Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

SCHEDULING AND SIMULATION IN WAFER FABS: COMPETITORS, INDEPENDENT PLAYERS OR AMPLIFIERS?

Peter Lendermann

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

Stéphane Dauzère-Pérès

Mines Saint-Etienne, Univ Clermont Auvergne
CNRS, UMR 6158 LIMOS, CMP
Dept. of Manufacturing Sciences and Logistics
F-13541 Gardanne, FRANCE

Leon McGinnis

School of Industrial Systems and Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
USA

Lars Mönch

Department of Mathematics and Computer Science
University of Hagen
Universitätsstraße 1
58097 Hagen, GERMANY

Georg Seidel

Infineon Technologies Austria AG
Siemensstraße 2
9500 Villach
AUSTRIA

ABSTRACT

This panel will discuss the inherent conflict between the application of (Discrete-Event) Simulation and Scheduling techniques to manage and optimise capacity and material flow in Semiconductor Frontend Manufacturing (wafer fabrication). Representatives from both industry and academia will describe advantages and shortcomings of the respective techniques, with a specific focus on challenges arising from the recent and anticipated future evolution of the nature of such manufacturing environments, and suggest solution approaches as well as research issues that need to be addressed.

1 INTRODUCTION

Discrete Event Simulation has been an important enabler for production planning and control in Semiconductor Manufacturing, in particular to study the dynamic behavior of wafer fabrication facilities (wafer fabs) which are subject to many random effects such as unscheduled equipment downs, or measurement outcomes, and where the interdependency between capacity and cycle time is an important consideration. At the same time – because of the pressure to utilize capital-intensive capacity as effectively as possible – the underlying paradigm for operating wafer fabs has typically been to load tools with lots from the queue in front of an equipment group as soon as possible when a tool becomes available according to “real-time dispatch rules”.

Over the past couple of decades Semiconductor Manufacturing has become more and more complex, especially in foundries, with more products, smaller volumes per product, faster-changing product mixes to be handled, and novel constraints on the product operations to be considered. In order to be able to manage the material flow of such complex, non-steady-state operations, dispatch rules would typically also have been adapted and become more elaborate over time.

At the same time, interdependencies in between different equipment groups such as queue time constraints or the potential availability of lots for batching have become more critical. However, if factors from immediate upstream or downstream equipment groups need to be taken into account, time horizons for decision-making beyond real time to at least a few hours need to be considered. As a result, the use of scheduling systems to run and optimise operations in wafer fabs has become widespread. Such systems have helped increase the utilization of critical capacity and meet due dates although any "optimal schedule" determined would still constitute a local optimum around the respective equipment group. Increasing automation efforts also have been an important driver and enabler of scheduling solutions for wafer fabs.

A finite scheduling horizon means that factors other than dispatch sequences also need to be considered for decision-making. Such factors comprise, for example, timing for Preventive Maintenance (PM) activities, setup changes, batch formation decisions, or qualification decisions for equipment. Since these factors could have equal or even more effect on KPIs such as capacity utilization, cycle time, or on-time delivery performance compared to dispatch sequences alone, they actually need to be considered as decision variables as well which in turn gives rise to the need to further increase the time horizon to be considered to up to several days.

In Semiconductor Manufacturing though, increasing the time horizon of a scheduling procedure is limited by frequent random events (such as the earlier mentioned) happening within the scheduling horizon that gradually invalidate the schedule. As such, a scheduling horizon of several days simply does not make sense. On the other hand, decisions such as PM timings also cannot be handled using conventional long-term simulation models because equipment downs are typically portrayed as random events in such a simulation model.

Where simulation is concerned, people have also been struggling because of the difficulty to validate a simulation model (a "Digital Twin") of a full wafer fab which is also partly due to the difficulty to portray human behavior and decision-making in such a simulation model. Moreover, scheduling and control systems are very difficult to portray in a simulation model with sufficient fidelity unless the simulation is slowed down to near real-time, which in turn limits the applicability of simulation-based decision making dramatically.

In the light of the challenges described, the question how the backbone of operations management in Semiconductor Manufacturing may look like in the future naturally arises:

- Either to schedule and reschedule operations at critical equipment groups on a regular basis with the objective to create local optima to squeeze more moves out of the available capacity around a particular equipment group. This would, however, make it very difficult if not impossible to run simulations with sufficient fidelity to make high-quality decisions, for example about PM timing, with potentially higher impact and to effectively manage the global performance of fab operations.
- Or to run fab operations with relatively simple dispatch and/or scheduling rules which would allow exploitation of more optimization potential from adjustment of decision variables such as PM timing through regular simulation-based optimization. This would, however, make it very challenging to make non-intuitive scheduling decisions against implemented dispatch rules at a particular equipment group even though it would increase the performance at another equipment group (and the global fab performance) significantly.

In the setting of these challenges, this panel will discuss how a compromise between the needs to operate a powerful scheduling system on the one hand side and "High Fidelity Digital Twins" on the other hand side could look like in a wafer fab, the solution approaches this would comprise, and research issues that would have to be addressed.

2 STÉPHANE DAUZÈRE-PÉRES, ECOLE DES MINES DE SAINT-ETIENNE

2.1 Limits of Optimized Scheduling and Simulation

Optimized scheduling requires advanced methods to be applied in complex equipment groups and to deal with the very large instances that can be found in semiconductor manufacturing with hundreds of machines and lots. Various types of hard and soft constraints must be considered, and multiple objectives must usually be optimized simultaneously, with various degrees of importance. As shown for instance in Knopp et al. (2017) and Klemmt et al. (2017), metaheuristics and constraint programming approaches have proved to be successful on instances of industrial sizes or in practical settings. However, even though the detailed schedule of lots can be optimized at equipment group level, it still remains unrealistic to determine a detailed schedule at fab level. Some approaches, such as the distributed one introduced in Mönch and Drießel (2005) for complex job-shop scheduling problems, have been proposed to schedule an entire wafer manufacturing facility. However, they usually consider a single objective function, whereas criteria can differ between equipment groups. Hence, the relevance of the proposed detailed global factory schedules still needs to be evaluated in industrial facilities.

Both simple dispatch rules or scheduling procedures embedded in simulation models allow the behavior of an entire semiconductor manufacturing facility to be evaluated. However, if advanced scheduling methods are actually used in the facility, at least in some equipment groups, then the evaluation might actually be skewed. As phrased in Dauzère-Pérès and Lasserre (2002), “The scheduling module should represent as closely as possible the actual capacity of the equipment group, and the scheduling techniques are part of that capacity. The better the jobs can be sequenced on the shop floor, the higher is the capacity.” Using industrial instances of the photolithography equipment group of a high-mix facility, Bitar et al. (2014) analyze the gain on the mean product cycle times of using an efficient optimization algorithm compared to a simple approach based on dispatch rules. The gain depends on the complexity and how heavily the equipment group is loaded. When the workload and the complexity are high, the gap on the mean product cycle times is larger than 20%, and is still larger than 5% if some constraints are removed.

2.2 Different Functions for Different Uses

With the availability of data and the increase of automated decision making in factories, optimized scheduling is becoming mandatory and critical to gain capacity, in particular in high-mix manufacturing settings, where simple dispatch rules may lead to significant capacity losses. Complex constraints are very difficult, or even impossible, to handle with dispatch rules, leading to unrealistic dispatch decisions. Moreover, short-term schedules determined in each equipment group on a planning horizon of several hours help to plan other activities, such as the transportation and storage of lots or of auxiliary resources. However, global fab scheduling approaches are required to control the production flows between equipment groups, such as the one proposed in Bureau et al. (2007), to avoid the WIP to be imbalanced.

Simulation is important to analyze the impact of critical decisions through what-if analyses and to validate the relevance of novel approaches at factory level. The trends provided by simulation can be used for various strategic and tactical decisions, but might not be relevant to support operational decisions. Vehicle dispatching policies in an Automated Material Handling System (AMHS) are studied using a simulation model in Ben Chaabane et al. (2013) and Schmaler et al. (2017). The design of an Automated Guided Vehicle (AGV) transport system for the main auxiliary resources in the photolithography equipment group is analyzed using a simulation model in Ndiaye et al. (2016). Another example can be found in Barhebwa-Mushamuka et al. (2019), where a simulation model of wafer manufacturing facilities is used to show the efficiency of a global scheduling optimization model to control the Work-In-Process

(WIP). Lima et al. (2019) are using a sampling-based approach, relying on a fast simulation of dispatch rules, to estimate the probability that a lot satisfies given time constraints before it is released.

2.3 Research Perspectives

I believe that a first challenge is to design and develop "Digital Twins" that would benefit from the advances of Discrete Event Simulation and/or agent-based simulation but also embed advanced scheduling algorithms. Because these algorithms are now available and are or will be used in factories, the Digital Twin will then represent the reality, in particular how lots are being processed on tools, and support various types of short-term decisions impacting the whole factory. Controlling the runtimes of the Digital Twin, which could be prohibitive, will be necessary.

However, I think that simulation using dispatch rules is sufficient for many types of decisions, in the short and medium term. Hence, a second idea is to study in which contexts and for which problems considering optimized scheduling in simulation models is relevant.

Generally speaking, the coupling of simulation and optimization is a very active field of research, with a specific track in the Winter Simulation Conference. Still, much remains to be done to combine simulation and optimized scheduling in the context of semiconductor manufacturing. For instance, as discussed for instance in Tamssaouet et al. (2018), a simulation model could be used to model the detailed behavior of a complex machine, and the resulting Digital Twin of the machine could be embedded in algorithms that schedule all the lots on all the tools in the equipment group.

3 LEON MCGINNIS, GEORGIA INSTITUTE OF TECHNOLOGY

Stuart Kauffman's book on complexity and chaos, "At Home in the Universe" (Kauffman 1995) makes a distinction between predicting and explaining, a distinction that is critical for this discussion. Because there are many sources of apparent randomness in a wafer fab, the terrible truth is that we will *never* be able to predict its future state with precision beyond a few minutes unless we can eliminate unpredictable or poorly predictable events. With that realization, what are the implications for how we should approach the problem of planning and control of wafer fabs? Having said that, some of the ideas presented in this section could be perceived as outrageous indeed!

Let us suppose for a moment that we could create an effective though deterministic Digital Twin of the wafer fab and thus could predict an *ideal* future, i.e., one where no randomness or unforeseen event upsets our plans and schedules. The value of this prediction is that it lays down a marker for where we want to be *and how we intend to get there*. Beyond that, it could allow us to analyze the state trajectory of the fab toward that future state, and in particular, to identify those potential contingencies that would most upset our plans, which could include variability in transportation, storage/retrieval or lot handoff operations, queueing delays at inspection, unanticipated machine downtime, etc. If we could correctly identify and characterize these contingencies (perhaps using our vast prior experience) we might be able to "war game" them, making (standardized) contingency plans to respond to events that "might" occur in the future. This approach is possible *if* we have the necessary knowledge about contingencies, *and* the technical and computational tools to support this "war gaming" approach – which by the way, could be done largely off-line. If not, there is a clear opportunity for useful research and development. Note that the technical and computational tools required would include at least a deterministic Digital Twin of the wafer fab incorporating *all* elements of the fab that influence state trajectory.

Simulation methodologists will no doubt be horrified at the notion of using a deterministic simulation as the baseline for war gaming possible contingencies, and insist that we must start from statistically valid predictions. But such outrage ignores the indisputable fact that no matter how many samples (simulation runs) we look at, we still do not know *exactly* which future will obtain. And the point of war gaming the contingencies is not to be prepared for *one* future, it is to be prepared for *any* future. Note also that every previously unencountered contingency becomes fodder for war gaming.

If we cannot predict the future state with precision, can we at least *explain* what we observe? For example, when a WIP bubble is observed, can we explain what caused it? When a non-bottleneck resource becomes a bottleneck because it has been idle for so long, can we explain why? Note that explaining such observations is much more than simply listing all the events that might have contributed. If one of those events does not *always* lead to the observed phenomenon, then we need to be able to explain that as well. Here is the key observation—the development of rules for "optimizing" scheduling in a process group or across a fab is founded on the intuition that certain observed behaviors can be explained, and those explanations are the basis for "rules" guiding decision-making.

Explaining observed phenomena requires a theory about the system that presents the phenomena, and a theory requires a model of the system in question. Newton's theory of universal gravitation was based on a model of two objects and posited a force of attraction between them; it was useful for explaining why apples fall toward the ground. What is the model of the wafer fab that we might use as the basis for positing theories that would help to explain the phenomena presented by the wafer fab?

So, whether we want to predict or to explain, we need a useful system model as a starting point. If we can neither predict nor explain, our efforts to plan and schedule will be forever frustrated.

3.1 On Digital Twins

The idea of the wafer fab Digital Twin is like the siren's call. "If GE can do it for gas turbines as in General Electric (2020), Lockheed can do it for an F-35 as in Mail Online (2020) and it can be done for integrated circuits as in Semiconductor Engineering (2020), then surely we can do it for wafer fabs." We should be asking ourselves why we have not been able to, so far (see the fascinating panel discussion from WSC 2019 as in Shao et al. (2019). To approach an answer, we will need to understand why and how Digital Twins in the produced system space have been possible.

Digital Twins for produced systems are the result of some very specific technological developments:

- Standards for specifying both components and systems: When the United States Department of Defense commissioned the development of the Hardware Description Language (HDL), it was to enable one vendor's product to be transitioned to another vendor for manufacturing in case the original supplier failed. But what HD, and later VHDL (Very high speed integrated circuit Hardware Description Language), created was the foundation for design tools – without a standard way of specifying the produced system, generic "editors" are not possible.
- Interfaces to standard analysis methods: Once a standard specification of a produced system is available, a specific system specification can be translated into the required input for canonical analyses, such as circuit analysis, and thus circuit simulation. As a result, produced system designers have very inexpensive access to analyses that support design decision making, and developers have incentives to create new, better, more comprehensive analysis tools.

Representation standards are the cornerstone for all successful Digital Twin applications. Imagine what would be the state of integrated circuit design without VHDL? If every design house had to create their own (*ad hoc*!) tools to support design? Do you think we would have integrated circuits with billions of transistors?

3.2 The Challenge of Representation

What all successful Digital Twin applications have in common and what modern engineering design automation tools have in common is dependence upon an underlying Analysis Agnostic System Model (AASM). Whether designing integrated circuits, automobiles or airplanes, the designers work directly with such an AASM. If we aspire to be able to routinely create Digital Twins of Semiconductor Production Systems (SPS), then we need an AASM for them.

An AASM of the physical aspects of the SPS is intricate, but relatively straightforward, and the Discrete Event Logistics (DELS) framework provides an excellent starting point. What is much more challenging is modeling the control system, because it is not directly observable and thus models of it are virtually impossible to validate. In concept, the only way to realize a validated Digital Twin would be to integrate a computational model of the physical elements of the SPS with the actual control system. This is not a practical approach when people are making control decisions.

Again, the fundamental challenge is the absence of a suitable generic framework for conceptualizing, modeling, developing and implementing control systems for SPS.

3.3 Scheduling AND Simulation

For all the reasons Peter Lendermann has mentioned, and more, there always will be limits to the predictive power of a wafer fab Digital Twin. Planning and scheduling always will be forced to manage contingencies. The fundamental question we should be addressing is “How do we get much smarter about and more capable in managing contingencies?” The answer must be built around an AASM of the wafer fab, and its use in the collection and post-mortem analysis of contingencies (using a simulation model!) to develop standard strategies for each class of contingency. If we can do that, then integrating factors like PM scheduling is at least conceivable. Otherwise, we will continue to see new and better *ad hoc* approaches, and discover their limitations.

4 LARS MÖNCH, UNIVERSITY OF HAGEN

4.1 Scheduling and Simulation as Amplifiers

The assignment of scarce resources to activities over time is called scheduling. Scheduling is a decision-making process that aims to optimize at least one performance measure while taking into account all relevant constraints (Pinedo 2016). On the one hand, scheduling is explicitly or implicitly associated with a finite scheduling horizon of positive length. Hence, a scheduling approach for a semiconductor manufacturing (wafer fabrication) facility typically considers several lots and tools at the same time. A scheduling technique is able to adapt to different system conditions. Most of the scheduling approaches are based on deterministic data. Therefore, it may be challenging to execute schedules in a wafer fab, a manufacturing system that is highly dynamic and stochastic.

Dispatching, on the other hand, selects a job that has to be processed next for a given available tool among the jobs waiting in front of this tool (or tool group) for processing. Dispatching can be considered as the limiting case of scheduling with a fixed resource assignment and a scheduling horizon which tends to zero. It becomes clear that dispatching is myopic by nature. The ability of dispatch rules to consider the situation on upstream and downstream tool groups is limited. Dispatching has a long tradition in semiconductor manufacturing (cf. Uzsoy et al. 1993; Mönch et al. 2013). It is still the tool of choice in many wafer fabs (Pfund et al. 2006). While at the beginning fairly simple dispatch rules are applied (Wein 1988), sophisticated rule-based dispatching systems are in place in wafer fabs over the years. Dispatching approaches are conceptually easy to understand, and they are robust to cope with changes in WIP, static priorities of the lots, and the machine environment of a wafer fab.

Until a decade ago, deterministic scheduling techniques seemed to be too computationally costly as compared to dispatching. However, with the recent dramatic increase in computer efficiency, scheduling methods for single tool groups or even certain work areas in a wafer fab have become more competitive (Mönch et al. 2011). It seems that especially time-limited mixed integer linear programming (MILP)-based approaches with a short scheduling horizon and frequent rescheduling are quite popular in wafer fabs (Ham and Cho 2014; Jung et al. 2015). Scheduling techniques are important to deal with highly-constrained problems, for instance, time constraints between the start of consecutive processing steps in semiconductor manufacturing (Klemmt and Mönch 2012) or for making batching- or setup-related decisions (Jung et al. 2014).

One open question is whether fab-wide scheduling approaches are practicable for a real-world wafer fab or not. Such a global scheduling approach has the advantage that factory-wide performance measures such as total weighted tardiness or cycle time can be optimized. While some of the fab-wide scheduling approaches, mainly based on disjunctive graphs and the shifting bottleneck heuristic are quite promising (Ovacik and Uzsoy 1998; Sourirajan and Uzsoy 2008), for instance, the one designed in the Factory Operations Research Center (FORCe) project (Pfund et al. 2008; Mönch and Zimmermann 2011), it seems that no full real-world implementation of such a system exists nowadays. In wafer fabs, it is often required to make scheduling decisions for processing of lots on tools in an integrated manner with other decisions, for instance for the automated material handling system (AMHS) or advanced process control (APC). While fully integrated scheduling approaches are designed for such problems in academic research labs (Driessel and Mönch 2012; Yugma et al. 2015), they often lead to high modeling and computational burden. In such situations, local scheduler for bottleneck equipment groups or work areas that are coordinated through a higher level decision support layer are more promising.

Coming back to the original question of this panel, in my opinion Discrete Event Simulation is an amplifier that might improve scheduling (and also dispatching) decisions and vice versa on the following levels (Fowler et al. 2006):

1. Simulation-based schedule generation, refinement, and optimization
2. Simulation used for parameter setting and problem instance generation for scheduling approaches
3. Simulation for emulation and evaluation of scheduling approaches.

On all levels, simulation is used to anticipate the behavior of the shop floor when making scheduling decisions. We will discuss the three levels in a point-by-point manner.

Simulation-based schedule generation, also known as simulation-based scheduling, refers to the situation where a simulation tool together with an up-to-date simulation model is applied to make scheduling decisions. The built-in dispatch rules of the simulator are used to compute a schedule with a horizon of several hours. Stochastic effects such as tool breakdowns are turned off due to the short horizon. An appropriate initialization of the simulation model is a non-trivial task. Once a schedule has been computed, it can be refined by taking into account additional resources, for instance secondary resources such as reticles. Their consideration requires some deterministic forward simulation. Simulation-based schedule optimization is based on the idea that the objective function of a scheduling approach is computed based on simulation by executing the schedule in a simulation model. The schedule is changed and the assessment via simulation is repeated. Note that this approach can also be used to discover dispatch rules via genetic programming (Hildebrandt et al. 2014).

Simulation can be used to set parameters in scheduling heuristics in a situation-dependent manner. In order to evaluate the impact of a concrete parameter setting, the schedules have to be executed in a simulation environment. This can be part of the training phase of Machine Learning approaches (Mönch et al. 2006). It is also possible to use simulation to generate problem instances for scheduling approaches that reflect the current load and due date setting in a wafer fab (Sourirajan and Uzsoy 2007).

Scheduling approaches for wafer fabs are mainly based on deterministic data. Simulation is used to allow for rolling horizon settings, i.e. instead of a single problem instance a series of interrelated instances is solved where the instances come from executing the schedule in a single scheduling epoch (Mönch and Zimmermann 2011). This allows assessment of the performance of a scheduling approach in a risk-free simulation environment before its deployment at the shop floor. Overall, we observe that scheduling and simulation can be seen as two sides of a coin.

4.2 Future Research Directions

A first important research direction is given by designing and assessing hierarchical approaches where the higher level is able to coordinate the lower level, i.e. the scheduling or dispatching level by appropriate instructions. Such an approach would allow the design of local schedulers. For instance, it would be interesting to explore the interface between production planning and shop-floor scheduling which is a significantly under-researched area. I also see the need to develop a better understanding of the feasibility of fab-wide, i.e. global scheduling approaches for wafer fabs. This includes a more in-depth investigation of integrated approaches with for instance, PM, AMHS, or APC. Another research avenue is given by designing and assessing scheduling approaches for local or global schedulers in wafer fabs where the considered machinery includes cluster tools. Cluster tools are special integrated machines for wafer processing in wafer fabs. Since wafers with different types of process steps can circulate in a cluster tool simultaneously, it can be regarded as a fully automated machine environment (Mönch et al. 2013). Cluster tools are used to maximize quality performance at the cost of very complex behavior. While there are many scheduling techniques known for scheduling of wafer movements inside a cluster tool, external cluster tool scheduling which aims for scheduling jobs waiting to be processed in front of a cluster tool is an under-researched area. This is a non-trivial task since several wafers are simultaneously processed in a cluster tool and the internal behavior of the cluster tool determines the completion time of these wafers. As a last direction, it seems desirable to better incorporate the stochastic behavior of lot arrivals into scheduling approaches for batch processing tools.

5 GEORG SEIDEL, INFINEON TECHNOLOGIES

5.1 Dispatching and Scheduling in Simulation and Reality

At Infineon front end wafer fab simulation models are used to support operations and planning (Seidel et. al. 2017) and to carry out scenario runs to address strategic capacity planning and material flow related questions. There is always the challenge to provide a simulation model as close as possible to reality but with reasonable effort to maintain and to validate it.

One crucial simulation modelling aspect are the dispatch rules that are used within the simulation model to determine which of the lots, waiting in a queue in front of equipment, will be processed next. Semiconductor fabs use typically many different dispatch rules, local and global ones, to optimize the fab performance. In the past years the importance of scheduling has been growing as well. More and more equipment or equipment groups are controlled through scheduling rather than dispatching. Mixed integer linear programming (MILP) and constraint programming (CP) based scheduling can be used to improve fab performance, at least locally (Klemmt et. al. 2017).

Infineon simulation models contain global dispatch rules as well as local ones. Not all local dispatch rules are used in the models for reasons of simplicity and because the impact on simulation results are considered to be low. Reasonable accuracy levels for simulation results have been achieved without incorporating all dispatch rules. The same holds true for equipment groups where scheduling solutions are used in reality. Instead of representing these scheduling solutions in the simulation, global dispatch rules are used instead.

5.2 Future Research Directions

Incorporating scheduling solutions into a simulation model can answer questions such as: What is the impact of local optimization of equipment or work center scheduling on the global fab performance? Can the fab improve further by using a global scheduling algorithm? How is the impact on simulation accuracy if global dispatch rules are used in the simulation instead of the scheduling solutions used in reality? It is desirable indeed to get answers to these very interesting and relevant questions.

5.3 Practical Limitations

Calculation of an executable and optimized schedule in a real fab is challenging. Data accuracy and availability must be extremely high. Optimized scheduling provides much better results compared to dispatching only at equipment groups where complexity is high. For example, equipment groups with high dedication level, comprising batch tools and/or cluster tools and equipment with different processing speeds are worthy targets for improvement by scheduling. Even with the increased computational power in the last years it takes some time to calculate an optimal schedule for a complex equipment group.

Generation of a simulation model for a whole frontend fab requires some sort of abstraction. I do not see any practical way to incorporate complex scheduling algorithms with the needed high data granularity (e.g. modelling of detailed cluster tool behavior, assumption of future lot arrivals to improve batching efficiency, etc.) in a fab model, and at the same time keeping the simulation runtime low and the model reasonably well maintainable. Reasonable simulation accuracy levels have been achieved on the fab level by using global dispatch rules in simulation instead of the real scheduling solutions. This indicates that the impact of existing scheduling algorithms is small, at least on the fab level. This may change in the future if more sophisticated scheduling solutions for combinations of several equipment groups or even the whole fab will be available.

REFERENCES

Barhebwa-Mushamuka, F., Dauzère-Pèrès, S., and Yugma, C. 2019. “Work-in-Process Balancing Control in Global Fab Scheduling for Semiconductor Manufacturing”. In *Proceedings of the 2019 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 2257–2268.

Ben Chaabane, A., Dauzère-Pérès, S., Yugma, C., Rullière, L., and Lamiable, G. 2013. “Analyzing the Impact of Key Parameters of Vehicle Management Policies in A Unified AMHS”. In *Proceedings of the 2013 Winter Simulation Conference*, December 08th-11th, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 3818–3828.

Bitar, A., Dauzère-Pérès, S., and Yugma, C. 2014. “On the Importance of Optimizing in Scheduling: the Photolithography Workstation”. In *Proceedings of the 2014 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 2561–2570.

Bureau, M., Dauzère-Pérès, S., Yugma, C., Vermariën, L., and Maria, J. B. 2007. “Simulation Results and Formalism for Global-local Scheduling in Semiconductor Manufacturing Facilities”. In *Proceedings of the 2007 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 1768–1773.

Dauzère–Pérès, S., and Lasserre, J. B. 2002. “On the Importance of Sequencing Decisions in Production Planning and Scheduling”. *International Transactions in Operational Research* 9(6): 779–793.

Driessel, R., and L. Mönch. 2012. “An Integrated Scheduling and Material Handling Approach for Complex Job Shops: A Computational Study”. *International Journal of Production Research* 50(20): 5966-5985.

Fowler, J. W., L. Mönch, and O. Rose. 2006. “Simulation and Scheduling”. In *Handbook of Production Scheduling*, edited by J. Hermann, 109-133, New York: Springer.

General Electric. 2020. https://tinyurl.com/wlyjhlf, accessed 7th May 2020.

Ham, A., and M. Cho. 2015. “A Practical Two-Phase Approach to Scheduling of Photolithography Production”. *IEEE Transactions on Semiconductor Manufacturing* 28(3): 367-373.

Hildebrandt, T., D. Goswami, and M. Freitag. 2014. “Large-scale Simulation-based Optimization of Semiconductor Dispatching Rules”. In *Proceedings of the 2014 Winter Simulation Conference*, 2580-2590.

Jung, C., D. Pabst, M. Ham, M. Stehli, and M. Rothe. 2014. “An Effective Problem Decomposition Method for Scheduling of Diffusion Processes Based on Mixed Integer Linear Programming.” *IEEE Transactions on Semiconductor Manufacturing* 27(3): 357-363.

Kauffman, S.A. 1995 *At Home in the Universe: The Search for Laws of Self-organization and Complexity*. Oxford University Press.

Klemmt, A., and L. Mönch. 2012. “Scheduling Jobs with Time Constraints between Consecutive Process Steps in Semiconductor Manufacturing”.In *Proceedings of the 2012 Winter Simulation Conference*, 2173-2182.

Klemmt, A., Kutschke J., and Schubert C. 2017. “From Dispatching to Scheduling: Challenges in Integrating A Generic Optimization Platform into Semiconductor Shop Floor Execution”. In *Proceedings of the 2017 Winter Simulation Conference*, edited by W.K.V. Chan, A. D’Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 3691-3702.

Knopp, S., Dauzère-Pérès, S., and Yugma, C. 2017. “A Batch-oblivious Approach for Complex Job-Shop Scheduling Problems”. *European Journal of Operational Research* 263(1), 50–61.

Lima, A., Borodin, V., Dauzere-Pérès, S., and Vialletetelle, P. 2019. “Sampling-based Release Control of Multiple Lots in Time Constraint Tunnels”. *Computers in Industry* 110, 3–11.

Mail Online. 2020. https://tinyurl.com/s2tnn4h, accessed 7th May 2020.

Mönch, L., and Drießel, R. 2005. “A Distributed Shifting Bottleneck Heuristic for Complex Job Shops”. *Computers & Industrial Engineering* 49(3), 363–380.

Mönch, L., J. W. Fowler, S. Dauzère-Pérès, S. J. Mason, and O. Rose. 2011. “A Survey of Problems, Solution Techniques, and Future Challenges in Scheduling Semiconductor Manufacturing Operations”. *Journal of Scheduling* 14: 583-599.

Mönch, L., J. W. Fowler, and S. J. Mason. 2013. *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and System*. New York: Springer.

Mönch, L. and J. Zimmermann. 2011. “A Computational Study of a Shifting Bottleneck Heuristic for Multi-Product Complex Job Shops”. *Production Planning & Control* 22(1): 25-40.

Mönch, L., J. Zimmermann, and P. Otto. 2006. “Machine Learning Techniques for Scheduling Jobs with Incompatible Families and Unequal Ready Times on Parallel Batch Machines”. *Engineering Applications of Artificial Intelligence* 19(3): 235-245.

Ndiaye, M. A., Dauzère-Pérès, S., Yugma, C., Rullière, L., and Lamiable, G. 2016. “Automated Transportation of Auxiliary Resources in A Semiconductor Manufacturing Facility”. In *Proceedings of the 2016 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 2587–2597.

Ovacik, I. M., and R. Uzsoy. 1997. *Decomposition Methods for Complex Factory Scheduling Problems*. Boston, MA: Kluwer Academic Publishers.

Pinedo, M. 2016. Scheduling: Theory, Algorithms, and Systems. 5th ed. Cham: Springer.

Pfund, M., S. J. Mason, and J. W. Fowler. 2006. “Semiconductor Manufacturing Scheduling and Dispatching”. In *Handbook of Production Scheduling*, edited by J. Hermann, 109-133, New York: Springer.

Pfund, M. E., H. Balasubramanian, J. W. Fowler, S. J. Mason, and O. Rose. 2008. “A Multi-criteria Approach for Scheduling Semiconductor Wafer Fabrication Facilities.” *Journal of Scheduling* 11:29-47.

Schmaler, R., Schmidt, T., Schoeps, M., Luebke, J., Hupfer, R., and Schlaus, N. 2017. “Simulation Based Evaluation of Different Empty Vehicle Management Strategies with Considering Future Transport Jobs”. In *Proceedings of the 2017 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 3576–3587.

Seidel, G., C.F. Lee, A.M. Kam, B.P. Gan, C.W. Chan, 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*, edited by W.K.V. Chan, A. D’Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 3670-3678.

Semiconductor Engineering, 2020. https://tinyurl.com/w3wu92z , accessed 7th May 2020.

Shao, G., Jain, S., Laroque, C., Lee, L. H., Lendermann, P., and O. Rose. 2019. “ Digital Twin for Smart Manufacturing: The Simulation Aspect”. In *Proceedings of the 2019 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 2257–2268.

Sourirajan, K., and R. Uzsoy. 2007. “Hybrid Decomposition Heuristics for Solving Large-Scale Scheduling Problems in Semiconductor Wafer Fabrication.” *Journal of Scheduling* 10: 41-65.

Tamssaouet, K., Dauzère-Pérès, S., Yugma, C., Knopp, S., and Pinaton, J. 2018. “A Study on the Integration of Complex Machines in Complex Job Shop Scheduling”. In *Proceedings of the 2018 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 3561–3572.

Uzsoy, R., C. Y. Lee, and L. A. Martin-Vega 1994. “A Review of Production Planning and Scheduling Models in the Semiconductor Industry Part II: Shop-Floor Control.” *IIE Transactions on Scheduling and Logistics* 26: 44-55.

Wein, L. M. 1988. “Scheduling Semiconductor Wafer Fabrication.” *IEEE Transactions on Semiconductor Manufacturing* 27(3): 357-363.

Yugma, C., J. Blue, and S. Dauzère-Pérès. 2015. “Integration of Scheduling and Advanced Process Control in Semiconductor Manufacturing: Review and Outlook.” *Journal of Scheduling* 18(2):195-205.

AUTHOR BIOGRAPHIES

PETER LENDERMANN is the Co-Founder and Chief Business Development Officer 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.

STEPHANE DAUZERE-PERES is a Professor at the Center of Microelectronics in Provence (CMP) of Mines Saint-Etienne in France and Adjunct Professor at BI Norwegian Business School in Norway. He received the Ph.D. degree from the Paul Sabatier University in Toulouse, France, in 1992; and the H.D.R. from the Pierre and Marie Curie University, Paris, France, in 1998. He was a Postdoctoral Fellow at the Massachusetts Institute of Technology, U.S.A., in 1992 and 1993, and Research Scientist at Erasmus University Rotterdam, The Netherlands, in 1994. He has been Associate Professor and Professor from 1994 to 2004 at the Ecole des Mines de Nantes in France. His research interests broadly include modeling and optimization of operations at various decision levels (from real-time to strategic) in manufacturing and logistics, with a special emphasis on semiconductor manufacturing. He has published 84 papers in international journals. He has coordinated multiple academic and industrial research projects. He was runner-up in 2006 of the Franz Edelman Award Competition, and won the Best Applied Paper of the Winter Simulation Conference in 2013. His email address is dauzere-peres@emse.fr**.**

LARS MÖNCH is a Professor of Computer Science at the Department of Mathematics and Computer Science, University of Hagen where he heads the Chair of Enterprise-wide Software Systems. He holds M.S. and Ph.D. degrees in Mathematics from the University of Göttingen, Germany. After his Ph.D., he obtained a habilitation degree in Information Systems from Technical University of Ilmenau, Germany. His research and teaching interests are in information systems for production and logistics, simulation, scheduling, and production planning. His email address is lars.moench@fernuni-hagen.de.

LEON MCGINNIS is Professor Emeritus in the Stewart School of Industrial and Systems Engineering at The Georgia Institute of Technology, where he has held leadership positions in facutly governance, academic programs and research centers. His research is focused on developing the methods and tools necessary for systems engineering of discrete event logistics systems, including design, planning, management and control issues. His email address is leon.mcginnis@gatech.edu and his work is featured on the website http://factory.isye.gatech.edu.

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.