Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

EVALUATING THE IMPACT OF DYNAMIC QUALIFICATION MANAGEMENT IN SEMICONDUCTOR MANUFACTURING

Antoine Perraudat
Stéphane Dauzère-Pérès

Mines Saint-Étienne, Univ Clermont Auvergne
CNRS, UMR 6158 LIMOS
CMP, Depart. of Manufacturing Sciences and Logistics
880 route de Mimet
Gardanne, 13541, FRANCE

Philippe Vialletelle

STMicroelectronics
850 rue Jean Monnet
Crolles, 38926, FRANCE

ABSTRACT

In semiconductor manufacturing, before executing any operation on a product, a machine must be qualified, i.e., certified, to ensure quality and yield requirements. The qualifications of machines in a work-center are essential to the overall performance of the manufacturing facility. However, performing a qualification can be expensive and usually takes time, although the more qualified the machines, the more flexible the production system. Qualification management aims at determining the right qualifications at the lowest cost. We first discuss the limitations of a single-period optimization model, in particular due to capacity losses and delays inherent to qualification procedures. Then, we motivate and briefly introduce a multi-period optimization model. Finally, we compare both optimization models in a computational study on industrial instances from a High Mix/Low Volume (HMLV) production facility with a high production variability.

1 INTRODUCTION

Semiconductor industry is a complex process industry that manufactures integrated circuits on silicon wafers. Silicon wafers are generally grouped by lots of 25 wafers. To complete the process flow from raw material, a lot must undergo hundreds of different physical and chemical processes performed by expensive machines in work-centers. Semiconductor production facilities are characterized by a very high degree of re-entrant product flows. However, before applying physical or chemical processes, machines must undergo recipe-to-machine qualification operations. A fruitful qualification operation certifies that a machine can execute a recipe (i.e., a fabrication process) and that this machine respects yield and quality requirements.

However, machine qualifications are dynamic. In other words, machine disqualifications can occur over time. In this case, machine qualifications are lost and the machine can no longer execute some recipes. Before being capable of re-applying the recipe on lots, machines must be qualified. Disqualification reasons are numerous. Machines can be disqualified because yield shrinks, a consumable is entirely consumed (i.e., a bottle of gas). Machines can also be disqualified after a failure, or due to WIP management policies, e.g., when two recipes are incompatible and cannot be qualified at the same time on the machine. There also exist qualification time windows (Kopp et al. (2016)), which define the amount of time after which machines must be disqualified if they did not process a given recipe. This is done for yield or quality reasons as the quality of a recipe is time-varying and influenced by other recipes. Efficient qualification procedures can be associated to better industrial performances (Johnzén et al. 2011; Rowshannahad et al. 2015; Kopp et al. 2018) in terms of throughput, cycle time, workload balancing and variability.

Literature often discusses qualification management in production planning or scheduling problems, in particular for the lithography work-center, which is often bottleneck in wafer fabs. Kopp et al. (2016)

consider a production planning problem. They consider that qualifications can be lost due to qualification time windows. Kopp et al. (2018) propose a mixed integer linear program (MILP) and a simulation approach to evaluate different qualification strategies. Kopp and Mönch (2018) introduce a three-level hierarchical approach to better manage qualifications of machines. Ignizio (2009) and Ignizio (2010) use a MILP and simulation to study the impact of qualifications on mean cycle time and cycle time variability. Kabak et al. (2013) use discrete even simulation to assess the impact of recipe restrictions and disqualifications on the mean cycle time. Aubry et al. (2008) study the impact of qualifications in a production scheduling problem. Similarly, Johnzén et al. (2008) analyze the effect of additional qualifications on cycle times using simulation models.

Fu et al. (2015) consider that the demand is stochastic and seek to minimize expected total production costs, i.e., production, inventory, backlogging and qualification costs. Johnzén et al. (2011) define flexibility measures to assess the qualification configuration of a work-center, and the impact of qualification decisions. Rowshannahad and Dauzère-Pérès (2013) include batching constraints in flexibility measures. Rowshannahad et al. (2014) propose another measurement to assess workload variability between machines in a work-center. Rowshannahad et al. (2015) define a mixed integer non linear programming (MINLP) model to find the best qualifications that optimize the time flexibility measure at finite capacity. Pianne et al. (2016) introduce ideal and potential flexibility measures.

In general, literature rarely considers in optimization models the fact that qualifications can be subject to lead times or can require maintenance operations. Chang and Dong (2017) consider a single-period approach where total priority moves must be maximized. The demand is stochastic. Moreover, a qualification induces stochastic capacity losses due to maintenance operations.

In this paper, we put ourselves in the shoes of a work-center manager in charge of meeting daily production targets. He/she knows that recipes are disqualified on some machines. He/she knows that qualifying recipes will probably improve capacity production so that daily production targets can be met. However, we cannot simply qualify all recipes on all machines because qualifications are expensive, can induce capacity losses, can be subject to lead times and available human resources are limited. Given a production plan, i.e., the quantity and period of arrival associated to each recipe, machine production capacities, and disqualifications, the problem consists in finding what are the best k qualifications to perform in order to improve work-center performances. In this paper, performances are defined in terms of moves out, i.e., the number of wafers processed by the end of the planning horizon. To solve this problem, we propose two new optimization models based on the capacitated time flexibility measure proposed in Rowshannahad et al. (2015). Special attention is to given to the operational dimension of this qualification management problem.

This paper is organized as follows. In Section 2, optimization models are presented. In Section 3, numerical results are shown on industrial data from a 300mm wafer fab located in Crolles, France. Finally, in Section 4, we conclude and give perspectives on this work.

2 MATHEMATICAL FORMULATIONS

We propose to model the studied problem with a bi-level optimization model. The lower-level optimization problem builds realistic queues of recipes in front of machines by using empirical observations of dispatching engines. Once queues of recipes are defined, moves out can be computed by the upper-level optimization problem. Considering dispatching rules in qualification management is relevant because they may affect the benefit of qualification decisions (Johnzén et al. (2008)).

2.1 Single-period Optimization Model

Indices and sets:

- m: Index for machines, $\in \{1,..,M\}$,
- r: Index for recipes, $\in \{1,..,R\}$.

Parameters:

- k: Number of qualification decisions to be made at the beginning of the planning horizon,
- $Q_{r,m}$: Is equal to 1 if machine m is qualified for recipe r, is equal to 2 if machine m is qualifiable for recipe r, is equal to 0 if machine m cannot be qualified for recipe r,
- $TP_{r,m}$: Throughput rate (in seconds) of recipe r on machine m,
- $C^{loss}_{r,m}$: Capacity loss generated (in seconds) by qualifying recipe r on machine m,
- C_m: Initial availability time (in seconds) of machine m over the planning horizon,
- D_r: Demand in number of wafers for recipe r over the planning horizon,
- γ: Workload balancing parameter strictly greater than one.

Decision variables:

- $OQ_{r,m} \in \{0,1\}$: Is equal to 1 if a qualification procedure is proposed for recipe r on machine m at the beginning of the planning horizon, and 0 otherwise,
- U_m: Capacity utilization of machine m,
- C^{eff}_m: Effective availability time (in seconds) of machine m over the planning horizon,
- $WIP_{r,m}$: Quantity of recipe r processed by machine m.

A bi-level qualification management optimization model is proposed to model the problem.

Upper-level optimization problem:

$$\max \quad MovesOut = f(\mathbf{U}, \mathbf{WIP}) \tag{1}$$

$$\text{s.t.} \quad \sum_{r,m} OQ_{r,m} = k \tag{2}$$

$$C^{eff}_m = \max(C_m - \sum_r C^{loss}_{r,m} OQ_{r,m}, 0) \qquad \forall m \tag{3}$$

$$U_m, WIP_{r,m} \in \arg\min LBP(\mathbf{OQ}, \mathbf{C^{eff}}) \tag{4}$$

$$OQ_{r,m} \in \{0,1\} \qquad \forall r, \forall m \tag{5}$$

Lower-level optimization problem:

$$LBP(\mathbf{OQ}, \mathbf{C^{eff}}) = \min \quad \sum_m U^{\gamma}_m \tag{6}$$

$$\text{s.t.} \quad \sum_m WIP_{r,m} = D_r \qquad \forall r \tag{7}$$

$$U_m \geq \sum_r \frac{WIP_{r,m}}{TP_{r,m} C^{eff}_m} \qquad \forall m \tag{8}$$

$$WIP_{r,m} \leq D_r \qquad \forall r, \forall m | Q_{r,m} = 1 \tag{9}$$

$$WIP_{r,m} \leq D_r OQ_{r,m} \qquad \forall r, \forall m | Q_{r,m} = 2 \tag{10}$$

$$WIP_{r,m} \leq 0 \qquad \forall r, \forall m | Q_{r,m} = 0 \tag{11}$$

$$WIP_{r,m} \geq 0 \qquad \forall r, \forall m \tag{12}$$

Upper-level optimization problem: Equation (1) defines the objective function that consists in maximizing the number of moves out that is computed from the workload balancing on the machines in the work-center (see Section 2.3). Constraint (2) sets to k the number of qualifications that must be performed at the beginning of the planning horizon. The parameter k is defined by decision makers, often with respect to limitations on time and human resources, and can be used to identify the k most blocking issues in terms of moves over the planning horizon. These blocking issues should then be managed in priority. Constraint (3) defines the effective availability time of machine m if there are capacity losses due to qualification procedures. Changeover times are considered by subtracting an average capacity loss to each machine production capacity. Constraint (4) links the upper-level and lower-level problems. Constraint (5) are the binary constraints for the qualification decisions.

Lower-level optimization problem: The lower-level optimization problem is used to simulate real time dispatching rules. In practice, dispatching engines try to balance the workload on the machines, i.e., to maximize the utilization of machines, as much as possible to maximize the moves out. Equation (6) defines the objective of the lower-level problem, i.e., the capacitated time flexibility measures (Rowshannahad et al. (2015)) that consists in balancing the workload on the machines. Constraint (7) defines the flow conservation on the planning horizon. Constraint (8) defines the capacity utilization rate of each machine in the work-center. Constraint (9)-(10) ensures that workload can only be assigned to machine m if recipe r is qualified on machine m. Constraint (11) ensures that if recipe r is not qualified and cannot be qualified on machine m, then the workload corresponding to recipe r is never assigned to machine m. Constraint (12) is the non-negativity constraint for variable $WIP_{r,m}$.

Note that if there exists a line stop for a recipe, i.e., if all qualified machines are down or if there is no qualified machine, it is not included in the model for feasibility reasons. In practice, as these recipes are often critical in operational management of qualifications, they are included in resolution approaches.

2.2 Multi-period Optimization Model

In complex industrial environments, like semiconductor manufacturing, where production variability is high, a single-period (static) approach can be insufficient to capture dynamic WIP quantities and capacities, product mix changes and disqualification over time. Moreover, as qualification procedures can require maintenance operations or can be subject to lead times for yield/quality verification, the single-period model seems to lose relevance. This is because the single-period implicitly averages the demand and the capacity over the planning horizon. In the multi-period optimization (dynamic) approach, we compute the capacity vector the same way we compute the capacity vector on a single-period planning horizon (see Constraint (3)). However, if the entire capacity loss due to the qualification cannot only be attributed to the first period, then the remaining capacity loss is attributed to the next period, until there is no capacity loss left. For instance, this can happen if the maintenance operations lasts 12 hours whereas a period lasts 8 hours. For lead times denoted L, we proceed in a similar way. As we perform qualifications at the beginning of the planning horizon, we change the qualification matrix at period t if $1+L=t$. From period t and for the rest of the planning horizon, the machine is newly qualified for the recipe. Finally, for the multi-period optimization model, the workload is balanced on the machines in the work-center for each period of the planning horizon. All wafers that cannot be processed at period t are backlogged at period $t+1$. The difference between the single-period and multi-period approaches is that, in the multi-period approach, each queue is reevaluated in each period to better consider priorities and backlogging.

2.3 Calculation Mode of Moves Out

Once queues of recipes in front of machines are defined through workload balancing in the lower-level optimization problem, two options are proposed to compute the number of moves out while considering dispatching rules: A first approach "average product mix", and a second approach based on recipe priorities.

Average product mix by machine: An average product mix means that all machines process equivalently any recipe in terms of completion percentage. For instance, consider that a machine is used at 120% of its capacity and assigned two recipes, the recipe A with a demand of 25 wafers, and the recipe B with a demand of 100 wafers. In an average product mix situation, the machine processes $\frac{100}{120} = 83.33\%$ of each recipe. If each machine processes an average product mix throughout the planning horizon, then it is possible to derive a closed-form solution for the throughput of a work-center:

$$MovesOut = f(\mathbf{U}, \mathbf{WIP}) = \sum_{m} \frac{1}{\max\{1, U_m\}} \sum_{r} WIP_{r,m} \tag{13}$$

Equation (13) shows that if the capacity utilization rate of a machine is below 100%, then the machine produces everything that is assigned to it. Otherwise, if its capacity utilization rate exceeds 100%, the machine produces $\frac{1}{\max\{1, U_m\}}$ % of each recipe. From an aggregated point of view, an average product mix can be seen as a First-In First-out (FIFO) dispatching rule.

Recipe priorities: Equation (13) may be limited in industrial environments because recipes, therefore lots, are supposed equivalent and do not consider priorities. In practice, lots also have priorities, if they contain new products, must be delivered to important clients, or are behind their schedule. Recipes are therefore run according to their priority. Using recipe priorities may be a better way to simulate real dispatching rules and therefore the effect of a qualification decision. Algorithm (1) is an empirical algorithm, based on the observations of dispatching rules, that includes priorities in the computation of the number of moves out. Algorithm (1) requires $O(MR)$ operations per period.

Algorithm 1 Estimation of moves out with recipe priorities

procedure MOVES OUT = $f(\mathbf{U}, \mathbf{WIP})$
    Moves out = 0
    for each machine m in the work-center with workload **do**
        Get allocated recipes on machine m with their quantity
        Sort allocated recipes by their priority in descending order (priority based dispatching policy)
        If two recipes have the same priority, process first the one with the most demand (setup avoidance policy)
        If two recipes have the same demand, process the fastest one (move maximization policy)
        Count the number of wafers by taking recipes in sorted order until there is no capacity left on the machine
        Add this count to Moves out
    end for
    return Moves out
end procedure

Note that both methods to compute the number of moves out are equivalent, i.e., give the exact same value if all machines in the work-center are loaded below 100% of their capacity (i.e., $U_{m,t} \leq 1.0$).

3 NUMERICAL EXPERIMENTS

In this Section, we want to study if considering more than one period, i.e., considering dynamic WIP quantities and capacities, affects the choice of qualifications. And if it does, to what extent. In the numerical experiments, we assume that no qualification is lost when a qualification is added.

3.1 Instance Generation

We propose to compare both optimization models on industrial data extracted from a 300mm High Mix/Low Volume (HMLV) wafer fab located in Crolles, France. The wafer fab is characterized by shifting bottleneck work-centers, short product life cycles, frequent product mix changes, a high production variability with frequent disqualifications, very high utilization rates of machines and strong tool dedication constraints.

Data were extracted on 15 different weeks in 2018 and 2019. Table 1 shows the size of each industrial instance in terms of number of recipes R, and in terms of number of machines M. 60 industrial instances are used to compare both optimization models on four different work-centers. Demand and capacity were extracted from historical data by using the realized demand and capacity. The qualifications are the ones at the beginning of the planning horizon. The "Implant" work-center is characterized by a high number of different recipes. A same recipe can have very different throughput rates from a machine to another.

The "Lithography" work-center is characterized by machines that process lots in cascade. A large number of lots can be processed at the same time (e.g., up to five or six lots). In the "Diffusion" work-center, certain machines process batches of lot (e.g., at most two to four lots processed at the same time). Other machines process lots wafer by wafer. In the "Diel" work-center, machines process lots wafer by wafer. In general, most machines have a sort of parallel mechanism that allows several lots to be processed at the same time. When extracting industrial data, we could not compute recipe priorities because this parameter is not saved in databases. Therefore, we generate recipe priorities by drawing random numbers between 10 and 1000 from an uniform distribution.

Table 1: Industrial instances used for the computational study (15 instances by work-center).

Work-center	#	R	M	Work-center	#	R	M	Work-center	#	R	M	Work-center	#	R	M
Diel	1	341	77	Implant	1	419	19	Lithography	1	368	23	Diffusion	1	387	132
	2	325	74		2	465	18		2	315	23		2	366	129
	3	328	76		3	494	18		3	338	23		3	367	129
	4	294	76		4	508	18		4	341	23		4	363	129
	5	344	76		5	461	18		5	353	23		5	377	129
	6	274	76		6	432	18		6	350	23		6	358	130
	7	322	76		7	484	18		7	336	23		7	377	130
	8	397	77		8	486	19		8	362	30		8	406	137
	9	353	77		9	455	19		9	373	30		9	400	137
	10	367	77		10	441	19		10	354	30		10	401	137
	11	414	77		11	490	19		11	368	30		11	432	137
	12	350	77		12	473	19		12	364	30		12	364	137
	13	328	77		13	494	19		13	356	30		13	408	137
	14	310	77		14	385	19		14	314	30		14	319	137
	15	318	74		15	489	18		15	346	23		15	398	129

3.2 Design of Experiments

Table 2 presents the design of experiments. We did not run experiments where qualification procedures simultaneously require maintenance operations and are subject to lead times. This is left for future research. We limit ourselves to $k = 1$ so that we can study and compare the optimal solution of both models. The problem is studied on a 24-hour planning horizon. To solve the lower-level workload balancing problem, we use a multi-cut cutting plane algorithm (Bazaraa (2013)) with $\gamma = 6$. The algorithm stops when a relative gap of 10^{-5} is reached. All experiments are run using Java 8 and CLP Java (Lougee-Heimer (2003) and Nils Löhndorf (2016)) as the linear solver for solving the multi-cut cutting plane algorithm. To search for the best qualification, the input qualification matrix $Q_{r,m,0}$ is modified, then the lower-level workload balancing is solved, and then the upper-level problem is solved.

Table 2: Design of experiments.

Parameters	Values
Lead time (in 8-hour shifts)	0, 1, 2
Capacity loss C^m_{loss} (in hours)	0, 4, 8, 12
T (in 8-hour shifts)	3
Number of qualifications k	1
Simulated dispatching	Average product mix, Priority
Work-center	Diel, Implant, Lithography, Diffusion
Optimization model	Single-period, Multi-period

3.3 Numerical Results

As the single-period model does not model lead times, to have a fair comparison between both optimization models, we solve the single-period model, get the qualification plan and compute the number of moves out with the multi-period optimization model. This way both approaches have the same base and can be compared. In the rest of this section, the single-period optimization approach is denoted SP, and the multi-period optimization approach is denoted MP. In addition, "base case" refers to the case where a qualification does not require maintenance operation or is not subject to a lead time. We do not present the detail instance by instance due to space constraints.

3.3.1 Capturing Dynamic WIP Quantities and Capacities

Table 3 compares the mean gap(%) $= 100 * \frac{MP-SP}{MP}$, in terms of moves out between the single-period (SP) and multi-period (MP) approaches. Table 3 enables us to assess if the single-period optimization approach is able to capture the dynamic WIP and capacity. Numerical results show that the single-period optimization model can lead to less relevant qualification decisions. The largest mean and maximum gaps are observed for the Implant work-center when recipe priorities are considered. Even without any lead time or capacity loss, the mean gap is of 1.50%. For the Implant work-center, this indicates that the single-period optimization model does not always capture dynamic WIP quantities. It proposes to qualify a recipe with higher demand on average whereas higher gains can be achieved by focusing on recipes with high peaks of demand on certain shifts. For the Diel work-center, the mean gap is of 0.42%. For other work-centers, mean gaps are closer. Nevertheless, maximum gaps are always greater than 0.6%.

When there are capacity losses and recipe priorities are considered, mean and maximum gaps are significant. The highest mean gap, 2.36%, is observed for the Implant work-center. The highest maximum gap, 5.02%, is also observed for the implant work-center. The maximum gap for the Lithography is about 4.36% when an 8-hour maintenance operation is required. The maximum gap for the Diel work-center is equal to 1.62%. The maximum gap for the Diffusion is equal to 1.39%. Overall, mean gaps are always greater than 0.46%. Mean and maximum gaps are smaller when qualifications are subject to lead time than when they require maintenance operations. This can be surprising because the single-period optimization model does not consider lead times. However, this can be explained by the fact that lead times do not interrupt production contrary to maintenance operations.

Table 3: Mean and maximum gaps (%) $= 100 * \frac{MP-SP}{MP}$, in terms of moves out between the single-period (SP) and the multi-period (MP) optimization models.

				Lead time (in shifts)				Capacity loss (in hours)					
		Base case		1		2		4		8		12	
Moves	Work-center	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Priority	Diel	0.42	1.58	0.23	0.76	0.16	0.67	0.66	1.95	0.82	1.78	0.63	1.62
	Implant	1.50	4.65	0.77	1.97	0.79	2.38	2.32	4.96	2.29	5.58	2.36	5.02
	Lithography	0.16	0.60	0.24	2.32	0.31	1.70	0.46	2.15	0.72	4.36	0.51	4.30
	Diffusion	0.23	0.68	0.08	0.41	0.11	0.27	0.53	1.67	0.46	1.35	0.48	1.39
Average product mix	Diel	0.00	0.01	0.05	0.37	0.09	0.40	0.01	0.09	0.01	0.19	0.01	0.15
	Implant	0.05	0.65	0.15	1.22	0.20	1.43	0.10	0.82	0.10	0.29	0.05	0.42
	Lithography	0.01	0.15	0.01	0.15	0.06	0.56	0.00	0.00	0.00	0.00	0.01	0.08
	Diffusion	0.00	0.04	0.02	0.22	0.01	0.10	0.00	0.04	0.00	0.18	0.00	0.06

In addition, we can observe that gaps are also smaller when an average product mix is considered. This can be due to the fact that the backlogged product mix is less variable contrary to when recipe priorities are considered. Table 4 reinforces this idea. When considering an average product mix, both optimization models propose much more frequently the same qualification than when recipe priorities are considered. For instance, for an 8-hour capacity loss, both optimization models propose eight times the same qualification plan when the average product mix is considered, and only twice when recipe priorities are considered. Recipe priorities are then a source of production variability for qualification management but must be considered.

Table 3 also shows that mean gaps are often far from maximum gaps. For example, for the lithography work-center, when there is a 4-hour capacity loss, the mean gap is equal to 0.46% whereas the mean gap is equal to 2.15%. This is something that we can observe for all work-centers, in particular when maintenance operations are required. Moreover, as on a non-negligible amount of instances, both optimization models propose the same qualification (see Table 4), this indicates that there exists, even among the same work-center, a large disparity between instances. There are instances where the gap between both optimization models is equal to zero or very small whereas other instances where the gap is very large.

Table 4: Number of identical qualification plans (out of 15) recommended by both optimization models.

			Lead time (in shifts)		Capacity loss (in hours)		
Moves	Work-center	Base case	1	2	4	8	12
Priority	Diel	7	4	2	3	2	0
	Implant	4	3	1	1	2	0
	Lithography	5	11	7	7	10	12
	Diffusion	8	8	3	5	2	1
Average product mix	Diel	10	8	6	10	7	8
	Implant	12	10	8	11	8	8
	Lithography	14	13	9	11	11	11
	Diffusion	11	10	8	11	8	9

Finally, Table 5 shows the mean gain (%) in number of moves out after performing a qualification. As mean gaps between both optimization models can be large (Table 3), in general, the multi-period optimization model more frequently proposes qualifications that capture dynamic WIP quantities. When recipe priorities are considered, the mean gain can be quite different between both optimization models. For instance, in the Implant work-center, when there is a 4-hour capacity loss, the single-period optimization model proposes a qualification plan that leads to a diminution of moves out by -1.60%. Instead, the multi-period optimization model proposes a qualification that leads to an augmentation of moves out by 0.74%! This situation is observed for most work-centers when there is capacity loss. There is only in the Diffusion work-center where the single-period optimization model with a 12-hour maintenance operation does not induce a negative mean gain. However, the mean gain is equal to 0.07%, which is very small, compared to the mean gain of the multi period optimization model that is equal to 0.56%. We also observe that for a 12-hour maintenance operation, the mean gain of the multi-period optimization model is negative for the Implant work-center. However, the mean gain in almost ten times worse with the single-period optimization model (-0.29% versus -2.65%). Overall, the multi-period optimization model proposes qualification plans that achieve better mean gain than the single-period optimization model. This means that the single-period optimization model can propose *wrong* qualification decisions. When qualifications are only subject to lead times, mean gains are closer. However, they remain significant for the Implant work-center with a difference greater than 0.7%. When an average product mix is considered, mean gains are *very* close.

Table 5: Comparison of the mean gain(%) in number of moves out after performing a qualification between the single-period (SP) and multi-period (MP) optimization models. Bold values are negative mean gain.

				Lead time (in shifts)				Capacity loss (in hours)					
		Base case		1		2		4		8		12	
Moves	Work-center	MP	SP	MP	SP	MP	SP	MP	SP	MP	SP	MP	SP
Priority	Diel	1.53	1.11	0.78	0.54	0.45	0.29	1.08	0.42	0.49	**-0.34**	0.39	**-0.24**
	Implant	2.04	0.51	1.30	0.52	1.14	0.33	0.74	**-1.60**	0.21	**-2.09**	**-0.29**	**-2.65**
	Lithography	0.88	0.72	1.02	0.78	0.89	0.57	0.55	0.09	0.50	**-0.22**	0.42	**-0.09**
	Diffusion	0.79	0.56	0.46	0.38	0.31	0.21	0.68	0.15	0.66	0.20	0.56	0.07
Average product mix	Diel	1.16	1.16	0.93	0.88	0.61	0.52	0.97	0.96	0.76	0.74	0.52	0.50
	Implant	1.57	1.52	1.38	1.23	1.01	0.81	1.32	1.22	0.94	0.91	0.55	0.50
	Lithography	0.63	0.62	0.63	0.62	0.52	0.46	0.61	0.61	0.53	0.53	0.47	0.47
	Diffusion	0.58	0.58	0.55	0.54	0.37	0.36	0.52	0.51	0.42	0.39	0.31	0.31

3.3.2 Influence of/on Dispatching Rules

Table 3 shows that although the capacity loss/lead time increases, the mean gap does not necessarily increases. Table 6 shows that for some instances, performing a qualification with a lead time greater than zero better optimizes the number of moves out than performing a qualification with no lead time. These results seems counter-intuitive. Actually, this effect is due to the way the number of moves are computed, and more generally, how the production system works with dispatching rules. When lots arrive in an work-center, a dispatching engine assigns lots on machines. The dispatching engine is shortsighted. It does not consider lots that arrive one or two shifts later. In addition, it does not necessarily challenge previous dispatching decisions made when a new lot arrives. This means that, if a recipe is qualified on a machine, the dispatching engine will take advantage of the new qualification and assign lots to the machine. Thus, if this qualification decision is taken right now for a recipe with longer process times than those already qualified on the same machine, the number of moves per shift slightly decreases due to the fact the average throughput rate on that machine decreases. The magnitude of this effect varies with WIP variability over time and if priorities are considered. This effect is also observed in Johnzén et al. (2008) where numerical experiments are run to assess the impact of new qualification on cycle time. After qualifying machines, the mean cycle time did not necessarily decrease. A similar explanation is also detailed in Johnzén et al. (2008). Therefore, qualifications influence dispatching rules decisions, and dispatching rules also influence qualification decisions.

3.3.3 Practical Insights

Numerical results highlight the fact that proposing the best qualifications is a complex procedure, and that performing the qualifications at the right time is critical to improve the number of moves out. Qualification decisions are influenced by WIP and capacity variability over time but also by decision maker preferences or dispatching rules and priorities. In addition, numerical experiments show that performing a qualification may lead to uncompensated capacity losses (e.g., due to required maintenance operations). Thus, after performing a qualification, the number of moves out can be lower than in the case where no qualification is performed. Instead of only considering the number of moves out, other indicators might be interesting to assess the quality of a solution by for example prioritizing lots with large priorities. For instance, although the overall number of moves out decreases, if the mean cycle time of priority lots also decreases, then a qualification can be acceptable. Since maximizing the number of moves out is not always the best option, qualification management could also therefore be modeled and solved as a multi-objective problem.

Table 6: Number of instances by work-center where performing a qualification with lead time gives a larger number of moves out than performing qualification with no lead time.

		Lead time (in shifts)			
		1		2	
Moves	Work-center	MP	SP	MP	SP
Priority	Diel	2	6	1	6
	Implant	5	7	5	6
	Lithography	5	6	5	4
	Diffusion	0	4	0	3
Average product mix	Diel	4	2	3	3
	Implant	0	1	1	2
	Lithography	1	1	0	0
	Diffusion	5	4	1	2

Numerical results highlight the fact that the single-period and multi-period optimization models can propose different qualification plans with respect to the demand profile of the recipes. Depending on the demand profiles, a model is more appropriate than the other. In general, in work-centers where lots come by wave, the multi-period optimization model should be more suited because it better captures demand peaks. This model is then useful to identify and fix short-term bottlenecks with cross qualifications. It is also more robust again highly variable demand and capacity profiles.

Numerical results also show that dispatching rules significantly impact the quality of a qualification plan. Therefore, how the lots are scheduled should be considered in qualification management, in particular in operational qualification management.

As we study a high-mix low-volume production facility subject to high production variability, the demand and capacity can be very uncertain. Therefore, it can be preferable to perform a qualification that requires no lead time or maintenance operation and looks sub-optimal, rather than perform an "optimal" qualification with a larger lead time or longer maintenance operation. If all qualification decisions are subject to lead times or maintenance operations, shortest ones should be preferred. In addition, uncertainty can be managed by using a rolling horizon approach (Clark and Clark 2000; Curcio et al. 2018). A qualification plan is determined at the beginning of the first shift of the planning horizon. At the beginning of the next shift, new information is revealed, the optimization model is solved and a new qualification plan is determined.

4 CONCLUSIONS AND PERSPECTIVES

In this paper, a single-period optimization model and a multi-period optimization model are studied to maximize the number of moves out with qualification plans. Dispatching rules are included and simulated in optimization models. The dynamic qualification optimization model is used to better consider lead times and maintenance operations. Numerical experiments on industrial data show the relevance of the dynamic qualification optimization model. In particular, numerical experiments show that the choice of the model can have a significant impact on the qualification plan, and therefore on the mean gain in terms of moves out. The mean gain is particularly affected when recipe priorities are considered and a maintenance operation is required to qualify a recipe on machine. However, the single-period optimization model remains relevant for some instances.

There are directions that are worth being investigated in the future. In this paper, we limit ourselves to $k = 1$. Efficient resolution approaches can be designed to propose qualification plans for large values of k for both optimization models. New methods to simulate the number of moves out can be proposed to be closer to the behavior of the dispatching engine, e.g., by including batching constraints (Rowshannahad and Dauzère-Pérès (2013)). Numerical experiments show that the single-period optimization model often proposes the same qualification plans as the multi-period optimization model on industrial data. It would be interesting to automatically identify when using the single-period optimization model is likely to suggest the same qualification plan. Doing this would save a lot of time when searching for qualifications.

ACKNOWLEDGMENTS

This work has been partly funded by the ANRT (Association Nationale de la Recherche et de la Technologie) through CIFRE funds (PhD number 2017/0750) and a cooperation contract between STMicroelectronics and Mines Saint-Etienne. This project has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737459 (project Productive4.0). This Joint Undertaking receives support from the European Unions Horizon 2020 research and innovation program and Germany, Austria, France, Czech Republic, Netherlands, Belgium, Spain, Greece, Sweden, Italy, Ireland, Poland, Hungary, Portugal, Denmark, Finland, Luxembourg, Norway, Turkey.

REFERENCES

Aubry, A., A. Rossi, M.-L. Espinouse, and M. Jacomino. 2008. "Minimizing Setup Costs for Parallel Multi-purpose Machines under Load-Balancing Constraint". *European Journal of Operational Research* 187(3):1115–1125.

Bazaraa, M. S. 2013. *Nonlinear Programming: Theory and Algorithms*. 3rd ed. New Jersey: Wiley Publishing.

Chang, X., and M. Dong. 2017. "Stochastic Programming for Qualification Management of Parallel Machines in Semiconductor Manufacturing". *Computers & Operations Research* 79(3):49–59.

Clark, A. R., and S. J. Clark. 2000. "Rolling-horizon Lot-sizing when Set-up Times are Sequence-dependent". *International Journal of Production Research* 38(10):2287–2307.

Curcio, E., P. Amorim, Q. Zhang, and B. Almada-Lobo. 2018. "Adaptation and Approximate Strategies for Solving the Lot-sizing and Scheduling Problem Under Multistage Demand Uncertainty". *International Journal of Production Economics* 202:81–96.

Fu, M., R. Askin, J. 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.

Ignizio, J. P. 2010. "The Impact of Operation-to-tool Dedications on Factory Stability". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Ycesan, 2606–2613. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

Johnzén, C., P. Vialletelle, S. Dauzère-Pérès, C. Yugma, and A. Derreumaux. 2008. "Impact of Qualification Management on Scheduling in Semiconductor Manufacturing". In *Proceedings of the 2008 Conference on Winter Simulation*, edited by S. J. Mason, R. R. Hill, L. Mnch, O. Rose, T. Jefferson, and J. W. Fowler, 2059–2066. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kabak, K. E., C. Heavey, V. Corbett, and P. Byrne. 2013. "Impact of Recipe Restrictions on Photolithography Toolsets in an ASIC Fabrication Environment". *IEEE Transactions on Semiconductor Manufacturing* 26(1):53–68.

Kopp, D., and L. Mönch. 2018. "A Hierarchical Approach to Qualification Management in Wafer Fabs". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 3514–3525. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 2610–2620. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kopp, D., L. Mönch, D. Pabst, and M. Stehli. 2018. "Simulation-based Performance Assessment of Tool Requalification Strategies in Wafer Fabs". In *2018 Institute of Electrical and Electronics Engineers, Inc 14th International Conference on Automation Science and Engineering (CASE)*, 638–641. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nils Löhndorf 2016. "Java Interface for the CLP Solver". https://github.com/quantego/clp-java. accessed 10^{th} May.

Lougee-Heimer, R. 2003. "The Common Optimization INterface for Operations Research: Promoting Open-source Software in the Operations Research Community". *IBM Journal of Research and Development* 47(1):57–66.

Pianne, A., L. Rivero, S. Dauzère-Pérès, and P. Vialletelle. 2016. "Ideal and Potential Flexibility Measures for Qualification Management in Semiconductor Manufacturing". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 2621–2632. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rowshannahad, M., and S. Dauzère-Pérès. 2013. "Qualification Management with Batch Size Constraint". In *Proceedings of the 2013 Conference on Winter Simulation)*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, , and M. E. Kuhl, 3707–3718. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rowshannahad, M., S. Dauzère-Pérès, and B. Cassini. 2014. "Qualification Management to Reduce Workload Variability in Semiconductor Manufacturing". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2434–2443. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rowshannahad, M., S. Dauzère-Pérès, and B. Cassini. 2015. "Capacitated Qualification Management in Semiconductor Manufacturing". *Omega* 54:50–59.

AUTHOR BIOGRAPHIES

ANTOINE PERRAUDAT is a Ph.D. student at the Center of Microelectronics in Provence (CMP) of Mines Saint-Etienne in France and an engineer at ST Microelectronics Crolles. He studied Industrial Engineering and Computer Science at CMP where he graduated in 2017. His email address is antoine.perraudat@emse.fr.

STEPHANE DAUZERE-PERES is 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 75 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.

PHILIPPE VIALLETELLE is a Senior Member of Technical Staff at STMicroelectronics Crolles, France. He received an Engineering degree in Physics from the INSA in 1989. He is responsible for the definition of advanced methodologies and tools for the Crolles 300mm production line of STMicroelectronics. Expert in Manufacturing Sciences for Process and Production Control items, he is in charge of driving collaborative projects at European level. His email address is philippe.vialletelle@st.com.