Enhancing Prediction Quality of Fab Simulation by Advanced Cycle Time Modelling

Maximilian Dilefeld, Sebastian Rank^(✉), and Thorsten Schmidt

Chair of Logistics Engineering, Technische Universität Dresden,
01069 Dresden, Germany
sebastian.rank@tu-dresden.de

Abstract. Simulation in the Semiconductor industry is an established method for planning greenfield projects and fab extensions as well as optimizing existing fabs and testing different scenarios for machine distribution with a constantly changing product mix. This paper discusses the importance of transportation and handling times within these simulations. Up to now, respective transportation and handling times are only considered marginally, oversimplified or are even neglected—mostly due to simulation run time performance. From experience this approach is known as error prone in respect to get valid simulation results. This paper discusses and presents strategies to increase the prediction quality of fab simulation models (e.g. for cycle time) through a more detailed modelling of transportation times. Data mining is considered as a possible approach for generating useful information with less effort than manual modeling.

Keywords: Material flow simulation · Fab simulation · Data mining · Knowledge discovery in databases · Cycle time · Transport system · AMHS

1 Motivation

Semiconductor fabrication plant (short: fab) simulation is an established approach for analyzing the capacity of an entire wafer fab and forecasting e.g. lot cycle time. In general simulation can be used for planning greenfield projects, existing fabs and fab extensions against the background of meeting new productivity requirements as well as testing different scenarios for machine distribution with a constantly changing product mix (Fowler et al. 2015; Rozinat et al. 2009).

This paper discusses the importance of transportation and handling times within fab simulation. Up to now, respective transportation and handling times are only considered marginally, oversimplified or are even neglected—mostly due to simulation run time performance. From experience this approach is known as error prone in respect to get valid simulation results. This paper discusses and presents strategies to increase the prediction quality of fab simulation models (e.g. for cycle time) through a more detailed modelling of transportation times.

2 State of the Art

Simulation is an established tool in semiconductor manufacturing with a wide variety of applications. In the semiconductor domain, usually two types of simulation are present: fab simulation and AMHS simulation (AMHS: automated material handling system). Fab simulation concerns about tools and their corresponding jobs which are to be processed on the tools in a specific order. AMHS simulation has the transportation and handling system in focus. So it realizes the transportation jobs between the tools.

Depending on the pursued goals the models representing the real-world systems are developed with specific aspects in mind (Fowler et al. 2015). Therefore, different models with a varying level of detail can exist to mimic the behavior of the whole system or to focus on specific parts of the system. Usually, these simulation models are created manually, which is time-consuming and prone to errors (Rozinat et al. 2009). Such models are based on human perception on reality rather than reality itself. The necessary meeting of assumptions and the selection of suitable key figures can be a challenging task.

In the semiconductor domain, usually discrete-event simulation models are applied. They are characterized by a sequence of events that mark changes of the examined system. The system is represented by a number of dynamic and static entities (resources). These resources typically have a limited capacity and the entities handled by these have to be ordered in accordance to specific strategies/rules. If a resource is occupied, entities join a queue to wait for processing. For realistic models the necessary times in which entities get transported from one resource to another are to be derived from the real-world transportation systems. These transportation times have to be considered within simulation studies (Fowler et al. 2015).

The automated material handling system (AMHS) for wafers gathered an essential role in semiconductor manufacturing as the increase in wafer size from 200 mm to 300 mm accompanied an increase of weight from 4 kg to 10 kg of the corresponding wafer carriers (Lin et al. 2003). The carriers used in AMHS are called front opening unified pods (FOUP) and usually have a capacity of 25 wafers. The most common system for transporting and handling FOUPs are overhead hoist transport (OHT)-systems, where fully automated vehicles travel along a unidirectional rail system built near the ceiling of the clean room/manufacturing plant. The different stations are accessed by lifting and lowering the FOUP. During this hoisting process the rail is blocked for other vehicles which can lead to delays and longer transport times. This is especially true when the OHT is working at high utilization.

In general, the AMHS should be designed to never become a limiting factor in semiconductor manufacturing (Jimenez et al. 2005). Therefore, transportation times should have a low impact on the lots' cycle times. In reality this cannot be achieved in every possible scenario due to the constantly changing product mixes (Arisha and Young 2005) and planning with the least amount of resources to meet the expected productivity (Jimenez et al. 2005). At high utilization or overload the impact of transportation and handling times rises significantly. This leads to the conclusion of giving transportation and handling times a significant consideration within the conduction of fab simulation. Surprisingly in traditional fab simulation the AMHS is not modelled at all, with less detail, or transport times are approximated by mathematical methods (Jimenez et al. 2005; Wang and Zhang 2016).

Combining or extending fab simulation models with a simulation model of the AMHS in detail would result in a sophisticated model with equipment in the form of physical models with a large scale, multiple machine types, and complex routing and dispatching strategies. These holistic models are challenging and time-consuming in terms of development, maintenance and application (Jimenez et al. 2008; Wang et al. 2018) and hence often limited by the capability of the used simulation tools. Therefore, very often individual models are used to represent the fab with its resources on the one hand and the AMHS on the other hand (Jimenez et al. 2008).

In this connection Jimenez et al. (2005) describe a common approach for reducing computation time of fab simulation models which tries to consider the AMHS: the application of from-to-matrices. These matrices hold the delivery times between source and destination locations. This method can produce accurate results for studies of fab capacities. The particular from-to-matrices can be obtained by detailed AMHS models that need to be created in advance and with a high expenditure of time. Other options with less effort, such as analytical AMHS models, lack of accuracy.

Considering runtime aspects, in a case study comparing a sophisticated holistic model (fab simulation with a fully integrated AMHS simulation model) with a simplified holistic model (fab simulation with mentioned from-to-matrices as representation for the AMHS) Jimenez et al. (2005) show a performance improvement by the factor 100 when using the mentioned from-to-matrices instead of a fully integrated model. The presented results are supposed to be of high accuracy but presuppose very precise transportation times held by the from-to-matrices as a basis for the simulation.

The time between the different process steps does not only consist of transportation and handling time, but also of the waiting times. Especially the latter varies greatly (Arisha and Young 2005) and depends on a wide range of factors and the dynamics of the system (Wang et al. 2018). These times have to be treated as randomly distributed values. Different statistical distributions can be used to represent these time variations. However, so far there are no universal solutions for the task.

3 Approaches for Enhancing the Accuracy of Transport Times Consideration in Fab Simulation

Even though there are efforts to provide “slim” AMHS models with comparatively high performance (e.g. Rank et al. 2016), holistic simulation models with a reasonable level of detail to represent the wafer fab accurately are often not able to be executed fast enough to allow efficient scenario testing (Jimenez et al. 2008). There are only a few attempts to improve prediction accuracy of fab simulation without complex AMHS modelling. Unfortunately, in general they do not perform very well in terms of accuracy and/or run time performance of the simulation model.

As mentioned, applying from-to-matrices can be a reasonable approach for modelling AMHS within fab simulation. In contrast to utilizing detailed AMHS simulation models, from-to-matrices can also be extracted from transport and waiting times based on historical datasets. When generalizing recorded data, the identification of relevant factors that influence the observed times is of great importance to avoid wrong interpretations and to be able to correctly estimate times for the fab simulation. Obviously,

historical data is only available for existing systems. Therefore, this approach is not applicable in the planning phase of new plants.

Besides the common/popular way of combining fab simulation models with AMHS simulation models, Rozinat et al. (2009) propose an alternative approach for generating simulation models which might be useful in the semiconductor domain. In detail they apply process mining techniques to (semi-) automatically generate simulation models (see Fig. 1). The step of a first simulation model which can later be further evaluated and modified can be reached much quicker following this path. Besides the time savings the authors are arguing with an improved representation of reality because the development is based entirely on objective information. Additionally, the steps for generating the simulation model can easily be repeated when the observed system changes. It is pointed out that good knowledge about the examined process is of great importance for drawing the right conclusions. Another requirement is the existence of comprehensive recordings of historical data itself. The completeness and quality of this information have major influence on the resulting simulation model.

Fig. 1. Manually and (semi-) automatically created simulation models (based on Rozinat et al. 2009)

The term process mining refers to a number of different techniques to extract useful information from event logs (Rozinat et al. 2009). The basis for this are information collected by information systems from the real-world process. These systems usually coordinate the required process steps and record important events related to the performed activities. Process mining is used to discover and represent the causalities between the activities. Besides the performed activity itself event logs usually also include additional information like performance, equipment identifiers, time stamps and more. Often the activities are logged from different perspectives like scheduling, start and completion, resulting in multiple entries with different time stamps or one entry with multiple time stamps. The more information is contained in the data the more factors can be incorporated into the simulation.

Rozinat et al. (2009) point out that the usefulness of simulation models generated by this approach depends on the level of detail. A simple model can approximate the behavior of a system (e.g. routing or waiting time) based on probabilities. But it is not suitable for making predictions based on changes in the system. When starting from historical data usually a specific condition of the system is modeled. Only if different

perspectives (e.g. control-flow, data, resources and time) are covered by the model, statements can be made how specific aspects influence the behavior of the whole system.

A similar approach within the scope of simulating transport systems is pursued by Wang and Zhang (2016). Big Data is deemed as a great opportunity to improve the prediction of relevant key figures in production systems. The necessity of a high level of automation and digitalization is met by the semiconductor industry. Within the approach itself, an artificial intelligence (AI) system treats the manufacturing system as a black box and makes predictions on key characteristics of the wafer lots. According to the authors, the applied neural networks can achieve higher prediction accuracy than conventional regression-based models. The considered characteristics used in AI and hybrid AI systems and their identification are presented as a major challenge. It is pointed out, that the quality of the input data of such systems has a great influence on the accuracy of the results. Collected raw data has to be pre-processed, especially when working with Big Data.

As already mentioned, the lot cycle time and hence the wafer fab's performance varies greatly depending on the manufacturing process itself as well as the transport between the different process steps. In this regard Wang et al. (2018) deal with the identification of candidate factors for cycle time forecasting based on historical data-sets. The correlation of these factors on each other and on cycle time is analyzed with a regression-based model. From 774 candidates 108 factors are deemed correspondent to the cycle time for the presented system. The factor selection is stepwise from highest correlation to lowest. The factors are integrated successively into the forecasting model. The prediction error is calculated. Factors are incrementally added to the model until the subset of factors is sufficient. A relevant improvement in mean relative deviation is shown in dependence of the number of factors.

In both presented papers, the authors mentioned are using data driven concept to build complete, holistic models. Similarly, these approaches can be used to consider only a specific aspect of the system. In this case it might be possible to generate the mentioned from-to-matrices that afterwards can be implemented into established simulation methods which may already exist.

4 Extracting Transport Times from Historical Data

So far, applying from-to-matrices with corresponding transportation times is the choice for extending fab simulation models in order to get more realistic simulation results. This is mostly because none of the presented alternative approaches have been applied in the semiconductor domain and hence their capabilities are rather unclear. So, it will be investigated to what extent from-to-matrices can be extracted from historical data. In this regard there are multiple existing procedures which can be used for data analysis and exploration. The procedure applied in regard to this paper is based on the knowledge discovery in databases (KDD)-process introduced in 1996 by Fayyad et al. (for additional information see Ristoski and Paulheim 2016). For better understanding the KDD-process and hence the extraction of the from-to-matrices is described in the following. It starts with raw data, from which valuable knowledge and insights are gained following five steps: selection, preprocessing, transformation, Data Mining and evaluation and interpretation (Fig. 2).

Fig. 2. KDD process (based on Ristoski and Paulheim 2016)

- **Selection:** In the first step an understanding of the application domain was developed, relevant prior knowledge was collected (which has been done in the preceding sections of this paper) and the goal of the end user was defined. In our case the goal was the generation of from-to-matrices containing the transport times faster than manually building an AMHS simulation model and with higher or at least similar accuracy. Based on this information the target data was chosen with reasonable sample size and relevant variables. Here the target data is a database containing transport logs from an existing AMHS. Knowledge about the different variables was collected in cooperation with experts on sight.
- **Preprocessing:** This step includes dealing with missing values, duplicates noise and error in the data. The quality of the dataset has to be analyzed. Incomplete or faulty entries have to be removed from the database or corrected if possible.
- **Transformation:** The data has to be converted into a form the later used algorithms can work on. The specific form depends on the used programs/the used programming language. In our case the data was already tabular, which in most cases is an appropriate form. The raw data as well as the results later are stored in a MySQL database for easy access. Python with the “pandas-package” for data analyses is a common choice as a programming environment. The DataFrame-object from pandas is a convenient structure for working with tabular data.
- **Data Mining:** In this step different methods will be applied to the data to achieve the goals defined in the first step. The expectations and goals of the end user have to be considered in selecting appropriate methods and algorithms. One example is the level of detail. More simple models are less accurate but easier to interpret and therefore preferred in some circumstances.
- **Evaluation and interpretation:** In the last step the finding of the data analysis will be examined with respect to their validity. The usefulness of the gained knowledge will be evaluated. Visualization and presentation of the results is also a typical component of this step.

5 First Findings

An extract of log data from the AMHS was provided by Infineon Dresden. This data will be the basis for our data analysis on a larger scale. The structure of the provided data was analyzed and will further be discussed with experts from Infineon. The log

files are available in tabular form with a total amount of 25 attributes. For the investigation particularly the different time stamps, source and destination location, the carrier-ID as well as the priority are of interest. Each row of the database represents a single transport operation.

One problem in respect to model from-to-matrices, which in a first analysis became apparent, is that many transport operations start or end at a storage unit. This phenomenon of low tool-to-tool ratios is well known in the semiconductor domain and has also been covered in a couple of other papers. Fischmann et al. (2008) assume a tool-to-tool-ratio as low as 20% to 40%. Jimenez et al. (2010) refer to a tool-to-tool-ratio between 25% and 47%. Heinrich et al. (2008) on the other hand assume a tool-to-tool ratio of 60%. The storage units have no identifiers in the database. Therefore, the location of the individual storage units is unknown. One possible solution to this issue is the combination of the transport operation tool-storage and storage-tool according to the pattern tool-storage-tool using the carrier-ID. This procedure will be discussed in cooperation with the experts from Infineon Dresden.

6 Results

The examined literature indicates that the quality of efficient prediction models is highly dependent on the determination of precise time tables for simulating the AMHS. It is shown that data driven concepts are likely to improve the prediction quality and are of interest of further research.

7 Discussion/Implications

A key finding of the research and the literature review is the identification of a lack of methods to accurately apply transportation and handling times within fab simulation. Even though transportation and handling times are essential to get high accuracy simulation results, several time and performance related challenges lead to the application of non-integrated models. So far, AMHS- and fab simulation models are treated as independent models. To overcome the issue, and as an alternative for holistic models, new approaches to apply transportation and handling times in fab simulation have to be found. Within the project iDev40, the ability of e.g. from-to-matrices and corresponding transportation and handling time distributions will be tested for application.

References

- Arisha, A., Young, P.: Simulation in semiconductor manufacturing facilities. In: Fifth International Workshop on Advanced Manufacturing Technologies, vol. 9 (2005)
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* **17**, 37–54 (1996)

- Fischmann, C., et al.: Buffer management for automated material handling systems in semiconductor industries. In: Louca, L.S., Chrysanthou, Y., Oplatkova, Z., Al-Begain, K. (eds.) *Proceedings of the 22nd Conference on Modelling and Simulation, ECMS 2008*, pp. 423–427 (2008). <http://www.scs-europe.net/dlib/2008/2008-0423.htm>
- Fowler, J.W., Mönch, L., Ponsignon, T.: Discrete-event simulation for semiconductor wafer fabrication facilities: a tutorial. *Int. J. Ind. Eng.* **22**, 5 (2015)
- Heinrich, H., et al.: Pursuing the increase of factory automation in 200 mm frontend manufacturing to manage the changes imposed by the transition from high-volume low-mix to high-mix low-volume production. In: *2008 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, pp. 148–155 (2008)
- Jimenez, J.A., Mackulak, G., Fowler, J.W.: Efficient simulations for capacity analysis and automated material handling system design in semiconductor wafer fabs. In: *Proceedings of the Winter Simulation Conference, Orlando, FL, USA*, pp. 2157–2161. IEEE (2005)
- Jimenez, J.A., Mackulak, G., Fowler, J.W.: Levels of capacity and material handling system modeling for factory integration decision making in semiconductor wafer fabs. *IEEE Trans. Semicond. Manuf.* **21**(4), 600–613 (2008)
- Jimenez, J.A., et al.: AMHS factors enabling small wafer lot manufacturing in semiconductor wafer fabs. In: *Proceedings of the Winter Simulation Conference, WSC 2010, Baltimore, Maryland*, pp. 2575–2585 (2010). <http://dl.acm.org/citation.cfm?id=2433508.2433826>
- Lin, J.T., Wang, F.-K., Wu, C.-K.: Simulation analysis of the connecting transport AMHS in a wafer fab. *IEEE Trans. Semicond. Manuf.* **16**(3), 555–564 (2003)
- Rank, S., et al.: The correct level of model complexity in semiconductor fab simulation—lessons learned from practice. In: *2016 27th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pp. 133–139 (2016)
- Ristoski, P., Paulheim, H.: Semantic Web in data mining and knowledge discovery: a comprehensive survey. *Web Semant. Sci. Serv. Agents World Wide Web* **36**, 1–22 (2016)
- Rozinat, A., et al.: Discovering simulation models. *Inf. Syst.* **34**(3), 305–327 (2009)
- Wang, J., Zhang, J., Wang, X.: A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system. *IEEE Trans. Semicond. Manuf.* **31**(1), 173–182 (2018)
- Wang, J., Zhang, J.: Big data analytics for forecasting cycle time in semiconductor wafer fabrication system. *Int. J. Prod. Res.* **54**(23), 7231–7244 (2016)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.