Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

SIMULATION-ENABLED DEVELOPMENT LOT JOURNEY SMOOTHENING IN A FULLY-UTILISED SEMICONDUCTOR MANUFACTURING LINE

Wolfgang Scholl
Matthias Förster

Infineon Technologies Dresden GmbH
Königsbrücker Strasse 180
01099 Dresden, GERMANY

Patrick Preuss
André Naumann

D-SIMLAB Technologies GmbH
Wiener Platz 6
01069 Dresden, GERMANY

Boon Ping Gan
Peter Lendermann

D-SIMLAB Technologies Pte Ltd
8 Jurong Town Hall Road
#23-04 JTC Summit
Singapore 609434, SINGAPORE

ABSTRACT

Technology and product development have high priority in an advanced semiconductor manufacturing facility such as the Infineon Dresden fab. From the perspective of line performance this means that short cycle times for development lots have to be guaranteed to enable the required learning cycles. Long-term simulation is used in dynamic capacity planning to find a compromise between short cycle times for the development corridor and high utilisation of the installed tool capacity. All products in the fab run with customer-specific due dates. As such, negative side-effects caused by the accelerated development lot corridor through increased dispatch priorities have to be minimised. In turn, for day-to-day operations short-term simulation is used for early detection of bottleneck situations and other sudden resource availability problems. With focus on the development corridor, a Lot Cycle Time Forecaster was realised. The aforementioned manifold applications of discrete-event simulation are described in this paper in more detail.

1 INTRODUCTION

Discrete Event Simulation (DES) is used at Infineon Dresden to enhance the flow of development lots within the flow of regular productive lots. To do this effectively, the capacity aspect has to be analysed with regard to the maximum agreeable number of accelerated lots from the perspective of line performance as well as the expected side-effects and how they can be mitigated. This is an important application for long-term simulation because it is not possible to assess them with static methods.

Apart from long-term capacity planning, DES can also be used for online short-term simulation to address issues arising from daily business operations (School 2011). For development lots specifically, such lots often run on new-defined routes. This requires additional engineering effort on many steps for recipe creation and qualification, for example for certain lithography steps for which pre-runners for fine-tuning of recipe parameters are required. In the early stage, many processes need special supervision by specialists including experiments that need to be carried out. Many of these activities are not preventable but they can have a considerable negative impact. Moreover, temporary capacity bottlenecks caused by non-availability of resources (e.g. reticle) or tool and recipe blockings result in process stops that lead to an increase in lot cycle time.

In this setting, to enable a smooth flow of development lots an early warning process for such situation can reduce the probability of process stops. From the development engineer's perspective the lot cycle forecast and arrival forecast at steps with experiments to be carried out is very relevant. The production engineer needs a lot arrival forecast for the work centers of his responsibility and the line control an overview of all anticipated deadlocked development lots in the line within the next days.

Satisfactory solutions to the above-described problem have proved very difficult to develop. In the past these problems have been addressed using queuing approaches, for example, Narahari and Khan (1997, 1998) and also by mathematical programming approaches (Kim and Uzsoy 2008, 2013), as well as using simulation by Ehteshami et al. (1992). In this setting, the hierarchical use of simulation models as described in this paper is considered a major step ahead on the path towards addressing this difficult challenge.

2 THE ROLE OF LONG-TERM SIMULATION

2.1 Capacity Planning and KPI Forecasting

The most important applications for long-term simulation are Dynamic Capacity Planning and KPI Forecasting. At Infineon Dresden this is achieved through a simplified, mostly stochastic fab simulation approach initialised with real line conditions. The forecast horizon is in the range of 3-6 months, the results are derived from ten replications. All the important input parameters to describe a semiconductor manufacturing line are considered in a more or less pragmatic way. Simplified modelling approach means that, for example, cluster tools are not modelled in detail, and products are mapped to representative routes. This results in high transparency and reduced data maintenance effort. Specific questions to be addressed are the assessment of product mix changes and the effect of additional wafer starts exceeding the maximum planned fab load as well as the optimisation of recovery scenarios after line incidents

2.2 Dispatching of Development Lots

Long-term fab simulation is also used for optimisation of dispatch rules, in particular for development lots. Productive lots are dispatched in different due date classes depending on delivery time commitments. They are ranked in the work center queue according to their delay with regard to the operations cycle time target. Development lots have to run with a significant shorter cycle time, typically in the range of 60% compared to productive lots. Acceleration of the development lot corridor results in negative side-effects on fab capacity and as a result increased cycle time of productive lots. The magnitude of negative impact depends on the level of acceleration. A simple due date shift to an earlier date leads to a classification of the respective lots into a higher queuing rank position at the work center. The resulting cycle time effect of this acceleration method depends on the local dispatch rules, the current fab load and several other input parameters. A high prioritisation results in breaks of running setups and preferred processing at batch tools and ultimately a significant local capacity reduction. The benefit will be a stable cycle time widely independent of the current fab situation. An important question to be answered by simulation is what kind of acceleration method to be applied. In our example a high prioritisation was required to accelerate the development lots to target cycle time. This is illustrated in Figure 1.

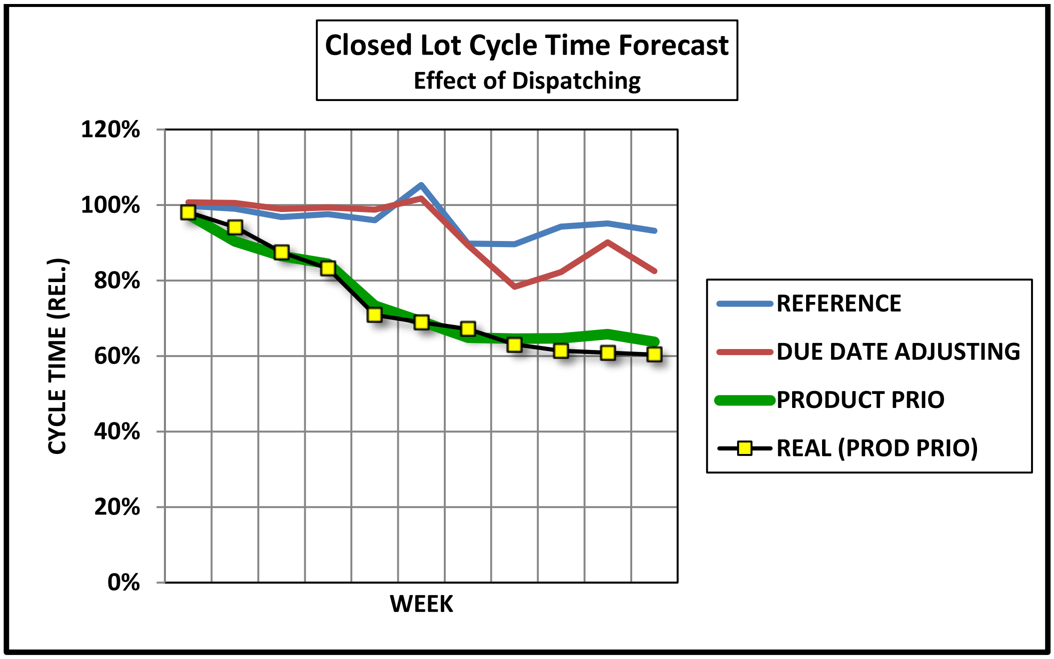

Figure 1: Long-term product cycle time forecast in dependence on dispatching.

The reference line represents the cycle time forecast of a selected product with usual dispatching for normal lots. The simulated scenario "Due Date Adjusting" comprises an acceleration by due date tightening, resulting in a 10% acceleration compared to the reference. The requested stronger and faster acceleration was only predicted with the prioritisation scenario "Product Prio". Ten weeks after changing the product dispatching to the latter setting, real data monitoring proves the confidence in long-term simulation for such use-cases.

2.3 Assessment of Side-Effects of Lot Prioritisation

An important question to be addressed by long-term fab simulation is the cost associated with a prioritised development corridor. A semiconductor manufacturing line with an expensive tool set has to be operated efficiently. Effects such as short cycle time and maximum number of prioritised lots have to offset against financial performance considerations. Simulation can support the definition of surcharges for lot prioritisations. In our example, an extreme scenario as assessed – an acceleration of 20% of the wafer starts by hard product prioritisation. Reference is a fab with typical number of development lots and lot prioritisations. The cycle time effect of an additional 20% accelerated capacity corridor and the compensation by step-by-step reduction of the wafer starts was to be demonstrated.

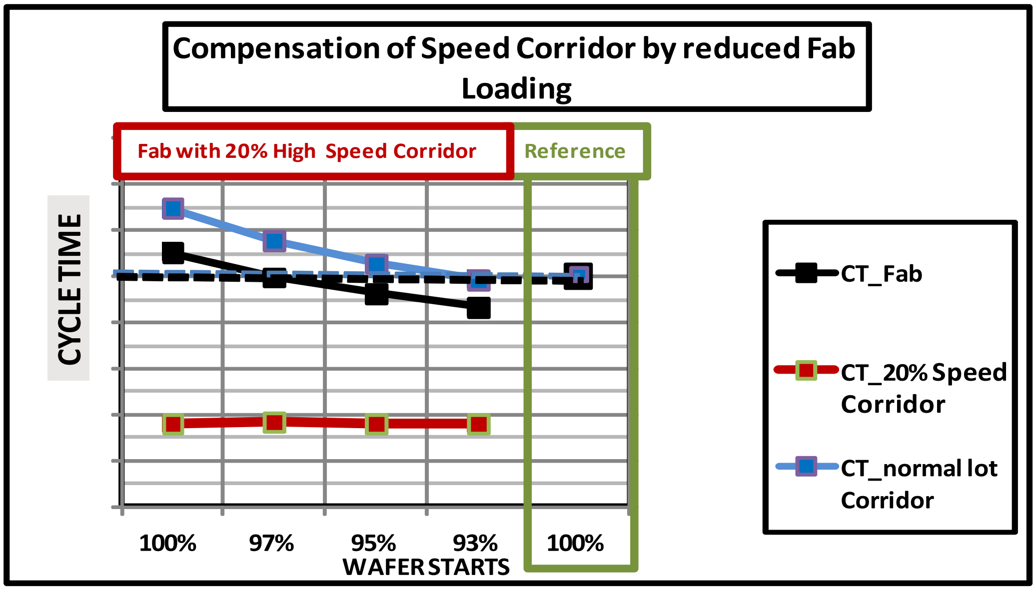

Figure 2: Cycle time effect of reduced fab loading.

In the hypothetical example shown in Figure 2 the implementation of a 20% prioritised corridor has to be compensated by approximately a 7% wafer start reduction to reach the same cycle time compared to normal due date dispatched productive lots. This result depends on actual performance conditions such as the fab load. The cycle time of prioritised lots is not dependent on the fab load in the analysed range because they jump queues at the respective work centers.

The side-effects on cycle time can be different for each product, depending on the length of the process flows, the number of qualified machines, the number of batching processes etc. Of course this has to be analysed by simulation in more detail.

3 DEVELOPMENT LOT FORECAST USING SHORT-TERM SIMULATION

Short-term simulation has also been established at Infineon Dresden for several years. An automated, highly detailed online simulation approach was developed together with D-SIMLAB Technologies. Main application is the lot arrival and WIP forecast on the work center level. With a forecast horizon of seven days it is used in daily operations for optimised scheduling of Preventive Maintenance activities in dependence on expected work center performance (Scholl 2012). Based on this, an additional use case to forecast lot cycle times with particular focus on a dynamic disaster check for development lots was implemented.

3.1 Relevance and Use-Cases

For a short-term lot cycle time forecast in addition to normal simulation input parameters, a lot of detailed static information about tool capacity, dispatching and process releases are needed. This has to be completed by more detailed dynamic run-time dependent information regarding actual resource availabilities (e.g. reticle), tool blockings and other temporary exclusions. The data scheme required for this use case is shown in Figure 3.

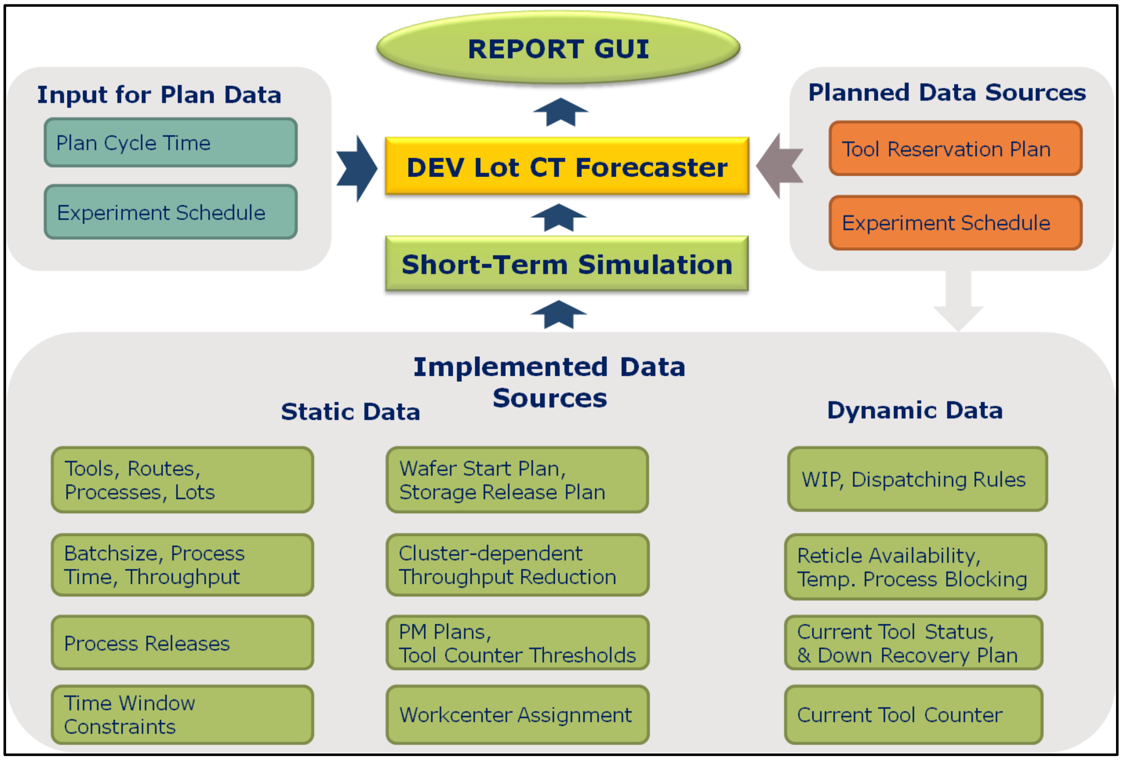

Figure 3: Data scheme of Development Cycle Time Forecaster extension.

The reporting of the simulation results has to cover a wide range starting from customers spanning to technology department, line control and production. This requires different customer-specific reports with focus on single lots as well as work centers, with detailed information on lot level as well as overview tables.

3.2 Single Lot Forecast

The chart for the single lot forecast as displayed in Figure 4 shows the anticipated arrival time (y-axis) at process steps (x-axis) within the next seven days.

Scope of this application is to forecast the cycle time of next seven days compared to target. In our example the forecast gives two important early warnings. The predicted cycle time (red dotted line) is higher as planned (green dotted line). If the expected cycle time violation is not acceptable, the lot has to be prioritised higher. After about six days the forecast drifts up due to a missing tool availability. Here the preventive action can be the proactive change of a scheduled Preventive Maintenance activity or the release of an additional equipment. The single lot cycle time forecast is the most important application for technology engineers. It is also used to enable the presence of specialists when a lot arrives at a process step with planned experiments.

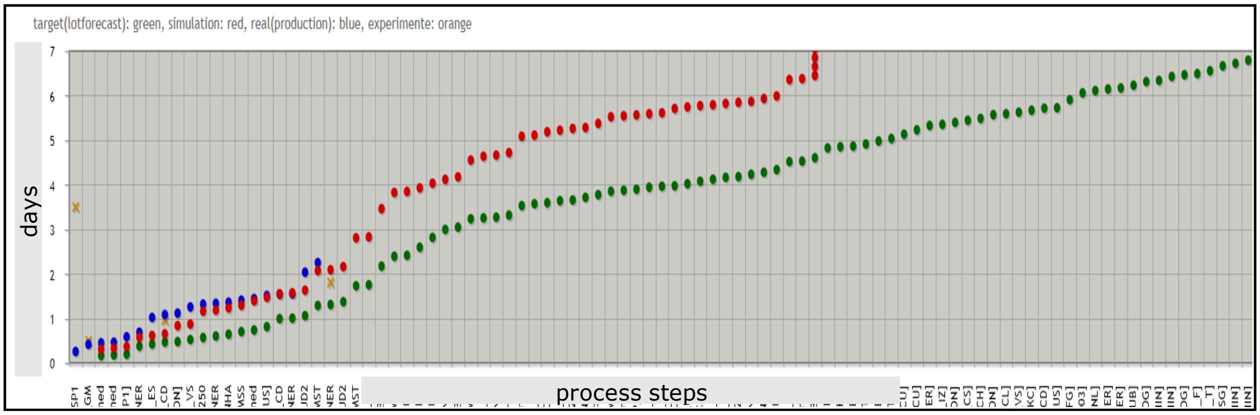

Figure 4: Single lot forecast.

Real values from history are also displayed (blue dotted line) for previous forecasts. This is an important information for the assessment of the forecast accuracy by the user.

3.3 Stuck Lots Summary

Another report is the stuck lot summary table, helpful for user from line control department and for model validation issues. All lots forecasted to be blocked within the next 7 days are listed. From the perspective of simulation it helps to find data mismatches (e.g. missing recipe release information). For the experts from line control it has an early warning functionality to initialise proactive activities. The detailed root cause analysis can be very difficult, so the challenge is to categorise and report the underlying reasons automatically.

3.4 Equipment Group Perspective

Another user group of the development lot cycle time forecast are the production engineers. They are responsible for the performance of the work centers in their area including scheduling of maintenance and engineering activities. Therefore they can use the established short-term simulation application with total lot arrival and WIP forecast. But they also have to ensure fast processing of development lots. If development lots run on new-created routes, usually there are only very few tool qualified for the respective recipes. As a result, even a single maintenance activity can cause a total processing stop for development lots. This shows the importance of an production area specific arrival forecast for development lots.

The forecast of the expected amount and time of arriving development lots for a selected equipment group also allows the production areas to carry out preparation activities accordingly.

3.5 Accuracy

Lot cycle time forecast accuracy is monitored continuously, therefore real data is available from previous reports. A forecast on single lot level is significantly more difficult, because there are no compensatory effects resulting from data aggregation on product or technology level.

For simulation validation issues a special report is created, analysing the deviation from reality on different selectable levels (e.g. product, dispatching class) depending on the forecast horizon. This is shown in Figure 5.

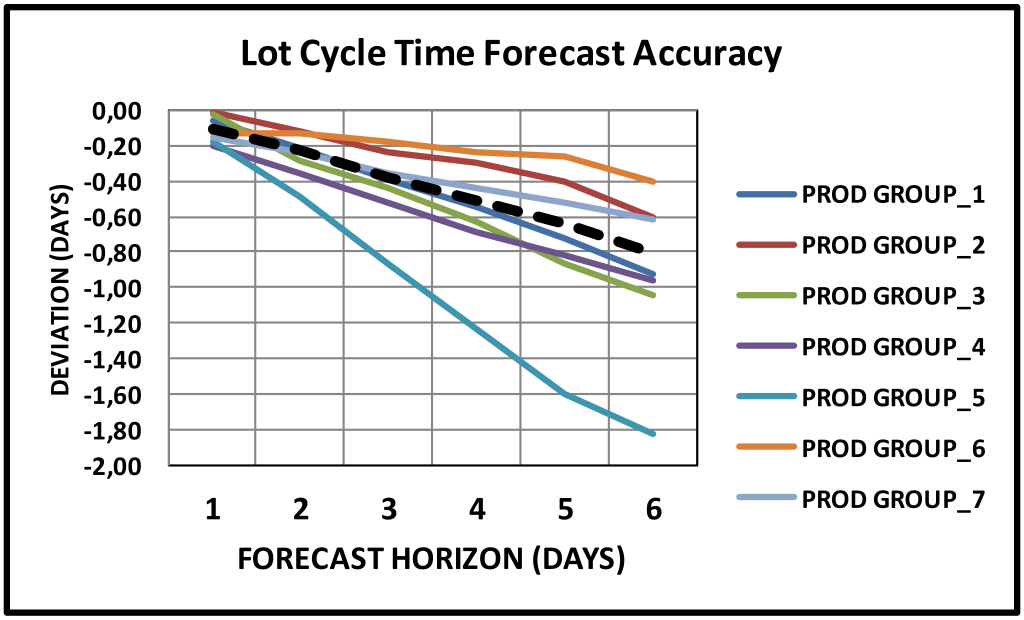

Figure 5: Forecast Accuracy by technology (product group).

Usually the forecasted cycle time is too optimistic and shows a strong dependence on the maturity of the technology. Currently the average accuracy is in the range of 90%, negatively affected by first-runners (first started lots for a new technology). In an early development stage of the technology, additional unpredictable effects at lithography and measurement steps are common. Often the routes are changed during run-time of the lot. On the other hand, we are just identifying potential for improvement of the forecast accuracy. One example is a better modelling of expected additional cycle time consumption at steps with experiments derived from inputs in experiment management database or derived from historical data analysis.

4 SUMMARY AND OUTLOOK

Discrete-event simulation has been established as an important method to support a short cycle-time of development lots. Dispatch rules can be optimised before the lots are started, expected side-effects on fab performance can be assessed and be reduced. In operational business, simulation helps to prevent bottleneck situations and makes needed additional engineering effort easier to schedule.

In future the implementation of extended functionalities is planned. This includes the possibility to start virtual lots for proactive resource checks or the rerun of simulation after a lot has reached a first stuck position. The main focus, however, is on the improvement of forecast accuracy.

REFERENCES

Ehteshami, B., R. Petrakian, and P. Shabe. 1992. "Trade-Offs in Cycle Time Management: Hot Lots." *IEEE Transactions on Semiconductor Manufacturing* 5(2): 101-106.

Kim, S., and R. Uzsoy. 2008. "Integrated Planning of Production and Engineering Process Improvement." *IEEE Transactions on Semiconductor Manufacturing* 21: 390-398.

Kim, S., and R. Uzsoy. 2013. "Modeling and Analysis of Integrated Planning of Production and Engineering Process Improvement." *IEEE Transactions on Semiconductor Manufacturing* 26(3): 414-422.

Narahari, Y., and L. M. Khan. 1997. "Modeling the Effect of Hot Lots in Semiconductor Manufacturing Systems." *IEEE Transactions on Semiconductor Manufacturing* 10(1): 185-188.

Narahari, Y., and L. M. Khan, 1998. "Asymptotic Loss of Priority Scheduling Policies in Closed Reentrant Lines: A Computational Study." *European Journal of Operational Research* 110: 585-596.

Scholl, W., B. P. Gan, P. Lendermann, D. Noack, O. Rose, P. Preuss, and F. S. Pappert. 2011. "Implementation of a Simulation-Based Short-Term Lot Arrival Forecast in a Mature 200mm Semiconductor Fab." In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 1932- 1943. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Scholl, W., M. Mosinski, B.P. Gan, P. Lendermann, P. Preuss, and D. Noack. 2012. "A Multi-Stage Discrete Event Simulation Approach for Scheduling of Maintenance Activities in a Semiconductor Manufacturing Line." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 2263-2272. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

WOLFGANG SCHOLL works as a Senior Staff Expert for modelling and simulation for Infineon Technologies in Dresden (Germany). He studied physics at the Technical University of Chemnitz (Germany) and graduated in solid-state physics in 1984. From 1984 to 1995 he worked as a process engineer for ZMD in Dresden. In 1996 he joined Infineon Technologies (former SIMEC) and worked in the field of capacity planning. Since 2003 he is responsible for fab simulation. He supervises development and roll-out projects and is also a member of the Supply Chain Simulation community. His email address is wolfgang.scholl@infineon.com.

MATTHIAS FÖRSTER works as a System Expert for modelling and simulation for Infineon Technologies in Dresden (Germany). He studied microtechnology at the University of Applied Sciences Zwickau (Germany) and graduated with a diploma degree. Upon joining Infineon Technologies in Dresden in 1996, he first worked in the Plasma Processes Department. In 2001 he changed to the Furnace Processes Department where he worked as a System Expert. Since 2014 is he responsible for short term simulation and supporting long term fab simulation in Infineon Dresden. His email address is matthias.foerster.drs@infineon.com.

PATRICK PREUSS is a Project Manager and the Deputy Manager Germany Operations of D-SIMLAB Technologies (Germany). He has been working in the development of simulation-based applications for Airbus, German Aerospace Centre and Infineon with focus on data analysis and heuristic optimisation methods since 2005. Patrick holds a M.S. degree in computer science from Dresden University of Technology. His email address is patrick@d-simlab.com.

ANDRÉ NAUMANN is a Software Engineer at D-SIMLAB Technologies (Germany). He studied computer science with focus on simulation-based optimisation at Dresden University of Technology and graduated with a diploma degree. Since 2012 André is working D-SIMLAB Technologies (Germany) in deployment projects for customers in semiconductor industry. His email address is andre@d-simlab.com.

BOON PING GAN is the CTO of D-SIMLAB Technologies (Singapore). He has been involved in simulation technology application and development since 1995, with primary focus on developing parallel and distributed simulation technology for complex systems such as semiconductor manufacturing and aviation spare inventory management. He was also responsible for several operations improvement projects with wafer fabrication clients which concluded with multi-million dollar savings. He is now responsible for managing the technology and product development at D-SIMLAB as well as execution of projects in the

semiconductor and engineering domains. He holds a Master of Applied Science degree, specializing in Computer Engineering. His email address is boonping@d-simlab.com.

PETER LENDERMANN is the Co-Founder and CEO of D-SIMLAB Technologies, a Singapore-based company providing simulation-based decision support solutions and services to Aerospace, Semiconductor Manufacturing and other asset-intensive industries. Prior to this he worked at the Singapore Institute of Manufacturing Technology where he led the simulation-related research activities until spinning them off into D-SIMLAB Technologies. He has been engaged in the simulation community since the early 1990's. Peter holds a PhD in Applied High-Energy Physics from Humboldt University in Berlin (Germany) and an MBA in International Economics and Management from SDA Bocconi in Milan (Italy). His email address is peter@d-simlab.com.