Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

SIMULATION-BASED SOLUTION OF LOAD-BALANCING PROBLEMS IN THE PHOTOLITHOGRAPHY AREA OF A SEMICONDUCTOR WAFER FABRICATION FACILITY

Lars Mönch

Institute of Information Systems
Technical University of Ilmenau
Helmholtzplatz 3, P.O. BOX 100565
D- 98684 Ilmenau, GERMANY

Matthias Prause
Volker Schmalfuss

X-FAB Semiconductor Foundries AG
Haarbergstraße 67
D-99097 Erfurt, GERMANY

ABSTRACT

In this paper we present the results of a simulation study for the solution of load-balancing problems in a semiconductor wafer fabrication facility. In the bottleneck area of photolithography the steppers form several different subgroups. These subgroups differ, for example, in the size of the masks that have to be used for processing lots on the steppers of a single group. During lot release it is necessary to distribute the lots over the different stepper groups in such a way that global targets like cycle time minimization, the maximization of the number of finished lots and due date performance are inside a certain range. We present a simulation model of a wafer fab that models the photolithography area in a detailed manner. By means of this simulation model it is possible to decide at release time on which stepper subgroup processing of the lots of a certain product is favorable.

1 INTRODUCTION

Discrete event simulation is an important tool for the analysis of complex manufacturing systems. The manufacturing of integrated circuits (IC) on silicon wafers is a complex production process. Between 250-500 process steps on 50-120 different types of equipment are required to produce a medium complexity circuit. A single product moves through the wafer fabrication facility (fab) in lots. Each lot consists of several wafers (in the case of the example company at maximum 24 wafers). Because of the customer-order orientation of production, there are a large number of lots with significantly fewer wafers than a typical lot (up to 25 percent of the lots).

The system analysis is complicated in the case of semiconductor wafer fabrication facilities due to the occurrence of reentrant process flows. These recursive flows are necessary for a layered structuring of the wafers (in our case between 15 and 25 layers are typical for the products). Another source of difficulties in the example company is the large and permanently changing technology and product variety. We refer to the monograph (Atherton and Atherton 1995) and to the review papers (Uzsoy et al. 1992, 1994, Schömig and Fowler 2000) for a more detailed description of the characteristics of the semiconductor production process from production control point of view.

In the example company, the photolithography area is a problem zone. The problems of the area are mainly caused by the fact that each lot passes through the photolithography area k times because the circuits are made up of layers as mentioned above. Here, we assume that the underlying technology requires a chip to consist of k layers. Furthermore, the purchase of new steppers caused a situation, where it was impossible to use the experience of the members of the production control department. Because of the described system characteristics, mainly the changing product mix, the use of static models is only partially possible. In order to solve the problems of the photolithography area of the fab it was decided to build a simulation model, that allows for a precise description of the dynamics of the underlying production process. During the course of the work, questions like collecting and analyzing data and verification, validation of the simulation model were quite important for the success of our work.

The paper is organized as follows. In the next section we describe the photolithography area of the wafer fab. We formulate the problem under consideration in section 3 in detail. Then we explain the developed simulation model and our methodology. We present the results of our simulation study in section 5.

2 PHOTOLITHOGRAPHY ISSUES

In the photolithography area, the structure of the circuits is mapped from the pattern on the mask to a wafer. The step

& repeat equipment for the exposure step is called a stepper. In the photolithography area, a single wafer is first coated with a thin film of a light-sensitive polymer. Then ultraviolet light is used to expose the wafer. This is done by projecting the light through a mask. For that purpose, a single wafer is first adjusted on the stage. Then the exposure step takes place. A single exposure step is required for each image (group of circuits) on the wafer. The stage is aligned after each single exposure step. During an exposure step the structures of a single circuit are mapped from the mask to the wafer. The ultraviolet light is responsible for changing the molecule structures of the light-sensitive polymer. After all wafers of a single lot are exposed, a development step takes place. Here, the polymer is removed from the non-exposed regions of the wafer. The development step is followed by a manual control step of the single wafers by an operator. Finally, if the wafer passes this control successfully, another control step on a raster electron microscope takes place.

We found the following typical restrictions for the steppers.

- Because masks are quite expensive, for every product p_i exactly one mask set $\{m_{i1}, m_{i2}, \ldots, m_{ik_i}\}$ with a fixed mask size exists, where k_i, the number of masks of the mask set, is less or equal than the number of layers of product p_i. For products with a large number of work in process lots, we find more than one mask set, which usually differ in mask size.
- In order to reduce the number of mask changes on a single stepper caused by consecutively processing lots of different products and layers trains are formed. A train is defined as a set of lots of the same product and same mask layer that are processed in a consecutive manner on the same stepper.
- It is necessary to consider send-ahead wafers for process control. After a certain amount of time (that ranges from one shift to two days) a new send-ahead wafer is necessary for a specific stepper, a specific product and a specific mask layer. Therefore, a first wafer is taken from the lot and then launched. This send-ahead wafer will be exposed, developed and process control steps will take place. If the process control does not fail, the remaining wafers of the lot can be exposed. Otherwise, certain exposure parameters have to be changed and a new send-ahead wafer has to be launched. Send-ahead wafers are stepper specific. Note, that forming large trains reduce the number of required send-ahead wafers because the wafers of the last lot of the train serve as send-ahead wafers for the next train with the same specification.
- In the example company, there are different stepper subgroups. One characteristic for grouping is the size of the masks that can be used on the steppers of a subgroup. Another property used for grouping is the technology of the lots that in principle can run on a certain stepper.

3 STATEMENT OF THE PROBLEM

One goal of the development of the simulation model was to build a tool, that allows for the production control manager to find out the best product dependent load distribution over the steppers of the different subgroups before the release of the lots into the fab. Note that all of the lots, intended for release at time t are actually released into the fab. We are not interested in deciding whether the release of a certain lot makes sense or not. However, we have a certain degree of freedom in choosing the stepper subgroup to process a lot. This decision is valid until the lot moves out of the fab. The type of mask is defined at lot release time, i.e., which mask size is required to process the lots. This decision influences the dispatching of the lots in the photolithography area of the fab because it is basically a decision on the assignment of lots to a stepper subgroup.

Assume that we have m stepper subgroups. We denote these subgroups by $G_1, \ldots, G_m$. Furthermore, we have different lot vectors

$$\bar{L}_i(t) := \left(L_{i1}(t), \ldots, L_{in_i}(t)\right), \quad i = 1, \ldots, n_p, \tag{1}$$

with release dates $t_0, \ldots, t_0 + r_i$. Here, we denote the number of lots of product i that are planned for release at time t by n_i. n_p is the number of products. The number of different lot vectors is $\sum (r_i + 1)$. Suppose that the lots of product i could be processed at the steppers of the subgroups $\{G_{j_1}, \ldots, G_{j_k}\}$ with $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, m\}$. In practice, we found $k = 2$ or $k = 3$ as typical values for the number of possible subgroups of a certain product. We are interested in calculating the entries of the product dependent proportion matrix

$$W_i := \begin{pmatrix} w_{ij_1}(t_0) & \ldots & w_{ij_1}(t_0 + r_i) \\ \vdots & \ddots & \vdots \\ w_{ij_k}(t_0) & \ldots & w_{ij_k}(t_0 + r_i) \end{pmatrix}. \tag{2}$$

Here, we denote by $w_{ij}(t) \in [0,1]$ the portion of lots of product i, that are released at time t that will use the subgroup G_j for processing. For fixed i and t, then, $\sum w_{ij}(t) = 1$.

In our research, we use the following global performance measures.

1. Cycle time per mask layer

$$CTML(p_i) := \frac{CT(p_i)}{NML(p_i)}. \quad (3)$$

Here, we denote by $CT(p_i)$ the average cycle time of the lots of product p_i. The notation $NML(p_i)$ is used for the number of mask layers of p_i.

2. Average tardiness

$$AT := \frac{\sum \max(0, c_i - d_i)}{N_{tardy}}, \quad (4)$$

where c_i denotes the completion time and d_i the due date of lot i. The number of lots completed after their due date is denoted by N_{tardy}.
Furthermore, we are interested in the more local performance measure given next.

3. Average waiting time in front of the steppers

$$AWT := \frac{\sum (e_i - b_i)}{N_{waiting}}. \quad (5)$$

Here, e_i is used as a notation for the finishing time of processing lot i on a stepper. In addition, the time when the lot i enters the queue in front of the steppers is denoted by b_i. In this case, $N_{waiting}$ is the number lots queuing in front of the steppers.

Note, that these performance measures are also used as performance measures for the production control department. We are interested in the development and test of a method, that can obtain better values for both $CTML(p_i)$, AT and AWT with respect to current practice.

4 MODEL AND METHODOLOGY

We built a simulation model of the full fab. For this purpose, we used the simulator AutoMod™/AutoSched™ 9.1/6.1 from AutoSimulations. The ASCII files, that specify the model, were automatically generated from the Manufacturing Execution System (MES) and the Enterprise Resource Planning (ERP) System and other company specific databases (see Mönch and Schmalfuss 2000a).

This facilitates work with models that give a correct picture of the current situation in the fab. For the machines other than the steppers, we used a company specific combined dispatch rule, that is based on the critical ratio (cf. Atherton and Atherton 1995 for a definition of the critical ratio of a lot). This rule includes additional priorities for single lots (for example for lots for prototyping) and for special technologies. As a third component of the rule, the slack is considered with a small weight. After deriving the priorities of the lot in the queue, we used the batch rules of AutoSched to form batches. We model sequence-dependent set-up times for a number of tool groups. For these tool groups we used a combination of set-up avoidance rules and our basic dispatch rule. Based on the weekly production plan of the fab (that considers capacity constraints), the release of the lots in the factory is evenly distributed over the week. More advanced lot release rules (cf. Fowler et al. 2001, Rose 2001) were not taken into account. In the fab under investigation, the photolithography area forms a (planned) bottleneck. Because the bottleneck of the factory has a significant influence on the performance of the whole fab (cf. Atherton and Atherton 1995), we model the stepper tools in a detailed manner.

4.1 Data

One main problem is to correctly determine the processing times of the lots on the steppers. The processing times vary as a result of the different number of wafers in the lots, the different, product and layer dependent, exposure times and because of the different, product dependent, number of exposure steps. We found that the coefficient of variance of the processing times on the steppers is (dependent on the product mix) in most cases greater than one. A special database was developed, in which all product dependent information in connection with photolithography issues is stored (exposure times, number of exposure steps, dedication of steppers). This information is used for the automatic generation of the simulation model. The full processing time of a train is given by the following formula:

$$t_g := t_l + n_w(t_w + n_s(t_{ex} + t_a)) + t_{ul}, \quad (6)$$

where

t_g :	Full processing time of the lots of a train,
t_l :	Time for choosing the recipe, loading the stepper with lots and loading the reticle,
t_{ul} :	Time, required to unloading the last lot of the considered train,
n_w :	Number of wafers in the train,
t_w :	Time, required to load a single wafer and for the alignment of the wafer,
n_s :	Number of exposure steps per wafer,
t_{ex} :	Exposure time for a single exposure step,
t_a :	Time for alignment of the stage after a single exposure step.

The used time models for processing lots on other equipment are described in detail in a recent paper (Mönch and Schmalfuss 2000a). The distribution of machine failures and repair times and the data for preventative maintenance are derived from historical data. The planned due dates of the lots and special lot priorities are taken from the ERP System. The model is initialized by using a work in process (WIP) distribution of the fab. This distribution was obtained from the MES of the fab.

4.2 Modeling of Restrictions

As described in section 2, the masks form one process restriction that have to be considered in the model. Because there is exactly one mask with a fixed mask size for a specific product and layer, it is not possible to process lots of the same product and the same layer on different steppers simultaneously. The possible masks, exposure times and the number of exposure steps are read from an ASCII file at the beginning of a simulation run in order to reduce the effort for making data available. Compared with changing all routing files of the model, it is quite easy to change this file. The information about mask availability is stored in data structures (lists and arrays) that are written in the C programming language. Using the quick sort algorithm to obtain sorted arrays, the access to these data structures via a binary search is quite efficient.

We used another ASCII file, containing information about the time of the last send-ahead wafer that occurs for a fixed product, layer and stepper. We implemented data structures in C that allow a dynamic modification of these times at simulation run time. We did not explicitly consider rework for the exposure steps in our model because the rework rate is small because of launching send-ahead wafers.

4.3 Dispatch Rules for the Stepper Equipment

We did experiments with different dispatch rules for the steppers (cf. Mönch and Schmalfuss 2000b). During the course of our work we were able to verify the used dispatch strategy. Special steppers are used to process lots with a very high priority and a small number of wafers. After processing a lot on a stepper we look in the queue for other lots with these characteristics (not necessarily lots from the same product and mask layer). In this case, we avoid forming trains with a high number of wafers and taking into account the drawback of changing the masks. On the other steppers, we try to choose the lots of the train with the largest number of wafers and at least one lot with a priority greater than a given threshold. If we find trains with the same number of wafers, we choose the lots of the train with the highest average lot priority. During the processing of the lots of a certain train, arriving lots of the same product and mask layer will automatically become a member of this train. The implemented dispatch rules consider the need of send-ahead wafers, otherwise a correct model of the capacity of the bottleneck stepper equipment is not possible. If it is necessary to launch a send-ahead wafer, the rule determines first whether there are steppers that do not require a send-ahead wafer for that train. If the resulting stepper set is not empty, the rule computes the remaining processing time for the trains, which are processed on the steppers of the set. The rule determines the stepper with the smallest remaining processing time. If this time is acceptable for the train in consideration, the train waits until the chosen stepper is free to process the train. Otherwise, it is necessary to launch a send-ahead wafer on the first stepper.

4.4 Using Forecast within the Model

In this study, we investigate how an assignment of certain products to certain stepper subgroups will influence the future behavior of the system. Therefore, to a certain extent, it is necessary to consider lots that will be released into the fab in future time periods. The production plan provides us information about the future lot releases for the next three months. We included these lots into our model. For this purpose, an interface to the ERP System was built. Based on this information, the required process flows for new products were also included into the model.

4.5 Verification and Validation of the Model

The verification and validation process of the model was carried out in an iterative manner. The model was first verified using a number of simplifying assumptions, for which the model's true characteristics were known. After this development stage we used tracing and animation in order to investigate whether the model works as intended.

In a first iteration, we checked the validity of formula (6). To do this, we compared the time for exposure (data from the MES) with the calculated time for certain lots. Then, we compared the real throughput of the steppers (as reported in the MES) with the throughput computed based on formula (6). The observed accuracy was 95%. During collection and analysis of the data from the fab for the simulation model, a number of data errors were detected and adjusted. In a second step, we used reports from the production control department in order to validate the model. The utilization of the steppers in the simulation model as compared with that in the fab shows an accuracy of 90%.

In a second iteration, it was necessary to include the send-ahead wafers into the model in order to improve the accuracy of the model. An experienced person compared cycle times for the lots of certain products observed in the model to those from real fab data from the production control department. We also found that our model was able to detect the dynamic bottlenecks of the real fab.

4.6 Algorithm for Determining the Proportion Matrix

To solve the load-balancing problem described in section 3, we suggest a greedy algorithm. We use an initial solution, i.e., a proportion matrix determined by an experienced person from the production control department. Then, based on this initial solution we obtain a sequence of improved solutions in an iterative manner by making local changes in the neighborhood of the initial solution. We use our simulation model to evaluate a concrete proportion matrix, i.e., a solution of the load-balancing problem. First, we have to introduce an appropriate neighborhood. We consider the case $k = 2$, i.e., two stepper subgroups and $i = 1$, i.e., lots of one product. For discretization of the load-balancing problem we choose the weights $\tilde{w}_{ij}(t) \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$, instead of considering weights $w_{ij}(t) \in [0,1]$. We choose a planning horizon of 16 weeks (for this time period we know approximately the lots, that have to be released into the fab). During this period, we release lots of product $i = 1$ at the beginning of each week into the fab. That means we consider only the discrete times $\{t_0, t_1, \ldots, t_s\}$ in our model. We define a neighborhood $\{\tilde{w}_1^{0+}, \ldots, \tilde{w}_1^{s+}\}$ of $\tilde{w}_1$ with

$$\tilde{w}_1 := \begin{pmatrix} \tilde{w}_{11}(t_0) & \tilde{w}_{11}(t_1) & \ldots & \tilde{w}_{11}(t_s-1) & \tilde{w}_{11}(t_s) \\ 1-\tilde{w}_{11}(t_0) & 1-\tilde{w}_{11}(t_1) & \ldots & 1-\tilde{w}_{11}(t_s-1) & 1-\tilde{w}_{11}(t_s) \end{pmatrix}$$

as follows:

$$\tilde{w}_1^{k+} := \begin{pmatrix} \tilde{w}_{11}(t_0) & \ldots & \min(1.0,\ \tilde{w}_{11}(t_k)+0.25) & \ldots & \tilde{w}_{11}(t_s) \\ 1-\tilde{w}_{11}(t_0) & \ldots & \max(0.0,\ 0.75-\tilde{w}_{11}(t_k)) & \ldots & 1-\tilde{w}_{11}(t_s) \end{pmatrix}.$$

Note, that this neighborhood makes sense in a situation where the steppers of the second group have an overload, i.e., we are interested in decreasing the entries of the second row. We define a change for a suitable objective function as follows:

$$\Delta G^{k-} := G(\tilde{w}_1) - G(\tilde{w}_1^{k+}). \qquad (7)$$

Here, we denote by G one of the performance measures introduced in section 3. The algorithm can be stated formally as follows:

Algorithm Greedy - Search (GS):

Step 1: Determine an initial $\tilde{w}_1$.

Step 2: Compute $\tilde{w}_1^{k+}$ and ΔG^{k-} for each $k = 0, \ldots, s$.

Step 3: Compute $\Delta G := \max(\Delta G^{k-} \mid k = 0, \ldots, s)$.

Step 4: If $\Delta G > 0$ holds set $\tilde{w}_1 := \tilde{w}_1^{k+}$ and go to Step 2, otherwise go to Step 5.

Step 5: Stop, $\tilde{w}_1$ is a favorable proportion matrix.

Our method is only applicable in situations where a good initial solution exists. Note that the GS-Algorithm does not necessarily guarantee the detection of (global) optima (under the assumption of the existence of an appropriate neighborhood). This behavior is caused by the fact, that we do not allow non-improvement steps in the algorithm. It is possible to avoid such problems by using more sophisticated local search methods like simulated annealing or tabu-search (cf. Glover and Laguna 1997). However, such advanced algorithms are computationally costly. At this point it seems to be useful to consider reduced simulation models as suggested by Rose (2001).

5 RESULTS

In this section we present the results of our simulation experiments. In Figure 1 we see the WIP obtained from the simulation model. We work with a stable system. In our simulation experiments we used a simulation time of 112 days. We did five replications of all simulation runs. At the beginning of each week we release 28 lots of product $i = 1$ into the fab, which should be distributed over the steppers of two different stepper subgroups. We denote the two subgroups by G_1 and G_2. The steppers of the two subgroups use masks with different mask sizes. The other existing stepper subgroup is denoted by G_0. Note, that the workload of G_0 is high at simulation start. We are interested in decreasing this level by adding new steppers to the subgroup G_2.

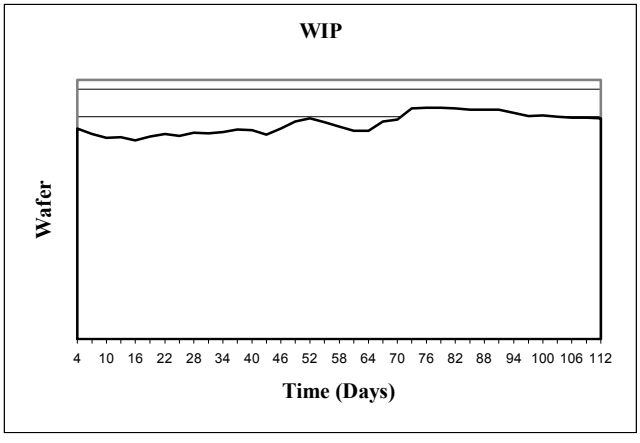

Figure 1: WIP Distribution Over Time

In Table 1 we find the description of certain scenarios with different proportion matrices. After week 6 we distribute the lots of product $i = 1$ in equal parts on G_1 and G_2 in the scenarios 1 to 3. In scenario 4 we distribute the lots

in proportion 1:2 on the subgroups G_1 and G_2 after week 5 until week 11. In the remaining weeks we distribute the lots in equal parts on G_1 and G_2. The released lots for subgroup G_0 are the same in all scenarios. In scenario 1 we assign more lots to G_2 because the workload of this subgroup is low at simulation start. In scenario 2 we distribute the lots in equal parts on the two subgroups. An assignment of all lots of product $i = 1$ to G_2 takes place in the third scenario. Scenario 4 is similar to scenario 1 in the first 6 weeks, but then we distribute more lots on G_2.

In Table 2 and 3 we present the resulting performance measures in terms of the ratio of the value of the actual scenario to the corresponding value of G_1 in scenario 1. From Tables 2 and 3 we conclude that the proportion matrix from scenario 1 is favorable because in this case we obtain the smallest (average) value for AT. The value for CTML(1) is almost equal for the lots processed on G_1 and G_2 in this scenario. Note, that we obtained this matrix with the help of the GS-algorithm after ten iterations by starting from the proportion matrix from scenario 2. In scenario 4 we get similar results as in scenario 1, but the value for CTML(1) and AT is slightly higher.

Table 1: Description of Different Scenarios

	Stepper Subgroups (Proportion)		
Scenario		G_1	G_2
1	Week1	0.25	0.75
	Week2	0.25	0.75
	Week3	0.25	0.75
	Week4	0.25	0.75
	Week5	0.25	0.75
	Week6	0.25	0.75
2	Week1	0.50	0.50
	Week2	0.50	0.50
	Week3	0.50	0.50
	Week4	0.50	0.50
	Week5	0.50	0.50
	Week6	0.50	0.50
3	Week1	0.00	1.00
	Week2	0.00	1.00
	Week3	0.00	1.00
	Week4	0.00	1.00
	Week5	0.50	0.50
	Week6	0.50	0.50
4	Week1	0.25	0.75
	Week2	0.25	0.75
	Week3	0.25	0.75
	Week4	0.25	0.75
	Week5	0.25	0.75
	Week6	0.33	0.66

Table 2: Number of Completed and Tardy Lots in the Different Scenarios

	Stepper Subgroups	
Scenario	G_1	G_2
Completed Lots		
Scenario1	65	165
Scenario2	91	142
Scenario3	35	163
Scenario4	56	161
Tardy Lots		
Scenario1	31	87
Scenario2	64	29
Scenario3	10	105
Scenario4	32	92

Table 3: Performance Measures for the Different Scenarios

	Stepper Subgroups	
Performance Measure/ Scenario	G_1	G_2
CTML(1)		
Scenario1	1.000	0.951
Scenario2	1.161	0.792
Scenario3	0.664	1.059
Scenario4	1.022	0.982
AT		
Scenario1	1.000	1.339
Scenario2	2.988	0.309
Scenario3	1.614	3.022
Scenario4	0.862	1.655
AWT		
Scenario1	1.000	1.216
Scenario2	1.243	0.513
Scenario3	0.621	1.811
Scenario4	0.865	1.243

We were also interested in the inventory in front of the three stepper subgroups. In Figures 2,3,4, and 5 we see these distributions for scenario 1 to scenario 4. Here, the data were collected during a period of 112 days with one observation every three days. As described we take five replications of a simulation run and then we calculate the average for the average number of lots during the periods of three days. We see in Figure 3, that the workload designated to the steppers of subgroup G_2 is too low. As a consequence the number of completed lots is small on G_2. In contrast, in scenario 3 the workload of subgroup G_2 is too high at the beginning of the simulation time (week 1 until week 4). From Table 2 we can verify, that the number of tardy lots is high on G_2 in this scenario.

The production control manager can use this simulation model in connection with a rolling horizon procedure in order to create a favorable distribution of the lots over the different stepper subgroups.

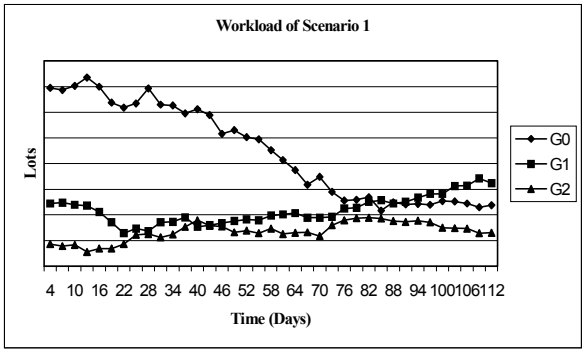

Figure 2: Workload of Scenario 1

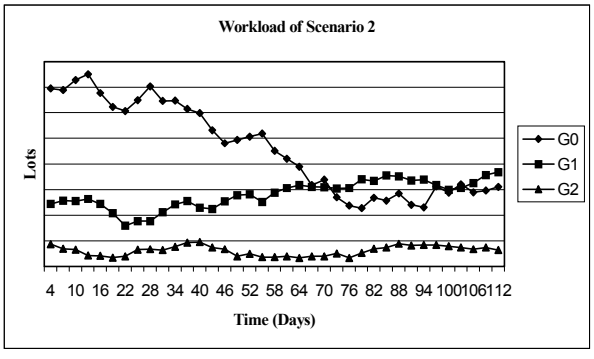

Figure 3: Workload of Scenario 2

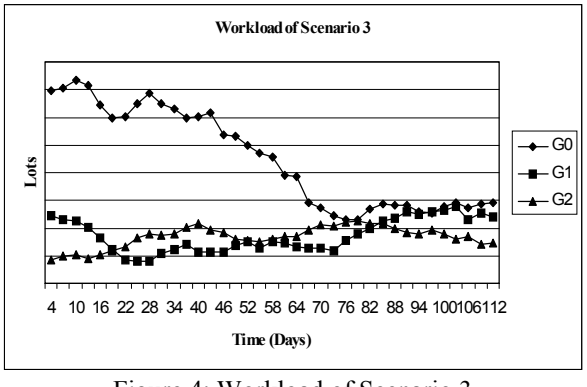

Figure 4: Workload of Scenario 3

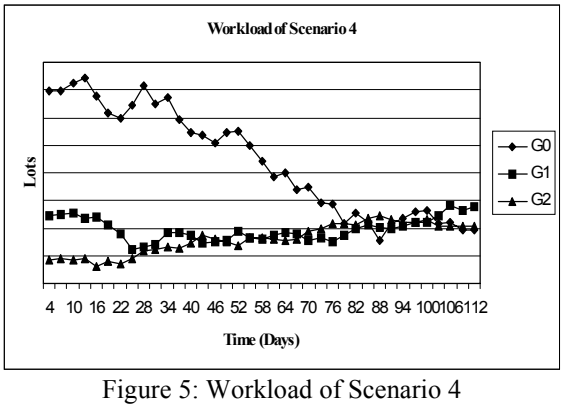

Figure 5: Workload of Scenario 4

6 CONCLUSION

In this paper, we presented a method of solving load-balancing problems for the photolithography area of a wafer fab. We built a detailed model of the photolithography area. We described the development of this model. By using expert knowledge, the simulation model and a local improvement method we are able to determine a situation dependent favorable distribution of the lots over different stepper subgroups. However, more research effort is needed to improve the performance of the local search algorithm by considering more appropriate neighborhoods and using more sophisticated search strategies.

REFERENCES

Atherton, L. F. and R. W. Atherton. 1995. *Wafer Fabrication: Factory Performance and Analysis*. Kluwer Academic Publishers, Boston, Dordrecht, London.

Fowler, J. W., G. L. Hogg, and S. J. Mason. 2001. Workload Control in the Semiconductor Industry. *Production Planning and Control. Special Issue on Workload Control.* Accepted for publication.

Glover, F. and M. Laguna. 1997. *Tabu Search*. Kluwer Academic Publishers, Boston, Dordrecht, London.

Kim, Y.-D., J.-U. Kim, S.-K. Lim, and H.-B. Jun. 1998. Due Date Based Scheduling and Control Policies in a Multiproduct Semiconductor Wafer Fabrication Facility. *IEEE Transactions on Semiconductor Manufacturing*, 11 (1): 155-164.

Mönch, L. and V. Schmalfuss. 2000a. Entwicklung von Simulationsmodellen zur Unterstützung der Produktionsplanung und -steuerung einer Halbleiterfabrik. *Zeitschrift für wirtschaftlichen Fabrikbetrieb* 95, 10: 502-509.

Mönch, L. and V. Schmalfuss. 2000b. Optimierung der Stepperbelegung in einer Halbleiterfabrik. In *Proceedings Symposium on Operations Research (OR 2000),* ed. B. Fleischmann, R. Lasch, U. Derings, W. Domschke, and U. Rieder, 333-338, Springer, Berlin, Heidelberg.

Peikert, A., J. Thoma, and S. Brown. 1998. A Rapid Modeling Technique for Measurable Improvements in Factory Performance. In *Proceedings of the 1998 Winter Simulation Conference,* ed. D. J. Madeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 1011-1015, Piscataway, New Jersey: Institute of Electronical and Electronics Engineers.

Rose, O. 2001. CONWIP-like Lot Release for a Wafer Fabrication Facility with Dynamic Load Changes. In *Proceedings of the 2001 International Conference on Semiconductor Operational Modeling and Simulation (SMOMS' 01)*, 41-45.

Schömig, A. and J. W. Fowler. 2000. Modelling Semicondutcor Manufacturing Operations. In *Proceedings of the 9 th ASIM Dedicated Conference Simulation in Production and Logistics*, ed. K. Mertins and M. Rabe, 55-64, Berlin.

Uzsoy, R., C.-Y. Lee, and L. A. Martin-Vega. 1992. A Review of Production Planning and Scheduling Models in the Semiconductor Industry, Part I: System Characteristics, Performance Evaluation and Production Planning. *IIE Transactions on Scheduling and Logistics*, 24: 47-61.

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

AUTHOR BIOGRAPHIES

LARS MÖNCH is an Assistant Professor in the Department of Information Systems at the Technical University of Ilmenau, Germany. He received a master's degree in applied mathematics and a Ph.D. in the same subject from the University of Göttingen, Germany. After his Ph.D. he worked for two years for Softlab GmbH in Munich in the area of software development. His current research interests are in simulation-based production control of semiconductor wafer fabrication facilities and applied optimization. He is a member of GI and GOR. His email address is <Lars.Moench@wirtschaft.tu-ilmenau.de>.

MATTHIAS PRAUSE received a master's degree in mathematics from the University of Chemnitz, Germany. He has over 20 years experience in the field of production control of semiconductor wafer fabs. Currently, he works in the production control department at X-FAB. In this position he is responsible for the line control. He has worked on different SMART FABRICATION projects. His email address is <Matthias.Prause@xfab.com>.

VOLKER SCHMALFUSS is the IT manager at X-FAB. In this position he is responsible for the MES and the ERP Systems of X-FAB. He received a master's degree in mathematics from the University of Leipzig, Germany. He has held various engineering and management positions since his graduation. His work included the design of statistical experiments and the application of operations research and simulation methods. From 1994 until 1998 he was a project leader in the joint development project SMART FABRICATION "CIM", which was founded by the German SC-Industry and the German government. His email address is <Volker.Schmalfuss@xfab.com>.