

COPING WITH TYPICAL UNPREDICTABLE INCIDENTS IN A LOGIC FAB

Wolfgang Scholl

Infineon Technologies Dresden
Koenigsbruecker Strasse 180
01099 Dresden, GERMANY

ABSTRACT

Within the last months the semiconductor plant of Infineon in Dresden has converted to a pure manufacturer of logic products. With it, premises for production control have changed to short and, first of all, predictable cycle times. Typical for a logic fab, intelligent performance monitoring and prediction methods have to be installed, coping scheduled and unscheduled disturbances. Most important method is transient fab simulation, giving a 3-months performance forecast based on actual line situation and planned changes in toolset and product mix. This continuous forecast has to be enriched with information to scheduled and unscheduled fab disturbances for just in time initiation of countermeasures in case of foreseeable excursions, e.g. changes in dispatching. In particular, breakdowns have to be analyzed from historical data to find typical patterns, for probability of occurrence and, especially, machine group specific behavior. This paper presents the procedure on the basis of three examples.

1 INTRODUCTION

In contrast to a manufacturer of commodity products, aspects like fab utilization and maximized machine group throughput play a tangential role in a logic fab. Main interesting topics are short, controllable and predictable product cycle times. Additionally, a fast speed corridor for the engineering lots is needed to provide necessary learning cycles and, in the end, a short time to market for new developed products. The whole wafer start volume is based on concrete delivery commitments. Now it is impossible to displace the effect of temporary fab performance problems or fluctuated demands on commodity products without strict cycle time targets. So in a multiproduct logic fab intelligent methods are needed to master these requirements. It already begins in capacity planning process. The long-term assessment of wafer start demand is done by static

planning based on predefined utilization limits. In short-term planning (up to 3 months) the focus is on cycle time statements and following necessary corrective actions, e.g. in dispatching. For the close range the current WIP distribution is more relevant than planned wafer starts. This is the part of tactical fab simulation. This is a transient discrete-event simulation with extraordinary attention to warm start conditions, e.g. modeling of real line WIP distribution, consideration of known down events and lot prioritizations and so on.

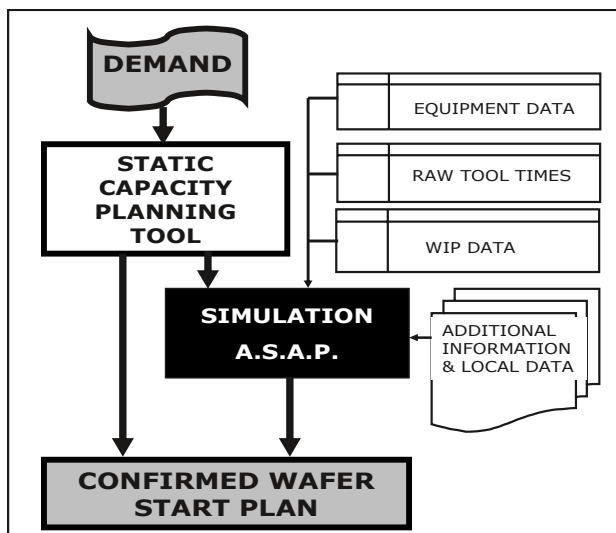


Figure 1: Capacity planning process with integrated fab simulation

“Real Time Dispatcher” with “On-time Delivery Due Date” fab rule is used for material flow control. Dispatch rules are evaluated by simulation, especially in case of local dispatching a compromise between optimal usage of machine group capacity and cycle time aspects of low volume products has to be found, e.g. simulation based batch size optimization. The executing of an intelligent perform-

ance monitoring is very important, e.g. move-weighted machine group uptimes correlate with dynamic fab cycle time, but often reported average fab uptime is not sufficient. Abnormal statistic trends of variability's of machine reliability parameters (e.g. MTOL) are early signs for cycle time drifting. But the key role in terms of fab performance mastery takes the fab simulation. Typical for a logistical network like high complex semiconductor manufacturing is the propagation and long-term building up of disturbances (Rose 1998). Figure 2 shows the time-shifted behavior of WIP, dynamic and static cycle time after a line incident, meanwhile daily reported indicators, like moves, instantaneously react on disturbances. So without simulation there is no chance for assessment of long-term effects of a line incident.

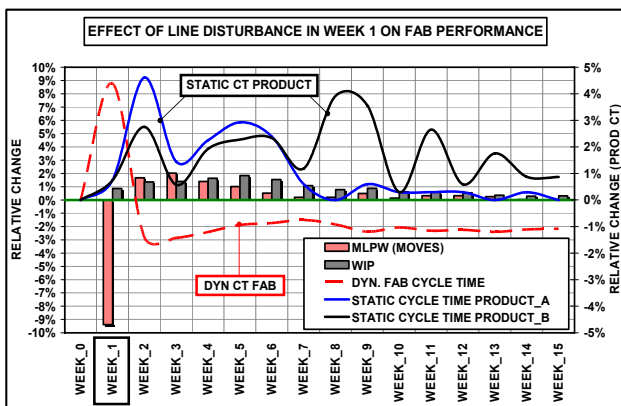


Figure 2: Long-term effect of a line incident

2 FACTORY SIMULATION

2.1 Model Architecture

For fab simulation the commercial software AutoSched AP is used. Basic data supply is realized from different sources, e.g. static planning tool, data warehouse and local Excel files. This transient simulation starts with real line conditions. Typical performance indicators (WIP, dynamic and static cycle time, uptime utilization, moves) on level fab or product group are generated from ten confidence runs with focus on weekly trend prediction of next quarter. On level machine group a qualitative bottleneck prediction is implemented. Planned future changes in toolset or process times are also loaded from the planning tool. In future a data link to Real Time Dispatcher (APF repository) will be created. Data from simulation result files are loaded into an Oracle database and have to be prepared for follow-up applications.

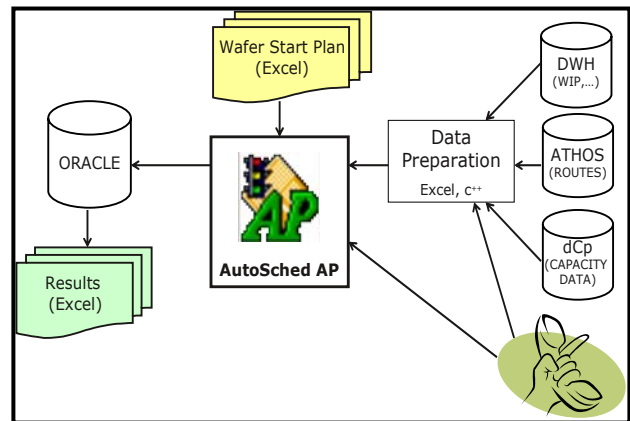


Figure 3: Fab simulation model architecture

2.2 Modeling Approach

There are different opinions, what model approach for tactical fab simulation is best. We prefer an application related model approach with focus on following aggregation levels fab, product group and machine group. This simplified approach for modeling can not be used e.g. for detailed analyses of cluster tool performance. But the benefits in opposite to the high-detailed all-in-one approach are a higher model transparency and better maintainability. Non-relevant details are modeled in simplified manner. So it is possible to conceal missing or failing basic data on a certain scale, e.g. by using of delay steps with real cycle time distributions in model. Of course, a sensitive model validation based on historical data has to be required. Limiting the huge data volume, more than 100 products are referred to about 20 process flow representatives. Most production areas are modeled on level machine group, with exception of production areas with permanent dedications (e.g. lithography). The model contains transport times (AMHS) in simplified manner. Real tool locations are not considered. Operator effect is modeled as adder on transport time, the absolute value results from model validation. Logistical constraints, like time windows between process steps, send ahead wafer procedures or rework are modeled from view of impact on material flow logistics. Global and local dispatch rules are built in with a high degree of precision.

2.3 Simulation Accuracy

Even though the fab simulation is mainly focused on performance trend prediction, the accuracy of weekly WIP forecast was analyzed over a period of one year. Figure 4 shows the absolute deviation of WIP forecast from reality in dependence on simulation horizon. For the usual simulation horizon of 13 weeks a long-term accuracy in the range of 95% can be ensured.

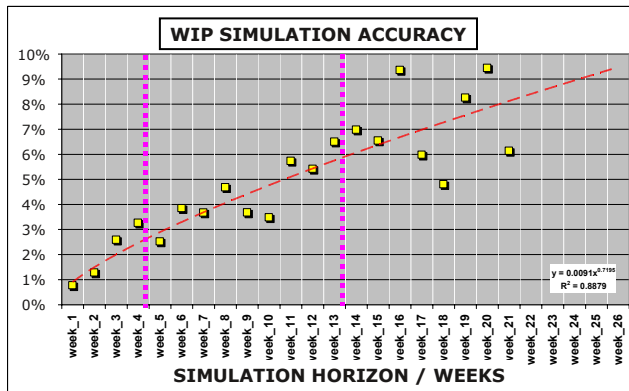


Figure 4: Fab WIP forecast accuracy

3 DISTURBANCES IN A LOGIC FAB

3.1 Changes in Number of Products

Multiproduct fabs have to get along with higher production flow variability's resulting in higher cycle times at same loading like a one-product memory fab. That means, they are disadvantaged in terms of productivity benchmark. Long-term capacity planning usually based on product groups, because it is not known in an early state, how many products will be generated from a new-developed master product. Often these products use most of the processes of the standard route. Only a couple of critical oxidation, implantation and lithography steps are product specific. In a case study a fab with 15 different standard products were simulated. From these standard products 15 and 30 child products are generated with typical differences in critical process steps only, but without consequences from view of static capacity planning (nearly constant uptime utilization). Results of the simulation study are shown in figure 5.

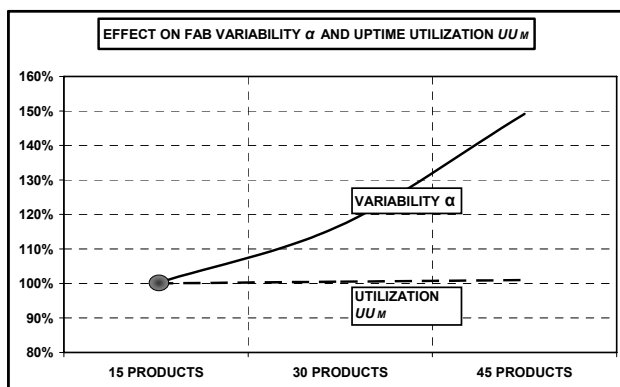


Figure 5: Effect of number of products on fab performance

Small individual impacts (e.g. selective reduced furnace batch sizes or stream sizes in lithography and implantation)

interfere with each other enforced by the consequent lot due date oriented dispatching. The following equation (1) describes the interdependency between cycle time, uptime utilization UU_M and variability α . The flow factor FF represents the normalized dynamic cycle time.

$$FF = \alpha \frac{UU_M}{1 - UU_M} + 1 \quad (1)$$

Consequently, the shown effect on variability of the production flow (figure 5) is manifested in increased cycle time in reality.

Simulation helps not to loose the sight on this effect and to force the process unification. But it also means, that differences within product families have to be considered exactly using route representatives for fab simulation.

3.2 Master Stock Effect

Some logic products run on a common route before metalization. The lots are intermediately stored in a Master Stock. The further processing will be made on customer requirements in terms of ROM code definition. From the vantage point of material flow logistics and capacity planning the missing forecast for demands from Master Stock is a big problem. So the Master Stock effect could not be considered in capacity planning. Apparently, it is not a problem, because over a long period additional wafer starts from Master Stock are compensated with reduced wafer starts in other weeks, the long-term average of the master stock WIP is nearly constant. But according operating curve theory, there is a nonlinear interdependency between cycle time and uptime utilization, cycle time loss due to additional wafer starts from master stock are not compensated completely by reduced wafer starts in following time periods.

This topic was fundamentally discussed by Rose (2006). We executed a steady-state simulation with real weekly master stock WIP fluctuation of one calendar year for estimation of average cycle time effect. The resulted distribution of daily fab cycle time (grey area in figure 6) was compared with a reference scenario without master stock. For a typical loading situation the average cycle time loss over one year was about 5% connected with a significant deteriorated variability.

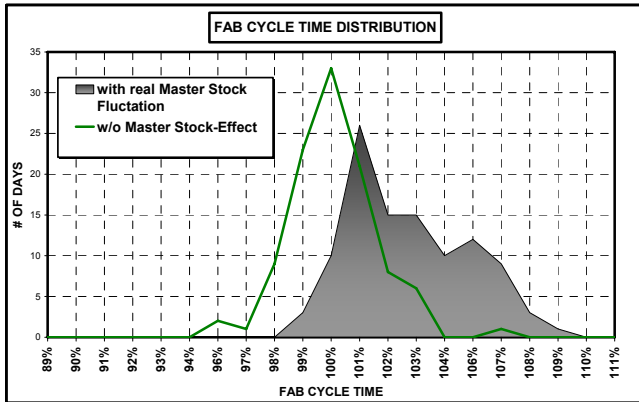


Figure 6: Effect of master stock fluctuation on fab cycle time

This additional need in capacity has to be considered in capacity planning. More interesting for cycle time forecast, the master stock fluctuation has a more or less critical occurrence on seasonal base. Based on long-term real data analyses a typical pattern was found (horizontal black lines in figure 7).

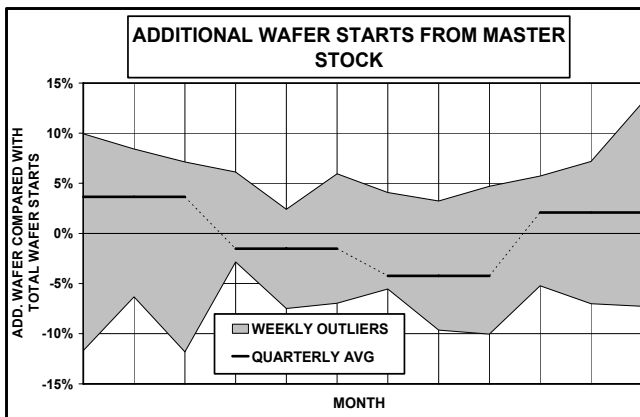


Figure 7: Pattern for typical master stock fluctuation

These results are used for cycle time forecast of critical products (worst case scenarios), if better knowledge on master stock fluctuation doesn't exist, or the assessment of maximum wafer start scenarios.

3.3 Production Breakdown

Unscheduled production breakdowns have to be examined by fab simulation with focus on long-term effects (fab WIP, closed lot cycle time and delivery impact). Changes in dispatching or, at the worst, wafer start reduction could be the countermeasures.

But effects of scheduled breakdowns, like a yearly maintenance of media supply, have also to be made more

predictable to compensate expected delivery problems in advance and to define realistic production targets for the affected time period. The specific activities and with it the duration of maintenance work is known. In point of fact, the staggered restart of the machines in dependence on process qualification effort is problematic. The theoretical approach is to stop the whole line until all machine groups are qualified. For avoiding WIP waves this is the correct procedure. But simulation has shown, this procedure would mean an extreme long total production stop exceeding the performance loss by WIP waves clearly. Based on historical data analyses, typical machine group specific durations for qualification were found.

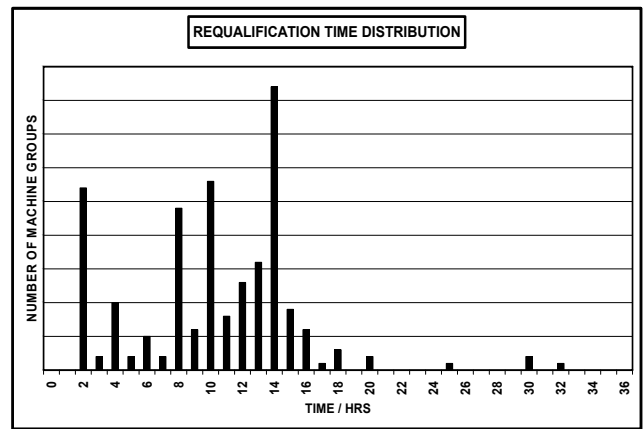


Figure 8: Typical requalification time distribution after a disturbance

Using this pattern and the schedule of the pure maintenance event, a realistic performance forecast can be created by simulation. Figure 9 shows the actual cycle time increase of different product groups due to fab maintenance event in comparison with prediction (trendlines).

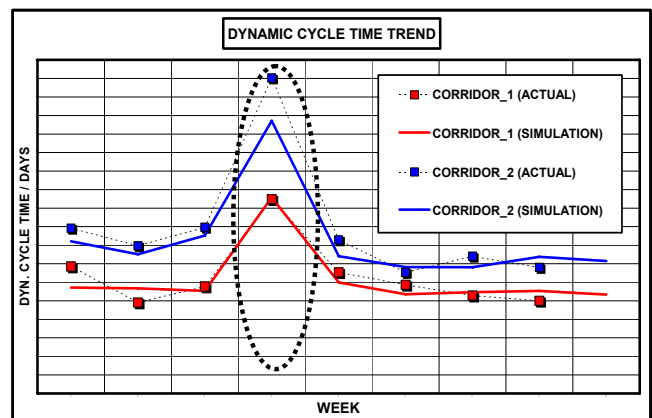


Figure 9: Cycle time forecast compared with actual data

4 CONCLUSIONS

In this paper the coping of typical disturbances in a multi-product logic fab is exemplary shown using intelligent capacity planning, performance monitoring and, in particular, prediction methods. The leading role plays the continuous mid-term forecast by transient fab simulation using typical failure patterns for modeling of expected disturbances. Consequently, future activities are focused on an extended use of prediction methods. That means, online simulation linked with Real Time Dispatcher for short-term prediction and short-term optimization of local dispatching are realized. Another topic is the validation of alternative mathematical methods for the rough, but very fast forecast in addition to established discrete-event fab simulation.

REFERENCES

- Rose, O. 1998. WIP evolution of a semiconductor factory after bottleneck workcenter breakdown. In *Proceedings of the 1998 Winter Simulation Conference*, 997-1003.
- Rose, O. 2006. Analysis of a semiconductor wafer fab with an intermediate storage for partially finished logic products. In *Proceedings of the Industrial Simulation Conference 2006*, 319-324.

BIOGRAPHY

WOLFGANG SCHOLL works as a Staff Expert for modeling and simulation for Infineon Technologies in Dresden (Germany). He has studied physics at the Technical University of Chemnitz (Germany) and graduated in solid-state physics in 1984. From 1984 to 1995 he worked as a process engineer for ZMD in Dresden. In 1996 he joined Infineon Technologies (former SIMEC) and worked on the field of capacity planning. Since 2003 he is responsible for fab simulation.