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

A HIERARCHICAL APPROACH TO QUALIFICATION MANAGEMENT IN WAFER FABS

Denny Kopp
Lars Mönch

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

ABSTRACT

We discuss a qualification management problem arising in wafer fabs. The steppers need to be qualified to process lots of different families. A qualification time window is associated with each stepper and family. The time window can be reinitialized as needed and can be extended by on-time processing of lots from qualified families. In this paper, we propose a hierarchical approach for this problem. The base-level is a dispatching strategy that takes into account qualification decisions. The medial-level consists of a mixed integer linear programming formulation for making qualification decisions, whereas the top-level is a linear programming (LP) formulation that computes target quantities for the steppers for a planning window taking fab-wide objectives into account. We present results of simulation experiments where the hierarchical approach is applied in a rolling horizon manner. The results demonstrate that the LP-based approach outperforms a heuristic to determine target quantities.

1 INTRODUCTION

Semiconductor manufacturing requires one of the most complex manufacturing processes existing today. It deals with producing integrated circuits (ICs) layer by layer on thin silicon discs, so-called wafers. The process complexity is caused by many factors including routes with several hundreds of process steps, a large number of very expensive tools that are often highly unreliable, multiple products with a changing product mix, and reentrant process flows (Chien et al. 2011; Mönch et al. 2013).

Lots are the moving entities in semiconductor wafer fabrication facilities (wafer fabs). A lot consists of up to 50 wafers. Lots are processed on tools that are organized in tool groups. Each tool group contains tools of similar functionality. A mix of different processes can be found in wafer fabs, for instance, we can observe single wafer processes and batch processes where several lots grouped into one batch to be processed at the same time on the same tool. Wafer fabs can be considered as highly reentrant job shops, i.e., a single lot can visit the tools of a certain tool group up to 40 times. Frequent tool failures are also common for wafer fabs.

Steppers in the photolithography work area of wafer fabs transfer the circuit pattern of a certain product layer from a mask onto the surface of a wafer using ultraviolet light exposure. They belong to the most expensive tools in wafer fabs. Therefore, it is very common that the stepper tool group is the planned bottleneck of a wafer fab that deserves special attention.

Yield, the fraction of the raw wafers launched into a wafer fab that completes the manufacturing process as salable devices at their original specification, is an important measure in semiconductor supply chains. In order to increase the yield, quality-driven qualification activities for tools are desirable. In the context of the present paper this means that certain parameters of the steppers are adjusted to obtain high-quality wafers for each mask layer from the steppers. Quality-driven qualifications have to be differentiated from principle qualifications. The latter require that an execution program associated with a

978-1-5386-6572-5/18/$31.00 ©2018 IEEE

process step, a so-called recipe, is performed on a tool to qualify it for the process step. While a tool without a principle qualification cannot be used by the corresponding process step, process steps can be performed on a tool with missing quality-driven qualification, however, this might result in poor quality, i.e., rework is necessary or wafers are scrapped.

In the present paper, we reconsider the quality-driven qualification problem from Kopp et al. (2016). More specifically, we embed the corresponding optimization formulation into a hierarchical three-level setting. The top-level of the hierarchy is responsible for determining demand targets for the qualification management of the stepper tool group while taking fab-wide objectives into account. Using a simulation-based rolling horizon approach, we show that this approach outperforms the approach proposed by Kopp et al. (2018) with respect to cycle time, throughput, tardiness, and qualification management-related performance measures.

The paper is organized as follows. The problem is described and analyzed in the next section. This includes a discussion of related work. The design of the three-level hierarchical approach is discussed in Section 3. This includes a presentation of the proposed LP formulation for the top-level. The applied simulation infrastructure is described in Section 4. The results of simulation experiments are discussed and analyzed in Section 5. Conclusions and future research directions are shown in Section 6.

2 PROBLEM SETTING AND ANALYSIS

2.1 Problem Description

We briefly recall the qualification management problem discussed by Kopp et al. (2016). A finite planning window of length τ divided into discrete periods of length Δ is assumed. Moreover, a new qualification plan is always determined after the first period of the planning window. Lot families are considered. A family is formed by all lots of a product that require the same reticle for processing them on the steppers. i.e., we obtain a family for each product and mask layer. We assume that we know the number of wafers, so-called demand targets, for each family that have to be processed on the steppers within a period of the planning window. We refer to these quantities as target quantities. The target quantities for family f in period t are d_t^f. Stepper dedications occur, i.e., a stepper might be only capable for processing wafers from specific families. The steppers must be qualified for a certain family to process wafers of this family. A qualification time window of length Δ_{fk} is associated with family f and tool k. We assume for the sake of simplicity that Δ_{fk} is an integer multiple of the period length Δ. If no wafers from a certain family are processed on a qualified stepper within the time window, the qualification of this stepper for the family will be expired. The qualification time window for a family on a stepper can be extended by on-time processing of wafers belonging to this family on the stepper, i.e., if at least one wafer of family f is processed on stepper k in period t the stepper will be qualified until the end of period $t + \Delta_{fk}$. Each stepper can be requalified for a family by performing a qualification activity. However, such an activity is expensive and time-consuming since it requires launching a send-ahead wafer (SAW). A SAW is taken from a mother lot, and an exposure step on the stepper and additional development and measurement steps are carried out for this wafer. The stepper is qualified for the family when the measurement step for the SAW is successful. The remaining wafers of the mother lot can be processed under these circumstances (Akçali et al. 2001).

A mixed integer linear programming (MILP) formulation is proposed for this problem by Kopp et al. (2016). The MILP has a cost-based objective function that considers qualification costs, backlog and inventory holding costs for the target quantities. The main decision variables are the number of wafers to be processed on the individual steppers and the qualifications to be performed on the different steppers in the periods of the planning window.

We assume that the current time is t. The next process step of lot j is $s \geq 1$. We denote the local due dates for process step $i \geq s$ by d_{ji}. The target quantities d_t^f for wafers of family f in period t are determined by the recursive equation

$$d_{ji} := d_{j,i-1} + \left(w_{ji} + 1\right)p_{ji}, \tag{1}$$

where w_{ji} is a lot- and process step-specific relative waiting time estimate. Moreover, we initialize $d_{j,s-1} := r_{js}$ where r_{js} is the release date for step s of job j. When a local due date of a lot belongs to any period of the planning window of the instance formed at time t the target quantity for the corresponding family is increased by the number of wafers that belong to the lot. The relative waiting times are determined by long simulation runs and by considering the slack of the lot. Due to space limitations we do not recall the details and refer instead of this to (Kopp et al. 2018). One obvious limitation of the sketched approach to determine target quantities is the fact that the finite capacity of the wafer fab is not fully taken into account. We call this fairly straightforward approach slack-based approach (SBA) in the remainder of this paper. Note that SBA-type approaches are applied in wafer fabs.

It is shown by simulation experiments that when the MILP is applied in a rolling horizon setting pure critical ratio (CR)-based dispatching (cf. Sarin et al. 2011) without any qualification planning approaches is outperformed by the MILP approach. Here, a special dispatching strategy is used at the steppers that exploits the qualification decisions made by the MILP. We refer again to (Kopp et al. 2018) for the details. The top-level is formed by the SBA scheme, the medial-level by the MILP, and the base-level by the dispatching strategy.

In the present paper, we strive for eliminating some of the limitations of the SBA by designing a LP model for the target quantity calculation. This approach is abbreviated by LPA. Note that the LPA procedure has to take into account the finite capacity of the different tool groups. Moreover, due dates have to be respected. We are interested in assessing the performance of the LPA within a rolling horizon approach using discrete-event simulation.

2.2 Discussion of Related Work

Next, we discuss related work with respect to qualification management and hierarchical approaches in semiconductor manufacturing. Ignizio (2009) analyzes a principle qualification management problem for the stepper tool group in a wafer fab. A binary LP is used to make qualification decisions. Substantial cycle time and qualification cost reductions are observed. Several flexibility measures for principal qualification are proposed and assessed by Johnzén et al. (2011) and by Rowshannahad et al. (2015). Fu et al. (2010) consider a principle qualification management problem for semiconductor backend facilities. A MILP approach is proposed. But because an entire backend facility is considered, the target quantities in the qualification problem in the present paper can be replaced by the demand for the backend facility. Since the assumption of deterministic demand is unrealistic in many situations, a stochastic programming approach that is able to deal with demand uncertainty is proposed by Fu et al. (2015). Qualification management problems for tool groups motivated by semiconductor manufacturing settings are formulated as two-stage stochastic programming problems by Chang and Dong (2017). In contrast to Fu et al. (2015), the uncertainty of the offered tool group capacity is taken into account. This uncertainty is caused by tool breakdowns or by uncertainty in qualification times. Lagrangean relaxation is used to obtain computationally tractable solution schemes.

Hierarchical approaches are useful to tackle difficult decision-making tasks (Schneeweiss 2003). Such approaches are applied to semiconductor manufacturing-related problems. For instance, a two-level hierarchical approach is proposed by Mönch and Drießel (2005) and Mönch et al. (2006) for scheduling lots in a single wafer fab. The top-level is formed by an operational planning approach that set target completion times for the lots with respect to the different work areas, while a distributed variant of the shifting bottleneck heuristic is used for each single work area to schedule the lots. A hierarchical approach is also proposed by Chen et al. (2011) that is based on the idea to use a production planning model to calculate targets for lot moves, whereas dispatching rules that exploit the planning results are used to decide which lot is processed next. A hierarchical approach to ensure consistency of global and local scheduling decisions in wafer fabs is described by Sadeghi (2017). Linear programming is used on the top-level to derive move targets for the base-level.

A multi-level hierarchical approach for decisions in the photolithography area is proposed by Klemmt et al. (2010). The top-level of the hierarchy deals with tool qualifications to maximize throughput while taking WIP projections for the next weeks into account. However, it is not described how these WIP projections are obtained.

The most pertinent papers are those by Kopp et al. (2016), (2018) discussed in Subsection 2.1. Based on the discussed literature it seems desirable to design a hierarchical approach for the qualification management problem discussed in (Kopp et al. 2016, 2018). The top level should ensure that the finite capacity of the shop floor is taken into account. Moreover, it should be an aggregated approach to reduce the computational burden.

3 HIERARCHICAL APPROACH

3.1 Overall Design Issues

We propose a three-level hierarchical approach for qualification management in wafer fabs. In this paper, we mainly focus on the design of the top-level. The top-level is required for setting family targets for each period of the planning window of the medial-level. The LPA is based on input from higher-level planning approaches such as release schedules from production planning and capacity-feasible demand from master planning. The main purposes of the medial- and base-level are already described in Subsection 2.1. The hierarchical setting is shown in Figure 1.

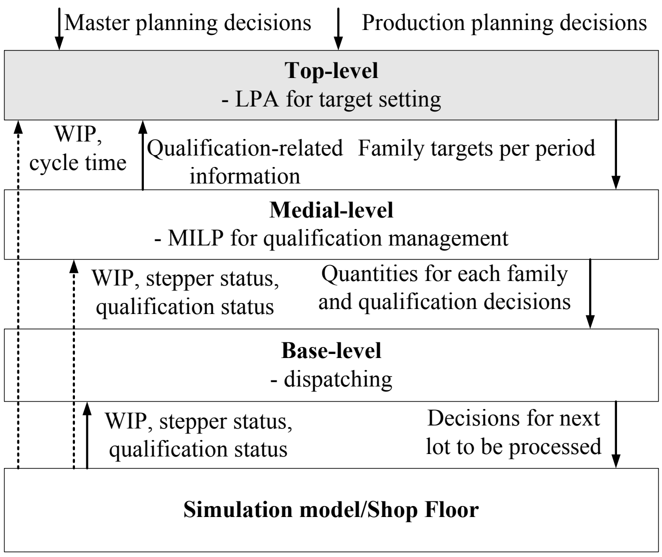

Figure 1: Hierarchical approach.

The execution level, i.e. the base system of the wafer fab, is represented by a simulation model that mimics the behavior of the base system. We will describe the rolling horizon setting in more detail in Subsection 4.1.

3.2 Top-Level

The top-level includes the LPA. Therefore, a finite time window of length T divided into discrete periods of equal length is considered. Each planning period has the duration of a single day. We assume that we know a release schedule and that demand information is available for the planning window. Moreover, cycle time information from the shop-floor is available that can be used to determine the number of operations that are completed within a planning period.

Due to the sake of simplicity, we start by presenting a model for an instance of the LPA that is generated and solved at the begin of each planning epoch of the rolling horizon approach. The following sets and indices are used in the LP model:

$g = 1,\ldots,G$	product index
$k = 1,\ldots,K$	tool group index
$l = 1,\ldots,n_g$	operation index for product g
$t = 1,\ldots,T$	period index.

The following parameters will be used within the model:

D_{gt}: demand for product g in period t (in wafers)
R_{gt}: (planned) quantity of product g released in period t (in wafers)
C_{kt}: capacity of tool group k in period t (in hours)
p_{gl}: processing time of a single wafer from product g on operation l (in hours)

$$d_{glk} = \begin{cases} 1, & \text{if the tools of tool group } k \text{ are able to process wafers of product } g \text{ at operation } l \\ 0, & \text{otherwise} \end{cases}$$

ω_{gtl}: WIP cost for product g at operation l in period t (per wafer)
b_{gt}: backlog cost for product g in period t (per wafer)
h_{gt}: inventory holding cost for product g in period t (per wafer)
s_{gt}: cost for shortfall of the quantity of product g to be released in period t (per wafer)
e_{gt}: cost for exceeding the quantity of product g to be released in period t (per wafer)
δ_{gl}: number of operations which can be completed in one period including operation l for product g.

The following decision variables are used within the LP:

X_{gtl}: quantity of product g released in period t to operation l
Y_{gtl}: quantity of product g completing operation l in period t
W_{gtl}: WIP of product g at operation l at the end of period t
B_{gt}: backlog quantity of product g in period t
I_{gt}: inventory quantity of product g in period t
S_{gt}: shortfall quantity of product g to be released in period t
E_{gt}: exceeding quantity of product g to be released in period t.

Now the problem can be formulated as follows:

$$\min \sum_{t=1}^{T}\sum_{g=1}^{G}\left[\sum_{l=1}^{n_g}\omega_{gtl}W_{gtl} + b_{gt}B_{gt} + h_{gt}I_{gt} + s_{gt}S_{gt} + e_{gt}E_{gt}\right] \tag{2}$$

subject to

$$W_{g,t-1,l} + X_{gtl} - Y_{gtl} = W_{gtl} \qquad g=1,\dots,G,\ t=1,\dots,T,\ l=1,\dots,n_g \tag{3}$$

$$Y_{gtl} = X_{g,t,l+1} \qquad g=1,\dots,G,\ t=1,\dots,T,\ l=1,\dots,n_g-1 \tag{4}$$

$$Y_{gt} + I_{g,t-1} - D_{gt} - B_{g,t-1} + B_{gt} = I_{gt} \qquad g=1,\dots,G,\ t=1,\dots,T \tag{5}$$

$$X_{gt} + E_{g,t-1} - R_{gt} - S_{g,t-1} + S_{gt} = E_{gt} \qquad g=1,\dots,G,\ t=1,\dots,T \tag{6}$$

$$\sum_{g=1}^{G}\sum_{l=1}^{n_g} d_{glk}\, p_{gl} Y_{gtl} \le C_{kt} \qquad k=1,\dots,K,\ t=1,\dots,T \tag{7}$$

$$\sum_{j=0}^{\delta_{gl}} W_{g,t-1,l-j} \ge Y_{gtl} \qquad g=1,\dots,G,\ t=1,\dots,T,\ l=1,\dots,n_g \tag{8}$$

$$X_{gtl}, X_{gt}, Y_{gtl}, Y_{gt}, W_{gtl}, B_{gt}, I_{gt}, S_{gt}, E_{gt} \ge 0 \qquad g=1,\dots,G,\ t=1,\dots,T,\ l=1,\dots,n_g. \tag{9}$$

The objective (2) is to minimize the sum of backlog and inventory holding costs for finished wafers and penalty costs for deviations of the number of released wafers from the intended quantity. The constraints (3) and (4) ensure that each lot has to complete all operations of their routes to finish the production process. The constraint set (5) represents inventory balance equations for lots completing their last operation while constraint set (6) balances the number of released wafers, i.e., wafers of lots entering the first operation. A capacity restriction for each tool group is modeled by constraints (7). Constraints (8) limits the lot movements along their routes. That means that a lot at operation $l-\delta_{gl}$ can finish in a single period at most the operations up to operation l. The constraint set (9) guarantees that the domain of the decision variables is respected. This LP is similar to the fixed lead time formulation for production planning described by Kacar et al. (2013). The δ_{gl} values are calculated in a preprocessing step using backward termination similar to expression (1). The processing times of the process steps and a fraction $\rho<1$ of the process step-specific relative waiting time estimates w_i are applied within the calculations.

Note that we are mainly interested in the Y_{gtl} quantities for those process steps l that belong to the operations that have to be carried out on the steppers. These quantities are used to derive values for the targets d_t^f. They are a major ingredient for the MILP formulation on the medial-level.

4 SIMULATION INFRASTRUCTURE

4.1 Simulation Framework

The simulation framework proposed by Mönch (2007) for production control and extended to planning approaches by Ponsignon and Mönch (2014) is applied to assess the performance of the hierarchical approach in a rolling horizon setting. A blackboard-type data layer that is located in the memory of the simulation computer contains the status of important objects such as lots, tools, and route information.

A stop and go approach is applied to implement the rolling horizon approach. This means that the simulation stops whenever a LPA instance or a MILP instance for qualification management have to be solved. This requires that the corresponding instances are generated based on data from the simulation. In this paper, the MILP model has a planning window of $\tau=12$ periods where the period length is $\Delta=4$

hours. A new qualification plan is computed every Δ hours. Since the generation of a new MILP instance requires instructions from the top-level, we generate a LPA instance every Δ hours. The LPA planning window is $T = 60$ days, while the length of a single period is one day. For the MILP solution, a maximum computing time of five minutes per instance is allowed. Moreover, a relative MIP gap of 5% percent is used to terminate the solution process before the end of the maximum computing time is reached. The simulation infrastructure is coded in the C++ programming language. AutoSched AP is used as simulation engine. The CPLEX libraries are applied to solve the LPA instances and the MILP instances for qualification management.

4.2 Simulation Model

The MIMAC I simulation model (cf. Simulation Data Sets 2018) is used within the rolling horizon experiments. It contains two products with over 200 process steps each. Over 200 tools are considered that are organized in 84 tool groups. Due to the reentrant process flows that lead to frequent stepper visits, we obtain 17 families. Each lot of the two products has 48 wafers. The simulation model contains batch processing tools and tools with sequencing-dependent setup times.

Moreover, it contains a fairly detailed model of the photolithography work area that includes reticles and SAW functionality. The time windows have a length of 1-3 days. They are randomly chosen according to a discrete uniform distribution. Existing qualifications might expire due to several reasons, for instance due to reaching the end of the qualification time window or due to tool breakdowns (see Kopp et al. 2018 for more details).

For the sake of simplicity, we assume that a release schedule exists for the entire simulation horizon. This release schedule is derived from normally distributed release quantities where we have to specify the mean that leads to a target bottleneck utilization (BNU) and the coefficient of variation (CV) of the release quantities. We artificially create demand that has to be fulfilled by the wafer fab by performing due date calculations for WIP lots and lots from the release schedule using the recursive approach from expression (1). Here, we use waiting time factors that result from the target BNU. A similar technique is used to set the due dates of the lots. We recall that in a real-world setting both the demand and the release schedule are provided by higher planning levels. Since the demand is also calculated based on WIP lots, the demand changes slightly along the simulation horizon. Note again that in a real-word setting forecast updates occur that lead to demand fluctuations over time.

5 SIMULATION EXPERIMENTS

5.1 Design of Experiments

The goal of the experiments consists in comparing the SBA and the LPA. We expect that the performance of the hierarchical approach depends on the target BNU level of the steppers, the demand variability represented by the CV level, and the qualification costs. Note that the remaining cost settings for the MILP approach on the medial-level are taken from (Kopp et al. 2016). The design of experiments is summarized in Table 1.

We are interested in assessing the quality of the hierarchical approach by means of global performance measure values. Therefore, we report the resulting throughput (TH), i.e. the number of lots completed over the entire simulation horizon, the average cycle time (ACT), and the average tardiness (AT). The tardiness of a lot with due date d_j and completion time C_j is defined as $T_j = \max(C_j - d_j, 0)$. We will also report the number of expired qualifications due to reaching the end of the time windows. This quantity is abbreviated by QE. In addition, we show the average number of existing qualifications, denoted by #Q.

Table 1: Design of simulation experiments.

Factor	Level	Count
BNU	70%, 95%	2
CV of the release quantities	0.1, 0.25	2
Qualification costs (for the MILP at the medial-level)	100, 800, 4000	3
Top-level approach	SBA, LPA	2
Number of independent simulation replications	5	
Total number of simulation runs		120

A simulation horizon of 26 weeks is used in the experiments. This includes a warm-up period of four weeks that is excluded from the statistics. Five independent simulation replications are performed for each factor combination. The average values of the corresponding performance measures are reported. The cost settings for the LPA on the top-level used in the experiments are shown in Table 2. These settings performed well compared to other settings in some preliminary experimentation. Similar cost values are also used by Kacar et al. (2016). Note that the costs e_{gt} and s_{gt} for exceeding and shortfall the quantity of product g to be released in period t are fairly high since the release schedule is already determined and has to be met by the LPA as much as possible.

Table 2: Cost setting of the LPA on the top-level.

Cost Setting	b_{gt}	h_{gt}	ω_{gtl}	e_{gt}	s_{gt}
	30	5	10	100	100

The simulation experiments are performed using a computer with an Intel Core i7-4790 CPU 3.60 GHz with 16 GB RAM. The simulation time for a single replication is on average 12 hours. Note that this includes the solution of 1092 MILP instances and 1092 LP instances.

5.2 Simulation Results

The simulation results for a target BNU level of 95% are shown in Table 3 where we compare the performance of the top-level approaches SBA and LPA. The results are grouped with respect to the different factor combinations. Average values are reported.

Table 3: Results of simulation experiments for 95% BNU.

CV	Cost	Top-level	TH	ACT (in days)	AT (in hours)	QE	#Q
0.1	**100**	SBA	1488	22.69	38.8	244	30.9
		LPA	1486	22.68	35.5	160	27.8
	800	SBA	1488	22.79	37.9	153	25.3
		LPA	1484	23.01	40.8	138	25.4
	4000	SBA	1479	24.15	63.6	113	20.3
		LPA	1480	23.79	54.4	108	22.5
0.5	**100**	SBA	1458	23.83	71.4	265	29.9
		LPA	1456	23.37	59.7	174	27.6
	800	SBA	1458	23.94	74.7	164	24.7
		LPA	1457	23.83	65.9	141	25.0
	4000	SBA	1457	25.39	96.7	120	20.0
		LPA	1458	24.21	70.8	100	22.3

The results for a BNU level of 70% are shown in Table 4. They are organized in a similar way as found in Table 3.

Table 4: Results of simulation experiments for 70% BNU.

CV	Cost	Top-level	TH	ACT (in days)	AT (in hours)	QE	#Q
0.1	100	SBA	1077	19.63	49.1	649	30.9
		LPA	1081	19.69	50.2	385	24.6
	800	SBA	1081	19.73	51.4	384	24.1
		LPA	1079	19.81	52.3	329	22.6
	4000	SBA	1083	20.18	59.9	242	19.1
		LPA	1082	20.05	56.9	223	19.0
0.25	100	SBA	1048	19.86	54.3	658	30.6
		LPA	1050	19.92	55.5	456	25.3
	800	SBA	1050	20.10	58.8	401	23.5
		LPA	1053	19.96	55.2	359	22.2
	4000	SBA	1051	20.38	64.3	267	19.2
		LPA	1050	20.18	59.4	258	19.1

The ACT values are depicted in Figure 2, while the number of expired qualifications is shown in Figure 3.

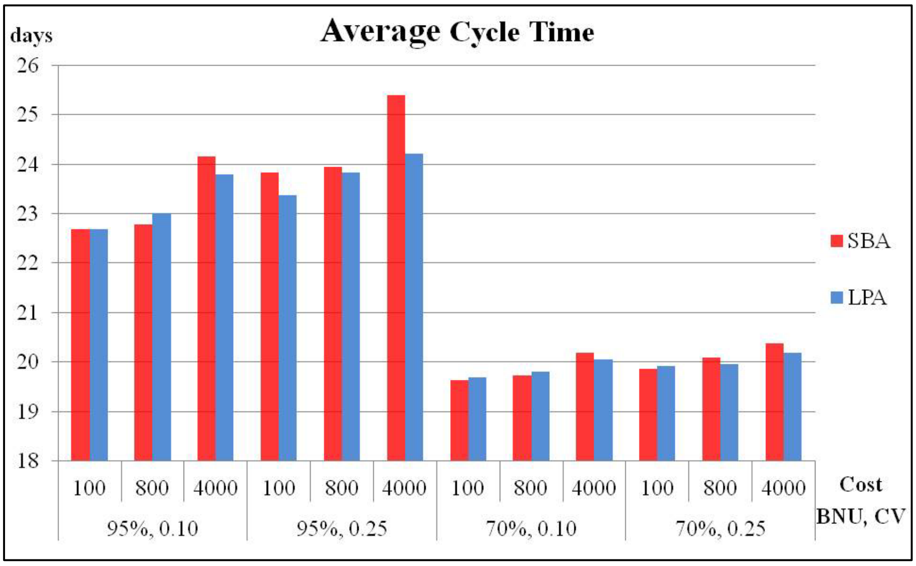

Figure 2: Average cycle times for all factor combinations.

5.3 Analysis and Discussion of the Results

We see from the Tables 3 and 4 that for both approaches low qualification costs result in a larger number of existing qualifications. In contrast, larger qualification costs lead to a smaller requalification effort due to a smaller number of expired qualifications. However, we see from Figure 2 that in this situation at the same time the reached performance measures such as ACT are worse.

We see from the Tables 3 and 4 that under some experimental conditions the SBA is outperformed by LPA with respect to TH, ACT, AT, and the qualification effort. Moreover, there are situations, for instance, at BNU = 70% and low qualification costs, where the SBA performs only slightly better with respect to the obtained ACT values but the number of expired qualifications is much larger. Under all experimental conditions, the LPA leads to a smaller number of expired qualifications.

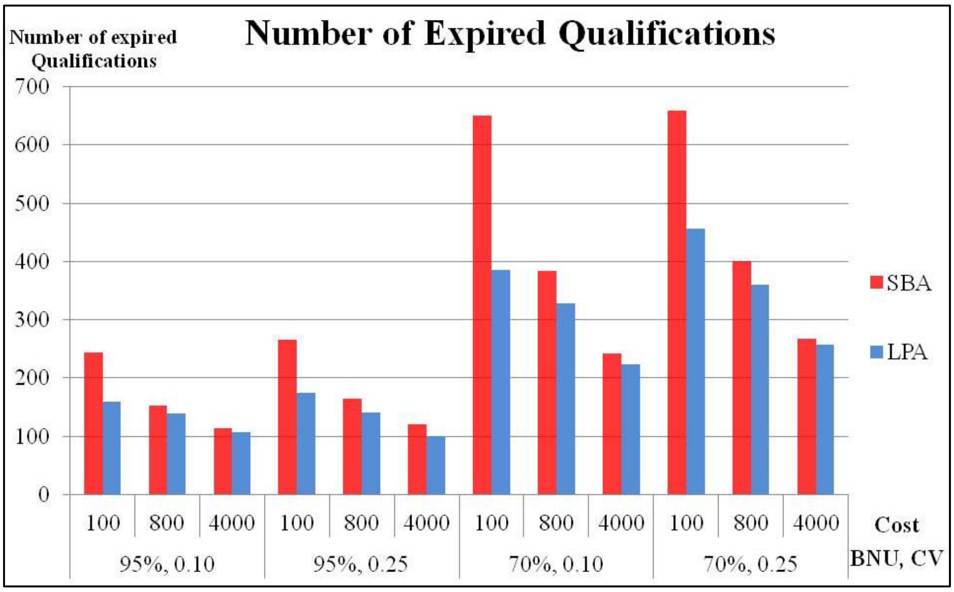

Figure 3: Expired qualifications for all factor combinations.

Next, we discuss the impact of the target BNU level and the demand variability. Analyzing the ACT and the AT values, both approaches achieve similar results for a BNU level of 70%, as we can see from Table 4. However, the LPA performs better for a BNU level of 95% since in this situation less room exists for waiting steppers due to lacking qualifications for proceeding available lots while the qualification effort is also smaller at the same time. We see from the Tables 3 and 4 that larger CV values lead to worse results with respect to the global performance measure values as well as the qualification effort. As expected, the LPA can better deal with a larger demand variability since its performance compared to that of the SBA is better.

We analyze the impact of the different qualifications costs for the SBA and the LPA. We observe from the Figures 2 and 3 that the differences for both the ACT and the number of expired qualifications are larger when using the SBA instead of the LPA. Therefore, the LPA is less sensitive to changes of the qualification costs.

6 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we discussed the design of a top-level in a hierarchical approach for qualification management. We proposed an LP-based approach that takes into account the finite capacity of the tool groups. Moreover, global performance measures such as ACT or TH are supported by this approach. We compared the new planning-based approach with a rule-based approach from previous research (Kopp et al. 2018) when both approaches are applied in a rolling horizon setting. Simulation experiments with a large-sized wafer fab model demonstrated that the LP-based approach outperforms the rule-based approach under almost all experimental conditions.

There are several directions for future research. First of all, we think that models with nonlinear clearing functions (cf. Missbauer and Uzsoy 2010) should be used instead of the present LP formulation that requires estimates with respect to the number of process steps to be performed within a period. We believe that this is especially important when demand that follows the multiplicative martingale model of forecast evolution (MMFE) (Heath and Jackson 1994) is considered within the rolling horizon. Here, experiments with the multi-product multiplicative MMFE scheme presented by Ziarnetzky et al. (2018) are interesting. Assessing stability issues within the rolling horizon setting is another avenue for future research.

ACKNOWLEDGMENTS

The authors would like to thank Detlef Pabst and Marcel Stehli, GLOBALFOUNDRIES, for fruitful discussions of the qualification management problem addressed in this paper.

REFERENCES

Akçali, E., K. Nemoto, K., and R. Uzsoy. 2001. "Cycle-time Improvements for Photolithography Process in Semiconductor Manufacturing. " *IEEE Transactions on Semiconductor Manufacturing* 14(1):48-56.

Chang, X. and M. Dong. 2017. "Stochastic Programming for Qualification Management of Parallel Machines in Semiconductor Manufacturing." *Computers & Operations Research* 79:49-59.

Chen, M., S. C. Sarin, and A. Peake. 2010. "Integrated Lot sizing and Dispatching in Wafer Fabrication." *Production Planning & Control* 21(5): 485-495.

Chien, C., S. Dauzère-Pérès, H. Ehm, J. Fowler, Z. Jiang, S. Krishnaswamy, L. Mönch, and R. Uzsoy. 2011. "Modeling and Analysis of Semiconductor Manufacturing in a Shrinking World: Challenges and Successes." *European Journal of Industrial Engineering* 5(3):254-271.

Fu, M., M. Haghnevis, R. Askin, J. W. Fowler, and M. Zhang. 2010. "Machine Qualification Management for a Semiconductor Back-end Facility." In *Proceedings of the Winter Simulation Conference*, edited by L. Yilmaz et al., 2486-2492. Piscataway, New Jersey: IEEE.

Fu, M., R. Askin, J. W. Fowler, and M. Zhang. 2015. "Stochastic Optimization of Product–machine Qualification in a Semiconductor Back-end Facility." *IIE Transactions* 47(7):739-750.

Ignizio, J. P. 2009. "Cycle Time Reduction via Machine-to-operation Qualification." *International Journal of Production Research* 47(24):6899-6906.

Johnzén, C., S. Dauzère-Pérès, and P. Vialletelle. 2011. "Flexibility Measures for Qualification Management in Wafer Fabs." *Production Planning & Control* 22(1):81-90.

Heath, D. C. and P. L. Jackson. 1994. "Modeling the Evolution of Demand Forecasts with Applications to Safety Stock Analysis in Production Distribution Systems." *IIE Transactions* 26(3):17-30.

Kacar, N. B., L. Mönch, and R. Uzsoy. 2013. "Planning Wafer Starts using Nonlinear Clearing Functions: a Large-Scale Experiment." *IEEE Transactions on Semiconductor Manufacturing* 26(4):602-612.

Kacar, N. B., L. Mönch, and R. Uzsoy. 2016. "Modeling Cycle Times in Production Planning Models for Wafer Fabrication." *IEEE Transactions on Semiconductor Manufacturing* 29(2):153-167.

Klemmt, A., J. Lange, G. Weigert, F. Lehmann, and J. Seyfert. 2010. "A Multistage Mathematical Programming based Scheduling Approach for the Photolithography Area in Semiconductor Manufacturing." In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson et al., 2474-2485. Piscataway, New Jersey: IEEE.

Kopp, D., L. Mönch, D. Pabst, and M. Stehli. 2016. "An Optimization Model for Qualification Management in Wafer Fabs." In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 2610-2620. Piscataway, New Jersey: IEEE.

Kopp, D., L. Mönch, D. Pabst, and M. Stehli. 2018. "Qualification Management in Wafer Fabs: Optimization Approach and Simulation-based Performance Assessment." Technical Report, Department of Mathematics and Computer Science, University of Hagen, Hagen, Germany.

Missbauer, H., and R. Uzsoy 2010. "Optimization Models for Production Planning." In *Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook*, edited by K. G. Kempf, P. Keskinocak, and R. Uzsoy, 437-508. New York: Springer.

Mönch, L. and R. Drießel. 2005. "A Distributed Shifting Bottleneck Heuristic for Complex Job Shops." *Computers & Industrial Engineering* 49(3):673-680.
Mönch, L., J. W. Fowler, and S. J. Mason. 2013. *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems*. New York: Springer.
Mönch, L., M. Stehli, J. Zimmermann, and I. Habenicht., 2006. "The FABMAS Multi-Agent-System Prototype for Production Control of Waferfabs: Design, Implementation, and Performance Assessment." *Production Planning & Control* 17(7):701-716.
Mönch, L. 2007. "Simulation-Based Benchmarking of Production Control Schemes for Complex Manufacturing Systems." *Control Engineering Practice* 15(11):1381-1393.
Ponsignon, T. and L. Mönch. 2014. "Simulation-based Performance Assessment of Master Planning Approaches in Semiconductor Manufacturing." *OMEGA* 46:21-35.
Rowshannahad, M., S. Dauzère-Pérès, and B. Cassini. 2015. "Capacitated Qualification Management in Semiconductor Manufacturing." *OMEGA* 54:50–59.
Sadeghi, R. 2017. *Consistency of Global and Local Scheduling Decisions in Semiconductor Manufacturing*. PhD Thesis, Ecole des Mines de Saint-Etienne, France.
Sarin, S. C., A. Varadarajan, and L. Wang. 2011. "A Survey of Dispatching Rules for Operational Control in Wafer Fabrication." *Production Planning & Control* 22(1): 4-24.
Schneeweiss, C. 2003. *Distributed Decision Making*, 2nd ed., Heidelberg: Springer.
Simulation Data Sets. 2018. "MIMAC Datasets." http://p2schedgen.fernuni-hagen.de/indexphp?id=242, accessed April 25, 2018.
Ziarnetzky, T., L. Mönch, and R. Uzsoy. 2018. "Rolling Horizon, Multi-product Production Planning for Wafer Fabs with Chance Constraints and Forecast Evolution " *International Journal of Production Research* https://www.tandfonline.com/doi/full/10.1080/00207543.2018.1478461.

AUTHOR BIOGRAPHIES

DENNY KOPP is a Ph.D. student at the Chair of Enterprise-wide Software Systems at the University of Hagen, Germany. He received a MS degree in Business Administration from the Otto von Guericke University Magdeburg, Germany. His research interests include applied optimization and simulation-based production control. His email address is Denny.Kopp@fernuni-hagen.de.

LARS MÖNCH is Professor in the Department of Mathematics and Computer Science at the University of Hagen, Germany. He received a master's degree in applied mathematics and a Ph.D. in the same subject from the University of Göttingen, Germany. His current research interests are in simulation-based production control of semiconductor wafer fabrication facilities, applied optimization and artificial intelligence applications in manufacturing, logistics, and service operations. He is a member of GI (German Chapter of the ACM), GOR (German Operations Research Society), and INFORMS. He is an Associate Editor for the *European Journal of Industrial Engineering, Journal of Simulation*, *Business & Information Systems Engineering*, *IEEE Robotics and Automation Letters*, *IEEE Transactions on Semiconductor Manufacturing*, and *IEEE Transactions on Automation Science and Engineering*. His email address is Lars.Moench@fernuni-hagen.de.