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.

MODELING SEMICONDUCTOR TOOLS FOR SMALL LOTSIZE FAB SIMULATIONS

Kilian Schmidt
Jörg Weigang

AMD Saxony LLC & Co. KG, MS I11-IE
Wilschdorfer Landstrasse 101
D-01109 Dresden, GERMANY

Oliver Rose

Department of Applied Computer Science
Technical University of Dresden
D-01062 Dresden, GERMANY

ABSTRACT

Short cycle times are critical to the success of semiconductor manufacturing. The addition of more and more mask layers leads to higher raw process times and makes short cycle times an increasingly challenging task. One cycle time reduction possibility semiconductor manufacturers now look at is lotsize reduction. A reduction in lotsize transfers directly into lower raw process times. Modeling and simulation are key to assess opportunities and risks of such an approach. This paper looks at the implications that follow from small lotsizes for tool models used for the assessment.

1 INTRODUCTION

Short cycle time is key to many operational success factors. It enables lean inventory, short time to market, fast yield learning, fast excursion finding and fast reactions to customer demand. Therefore the increase in raw process time caused by the ever increasing number of mask layers and the subsequent increase in cycle time presents an obstacle to operational excellence. Traditional cycle time reduction efforts targeting mainly at variability reduction have led to cycle time improvements in the past but seem to reach their limit as they can not influence raw process times. The introduction of cluster tools in the early 90s reduced raw process times for single wafer tools by processing wafers of a lot in parallel in different chambers. Adding even more chambers to cluster tools to further increase the parallelism of processing is unrealistic. Apart from practicability issues it would lead to fewer tools per tool group and subsequently to an increase of the negative impact by variability.

Therefore semiconductor manufacturers look at the second architectural factor that drives the raw process time of single wafer tools. It is well known, that raw process time of single wafer tools increases with lotsize (see Wood et al. 1994 and 1996 or Perkinson et al. 1994). After a long period of stable standard lotsizes at 25 wafers per carrier the reduction of lotsize becomes a tempting approach to reduce raw process times. Initial simulations have been done by Wakabayashi et al. (2004) and Bonnin et al. (2003). They show a significant reduction in fab cycle time for a 13 wafer lotsize.

The quality of fab simulation modeling output depends heavily on the accuracy of the tool models used. The tool modeling methods used for simulation and their accuracy have been studied in detail in a number of publications, but always under the premise of complete 25 wafer lots. This paper evaluates current modeling methods for small lotsizes and proposes a new tool modeling method for cluster tools with parallel chambers.

The paper is organized as follows. Different semiconductor tool types are presented in Section 2. In Section 3 we outline applications of tool models, before Section 4 introduces standard tool models. The comparison of modeling types is provided in Section 5. In Section 6 we suggest a different tool model type for small lotsize simulations.

2 DIFFERENT SEMICONDUCTOR TOOL TYPES

Semiconductor tools are usually grouped into three tool types:

- **Batch tools** process batches of one or multiple lots.
- **X-piece tools** process batches of x wafers (x < standard lotsize).
- **Single wafer tools** process single wafers.

Figure 1 illustrates the lotsize dependence of raw process times for these tool types in principle. It shows that different lotsizes do not influence batch raw process times. There is an influence on batch equipment productivity, which is both simple and well known. Batch tools already have a much longer process time than corresponding single wafer equipment and this gap increases with lotsize reduction.

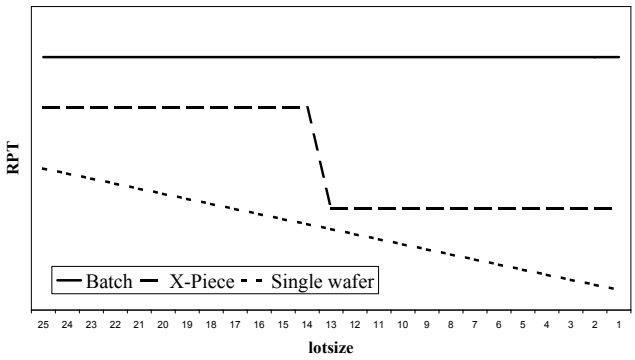

Figure 1: Raw Process Time of Different Tool Types Dependant on Lotsize (X=13 for X-Piece RPT)

Two possible ways exist for closing that gap:

- The change to mini-batch tools which have a much lower batch size and a significantly reduced RPT.
- The change to single wafer tools.

The RPT of x-piece tools depends on the number of x-piece runs necessary to process a lot. Hence, there is a large difference in RPT between lotsizes of $n \times x$ and $n \times x + 1$. Between these jumps RPT stays constant, assuming that lots are not mixed in x-piece runs. X-Piece tools have lost importance during the last years. Disk implanters used to be x-piece tools but for process reasons most were replaced by single wafer implanters. CVD tools remain important x-piece tools with an x-piece size of 2.

The RPT of single wafer tools decreases with lotsize reduction for non-cluster single wafer tools (The reduction is not always regular - see Section 5). Single wafer tools are already the predominant type of tools. Their share will probably increase further as they feature shorter raw process times than the corresponding batch tools and this advantage will increase with smaller lotsizes. Therefore the semiconductor industry is expected to adopt a uniform or near uniform single wafer processing strategy for all production equipment in an effort to reach fast cycle times (see Pettinato et al. 2004). In view of this development this paper will focus on single wafer tools in the following. The applicability and importance of models and insights on the other tool types will be pointed out though.

Apart from the distinction into batch, x-piece und single wafer tools there is a further tool type distinction which is often neglected in simulation models. Tools can be differentiated by how processing of consecutive wafers overlaps. Overlapping can be categorized in three ways:

- No Overlapping
- Constant Overlapping
- Varying Overlapping

Very simple tools like, e.g., many metrology tools which have only one process resource connected to a equipment front-end module usually fall into the first category. These tools place wafer n back into the carrier before wafer n+1 is retrieved from it.

Most tools fall into the second category with constant overlapping. Wafer n+1 is retrieved from the carrier before wafer n is placed back into it and retrievals and placements repeat itself in a constant interval.

Tools of the third category also retrieve wafer n+1 before wafer n is placed back, but the interval of it is not constant. Cluster tools with parallel chamber configuration usually fall into this category provided they do not run at the mechanical limit.

A third distinction of tool types is the lot operation mode. Some tools run in parallel lot mode. They mix wafers of different lots in the operational sequence, usually because process resources are dedicated for specific steps in the factory flow. However these tool configurations are very susceptible for typical problems of small lotsize manufacturing and it is expected that they will not play an important role in such an operational environment. Therefore these tools are not discussed further in this paper.

3 TOOL MODELS AND ITS APPLICATIONS FOR SIMULATION ANALYSES

For performance and clarity reasons, fab simulations usually do not model the individual machines in detail. Instead of modeling the machines with their robots, process and interface resources, they use abstract tool models based on mathematical formulas. These tool models have different applications. The apparent ones are the calculation of tool throughput (THP) and raw process times dependant on the input. A third type of output becomes more important especially with small lotsizes: How long is the carrier exchange by the AMHS allowed to last, without affecting the performance of the tool. This duration is defined as required carrier exchange time (req. CET). Figure 2 illustrates the different factors which influence the req. CET:

- The RPT (1)
- The Overlapping (OL) of Processing of two consecutive lots (2)
- The number of load ports (ΣLP)(3)

From these influencing factors we derive the following formula for the req. CET

$$\text{Req.CET} = \left(\sum \text{LP} - 1\right) \times \text{RPT} - \sum \text{LP} \times \text{OL} \,. \qquad (1)$$

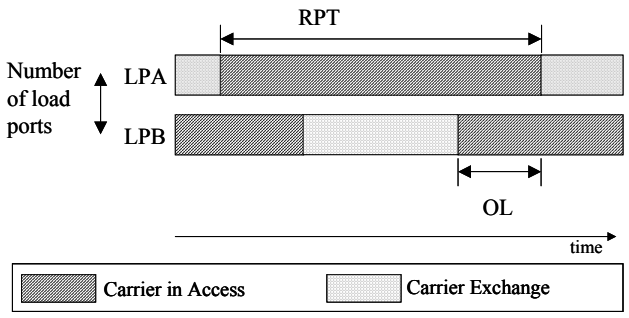

Figure 2: Illustration of the Factors Influencing the Required Carrier Exchange Time

If the req. CET is not or not always met by the AMHS, then THP will be affected of course. This means that tool throughput is depending on external factors that are not related to the tool itself. Therefore tool model applications are divided into direct and indirect applications. The calculation of RPT and req. CET are direct application whereas THP is now an indirect application that also requires other input. It is not within the scope of this paper to assess this indirect effect but this dependency drives the tool productivity impact of small lotsize manufacturing. The throughput assessment can be done in an linked simulation model or with analytical methods based on the figures presented here.

4 STANDARD MODELS

The simplest way to model tools is with their process rate only. The derived formula for RPT is

$$RPT = n \times \frac{1}{PR} \tag{2}$$

where PR represents the process rate and n the lotsize. Given the focus on RPT instead of THP, we will further use intervals instead of rates in the formulas for the sake of simplicity. This transforms Equation (2) into

$$RPT = n \times PI \tag{3}$$

with PI representing the Process Interval. There is no Overlap in this case. Thus no formula is given for it. This "No Overlap" approach leads to correct results if the processing of consecutive wafers (x-pieces, batches respectively) does not overlap.

To account for constant overlapping the formula has to be modified for RPT into

$$RPT = OL + n \times PI\ . \tag{4}$$

The RPT contains a fixed part, the overlapping, and an incremental part that is added once for each wafer.

Usually the size of overlapping can be obtained easily out of MES-event-logs. Alternatively it can be calculated from PI and the process time of the first wafer PT1 as

$$OL = PT1 - PI\ . \tag{5}$$

This "Constant overlap" approach leads to exact results for most tools with the exception of tools with varying overlap. The accuracy of this approach has been classified as sufficient for these tools by Wood 1996, however, assuming 25 wafer lots and THP as the primary application.

These RPT formulas can easily be adjusted for x-piece and batch tools. PI is then replaced with XPI (x-piece Process Interval) or BPI (Batch Process Interval) and n with $\lceil n/x \rceil$ or 1 respectively. For the OL formula, PT1 is in addition replaced by XPT1 (X-piece process time – X-Piece 1) or BPT1 (Batch process time – Batch 1), respectively. If X-pieces contain wafers of different lots additional adjustments have to be made.

5 COMPARISON OF MODELING TYPES

These two standard approaches are now compared to reality in an example. The layout of the example tool is shown in Figure 3. The cluster tool is configured with four parallel process chambers, three load ports and an equipment front-end module (EFEM) with a single blade robot. Each wafer is moved from the carrier on the load port to one process module by the robot. After processing the robot moves the wafer back into the carrier.

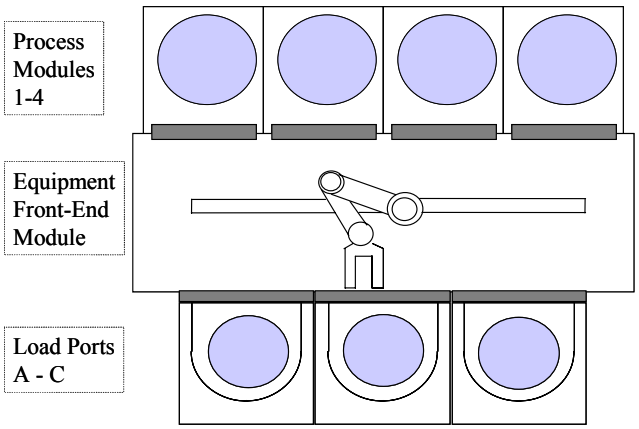

Figure 3: Cluster Tool with Parallel Chambers as an Example

In this example, the processing in the process module lasts 140 seconds and a move lasts 10 seconds regardless of source and destination.

With the physical configuration as cluster tool with parallel chambers and the logical configuration far beyond

the mechanical limit, this example tool features a varying overlap.

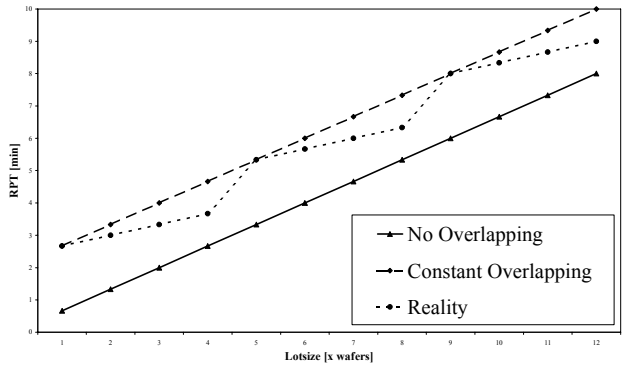

Figure 4: RPT Comparison of Modeling Approaches for First Lot

Figure 4 compares the RPTs calculated by the tool models to reality for the first lot of a lot cascade. The two RPT models form a linear band, which is not surprising as their RPT formula differs only by a constant. The reality numbers stay within the bandwidth but show a non-linear pattern. Some wafer numbers cause the RPT to jump while for others RPT stays nearly constant. Figure 5 illustrates the reason for this behavior. The four chambers of the example tool are loaded within short intervals and their unloading is equally short. However, processing in the chambers lasts relatively long. Therefore the differences in RPT are small between one and four wafers but large between four and five wafers.

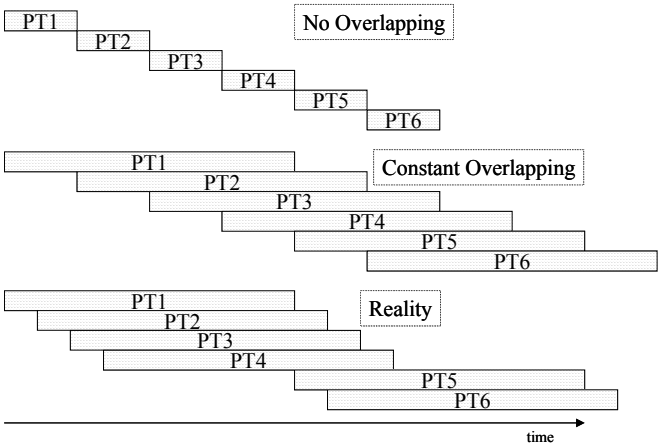

Figure 5: Illustration of Underlying PT-Sequence in Gantt-Chart

The RPT comparison in Figure 4 is a special case where only the first lot is considered to illustrate the basic concept. In continuous production it is also important how often lot changes happen at the long intervals. Figure 6 compares average RPT of tool models and reality under this premise.

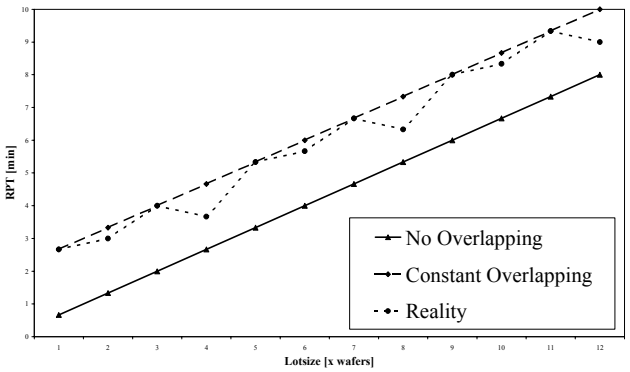

Figure 6: Average RPT Comparison of Modeling Approaches for Continuous Lot Processing

It can be seen that lotsizes that are multiples of the number of chambers make good use of lot changes at long intervals and their RPT differs significantly compared to both models. Lotsizes in between make no or scarce use of lot changes at long intervals and their RPT is pretty close to the values calculated with the "Constant Overlapping"-Model.

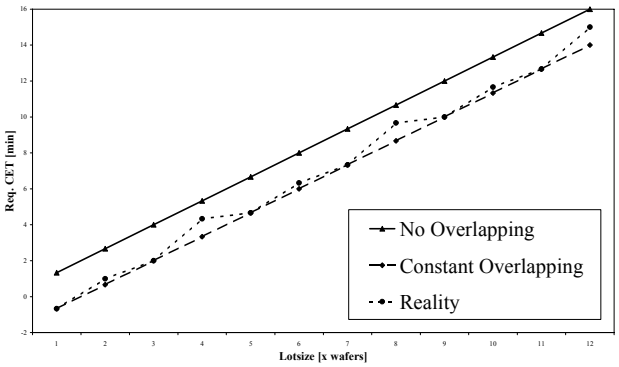

Figure 7: Average Req. CET Comparison of Modeling Approaches for Continuous Lot Processing

Figure 7 compares the second application, the required CET, for the tool models and reality. The same behavior can be seen here. Lotsizes that are multiples of the number of chambers make good use of lot changes at long intervals and allow for long CETs performed by the AMHS. Again the real CET differs significantly compared to both models. For all other lotsizes the calculated CET is close or matches the CET calculated with the "Constant Overlap" model. As the quantitative comparison might be difficult to perform with the charts only, Table 1 and Table 2 give model output, reality and relative error for a selection of lotsizes and the two direct model applications.

Table 1: "No Overlap"-Model Output

Lotsize	Output [min]	Model [min]	Reality [min]	Error (%)
25 wfr	RPT	16.67	18.67	11
	Req. CET	33.33	31.33	6
12 wfr	RPT	8	9	11
	Req. CET	16	15	7
4 wfr	RPT	2.67	3.67	27
	Req. CET	5.33	4.33	23
3 wfr	RPT	2.0	4.0	50
	Req. CET	4.0	2.0	100

Table 2: "Constant Overlap"-Model Output

Lotsize	Output [min]	Model [min]	Reality [min]	Error (%)
25 wfr	RPT	18.67	18.67	0
	Req. CET	31.33	31.33	0
12 wfr	RPT	10.00	9	11
	Req. CET	14	15	7
4 wfr	RPT	4.67	3.67	27
	Req. CET	3.33	4.33	23
3 wfr	RPT	4.0	4.0	0
	Req. CET	2.0	2.0	0

Three conclusions can be drawn from the above example:

- Lotsizes that are multiples of the number of parallel chambers are beneficial in terms of RPT and req. CET.
- The "No Overlap"-model is generally insufficient for small lotsizes.
- While the "Constant Overlap"-Model produces good results for most lotsizes it is insufficient for beneficial small lotsizes that make use of the tool configuration.

6 NEW TOOL MODEL TYPE FOR SMALL LOTSIZE SIMULATION

Based on the conclusions of Section 5 there is the necessity for a tool model that correctly accounts for the varying overlapping. Equation (6) gives a RPT formula that accounts for this distinction with c denoting the number of parallel chambers and SI denoting a staged interval that accounts for small intervals illustrated in Figure 5. Similar to the illustrations the RPT formula is first given only for the first lot for better understanding as

$$RPT = PT1 + c \times \left\lfloor \frac{n-1}{c} \right\rfloor \times PI + ((n-1) \bmod c) \times SI . \quad (6)$$

Figure 8 illustrates the two different intervals used in the formula. The process interval of the constant overlap approach is still used and the only relevant interval for RPT calculations with $n \times c + 1$ wafers. In between the staged interval adjusts the RPT to correctly represent reality.

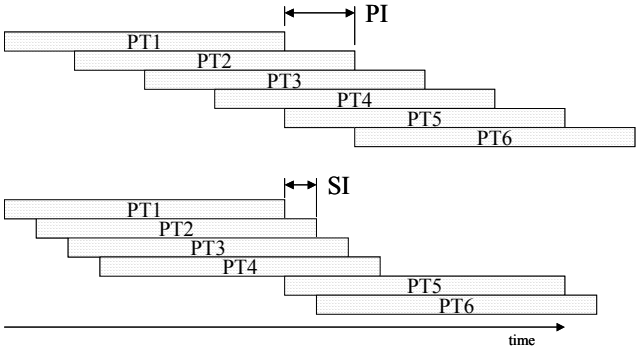

Figure 8: Illustration of Different Intervals Used for RPT Calculation in Gantt-Chart

For continuous lot processing the RPT can vary. E.g., assuming a lotsize of 3 the RPT of the first lot is different than the RPT of the second lot. The reason is illustrated in Figure 8. The RPT of the second lot encompasses the long time span between the process starts of wafer 4 and wafer 5. Therefore, the RPT formula for continuous lot processing given as

$$RPT = PT1 + \sum_{m=1}^{c} \left(\left\lfloor \frac{m \times n - 1}{c} \right\rfloor - \left\lfloor \frac{(m-1) \times n}{c} \right\rfloor \right) \times PI + \frac{1}{c} \times \sum_{m=1}^{c} \left(((m \times n - 1) \bmod c) - (((m-1) \times n) \bmod c) \right) \times SI \quad (7)$$

averages over the RPTs of the first c lots to correctly include all possible RPTs. Clearly these formulas are more complicated than the ones given in Section 4 but they correctly represent the real behavior of cluster tools with parallel process chambers. Applied to a typical semiconductor process flow with the majority of etch and wet tools having a four chamber configuration and a lotsize of four wafers, the cumulated RPT of the flow is 4-8% less than calculated with the "Constant Overlapping"-Model.

In Equation (8) we modify the formula for Overlapping in the simplest possible way into

$$OL = RPT - n \times PI \quad (8)$$

with OL depending on RPT given for the "Varying Overlap"-Model in Equation (7). The exactness of the overlapping determination is important as it drives the productivity impact by late delivery. If the impact assessment is performed with one of the conventional models for the example tool with a four wafer lotsize then the error is in the double digit percentage area. As Equations (7) and (8) cor-

rectly depict reality, there is no need for another model-reality comparison like in Tables 1 and 2. The table would show the same numbers for model output as for reality and an error of 0%.

7 SUMMARY

This paper evaluated current modeling methods for small lotsizes and proposed a new tool modeling method for cluster tools with parallel chambers. It has been shown that the "No Overlap"-Modeling approach is clearly insufficient for small lotsizes and the "Constant Overlap"-Modeling approach is insufficient for some beneficial lotsizes. Albeit regrettably more complex the proposed "Varying Overlap" approach offers the necessary exactness for a simulation assessment of opportunities and downfalls of small lotsize manufacturing.

Furthermore the paper showed that there is benefit in matching lotsize and cluster tool configuration.

Further studies will focus on the indirect application of tool models, the throughput dependant on tool model and AMHS performance.

ACKNOWLEDGMENTS

The authors want to thank AMD Dresden's Industrial Engineering team for their helpful contributions.

APPENDIX: DEFINITIONS

- Overlapping (OL):
  Time difference between last wafer back into carrier of lot n and first wafer out of carrier of lot n+1
- Process Time of wafer n (PTn):
  Time difference between wafer n back into carrier and wafer n out of carrier
- Raw Process Time (RPT):
  Time difference between last wafer back into carrier and first wafer out of carrier
- Process Interval (PI):
  Average time difference between consecutive wafer process ends
- Staged Interval (SI):
  Short Time differences between consecutive wafer process ends of parallel cluster tools

REFERENCES

Bonnin, O., D. Mercier, D. Levy, M. Henry, I. Pouilloux, and E. Mastromatteo. 2003. Single-Wafer/Mini-Batch Approach for Fast Cycle Time in Advanced 300-mm Fab. *IEEE Transactions on Semiconductor Manufacturing*, 16(2), pp. 111-120.

Perkinson, T. L., P. K. McLarty, R. S. Gyurcsik, and R. K. Cavin III. 1994. Single-Wafer Cluster Tool Performance: An Analysis of Throughput. *IEEE Transactions on Semiconductor Manufacturing*, 7(3), pp. 369-373.

Pettinato, J., and D. Pilai. 2004. Technology decisions to Minimize 450mm Wafer Size Transition Risk. In Proceedings of the 2004 ISSM Conference.

Wakabayashi, T., S. Watanabe, Y. Kobayashi, T. Okabe, and A. Koike. 2004. High-Speed AMHS and Its Operation Method for 300mm QTAT Fab. *IEEE Transactions on Semiconductor Manufacturing*, 17(3), pp. 25-28.

Wood, S. C., S. Tripthi, and F. Moghadam. 1994. A Generic Model for Cluster Tool Throughput Time and Capacity. In *Proceedings IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, pp. 194-199.

Wood, S. C. 1996. Simple Performance Models for Integrated Processing Tools. *IEEE Transactions on Semiconductor Manufacturing*, 9(3), pp. 320-328.

AUTHOR BIOGRAPHIES

KILIAN SCHMIDT is an Industrial Engineer at Advanced Micro Devices' Dresden site. He is responsible for throughput analysis and optimization of tools as well as other industrial engineering efforts. He holds a M.S. in Mechanical Engineering from the University of Stuttgart, Germany. In his Ph.D. research at Dresden University of Technology he analyses equipment and AMHS strategies for fast cycle times with modeling and simulation. He is a member of the ISMI 450mm Factory Architecture Focus Team. His e-mail address is <kilian.schmidt@amd.com>.

OLIVER ROSE holds the Chair for Modeling and Simulation at the Institute of Applied Computer Science of the Dresden University of Technology. He received a 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. His e-mail address is <oliver.rose@tu-dresden.de>.

JÖRG WEIGANG received a M.S. in Computer Science from Dresden University of Technology, Germany. Since November 1997 he is working with AMD Fab30 in Dresden. In his current position as Member Technical Staff he is responsible for WIP management and simulation. In this matter he has mainly developed the dispatching system of AMD's FAB30 and FAB36 using Brooks Real Time Dispatcher. His e-mail address is <joerg.weigang@amd.com>.