SCHEDULING MAINTENANCE TASKS WITH TIME-DEPENDENT SYNCHRONIZATION CONSTRAINTS BY A CP MODELING APPROACH

Jan Lange
Gerald Weigert

Andreas Klemmt
Peter Doherr

Electronics Packaging Laboratory
Technische Universität Dresden
D-01062 Dresden, GERMANY

Infineon Technologies
Königsbrücker Str. 180
D-01099 Dresden, GERMANY

ABSTRACT

Ensuring a high uptime for manufacturing machines is crucial for efficient and cost-effective production. In opposite, preventive maintenance tasks (PMs) are necessary to assure the reliability of manufacturing processes. This also prevents serious and unpredictable machine crashes, which affect uptime and process scheduling. However, PM tasks themselves lower the uptime and are to be smartly scheduled within their given domain considering some requirements as ensuring the availability of a sufficient number of qualified engineers for the concerning period. This work investigates a PM scheduling problem with time-dependent synchronization constraints for a lithography work center. For this, a constraint programming (CP) modeling approach including decomposition is used. Multiple objectives are considered. For example, the minimization of crew backup violations or scheduling PMs according to the work in process (WIP) for embedding upcoming PMs smoothly into the system work load. This minimizes negative effects on throughput and tardiness.

1 INTRODUCTION

In the modern semiconductor manufacturing industry, the control of complex manufacturing processes come along with time and pricing pressure. Under this conditions the question arises, how to schedule preventive maintenance tasks (PM tasks) best. Any down-time of producing machines, whether planned or unplanned, reduces production capacity and may have negative effects on tardiness and throughput. On the other side PMs are necessary to achieve a reliable and stable uptime of the machines.

This paper presents an approach based on constraint programming for scheduling a number of preventive maintenance tasks for a group of machines in given domains. Each maintenance task needs a crew of engineers covering multiple competences. Furthermore, there are additional constraints and objectives have to be fulfilled. Considered objectives are oriented on the main semiconductor issues as reducing down-time and smart integration into the projected work in process. Objectives are considered separately but also combined for different data sizes. To improve the solution process for big data size a decomposition approach will be used.

The underlying machine group is a lithography work center, which is generally the bottleneck of a semiconductor facility. This means that the scheduling of PM especially in this area has an high influencing potential to the whole manufacturing site. Consequently, factors as WIP-optimality are of special concern and interest.

The investigated problem is generally a rostering or personnel scheduling problem and according to Baker (1976) a tour scheduling problem. A comprising literature review for personnel scheduling is presented in Van den Bergh et al. (2013). Other reviews can also be found in Ernst (2004) and Alfares (2004), where the literature published since 1990 is classified for the tour scheduling problem. Abdennadher and Schlenker (1999) proposed a constraint logic programming approach for a nurse scheduling problem that involves three shifts per day. He, F. and Qu, R. (2012) present a hybrid constraint programming based column generation approach to nurse rostering problems. In Laporte and Pesant (2004) a CP-based approach is presented for a manpower planning scheduling problem with multi-shift rotating schedules. A crew composition problem is considered in Li and Womer (2009). Another publication about maintenance service and staff planning is presented by Lilly et al. (2007), which addresses a long term capacity planning problem for evaluating the cost-effectiveness of a 4 day work schedule for power-generating stations.

This paper deals with a solution for a synchronization problem of basically PMs on machines and PM engineering crews by means of constraint programming. The crew composition includes multiple and partly combined personal qualification types and fixed individual personal schedules. Multiple objectives are considered separately but also combined. Furthermore, a basic decomposition approach is applied.

2 PROBLEM DESCRIPTION

Figure 1 shows the general problem, where PM tasks are synchronized with availabilities and qualification of engineers and further constraints. The problem has distinct personal shifts with fixed shift lengths and fixed shift start times. An engineer can take part in multiple PM tasks per shift, but only at one task simultaneously. The provided result of the detailed model is a PM tasks schedule and engineer utilization schedule with a 4 hour granularity. This can be used for short term re-planning in case of additional unscheduled PM tasks or absent engineers.
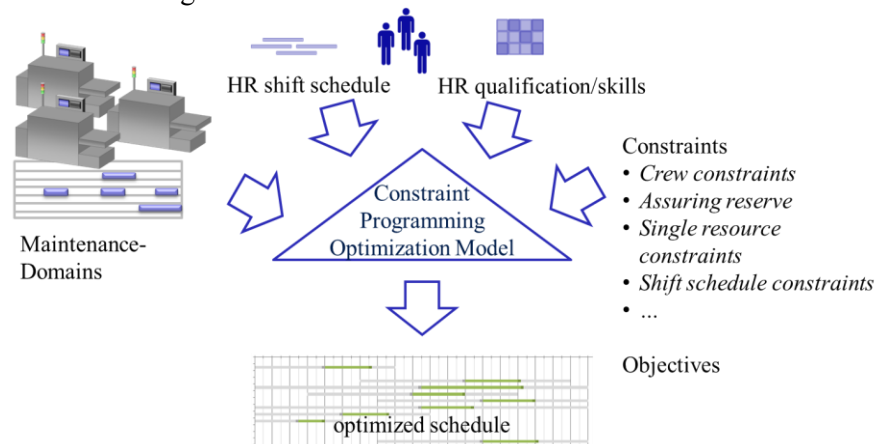

Figure 1: General Problem Description

There a two types of PM tasks, a short and a long variant. The duration of a short PM is planned with 4 hours, whereas the duration of a long PM task is about 8 hours. There are three types of engineer qualifications: A, B and C. An engineer can have up to all three of these qualifications. Furthermore, engineers have individual shift schedules. An engineer shift lasts 12 hours. Consequently there are 2 shifts per day, this means a day and a night shift. Each PM task requires a simultaneously working crew of 2 engineers of type A and 2 engineers of type B. Furthermore, if one or more tasks are performed, one engineer of type C is required for this shift. Additionally, a complete crew is required as back-up crew, which is available in a case of unscheduled downs. The planning horizon is about a month.

3   CP MODEL

The software used for modeling was the IBM ILOG CPLEX Optimization Studio 12.4. The evaluation model comprises a basic OPL model, which contains the whole CP model, and a Java implementation responsible for starting and controlling the OPL model. A general mathematical description is utilized in this paper to explain the principle of the model.

A number of p PM tasks $P_l (l = 1,...p)$ is to be scheduled within their domains $D_l (l = 1,...p)$. Each PM task is assigned to one of m machines $M_k (k = 1,...m)$ of the work center. There are t timeslots $T_o (o = 1,...t)$. Thereby, the width of a time slot is about 4 hours. The duration of a PM task is $P_l^{Duration} \in \{1,2\}$. There are e engineers $E_n (n = 1,...e)$, which have individual fixed shift schedules $S_{n,o} (n = 1,...e ; o = 1,...t) \in \{0,1\}$ and individual sets of qualification $Q_n (n = 1,...e)$. Thereby, $Q_n \subseteq Q$ where $Q = \{A,B,C\}$ and A, B and C are different qualification types. Finally, a set of decision variables $f(l,n,o) \in \{0,1\}$ describe that a PM task P_l is processed by an engineer E_n at time slot o.

3.1   Hard Constraints

The model contains the following constraints, which are to be satisfied:

PM task domain (C1)

PM tasks are to be scheduled inside the bounds of their domains. In the following, D_l^{start} and D_l^{end} are the start and the end of a PM task domain D_l.

$$\forall l,n,o: \quad \left(\left(o < D_l^{start} \right) \vee \left(o > D_l^{end} \right) \right) \rightarrow \left(f(l,n,o) = 0 \right) \tag{1}$$

Engineers shift schedule (C2)
Engineers can perform tasks only in their individual shift schedule.

$$\forall l,n,o: \quad \left(f(l,n,o) = 1 \right) \rightarrow \left(S_{n,o} = 1 \right) \tag{2}$$

Engineers are single resources (*C3*)
Engineers can perform not more than a single task simultaneously.

$$\forall n,o: \sum_{l=1}^{p} f(l,n,o) \leq 1 \tag{3}$$

Machines are single resources (C4)

Only one PM-task can be performed on a machine simultaneously. The function $g(l,k,o) \in \{0,1\}$ describes that a PM task P_l on machine M_k is active at time (slot) T_o.

$$\forall k,o: \sum_{l=1}^{p} g(l,k,o) \leq 1 \tag{4}$$

Crew constraint (C5)

Formulas 5.1, 5.2 and 5.3 describe that a preventive maintenance task requires two exclusive engineers with qualification A and two exclusive engineers with qualification B for the whole duration of the PM task. Additionally, a crew cannot change within performing a PM task. With $n_A = \{n \mid A \in Q_n\}$ and $n_B = \{n \mid B \in Q_n\}$:

$$\forall l,o: \qquad \sum_{n=n_A} f(l,n,o) \geq 2 \cdot \left(\sum_{n=1}^{e} f(l,n,o) \geq 1 \right) \qquad (5.1)$$

$$\forall l,o: \qquad \sum_{n=n_B} f(l,n,o) \geq 2 \cdot \left(\sum_{n=1}^{e} f(l,n,o) \geq 1 \right) \qquad (5.2)$$

$$\forall l,o: \sum_{n=n_A \cup n_B} f(l,n,o) = 4 \cdot \left(\sum_{n=1}^{e} f(l,n,o) \geq 1 \right) \qquad (5.3)$$

3.2   Soft Constraints

Reserve available soft constraint (SC1)
For every time slot of the planning horizon there should be an available reserve or back-up team to cover incidents. This team is a full maintenance team comprising two engineers with qualification type *A*, two engineers with qualification type *B* and one engineer with qualification type *C*. The team has to be additionally available also in the time of scheduled PM tasks. This soft constraint is also modeled in OPL, but is not shown as a mathematical definition here. $SC1_o$ attains a value out of $\{0,1\}$ and shows if *SC1* is valid at time o.

3.3   Objectives Modeling

3.3.1 Minimizing Parallel PM Tasks

One of the basic objectives is the minimization of simultaneously performed PM tasks to assure low workforce utilization. The principle of the objective is illustrated in Figure 2.
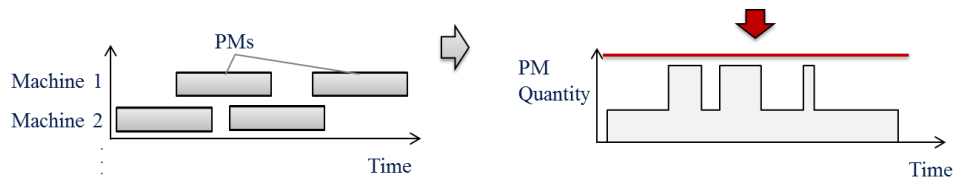

Figure 2: Minimizing simultaneously performed PM tasks

For this, a function is defined over the whole time of the planning horizon cumulating the number of active maintenance tasks at time *o* (7). Consequently, the upper bound of this function is minimized (8) in order that the number of active maintenance tasks does not exceed the maximum $simPM_{MAX}$ for the whole planning horizon.

$$simPM_O = \sum_{l=1}^{p} \left(\sum_{n=1}^{e} f(l,n,o) \geq 1 \right) \qquad (7)$$

$$\max_{o=1}^{t} (simPM_O) = simPM_{\max} \to \min \qquad \textbf{(8)}$$

3.3.2 Minimize Back-Up Violation

For the whole planning horizon there should be a reserve team for emergencies. However, this constraint may not be satisfied due to the variation of engineers' presences and unpredicted influences in reality.

This objective represents a soft constraint with the aim of maximizing the number of time slots where the reserve condition is fulfilled over the planning horizon (9).

$$ReserveFit = \sum_{o=1}^{t} SC1_o \to \max \tag{9}$$

3.3.3 Minimize Utilization of Engineers with Multiple Qualifications

As introduced before, engineers have qualifications. Furthermore, engineers can also have multiple qualifications. As multi-qualified engineers are more flexible than single-qualified engineers, the utilization of those multi-qualified engineers is to be minimized (10). The value of $|Q_n|$ describes the number of qualification types of an engineer E_n.

$$CrossTrained = \sum_{l=1}^{p}\sum_{n=1}^{e}\sum_{o=1}^{t}\left(|Q_n| \cdot f\left(l,n,o\right)\right) \to \min \tag{10}$$

3.3.4 Best fit to Work in Process

The work in process (WIP) is implemented by a function $W\left(o\right)$ over the planning horizon. The aim is to schedule the maintenance tasks at the minimums of this function. The simplest implementation would just use starting dates of the tasks as attribute for the function. However, this may lead to incorrect placement, especially for highly dynamic WIP-cost functions. Thus, the chosen approach evaluates the WIP-cost for the whole duration of each of the performed maintenance task. The principle is illustrated in Figure 3. The WIP cost values are aggregated for all PM tasks with the aim to be minimized, as shown in (11). Thereby, P_l^{start} is the scheduled start and P_l^{end} is the scheduled end of a PM task.
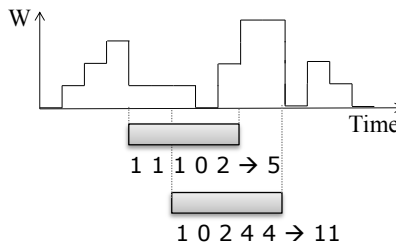

Figure 3: Placing PM tasks s according to WIP-cost function

$$WIPCost = \sum_{l=1}^{p}\sum_{o=P_l^{start}}^{P_l^{end}} W\left(o\right) \to \min \tag{11}$$

3.3.5 Maximizing the Ideal Delay of Maintenances

Every PM task has a predefined domain in which this task is to be performed. Delaying those maintenances inside their domains means also delaying future maintenances of concerning machines. However, to avoid domain violations through uncertain influences in reality, maintenances should not be performed at the very end of their domains. Consequently, maintenance tasks are to be performed preferably between the third and fourth quarter of their domains. In the model this is implemented by triangular functions defined over each domain of maintenance task. This is shown in Figure 4. The function $V_l\left(o\right)$ is the

triangular function for a PM task *l* and is described in equation (12). For every PM task, its start date is used as input for its triangular function multiplied. Thus the sum of these functions is to be maximized (13).
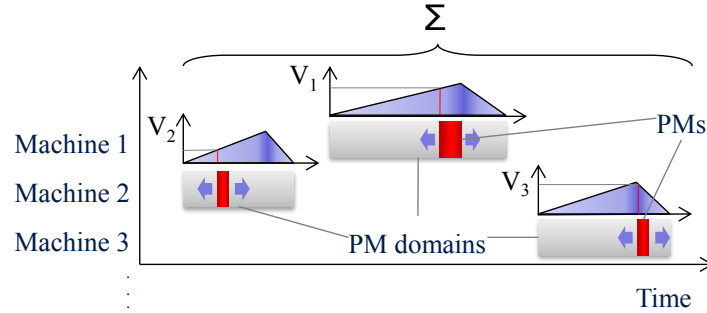

Figure 4: IdealDelay function over each PM task domain

$$V_l(o) = \begin{cases} \dfrac{4}{3} \cdot \dfrac{o - D_l^{start}}{D_l^{end} - D_l^{start}} & D_l^{start} \le o \le \dfrac{3}{4} D_l^{end} \\[2ex] 1 - \left(4 \cdot \dfrac{o - D_l^{start}}{D_l^{end} - D_l^{start}} - 3 \right) & \dfrac{3}{4} D_l^{start} < o \le D_l^{end} \\[2ex] 0 & o < D_l^{start}, o > D_l^{end} \end{cases} \qquad (12)$$

$$IdealDelay = \sum_{l=1}^{p} V_l(P_l^{Start}) \to \max \qquad (13)$$

4   EXPERIMENTAL RESULTS

This section provides model results for two input data sets. After showing the model behavior for separate objectives, the performance for combined objectives is investigated.

4.1   Data Set Description

Two data sets, DS1 and DS2, were defined for calculation. Both are generated based on the complexity of a real environment.

The model comprises more maintenance tasks on about 40 machines. All in all, the number of engineers for performing the tasks is about 50. The model covers a planning horizon of 32 days with a granularity of 4 hours. In the data set DS1, there are 46 PM tasks to be scheduled. Data set DS2 has a duplicated PM data set, that means a total number of 92 PM tasks. However this is performed with the same configuration of engineers and machines.

4.2   Objectives Behavior

The model performance for separately investigated objectives is shown in Figures 5 and 6 for the data sets DS1 and DS2. The figures show the normalized results of all objectives, which were introduced in section 3. For this, the normalization bases on the first available value within the time limit raster. The values of the horizontal axis show the maximum calculation time, this means the calculation abort time limit.

For calculation, a system with a 2.53 GHz Intel Core 2 Duo (P9500) CPU was used. Thereby, calculations were restricted to use only a single core. With this given computational environment, first results are available by a time of a few seconds. The results for the smaller data set DS1 at Figure 5, show that the objectives *SimPM*, *ReserveFit* and *CrossTrained* are easy to solve, since these are already optimal at very low calculation times. *WipCost* and *IdealDelay* are more complex. By 180 seconds *IdealDelay* is close to its optimum. A similar behavior is to see for the objective *WipCost*.
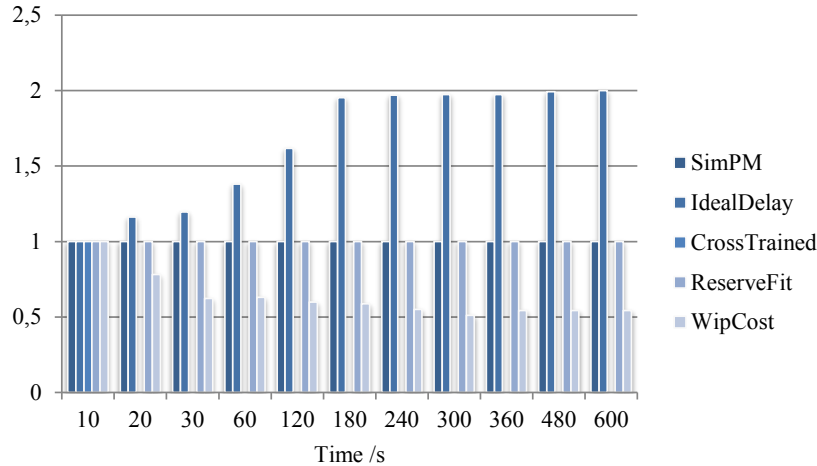

Figure 5: Separate objectives for data set DS1 (normalized)

The results for the bigger data set DS2 are shown in Figure 6. Again, the objective *SimPM* is easy to solve. For the objective *CrossTrained,* the time needed to get an optimal solution has risen to more than 60 seconds, compared to about 10 to 20 seconds for the DS1 scenario. The same is to say about *IdealDelay* and *WipCost*, but also *ReserveFit* which reaches optimality late.
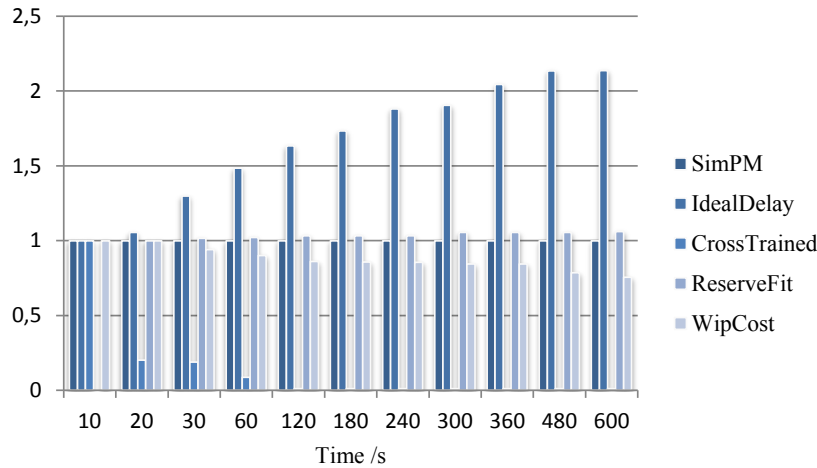

Figure 6: Separate objectives for data set DS2 (normalized)

4.3   Combined Objectives Behavior

For the purpose of using the model for real-world applications, the model behavior is considered for a combined objective. The tested objective is a combination of the objectives *IdealDelay*, *ReserveFit* and *WipCost*. Function (14) describes the combined objective. As ranges are of similar size for all objectives of the combined objective, the weights ω_I, ω_R and ω_W are set to 1.

$$CombinedObjective = (\omega_I \cdot IdealDelay + \omega_R \cdot ReserveFit - \omega_W \cdot WipCost) \to \max \qquad (14)$$

Figure 7 shows the achieved values for the combined objective for both data sets DS1 and DS2. The results show, that the model provides good results in a reasonable time. So, for the smaller data set DS1 the values do not improve significantly after 3 to 4 minutes, which is an adequate time for real world application.
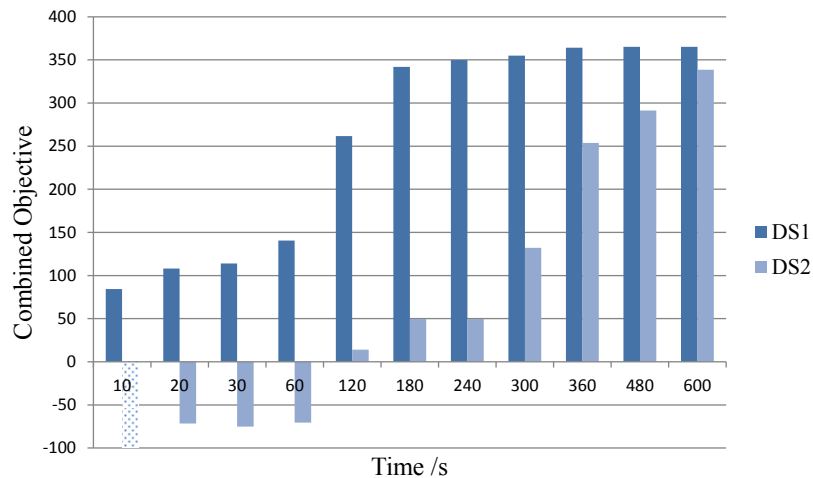

Figure 7: CombinedObjective Results

For DS2, results are available later than for DS1. There is no solution available for 10 seconds. The result curve has a high slope for the whole period of the observed calculation time. So, good results may be achieved after 10 minutes. This higher complexity of the problem may be tackled by a decomposition approach as shown in the following section.

5   DECOMPOSITION APPROACH

Decomposition approaches are widely used to handle complex problems. Qu and He (2009), for example, decompose a problem into weekly sub-problems for a hybrid constraint programming approach for nurse rostering problems. A decomposition approach for a personal scheduling problem at a call center is presented by Canon (2007). Also, Li and Womer (2009) present a three-phase hybrid approach for a shipboard man power scheduling problem utilizing constraint programming and tabu search. Klemmt et al. (2010) present a multistage mixed integer programming based decomposition approach for optimizing manufacturing flows.

Figure 8 schematically shows the principle of the applied decomposition approach. Thereby, the complete group of PM tasks is divided into smaller groups, that may be calculated more easily. For this, the following procedure is passed:

1. Tasks are ordered by 1st level: domain start and 2nd level: domain end
2. PM tasks are grouped sequentially (number of tasks in a group is step size)
3. First PM group is optimized
4. Results of first PM group are fixed
5. Next PM group is calculated including fixed previous PM groups
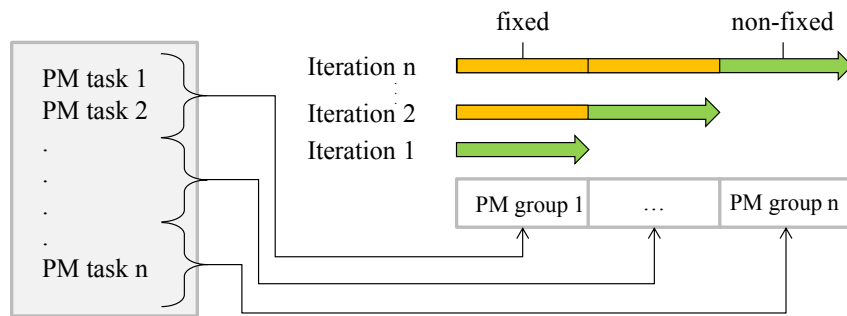6. Repeat 5. Until there is no remaining PM group

Figure 8: Decomposition approach

With this approach the size of the model increases step by step to the size of the whole model. However, since PM domains are fixed for previous PM groups, the complexity for each step is relatively low. An issue is that PM tasks of different groups can overlap in their domains. So, fixed PM tasks of a preceding PM group may directly influence PM tasks of the following groups, which may result in an unsolvable problem for this following groups. This can be prevented by an overlapping window of PM groups. This means that for the calculation of a PM task group, an additional number of succeeding PM tasks is included. However, only the tasks of the concerning PM group are fixed. The upper bound of this overlapping window is problem-specific and can be defined as the maximum number of parallel PM task domains of the whole problem. This was also verified by the results.

5.1 Step Size Optimization

Different step sizes can be implemented for decomposition. The approach is expected to perform differently for different step sizes. For this, the optimization to the objective *IdealDelay* (maximization) was used to compare different step sizes of the decomposition approach to the approach without any decomposition. The two data sets DS1 and DS2 were used for this.

Figure 9 shows the results for the data set DS1. Obviously, up to a calculation duration of about 180 seconds the decomposition provides better results compared to the non-decomposition approach for most step sizes. After this time there is no benefit from decomposition, which may result from its reduced solution space. Consequently, provided a calculation time of more than 180 seconds, the problem is small enough to achieve good results without applying decomposition.

For the bigger data set DS2, Figure 10 shows a comparison between the results achieved for the objective ideal delay, without decomposition in contrast to decomposition approaches with different step sizes. Because of the chosen implementation of the decomposition approach, there are no results for this at small step sizes with short overall time limits. The approach without decomposition provides results for all observed time limits. However, especially after the calculation time limit of 60 seconds, the decomposition approach performs better for this scenario. After this point, first small step size are best performing. By increasing the calculation time limit, bigger step size has continuously better results. So, the difference between the decomposition approach and the approach without decomposition gets smaller.
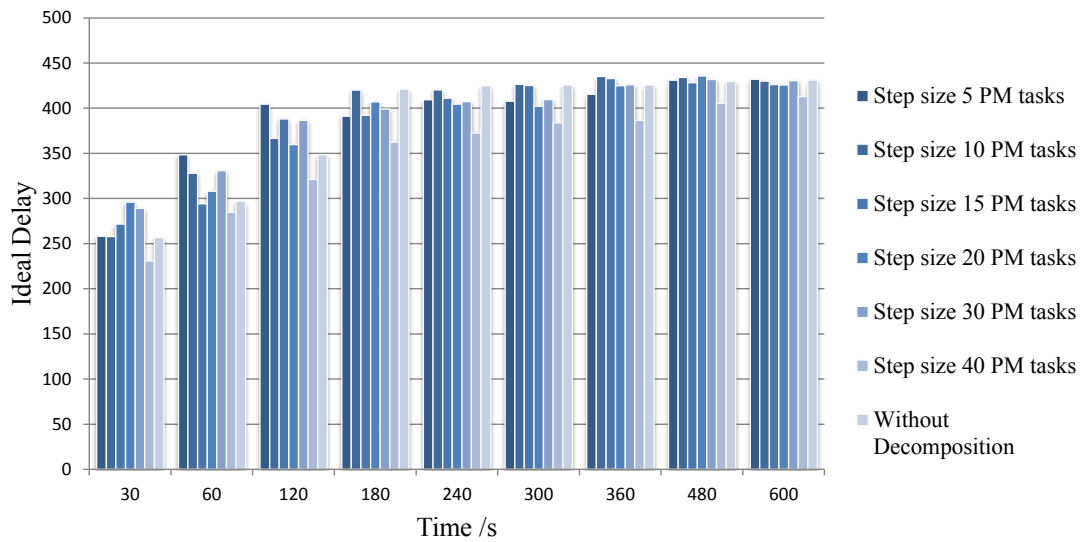
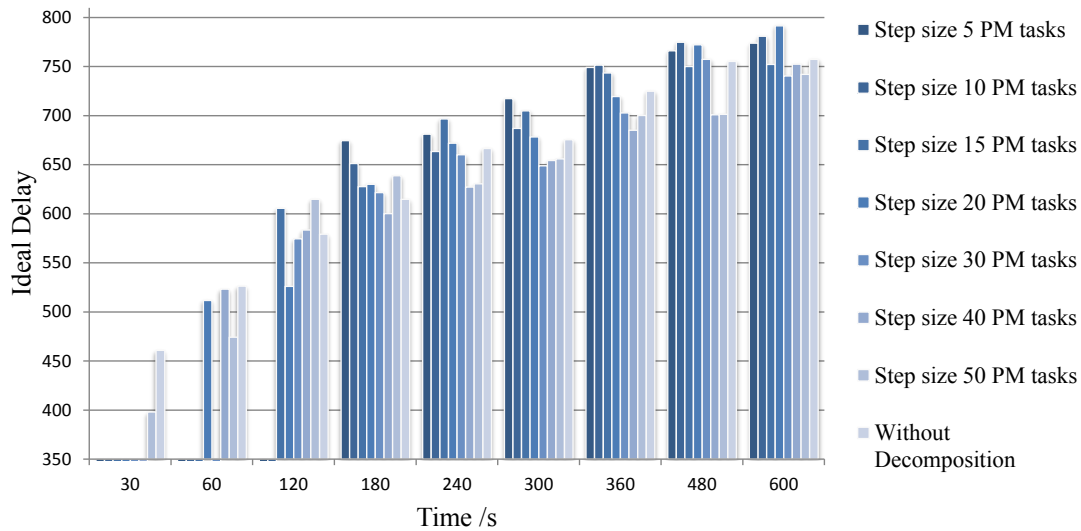Figure 9: Results for ideal delay with decomposition, DS1

Figure 10: Results for ideal delay with decomposition, DS2

6   CONCLUSION

This paper presented a CP-based approach for scheduling PMs for a lithography work center. The problem contains hard and soft constraints and has distinct personal shifts with fixed shift lengths and fixed shift start times.

Multiple objectives were considered which optimize toward the task force utilization as well as the positioning of the PM tasks within their domains. Thereby, the positioning is performed according to an ideal delay for the reduction of PMs over time and optimal according to the expected WIP for the work center. Also, a combination of these objectives was considered.

For most of the objectives, the model provides good results in a reasonable time. However, for higher problem complexities, a decomposition approach is advisable to achieve good results within a restricted calculation time.

Tests with different step sizes showed, that there is an ideal step size, which depends on the available time for calculation (calculation time limit) and the problem complexity. Results also show that, for an unrestricted calculation time, achieved values are better for the approach without decomposition.

Further research may include the individual skills of engineers. So, additionally to their qualification, engineers could also have an indicator showing skills for available qualifications, which could influence the speed or quality of maintenance tasks. Another concern may be in the field of social scheduling. This means teams could be created under consideration of interhuman relationships in order to group well attuned crews.

7 ACKNOWLEDGMENTS

REFERENCES

Abdennadher, S., Schlenker, H. (1999). Nurse scheduling using constraint logic programming. In Proceedings of the National Conference on Artificial Intelligence, 838-843.

Alfares, H. K. (2004). Survey, categorization, and comparison of recent tour scheduling literature. Annals of Operations Research, 127(1-4), 145-175.

Baker, K. R. (1976). Workforce allocation in cyclical scheduling problems: A survey. Operational Research Quarterly, 155-167.

Canon, C. (2007). Personnel scheduling in the call center industry. 4OR: A Quarterly Journal of Operations Research, 5(1), 89-92.

Ernst, A. T., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. European journal of operational research, 153(1), 3-27.

Klemmt, A., Lange, J., Weigert, G., Lehmann, F., & Seyfert, J. (2010, December). A multistage mathematical programming based scheduling approach for the photolithography area in semiconductor manufacturing. In Proceedings of the Winter Simulation Conference (pp. 2474-2485). Winter Simulation Conference.

Laporte, G., Pesant, G. (2004). A general multi-shift scheduling system. Journal of the Operational Research Society, 55(11), 1208-1217.

Li, H., Womer, K. (2009). A decomposition approach for shipboard manpower scheduling. Military Operations Research, 14(3), 67-90.

Lilly, M. T., Emovon, I., Ogaji, S. O. T., & Probert, S. D. (2007). Four-day service-staff work-week in order to complete maintenance operations more effectively in a Nigerian power-generating station. Applied energy, 84(10), 1044-1055.

He, F., Qu, R. (2012). A constraint programming based column generation approach to nurse rostering problems

Qu, R., He, F. (2009). A hybrid constraint programming approach for nurse rostering problems. Applications and Innovations in Intelligent Systems XVI, 211-224.

Van den Bergh, Jorne & Beliën, Jeroen & De Bruecker, Philippe & Demeulemeester, Erik & De Boeck, Liesje, 2013. "Personnel scheduling: A literature review," European Journal of Operational Research, Elsevier, vol. 226(3), 367-385.

AUTHOR BIOGRAPHIES

JAN LANGE received his master's degree in Information Systems Technology at the Dresden University of Technology in 2008. He is now a Research Assistant at the Electronics Packaging Laboratory of the Dresden University of Technology and works in the field of production control, simulation and optimization of manufacturing processes. His email is Jan.Lange@tu-dresden.de.

GERALD WEIGERT is an Assistant Professor at Electronics Packaging Laboratory of the Dresden University of Technology. Dr. Weigert works on the field of production control, simulation & optimization of manufacturing processes, especially in electronics and semiconductor industry. He was involved in development of simulation systems as well as in their application in industrial projects for scheduling. His email is Gerald.Weigert@tu-dresden.de .

ANDREAS KLEMMT received his master's degree in mathematics in 2005 and Ph.D. in 2011 at the Dresden University of Technology. He is employed as staff engineer in the operations research and engineering group of Infineon. His current research interests are capacity planning, production control, simulation & optimization. His email is Andreas.Klemmt@infineon.com .

PETER DOHERR received his master's degree in electrical engineering (1996) at the University of Technology Leipzig. He works as engineering expert at Infineon Dresden in the field of maintenance strategies, stability and planning. His email is peter.doherr@infineon.com.