Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

A FRAMEWORK FOR EFFECTIVE WIP FLOW MANAGEMENT IN SEMICONDUCTOR FRONTEND FABS

Mathias Duemmler

Infineon Technologies AG
Am Campeon 1-12
D-85579 Neubiberg, GERMANY

Juergen Wohlleben

Infineon Technologies AG
Wernerwerkstrasse 2
D-93049 Regensburg, GERMANY

ABSTRACT

Automated WIP Flow Management (WFM) is an essential factor for semiconductor fabs to maintain a competitive position in a cost-, time-, and quality-sensitive market. WFM comprises capabilities like dispatching, scheduling, material flow prediction, capacity planning and optimization. In this paper, we present a framework for setting up effective WFM mechanisms in a frontend fab. The framework assures the interaction of the individual WFM components and that the overall goals of WFM (e.g. cycle time reduction, capacity optimization) are achieved as a concerted effort of these components.

1 INTRODUCTION

Managing the material flow, or WIP flow, in a semiconductor fab is a task of high complexity. WIP stands for Work In Progress and refers to material that has entered the production process but is not yet a finished product. Different targets, like achieving cycle time or delivery targets while keeping the machine utilization high have to be achieved simultaneously. As indicated by Schoemig and Fowler (2000), improving these operational processes is one of the most promising opportunities to reduce cost in semiconductor fabrication facilities. This task becomes even more challenging in fabs that produce a high number of different products – sometimes several hundred – with different routes and processing requirements. Typically, this task can only be accomplished with the help of automated WIP Flow Management (WFM) mechanisms like automated lot dispatching, scheduling, simulation methods to predict material flow, advanced capacity planning and capacity optimization.

In this article, we present a framework for setting up effective WIP Flow Management mechanisms in a frontend fab. The goal is to align the different mechanisms in such a way, that they all support the overall fab goals, which need to be defined as a first step when introducing WIP Flow Management.

The article is organized as follows: we first introduce the overall automation framework used at Infineon's frontend fabs, which also comprises the automated WFM solutions. Then, we specify in detail the framework for WIP Flow Management and how the mechanisms interact to achieve the overall goals. We will also focus on a specific success criteria when introducing WFM, namely reliable sources for common master data. We conclude the paper by summarizing the findings of this article and providing an outlook on further activities in this field.

2 OVERALL AUTOMATION STRATEGY FRAMEWORK

In 2011, Infineon has introduced the Factory Integration Roadmap approach for the automation activities in its fabs. Infineon's Factory Integration (FI) comprises the 7 aspects of automation as depicted in Figure 1: Equipment Automation, consisting of e.g. equipment integration and recipe management; Manufacturing Data Management, which includes, among others, reporting and data analysis capabilities; Material

Handling Automation, which covers the physical automation like robots and transport systems; Material Identification & Tracking, consisting e.g. of MES functionalities and cassette and wafer tracking capabilities; Process Control Automation, including Advanced Process Control (APC) tools like Fault Detection and Classification (FDC), Run-to-Run control, etc.; WIP Flow Management, which will be explained in detail in Section 3; and finally Workflow Automation, covering capabilities like automated decision making, automated execution of experiments, etc.

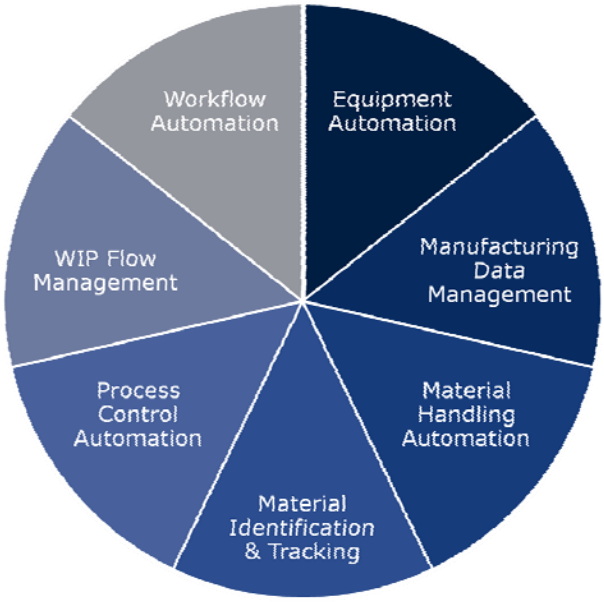

Figure 1: Different aspects of automation combined in FI Roadmap approach

The goal of the FI Roadmap is to define the long-term vision for these 7 aspects of automation and to achieve a common understanding of the roadmaps required to implement these visions. In order to do so, the FI Roadmap provides the framework to prioritize, align, and harmonize the automation activities within the fabs and therefore allows to allocate the available resources to the activities with the highest benefits. For Infineon, Factory Integration is a key enabler to maintain a competitive manufacturing position for its fabs and to handle the increasing complexity in a high-quality environment.

The FI Roadmap approach works as follows. First, the current status of the automation capabilities is assessed for each fab and capabilities which are required in the next five years are defined. Based on this gap analysis, the roadmaps to achieve these functionalities within the next five years are derived. A prioritization of the activities is done based on the expected benefits like productivity increase, quality assurance, yield improvement, cycle time reduction, etc. By aligning the individual roadmaps for each category, it is assured that the functionalities work together in the manufacturing environment and that a consistent strategy is derived.

3 FRAMEWORK FOR WIP FLOW MANAGEMENT

This article focuses on a specific category within the Factory Integration framework, namely WIP Flow Management. For frontend fabs, WIP are the wafers which are released to the production floor, and either are being processed or measured, being transported, or are waiting for the next activity. WIP Flow Management is the combination of all activities performed to improve the material flow in the fab in order to increase productivity as well as to manage fab performance proactively.

Figure 2 depicts the four main components of WIP Flow Management, namely global fab rules, operational optimization, capacity planning and forecasting, and simulation, which will be described in this section.

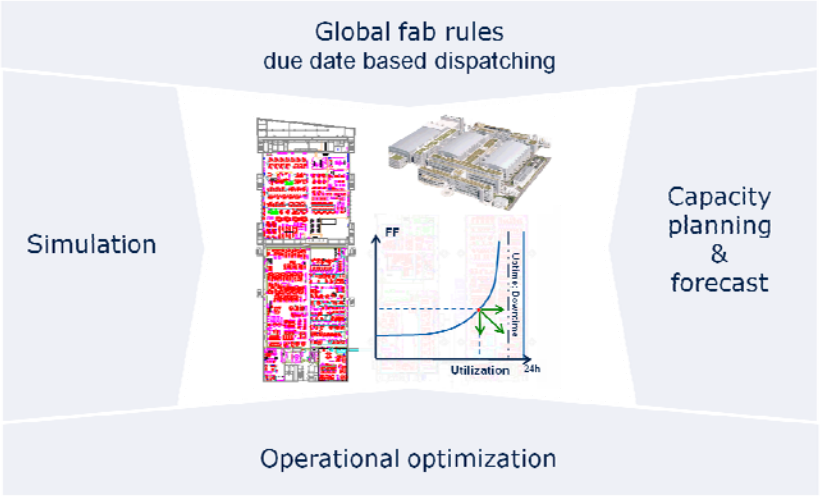

Figure 2: WIP Flow Management is a combination of four key components

The global fab rules cover the basic strategy how to control material flow in a fab, i.e. the definition of the ground rules for logistics on the shop floor. Within Infineon, the material flow is steered based on the operational due date of the lots. For each product and its respective process flow (route), a target flow factor is calculated. Flow Factor (FF) is defined as the cycle time (CT) divided by the raw processing time (RPT). The target CT of a route consists of the sum of physical CT of each process step and queuing time, to consider the impact of incoming WIP variability. Typically, the target FF is set to values leading to a good tradeoff between fab utilization and cycle time. The due date for each lot is calculated as the start date of the lot plus the target cycle time. The prioritization of production units on the dispatching user interface is equivalent to the delay of the lot versus the due date. This basic fab ground role leads to a significant reduction of the variability of cycle times for closed production units.

Whereas the global fab rule is applied to the whole fab, the next element, operational optimization, focuses on improving the performance of selected individual groups of tools, or work centers. Activities within this element comprise the so-called PushPull management of the material flow in a work center, which targets on the one hand at reducing the variability of the incoming and outgoing WIP flow of a work center, and on the other hand on ensuring a 100% utilization of the bottleneck work centers. Another activity aims at optimizing the capacity utilization of the tools by optimizing the assignment of lots to tools or, in case of cluster tools, even to individual process chambers. According to the SEMI E21–96 standard, a cluster tool is an ”integrated, environmentally isolated manufacturing system consisting of process, transport, and cassette modules mechanically linked together”. Especially in areas where set-up-times and maintenance intervals are critical for the overall work center performance, online scheduling based on real-time optimization algorithms are applied to dynamically assign lots to tools over a specific scheduling horizon.

The third component, capacity planning and forecasting, focuses on the well-balanced loading of the tools in the fab to reduce variability effects, e.g. WIP bubbles, especially in fabs with a huge number of parallel running technologies and products. Infineon’s capacity planning & forecasting toolset includes

tools for static forecasting of bottlenecks, analyzing and statically balancing production volume within the given total fab loading, and "golden mix" loading proposal. Golden mix loading refers to the capacity optimized loading of the fab, e.g. the product mix and volume scenario that leads to the optimal demand fulfillment for a given capacity.

Finally, fab simulation is applied to perform WIP, CT and utilization fab prognosis to enable decisions concerning investment in additional equipment, volume scenarios and impact of e.g. critical change of product mix. Scholl (2008) shows how simulation is used e.g. for recovery prognosis after severe unscheduled fab events (e.g. extended downtime in the bottleneck work center). Noack et al. (2011) describe how short-term simulation approaches are applied for operational decision making. Specialized simulation models are used for design, engineering and analysis of tools, work centers, as well as automation and transport systems. Another example for the use of simulation is staffing analysis to assess the impact of operator availability on work center and fab performance. For these applications, stochastic and deterministic discrete event simulation are applied and the application of fluid flow models is investigated.

These four core components of Infineon's WIP Flow Management provide a substantial toolset to improve productivity, reducing investments by optimization of equipment and fab usage as well reducing all investment related material and personal costs. Figure 3 summarizes the methods applied in each area as well as the targeted performance indicator.

	targets	methods
Global fab rules	▪ ↑ delivery reliability ▪ ↓ CT spread	▪ due date based dispatching
Capacity planning & forecast	▪ ↑ Utilization (Fab) ▪ assure FF/CT ▪ ↑ effective loading process	▪ capacity check ▪ portfolio optimization ▪ bottleneck forecast
Operational optimization	▪ ↑ Utilization (work center/tool) ▪ ↑ throughput (tool)	▪ active WIP control ▪ work center load balancing
Simulation	▪ proactive controlling (FF/CT) ▪ performance prognosis (dynamic) ▪ staff optimization & engineering	▪ dynamic simulation ▪ scenario evaluation

Figure 3: WIP Flow Management components and related targets and methods

Each core component and method is clearly focused on specific productivity targets or key levers. Even one component in isolation is very efficient. As they are all interrelated and influencing each other, the combination of the components provides a powerful toolset for real effectiveness and productivity management.

4 INTERACTION OF WIP FLOW MANAGEMENT ELEMENTS

Figure 4 depicts how the individual elements of WIP Flow Management are interlinked and depend on each other. These dependencies will now be described in detail.

Infineon's advanced capacity planning is using the predicted lot starts for the upcoming week to derive an optimal allocation of load to each of the work centers in a fab. Romauch et al. (2009) describe

how the planning tool minimizes the maximum load for the tools in a work center, thereby minimizing the loading variability in the work center and assuring optimal utilization of the existing capacity. This detailed distribution of the static load is then used by local optimization rules that dynamically control the incoming and outgoing material flow at selected work centers to achieve an optimal balancing of the available WIP and at the same time assuring high utilization of the bottleneck work centers.

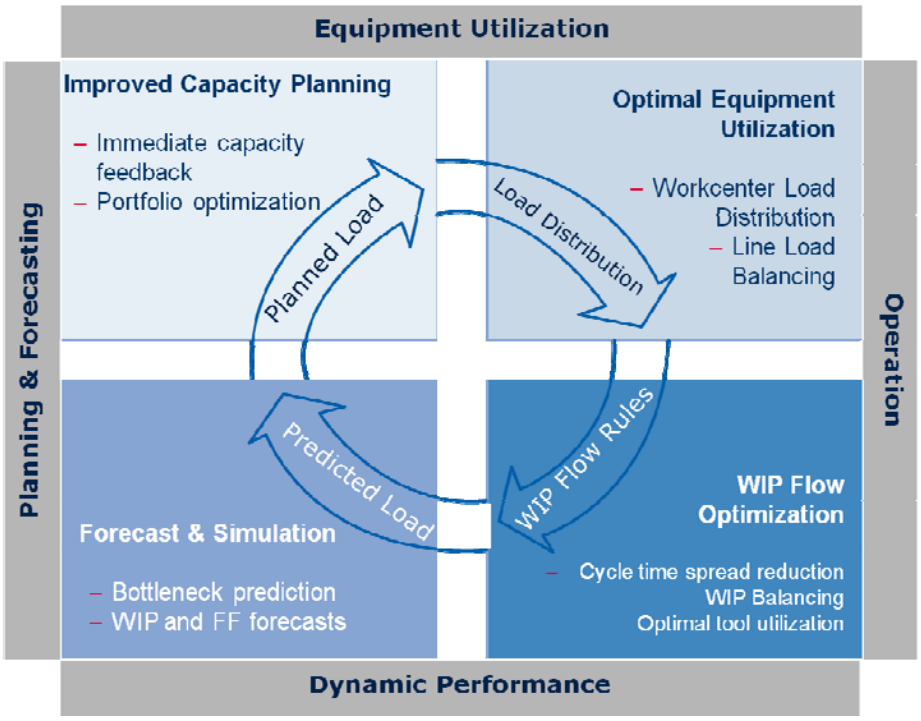

Figure 4: Interaction and dependencies of WFM components

The WIP Flow Optimization, on the other hand, is using the global dispatching rules mentioned in Section 3 and is combining them with the locally derived load balancing rules. Local optimization is only applied at selected work centers, to make sure that the global dispatching rules are not cancelled out by creating too many local optima.

In order to predict the evolution of WIP and cycle times using discrete event simulation, the aforementioned WIP Flow Management rules need to be incorporated into the simulation model. Also, it is critical for the validity of the simulation model that the optimal distribution of load to the tools in a work center, as achieved by operational optimization, is correctly reflected in the simulation model, in order to have a reliable prediction of the dynamic WIP and cycle time evolution.

Finally, the loop is closed by providing forecasted WIP distributions out of the forecasting and simulation component to the advanced capacity planning component, which combines this information with the predicted loading profile to derive the expected load for each work center for the upcoming week.

It is essential for the seamless interaction of these components, that all solutions are using the same input data for their calculations. This input data consists of, e.g., process routes, number of tools, tool capabilities, tool speeds, sampling algorithms, etc. Providing this input data reliably with a high quality level is an essential factor for the success of all WIP Flow Management activities.

5 SUMMARY AND OUTLOOK

In this article, we have presented Infineon's Factory Integration Roadmap as the framework for automation activities within Infineon's frontend fabs. We have described the WIP Flow Management activities

as an essential element of this roadmap and have presented the different components of the WFM toolset. It was described how the components are used to improve specific performance indicators in the fabs and how these components interact. Finally, we have highlighted the importance of reliable and high-quality input data as a success criteria for all WIP Flow Management activities.

Infineon will continually improve the tools within the WIP Flow Management framework. As one example, the availability of fast mathematical solvers opens opportunities for many additional applications of mathematical optimization, which will be evaluated and implemented in the near future. Furthermore, with the increasing accuracy of the WIP predictions out of simulation, it will be possible to extend the scheduling solutions to more than a few selected work centers, allowing for bigger levers for optimization algorithms. With the advancements in simulation modeling, the impact on overall factory performance can be assessed before the actual implementation of such concepts in the fab.

REFERENCES

Noack, D., M. Mosinski, O. Rose, P. Lendermann, B. P. Gan. 2011. "Challenges and Solution Approaches for the Online Simulation of Semiconductor Wafer Fabs". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, 1845–1856. Phoenix, USA.

Romauch, M., G. Seidel, H. Gold, W. Laure. 2009. "Plansim: Capacity Planning Based on Linear Programming". In *ISMI Symposium on Manufacturing Effectiveness 2009*, Volume 1, 646-662. Austin, Texas, USA.

Schoemig, A. and J. Fowler. 2000. "Modelling semiconductor manufacturing operations". In *Proceedings of the 9th ASIM Dedicated Conference Simulation in Production and Logistics*, 55–64.

Scholl W. 2008: "Coping with typical unpredictable incidents in a logic fab". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 2030-2034. Piscataway, New Jersey.

SEMI (1996). "Standard for Definition and Measurement of Equipment Reliability, Availability, and Maintainability (RAM)". SEMI.

AUTHOR BIOGRAPHIES

MATHIAS DUEMMLER studied computer science at the University of Wuerzburg, Germany. From 1997 to 2001 he was a Research Assistant at the working group "Performance Evaluation of Manufacturing Systems" at the same university, where he also received his Ph.D for his work on modeling and optimization of semiconductor manufacturing systems. He joined Infineon Technologies in 2003 where he held various positions in the area of Industrial Engineering and Operational Planning and Control. Since October 2010, he is head of the department Manufacturing IT Frontends, in charge of automation solutions for Infineon's frontend fabs. His email address is mathias.duemmler@infineon.com.

JUERGEN WOHLLEBEN studied Industrial Engineering at the University Munich, Germany. He joined Infineon Technologies in 1994 where he held various positions in the area of equipment engineering, quality control, productivity and production management. Since October 2010, he is head of the department Industrial Engineering Frontends, in charge of all aspects of productivity management for Infineon's frontend fabs. His email address is juergen.wohlleben@infineon.com