in a target FF of 1.6 = 2.6 - 1.0) we have to add 0.3 to the reference FF of the remaining product types (resulting in a target FF of 2.9 = 2.6 + 0.3).

We observe that for a considerable range of FF decrease for the accelerated product, up to 0.9 in this case, there is only a very small increase in FF required for the remaining products. There are only two FF values, 1.0 and 1.1, where the increase for the rest of the products is tolerable, then the FF has to be larger than 5.0 for all other products.

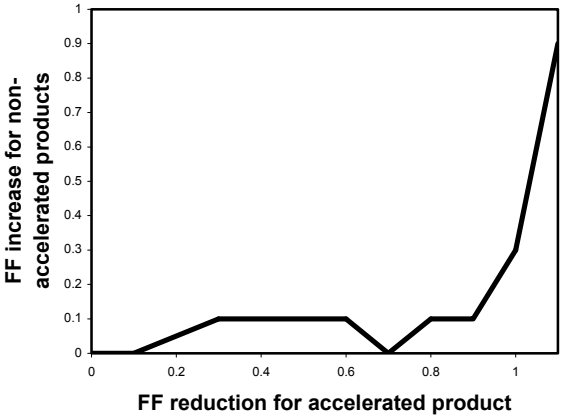

Figure 2: FF Function for Product 3, Fab 6 @ 98%

For practical purposes it is important to know the range of FF decrease for the accelerated product type with little impact on the other products, e.g., with an FF increase of less or equal than 0.3. The following figures provide insight in the FF relationship of accelerated and non-accelerated product for the following scenarios.

Figure 3 shows product 3 of fab 6 at 91% load, Figure 4 depicts product 1 of fab 1 at 91% load, and Figure 5 at 98% load.

In Figure 3, the range for accelerating product 3 is reduced compared to Figure 2 from 0.9 to 0.6. In general, there range for accelerating a product without increasing the FF values of the remaining product for more than, say, 0.2 or 0.3 is larger at fab loads of 98% than at 91%. A rather extreme case is shown in Figures 4 and 5. At 91% it is only possible to decrease the FF by 0.1 from 2.1 to 2.0 for product 1. For smaller FF targets the remaining products must have FF values larger than 5.0. It is worth noting that for a FF of 2.1 all products meet the 95% OTD percentage objective. In the 98% case it is possible to reduce the FF target by as much as 0.4 before the situation becomes critical.

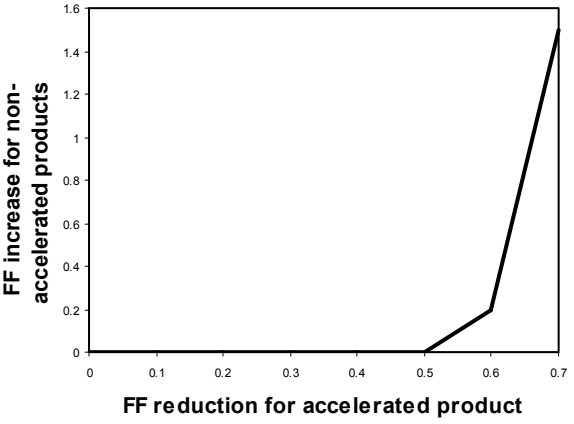

Figure 3: FF Function for Product 3, Fab 6 @ 91%

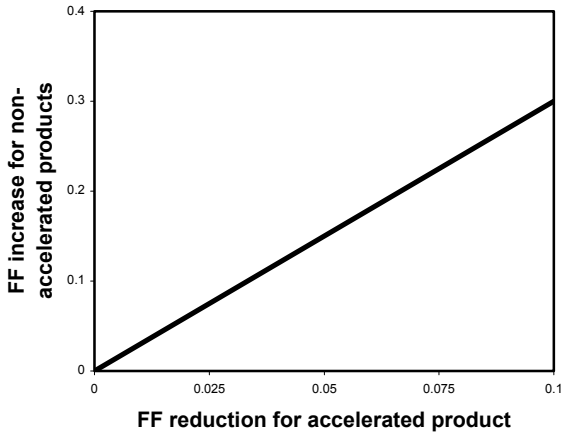

Figure 4: FF Function for Product 1, Fab 1 @ 91%

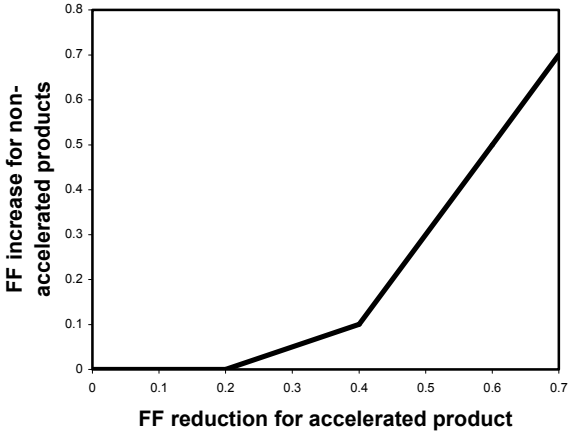

Figure 5: FF Function for Product 1, Fab 1 @ 98%

In Table 2, the FF reference points and the maximum FF reduction range for the products are given where the required FF increase for the non-accelerated products is less or equal 0.3.

Table 2: FF Reference Points and Ranges

Fab & Load	FF reference point	Product	FF range
Fab 1 @ 91%	2.2	0	0.2
		1	0.1
Fab 1 @ 98%	2.8	0	0.4
		1	0.5
Fab 6 @ 91%	2.1	0	0.6
		3	0.6
		8	0.6
Fab 6 @ 98%	2.6	0	0.9
		3	0.9
		8	0.9

It is important to take into consideration that we only used discrete points for the FF values on a grid with 0.2 or 0.1 distances for our experiments. If it would be possible to continuously change the FF values instead of our stepwise increase, the curves would look much smoother. However, the amount of generated simulation output data and the run length of our simulation study leads to the assumption that with current computer equipment FF increments of less than 0.05 are practically not tractable.

4 CONCLUSION

Based on our simulation results we conclude that data from two fab models are not sufficient to find reasonable heuristics to determine the functional relationship between flow factor (FF) decrease of the accelerated product and required FF increase of the rest. The acceptable ranges for FF decrease depend on the fab, its load, and the accelerated product. At the moment it is not clear how to find a practical approach for estimating the FF ranges apart from a detailed and time-consuming simulation study.

It is promising, however, to see that for a lot of products the range is rather large for reducing the FF target while keeping the FF increase for the other products at a slightly increased level. For example, in fab 6 at 91% load the FF for product 3 can be decreased from 2.1 to 1.6 without increasing the FF values for the remaining products and still achieving 95% on-time delivery performance.

ACKNOWLEDGMENTS

The author would like to thank Stefan Mühleck for his valuable simulation efforts and fruitful discussions.

REFERENCES

Atherton, L. and R. Atherton. 1995. *Wafer Fabrication: Factory Performance and Analysis*. Boston: Kluwer.

Fowler, J. and J. Robinson. 1995. Measurement and improvement of manufacturing capacities (MIMAC): Final report. Technical Report 95062861A-TR, SEMATECH, Austin, TX.

Menth, S. 2003. Simulationsstudie zur Abfertigungsdisziplin Operation Due Date (Simulation study about the Operation Due Date dispatch rule, in German). Technical Report, Institute of Computer Science, University of Würzburg.

Rose, O. 2001.The Shortest Processing Time First (SPTF) Dispatch Rule and Some Variants in Semiconductor Manufacturing. In *Proceedings of the 2001 Winter Simulation Conference*, pp. 1220-1224.

Rose, O. 2002. Some Issues of the Critical Ratio Dispatch Rule in Semiconductor Manufacturing. In *Proceedings of the 2002 Winter Simulation Conference*, pp. 1401-1405.

Rose, 0. 2003. Comparison of Due-date Oriented Dispatch Rules in Semiconductor Manufacturing. In *Proceed-*

ings of the 2003 Industrial Engineering Research Conference, May 18-20, Portland, OR.

Wein, L.M. 1988. Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1(3):115-130.

AUTHOR BIOGRAPHY

OLIVER ROSE is assistant professor in the Department of Computer Science at the University of Würzburg, Germany. He received an M.S. degree in applied mathematics and a Ph.D. degree in computer science from the same university. He has a strong background in the modeling and performance evaluation of high-speed communication networks. Currently, his research focuses on the analysis of semiconductor and car manufacturing facilities. He is a member of IEEE, ASIM, and SCS. His web address is <www3.informatik.uni-wuerzburg.de/~rose>.