Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

ECONOMY OF SCALE EFFECTS FOR LARGE WAFER FABS

Oliver Rose

Institute of Applied Computer Science
Dresden University of Technology
Dresden, 01062, Germany

ABSTRACT

In this paper, we present the results of a simulation study for semiconductor wafer fabrication facilities (wafer fabs) where we multiplied the number of tools per tool group and the number of operators. We were interested in the effects on the product cycle times when we keep the fab utilization constant while increasing the size of the tool groups by constant factors, i.e., forming so-called giga fabs. It turns out, that the drop in cycle time is considerable.

1 INTRODUCTION

In semiconductor manufacturing, there is an ongoing effort to build new fabs in order to increase the capacity to produce even more wafers. There are a number of factors which play an important role in this decision. Here, we restrict ourselves to multiplication effort, flexibility and product cycle times. Multiplication effort is the planning, building, and ramp-up effort for these new fabs. Flexibility means the freedom to choose another tool if the planned tool cannot be used, e.g., because of a breakdown or maintenance work.

The first approach which is used to increase capacity, is the "copy exactly" paradigm. In this case, the blueprints of an existing fab are used to build a clone of this factory both with respect to hardware and software. As a consequence, the multiplication effort is small, the capacity doubles but flexibility and cycle time stay the same.

The next opportunity is to build a number of new fabs with similar capacities from scratch. Then, the multiplication effort is large, the cycle times will be similar, and the flexibility will also stay on about the same level. The flexibility in the first two cases does not improve because we will have to move blocked lots to another fab to continue processing. This will only work if this fab is close to the original fab like in so-called "borderless" fabs (Lendermann et al. 2004).

The third way to plan capacity extension, is to build huge fabs, sometimes called giga fabs, which have multiples of the capacity of traditional fabs just by multiplying the number of tools in a tool group and the number of operators. The multiplication effort is similar to building a new fab of traditional size but there will be a boost in flexibility. For instance, if the factory has 4 tools where the traditional fab had just 1, it will happen rarely that all tools break down at the same time. As a consequence there is almost always a tool available for processing. In addition to flexibility improvements, we expect considerable reductions in product cycle times. This assumption is based on the queuing theory findings for parallel tools (Kleinrock 1975). For the same utilization a system with more parallel servers leads to smaller cycle times. Due to the fact that a wafer fab is a network of parallel tools, we expect the same cycle time reduction effects as for the simple one stage case. In addition, we expect effects on batch tools and tools with sequence-dependent setups.

To quantify the amount of these multiplication effects on a giga fab, we performed a simulation study with several factory models under a number of scenarios.

The paper is organized as follows. In the next section, we give an overview of the wafer fab models used and the simulation environment. Then we discuss the cycle time effects for a variety of scenarios. The paper concludes with an outlook on future studies.

2 SIMULATION EXPERIMENTS

As test models we used the MIMAC (Measurement and Improvement of MAnufacturing Capacities) test bed datasets 1 and 6 (Fowler and Robinson 1995). 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 2.9 from WWK. We simulated 6 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.

The default dispatching rule for all tools was FIFO (First In First Out) to avoid side effects due dispatch rule parameters such as due dates. We simulated fab (bottleneck) utilizations of 75%, 85%, and 95%. For both fab models, we increased the number of tools and operators by factors of 1, 2, 4, and 8.

3 EXPERIMENTAL RESULTS

First, we show some global results for the complete fabs. Then, we outline the effects on specific tool groups like batching equipment and tools with setups.

3.1 Overviews

Table 2 and 3 show typical fab performance measures for fab model 1 and 6 at a utilization of 85%.

Table 2: Overview Fab 1

Size Factor	Exit Rate (Units Per Week)	Exit Rate Normalized	Avg. Cycle Time (Days)	Cycle Time Over RPT	Cycle Time 95% Quantile (Days)
1	3234.2	3234.2	27.1	2.0	34.7
2	6468.3	3234.2	17.8	1.3	21.3
4	12936.7	3234.2	15.4	1.1	18.0
8	25873.4	3234.2	14.6	1.1	16.7

By multiplying the size of the fab, the cycle time averages and the upper percentiles decrease considerably, even for moderately loaded fabs.

Table 3: Overview Fab 6

Size Factor	Exit Rate (Units Per Week)	Exit Rate Normalized	Avg. Cycle Time (Days)	Cycle Time Over RPT	Cycle Time 95% Quantile (Days)
1	1066.3	1066.3	24.7	1.8	32.7
2	2132.6	1066.3	18.3	1.4	24.0
4	4265.3	1066.3	15.6	1.2	20.3
8	8530.5	1066.3	14.4	1.1	19.0

In fab 1, the variability of the cycle time is also reduced. For the original fab (Size Factor of 1), the 95% quantile is 28% larger than the average whereas for a Size Factor of 8 the quantile is 14% larger. This is not the case for fab 6.

3.2 Cycle Times

In this section, we examine the top cycle time contributors in more detail. Figure 1 shows that multiplying the factory size leads to enormous queue time reductions.

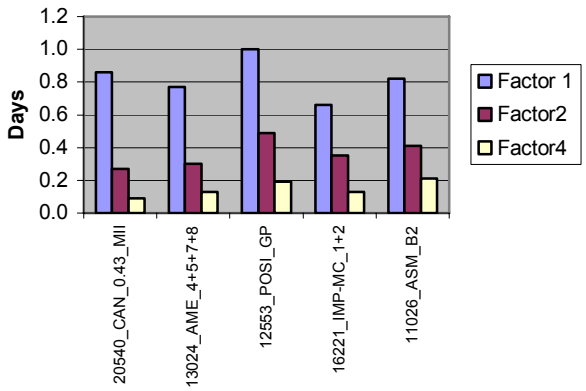

Figure 1: Total Queue Time for Top Contributors (Fab 6)

The improvements are larger for the top contributors than for the tool groups that have less influence on the cycle time (not shown here). The largest total amount of queue time reductions is achieved when doubling the original size (Factor=1). It becomes smaller and smaller when the size is further increased.

3.3 Batch Tools

In semiconductor manufacturing, most batch operations are very time consuming oxidation processes. To achieve low cycle times, i.e., to avoid long queue times at these tools, a variety of dispatching approaches is used, e.g., starting the batch processing with incomplete batches (Atherton and Atherton 1995). In a giga fab, with the same load and product mix as the default (Factor=1) fab, we expect considerable queue delay reductions because the volume of lots that are available for batch building is larger. Therefore the probability to form full batches becomes larger, too. In addition, it is more likely to find a free batch tool.

Table 4 shows a ranking of the average per operation queue delay (QD) of fab 6. The first five tool groups are batch tools. Only the last tool group (in brackets) is a single tool.

If we double the tool and operator counts the queue delays at the first 5 batch tools in the list is reduced by half. For the first single tool in the list, however, the queue delay is about a quarter of the default case (Factor=1).

Table 4: Per Operation Queue Delay (Fab 6)

Tool Group	QD [h] Factor=1	QD [h] Factor=2	QD [h] Factor=4
11026_ASM_B2	16.7	8.7	4.5
12553_POSI_GP	8.4	3.7	1.5
17421_HOTIN	6.0	3.1	1.4
11024_ASM_A4_G3_G4	5.8	3.1	1.6
11027_ASM_B3_B4_D4	5.1	2.6	1.3
(17221_K-SMU236)	5.0	1.3	0.1

Due to the prominent position of the batch tools in the queue delay ranking, we were interested whether how the batch tool improvement for larger fab sizes relates to the improvements of the other tools. Table 5 shows this comparison.

Table 5: Average Queue Delays (Fab 6)

	Factor=1	Factor=2	Factor=4
Avg. QD Single Tools [h]	1	0.29	0.075
Avg. QD Batch Tools [h]	1.76	0.67	0.26
Avg. QD All [h]	1.33	0.47	0.17

It turns out that for fab 6 the relative queue time reduction is approximately the same for single and for batch tools.

3.4 Setup Tools

Another important source for delay and capacity losses are tools with setups. In semiconductor manufacturing, the dispatching systems try to avoid setups with a variety of approaches, e.g., by dedicating tools to a certain product or sorting the waiting lots according to their setup requirements (Atherton and Atherton 1995). As a consequence, we expect less setups in larger fabs because it is more likely that the required setup state is already available and processing of the lot can continue without spending time for setups. Figure 2 shows the setup percentages for the top setup tools of fab 6.

Again, the giga fab approach leads to reductions in setup times. The effect, however is much smaller than we expected. For instance, for the left tool group, doubling the number of tools reduces the setup time percentage only by 3.4%. One reason for that is that we already used a setup avoidance rule for the default fab. As a consequence, the setup percentage was already rather low.

4 CONCLUSION AND OUTLOOK

In this paper, we presented the results of a small pilot study on the effects of multiplying the number of tools and operators in semiconductor manufacturing. It turns out that there is a considerable potential to reduce cycle times in such huge fabs, sometimes called giga fabs.

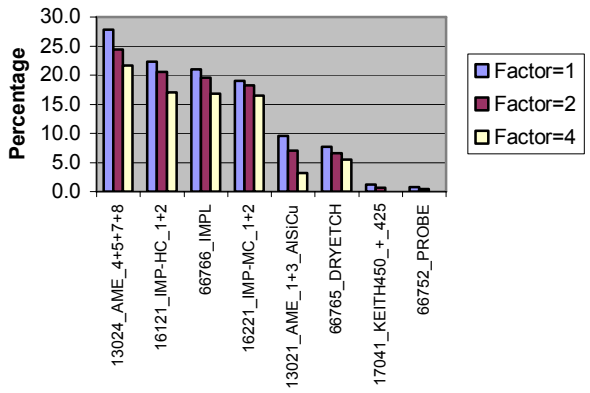

Figure 2: Setup Percentages (Fab 6)

We show results which indicate that some tools play a more prominent role in this reduction than others. In our case, batch tools were more important than setups.

The promising results motivate us to perform further studies with giga fab models that will include the following aspects.

- Effects on different products,
- Comparison of effects on high mix / low volume foundry type fabs vs. low mix / high volume memory or processor fabs,
- Effect of different factors for different tool sets.

ACKNOWLEDGMENTS

The author would like to thank Uwe Gosda for his valuable simulation and data analysis efforts.

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.

Kleinrock, L. 1975. *Queueing Systems. Volume 1: Theory*. Wiley.

Lendermann, P., B.P. Gan, Y.L. Loh, H.K. Tan, S.K. Lieu, J. Fowler, and L. McGinnis,. 2004. Analysis of a Borderless Fab Scenario in a Distributed Simulation Testbed In *Proceedings of the 2004 Winter Simulation Conference*, pp. 1896-1901.

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, O. 2003. Comparison of Due-date Oriented Dispatch Rules in Semiconductor Manufacturing. In *Proceedings of the 2003 Industrial Engineering Research Conference*, May 18-20, Portland, OR.

AUTHOR BIOGRAPHY

OLIVER ROSE holds the Chair for Modeling and Simulation at the Institute of Applied Computer Science of the Dresden University of Technology, 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. Web address: <www.simulation-dresden.com>

.