Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

OPERATOR RESOURCE MODELLING IN A MULTIPLE WAFER SIZES FABRICATION ENVIRONMENT USING DISCRETE EVENT SIMULATION

Georg Seidel

Infineon Technologies Austria AG
Siemensstraße 2
Villach, 9500, AUSTRIA

Patrick Preuss
André Naumann

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

Soo Leen Low
Chew Wye Chan
Boon Ping Gan

D-SIMLAB Technologies Pte Ltd
8 Jurong Town Hall Road
#23-05 JTC Summit
Singapore, 609434, SINGAPORE

ABSTRACT

Infineon Villach wafer fab facility has a unique characteristic of producing products of different wafer sizes with shared resources, both operators and tools. Daily adaptation of resource allocation to manage changing workloads is done dynamically and imposes challenges in providing high accuracy simulation forecast. In this paper, we discuss the modelling approach that we have taken to represent this dynamic feature, where resources are adjusted in simulation based on WIP situation. The concept is similar to acquiring KANBAN prior to running a production lot. KANBAN quantity is kept at a minimum number in low WIP situation, while the quantity is kept at maximum number at high WIP situation. These values were adjusted by studying the historical moves observed for a period of 3 months. Using this approach, we managed to improve the forecast quality by 7 percentage points in moves and 64 percentage points in WIP.

1 INTRODUCTION

Infineon's front end wafer fab in Villach, Austria, is very complex. The production has been started in the 1970s. Fab capacity has been increased gradually over the years. Therefore there are differently constructed production buildings, producing three different wafer sizes. Work centers are split over these buildings. New buildings are used for highly automated production, older ones still rely on operators for transport and loading. Some production tools are shared among different wafer sizes with costly setup changes to switch between the wafer sizes. Hundreds of different products are processed with different routes. Operator responsibility covers different work center and tasks like loading, unloading and transport of lots. Responsibilities can even change from shift to shift. All this boundary conditions lead to a very complex production environment.

For this complex manufacturing environment it is crucial to forecast problematic production areas for at least 7 days. For example, knowing in advance which production area could potentially have WIP build-up that requires additional resource allocation to mitigate the issue. Discrete-event simulation is the best approach to provide such a forecast, and similar works have already been done by our colleagues at other Infineon sites, such as Infineon Dresden (Scholl et al. 2011) in Germany and Infineon Kulim in Malaysia (Seidel et al. 2017).

Though the application of simulation forecast has been done before, we faced different challenges in the Villach fab due to its unique characteristics. One of these characteristics is the sharing of tool and operator resources from production lines of 150mm, 200mm, and 300mm. Some tools require a high setup effort to convert between different wafer sizes, which means any switch over may be associated with a critical WIP build-up situation. Also, operators are assigned to handle certain tools. These tools are sometimes running different wafer sizes. Often, priority of production is on 200mm wafers and leads to WIP build for 150mm wafers due to lack of operators. Production control department manages this situation by dynamically allocating limited operator resources to clear the WIP, when the WIP reaches a certain level. Though this maybe not the best way to manage the production line, it is at this point in time the best available option to maximize utilization of the available operator resources.

The issues described above post challenges to achieve high accuracy simulation forecast. In this paper, we discuss our approach in addressing the challenge of dynamic operator resources allocation. We provide an overview of our simulation model in Section 2. This is then followed by a discussion of the modelling approach to address the dynamic operator allocation in Section 3. In Section 4, we share the forecast comparison with reality, and conclude the paper in Section 5.

2 SIMULATION MODEL

The key modelling elements for the Villach simulation model are: process flows, tools, tool dedications, process time and throughput, wafer starts, and dispatch rules as discussed in Georg et. al. 2017. However, these modelling elements are not sufficient to represent the complexity of the Villach fab. Two additional modelling elements are required to achieve usable forecast quality, namely bridge tools and dynamic operator resource modelling.

The bridge tools are defined as production tools that can process lots of different wafer sizes. Usage of bridge tools reduces capital investment of the fab, and if setup changes are managed appropriately, will have minimal impact to the production line performance. A typical condition to switch a tool from one wafer size to another is an anticipated high WIP situation of one wafer size and a simultaneously low WIP situation of another wafer size. To model this behavior, we introduced a new modelling feature in D-SIMCON Simulator (D-SIMCON 2018) where for each bridge tool and wafer size a low and a high WIP threshold can be defined. The tool is considered for switching from its current wafer size to another if the corresponding wafer size WIP level is below the low WIP threshold, and the target wafer size WIP level is above the high WIP threshold. The challenge is setting appropriate WIP levels. Necessary information is derived from historical data analysis, by observing typical low and high WIP situations at the bridge tool group. Besides the condition to switch between wafer sizes, we also need to consider the fact that switching the tool requires substantially high setup times in some cases. This setup times are basically capacity loss, and there is no chance of regaining the capacity. Typically, a bridge tool will stay at the same setup for a minimum duration of time before switching to another wafer size again. The simulation model enables this behavior by using a parameter called minimum duration before setup change. One question was whether this duration should be calculated by cumulating the time the tool run production lots, or by using the elapsed time the tool has stayed in the current wafer size setup. We have decided for the latter because the accumulation of productive run time might extend the time a bridge tool stays on the same wafer size setup unnecessarily.

As discussed earlier, within production line operator resources are limited. Each operator is responsible to run more than one tool, and at times, this results in a WIP build up situation due to focusing on production of a specific wafer size. The number of operators is dynamically adjusted to mitigate such situations. Detailed operator modelling can be an option to portrait this characteristic in simulation. However, it is challenging to set this up because operator data like load/unload time, and data associated with an operator skillset that defines what they are capable of doing is not available in a structured form. As such, we have decided to take a different approach, which is to borrow the KANBAN concept to model the dynamic operator assignment. This is discussed in details in the next section.

3 DYNAMIC KANBAN

Traditional KANBAN system limits the WIP level in a certain work area (Ohno 1988). The system was originally designed to control the WIP level to reduce excessive wait time within the KANBAN steps. Due to the nature of the KANBAN as a resource, it is a good alternative to represent operator resources in the simulation model. However, due to the dynamic nature of operator assignment in reality, the KANBAN system is modified to fit our modelling requirement.

The objective of the dynamic KANBAN is to model a varying throughput for selected work areas, whereby we try to keep the WIP in work areas between two levels. KANBAN is assigned to a specific work area. Once the WIP at the work area exceeds the upper limit we allow additional KANBAN to be requested, therefore we increase the work area throughput and reduce the WIP level. When the WIP drops below a threshold, KANBAN resource is removed from the work area, reducing the throughput and therefore increases the WIP.

In order to adapt the KANBAN to model dynamic operator assignment like in reality, the lower and upper WIP thresholds are required to determine the low and high throughput KANBAN values. This is needed to model the dynamic nature of the operator resource in a particular work area. In reality, when WIP exceeds the limit in a particular work area, more operators will be assigned to process the WIP and this reduces the WIP at the work area. We analyzed the hourly WIP level of a work area over the period of 60 days and the lower and upper WIP thresholds were determined by the 20th and 80th quantile as shown in Figure 1.

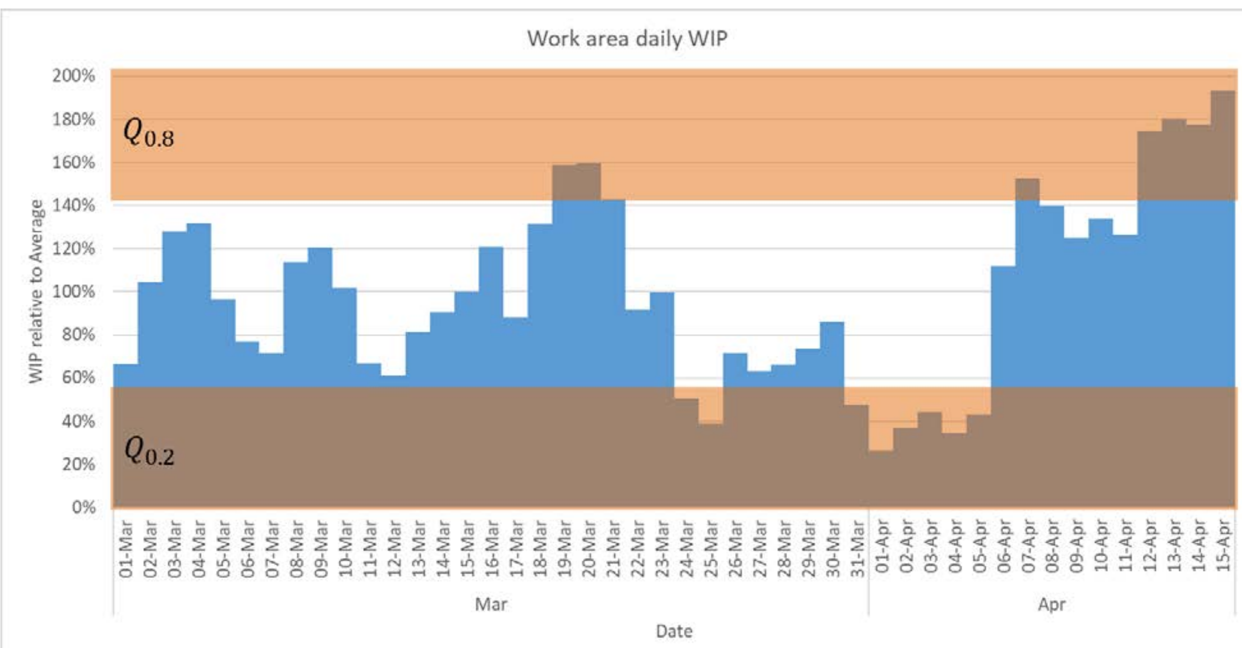

Figure 1: 20th and 80th quantile of work area WIP.

To identify the low throughput KANBAN value, we first measured the moves typically done when the WIP level is increasing. Accordingly we checked the WIP dropping situations for the high throughput KANBAN value. Figure 2 (segment of chart in purple color) shows an example of timeframes when the WIP was dropping. The value for average moves within these periods was calculated.

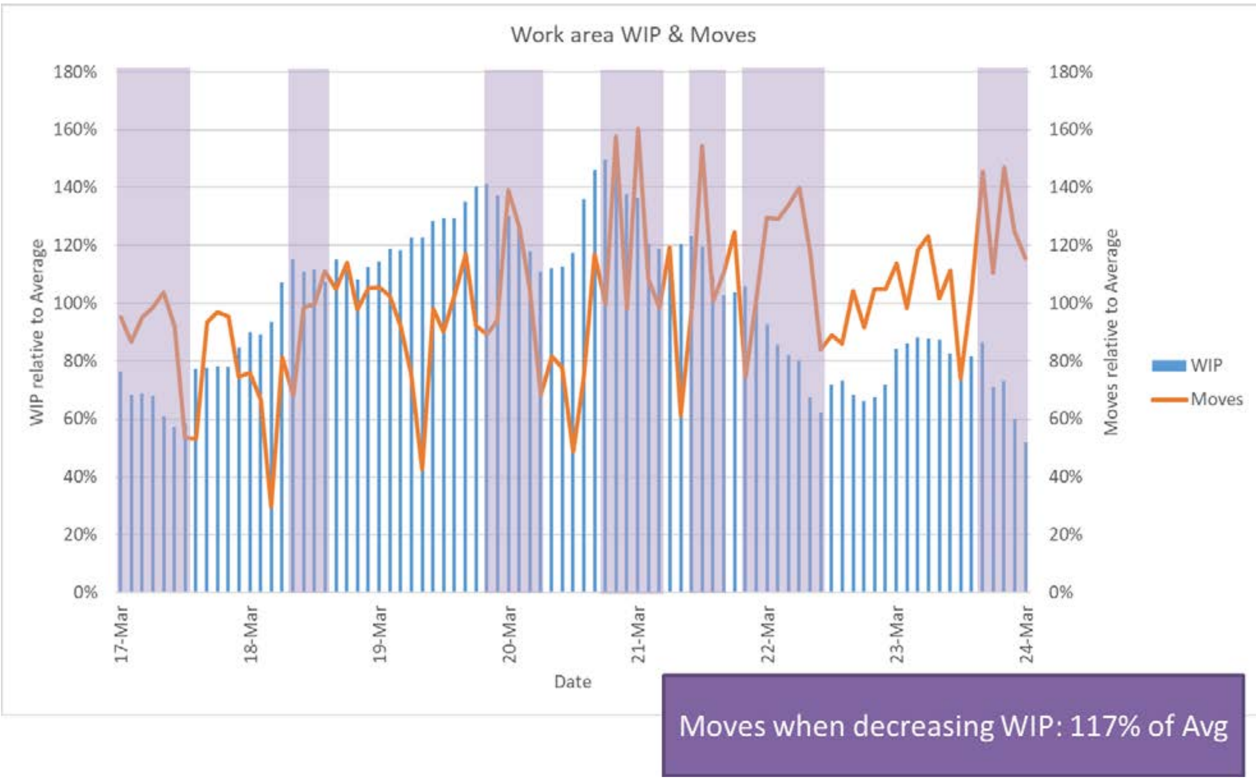

Figure 2: Work areas hourly moves and WIP in WIP decreasing situations.

We then adapted Little's law ($L=\lambda W$) and set these WIP increasing/decreasing moves as arrival rate λ with the average weighted process time as service rate W. The average number of customers in the system (L) will then be used as the upper/lower KANBAN value for the modelling.

During simulation run time, the WIP level is checked to determine if it falls below or above the WIP threshold value and it will trigger the usage of the corresponding KANBAN limits. Currently the check is done once per day at 6 AM as to mimic planning meetings at the start of the day shift where it is decided to speed up or slow down certain work areas. The frequency of switching and its impact to the forecast quality is still subject for further analysis.

However there are limitations to this modelling approach. Operator assignment in reality is dynamic and affected by multiple factors, such as the total manpower available in a particular shift which the dynamic KANBAN does not take into account. Furthermore, the dynamic KANBAN feature is used to increase or decrease KANBAN resources to mimic the reality of operator assignment based on a changing WIP level. The WIP level could change due to multiple factors others than the variance of operator resources, such as tool downs, and tool dedications that could result in changing WIP levels too. This would also trigger the KANBAN to change. In these cases, changing of KANBAN value might not have the intended effect.

4 COMPARISON TO REALITY

For verification we used a set of equipment in lithography area. In the original scenario the work area throughput was only limited by the equipment's capacity. Like shown in Figure 3 the simulation model will fully use capacity and clear all waiting WIP within the first two to three days. The initial high simulation moves decreased due to missing available WIP in the later part of the forecast period. In reality however, the work area was not running at its full capacity for the first four days. During this time period WIP significantly increased. Only at day 4 more operating resources were assigned, allowing the tools to increase their throughput and consequently reduced the WIP.

Figure 3: Base simulation model with high capacity.

Note that the maximum of daily moves in simulation is approximately the same as in reality. This implies that the tool capacity is modelled correctly, but the forecast quality for work center moves is low. In the first days of the simulation run moves were typically higher than in reality, followed by a period with lower moves in simulation compared to reality. This has a significant negative impact on the WIP forecast quality for work areas because simulation WIP is constantly too low.

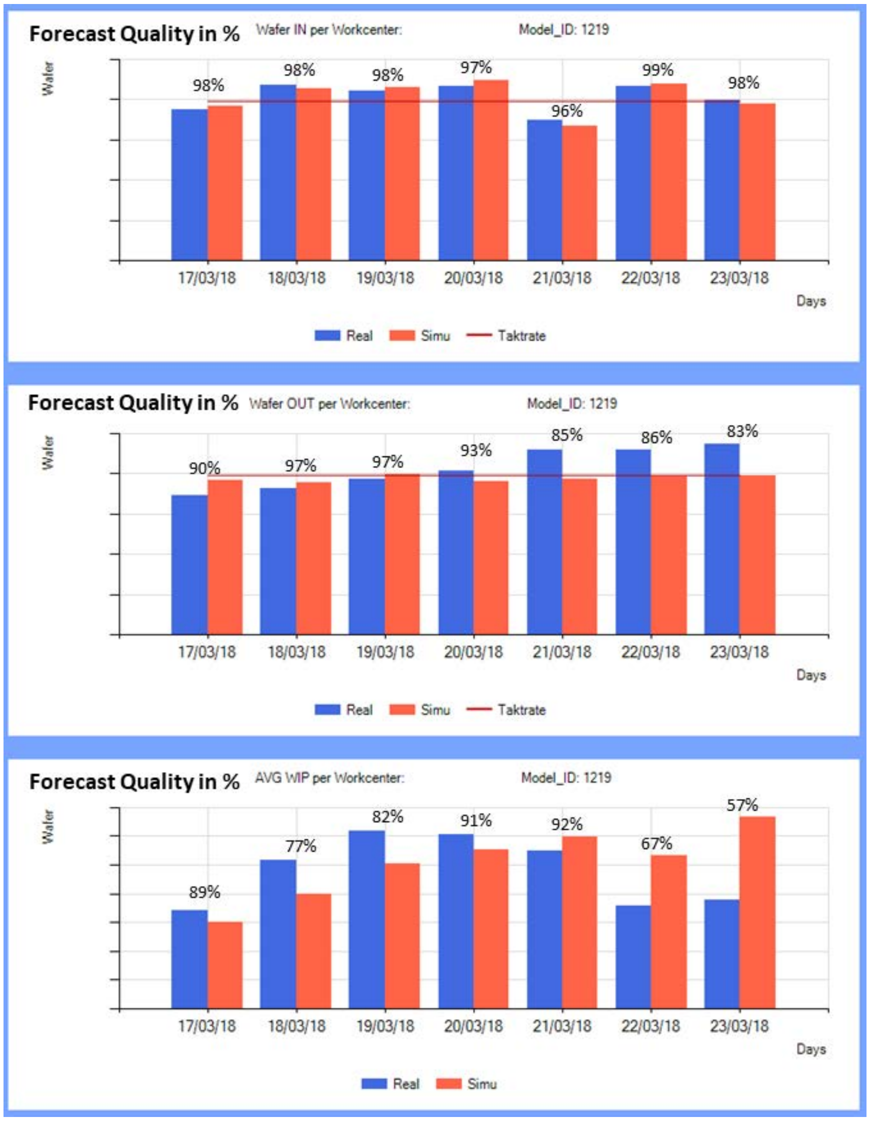

Figure 4: Model with static KANBAN as constant capacity limiter.

Our first approach to solve that issue used a constant throughput limit based on the work areas typical capacity shown in Figure 4. Moves were reduced by setting a static KANBAN value acting as a limit for excessive moves on the first days of simulation. As a result simulation WIP increased in the same manner as in reality. However during the second part of the forecast period simulation moves were now considerably less than in reality, as our model was not able to allocate additional resources after day 4. Consequently the lower simulation moves led to a higher WIP.

Figure 5 shows the forecast with the dynamic KANBAN approach, which improved the modelling of such a dynamic situation. Initial days of the forecast period are comparable to the previous approach, resulting in lower simulation moves and increasing WIP. By day 4, the upper WIP threshold was reached or exceeded and additional KANBAN was provided for the work area. As a result simulation moves increased and subsequently WIP dropped to a lower level.

Figure 5: Model with dynamic KANBAN.

The dynamic KANBAN approach has significantly improved the forecast quality for both moves and WIP, in comparison to the original base model shown in Figure 6. The method how we measure forecast quality can be found in Seidel et. al. 2017.

5 CONCLUSIONS

As seen in the earlier discussion, the KANBAN approach has helped us to improve the forecast quality significantly, and enables us to use the WIP forecast for production management and control. We are extending this modelling approach to all work areas where we observe oscillating WIP, and where the capacity and uptime are not the limiting factors. Some might argue that it is probably more straightforward to incorporate operator modelling, using actual operator schedules and allocations as part of the simulation model. But the reality is, the operator assignments do not always follow a plan, and the plan is always adapted based on production line situation.

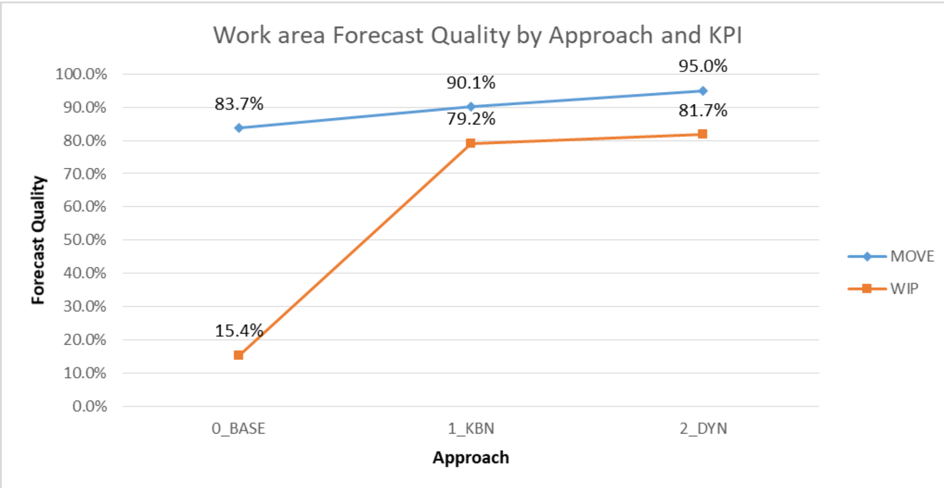

Figure 6: Move and WIP forecast quality for base model, static KANBAN and dynamic KANBAN.

The enhanced forecast quality with dynamic KANBAN modelling enables us to use the simulation model to provide the next 7 days forecast to production line, to facilitate a better planning of resource allocation, preventive maintenance planning, dedication releases, and priority lot journey forecast. This helps to move the production control away from a fire-fighting approach to a pro-active operation management approach.

The purpose of simulation can be to predict WIP and cycle times, but also to study the impact of certain influencing factors. Looking at the results, it becomes obvious that the operator allocation has a significant impact on both, WIP and cycle times. This allows production management to optimize operator allocation and the establishment of business procedures to reduce WIP and cycle times of work centers.

REFERENCES

Scholl, W., D. Noack, O. Rose, B. P. Gan, P. Lendermann, 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. Jan et al., 1932-1943. Piscataway, New Jersey: IEEE.

Seidel, G., B. P. Gan, C. W. Chan, C. F. Lee, A. M. Kam, A. Naumann, and P. Preuss. 2017. "Harmonizing Operations Management of Key Stakeholders in Wafer Fab using Discrete Event Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edit by W. K. V. Chan et al.. Piscataway, New Jersey: IEEE.

Ohno, Taiichi. 1988. "Toyota Production System - beyond large-scale production". Productivity Press. pp. 25–28.

D-SIMLAB Technologies. 2018. Forecaster and Scenario Manager. http://www.d-simlab.com/category/d-simcon/products-d-simcon/forecaster-and-scenario-manager, accessed August 24, 2018.

AUTHOR BIOGRAPHIES

GEORG SEIDEL is Senior Staff Engineer of Infineon Technologies Austria AG (Villach, Austria). He has been involved in simulation, WIP flow management and Industrial Engineering topics since 2000. He was responsible for WIP flow management, especially for Lot dispatching at Infineon's site in Kulim (Malaysia) from 2012 until 2015. He is now responsible to rollout Fab Simulation in Kulim and Villach. He holds a Master degree of Technical Mathematics. His email address is georg.seidel@infineon.com.

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

SOO LEEN LOW is a Senior Software Engineer at D-SIMLAB Technologies (Singapore). She is responsible for simulation modelling and analysis of Wafer Fabrication plants. She earned a Bachelor of Engineering in Computer Engineering from National University of Singapore (NUS) in 2014. Her email address is soo.leen@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 holds a Master of Applied Science degree, specializing in Computer Engineering. His email address is boonping@d-simlab.com.

CHEW WYE CHAN is a Senior Software Engineer of D-SIMLAB Technologies (Singapore). He has been working in the development of simulation-based applications for logistics and semiconductor industry with focus on simulation-based optimization since 2006. Chew Wye holds a Master of Computing degree from National University of Singapore. His email address is chew.wye@d-simlab.com.