A model for master planning in semiconductor manufacturing

Conference Paper · January 2008

Thomas Ponsignon
Infineon Technologies

Lars Mönch
FernUniversität in Hagen

Proceedings of the 2008 Industrial Engineering Research Conference
J. Fowler and S. Mason, eds.

A Model for Master Planning in Semiconductor Manufacturing

Thomas Ponsignon
Infineon Technologies AG
85579 Neubiberg, Germany

Christoph Habla, Lars Mönch
Department of Mathematics and Computer Science
University of Hagen, 58097 Hagen, Germany

We present a model for making master planning decisions in semiconductor manufacturing. The problem consists in determining appropriate wafer quantities for certain products, facilities (locations), and periods of time. Different demand types, i.e., confirmed orders or forecasts are considered. The reentrant flow in semiconductor wafer fabrication facilities is taken into account for capacity modeling. We use a combined objective function that considers production costs for in-house locations, sub-contracting costs, inventory costs, and costs for unmet demand. Demand fulfillments and capacity constraints are used. We suggest a mixed integer programming (MIP) formulation for the master planning problem. The results of some computational experiments with a commercial MIP solver are presented.

Keywords
Master Planning, Routing, Mixed Integer Programming, Semiconductor Manufacturing, Computational Experiments

1. Introduction
This research is motivated by planning problems found in semiconductor wafer fabrication facilities. Given a set of confirmed orders and demand forecasts, we are interested in determining an appropriate number of wafers to be completed within a certain period of time either in-house or in sub-contractor facilities taking capacity and demand fulfillment constraints into account. This information can be used to determine an appropriate number of wafer starts for each facility, product, and period of time. The result of this planning effort is called a Master Production Schedule (MPS). The MPS transforms aggregated production, sales and operations plans that consider product families and a long time horizon into disaggregated plans for individual final products, usually for the next half year in weekly time buckets [8]. A MPS is an important entry point for the lower production scheduling and control level within the different in-house facilities and for the sub-contractors. Master planning is sometimes called supply network planning [3].
Master planning problems for semiconductor manufacturing are rarely discussed in the literature. Some papers discuss questions of capacity planning for semiconductor wafer fabrication facilities [1], [2]. However, the planning horizon is longer than for master planning, usually one or two years, and it is worked on an aggregated level based on product families for only one semiconductor wafer fabrication facility. An enterprise-wide semiconductor manufacturing strategic resource planning approach based on mixed-integer programming is presented in [7]. An aggregated model is used because the product family level of detail is applied. The suggested model shows some similarity to our model because it explicitly considers the network structure of front-end and back-end facilities in semiconductor manufacturing. A more detailed model for one front-end facility is described in [6]. A linear programming formulation and iterative discrete-event simulation are used to determine start rates for wafers of the different products
Based on this literature review, to our best knowledge master planning problems are not addressed for supply networks in the semiconductor industry. This is surprising because developing master planning procedures is an active field of research for other type of manufacturing systems (cf. [9], [10], [3] amongst others). We present the results of modeling and solving the master planning problem for a set of in-house and sub-contractor front-end facilities.

The paper is organized as follows. In Section 2, we describe the problem. We suggest a MIP for solving the problem in Section 3. Finally, the results of computational experiments are discussed in Section 4. Some ideas for future work, especially for developing heuristics, are presented in Section 5.

2. Problem Description and Assumptions

In this section, we describe the researched problem, while we present a mathematical model formulation in Section 3. We are interested in determining appropriate wafer quantities for different locations and periods of time. Semiconductor manufacturing consists of front-end and back-end operations. Front-end operations are performed in semiconductor wafer fabrication facilities, called wafer fabs, while back-end operations are performed in back-end facilities.

We consider only front-end operations. Usually, they can be performed within several in-house facilities or they can be outsourced to sub-contractors. In-house facilities have to be modeled in more detail than sub-contractors. We assume that the demand is given weekly. It consists of confirmed orders and of forecasts for products to be manufactured. Confirmed orders are fulfilled prior to forecasts. We consider backlog quantities for unmet confirmed orders. Forecast is called additional demand. When capacity is available, we are interested in fulfilling the forecasts. It is assumed that we have an inventory where finished products are stored for later fulfillment of demands. The sales quantity related to confirmed orders does not exceed the number of wafers of the confirmed orders. The additional sales quantity related to forecast has to be smaller than the forecast quantity.

Capacity modeling is crucial for master planning. In our model, we assume fixed average cycle times of the products. Given the completion time of a wafer, we can compute when a certain wafer will arrive at the bottleneck work centers. We accumulate the time that the wafer spends on processing on the machines of the bottleneck work centers for each period of time. This allows us to take the reentrant flows into account. This method is similar to capacity representation approaches used for capacity planning in semiconductor manufacturing (cf. [1], [2]). This approach is not appropriate for sub-contractors, because we do not know the bottleneck work centers of the sub-contractors. Therefore, we simply measure the number of wafers processed within a sub-contractor facility. We do not allow that the number of wafers exceed certain bounds within a sub-contractor facility.

We are interested in determining the number of wafers of product p to be completed at the end of period t in facility m. We use one week as length of a period. The master plan has a horizon of six months.

3. Mixed Integer Programming Model

In this section, we present a mixed integer programming formulation of our master planning problem.

3.1 Decision Variables, Parameters, and Objective Function

First, we introduce the necessary index sets. Therefore, the following index sets are considered:

$t \in T$:	time buckets (periods),	$1,\ldots,t_{max}$,
$p \in P$:	products,	$1,\ldots,p_{max}$,
$m^{(ih)} \in M^{(ih)}$:	in-house facilities,	$1,\ldots,m^{(ih)}_{max}$,
$m^{(s)} \in M^{(s)}$:	sub-contractors,	$1,\ldots,m^{(s)}_{max}$,
$m \in M := M^{(ih)} \cup M^{(s)}$	set of all in-house facilities and sub-contractors (set of locations),	$1,\ldots,m_{max}$,
$b^{(m)} \in B^{(m)}$	set of all bottleneck work centers for one specific location $B^{(m)}$,	$1,\ldots,b^{(m)}_{max}$,
$b \in B$	set of all bottleneck work centers,	$1,\ldots,b_{max}$,
$k \in K$	time index for capacity consumption,	$0,\ldots,k_{max}$.

Note that we obtain the quantity k_{max} by $k_{max} := CT_{max} - 1$, i.e., we reduce the maximum expected cycle time of all products by one period of time. We introduce the following decision variables into our model:

x_{pmt} :	number of wafers of product p to be completed at the end of period t in facility m,	$p \in P, m \in M, t \in T$,
I_{pt} :	inventory of product p at the end of period t,	$p \in P, t \in T$,

$s_{pt}^{(o)}$:	sales quantity of product *p* in period *t* referring to confirmed orders,	$p \in P, t \in T$,
$s_{pt}^{(fc)}$:	sales quantity of product *p* in period *t* referring to additional forecasted demand,	$p \in P, t \in T$,
B_{pt}:	backlog of product *p* at the end of period *t* referring to confirmed orders,	$p \in P, t \in T$.

We also use the following more technically motivated decision variable:

$$u_{pmt} := \begin{cases} 1, & \text{when product } p \text{ is processed in period } t \text{ in location } m, \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

The suggested model contains the following parameters:

$d_{pt}^{(o)}$:	confirmed orders for product *p* at the end of period *t* (in wafers),	$p \in P, t \in T$,
$d_{pt}^{(fc)}$:	additional forecasted demand for product *p* at the end of period *t,* (in wafers),	$p \in P, t \in T$,
I_{p0}:	initial inventory level of product *p* at the beginning of the first period,	$p \in P,$
B_{p0}:	initial backlog of product *p* at the beginning of the first period,	$p \in P,$
x_{pmt}^{i}:	number of wafers of product *p* to be completed at the end of period *t* in facility *m* (started before the first period),	$p \in P, m \in M, t \in T$,
C_{bt}^{min}:	minimum utilization of bottleneck *b* in period *t*,	$b \in B, t \in T,$
C_{bt}^{max}:	maximum capacity offer at bottleneck *b* in period *t*,	$b \in B, t \in T,$
mc_{pm}:	cost to produce one wafer of product *p* in location *m,*	$p \in P, m \in M$,
lc_{pm}:	location cost per period when location *m* is used to produce product *p* (fixed production costs),	$p \in P, m \in M$,
hc_{p}:	inventory cost for holding one wafer of product *p*,	$p \in P,$
udc_{p}:	cost due to unmet confirmed orders for one wafer of product *p*,	$p \in P,$
rev_{pt}:	expected revenue per wafer for satisfying additional demand of product *p* in period *t,*	$p \in P, t \in T$,
cc_{pmbk}:	capacity consumption of one wafer of product *p* at bottleneck *b* when this product is produced in location *m* and the completion period is *k* periods ahead,	$p \in P, m \in M, b \in B, k \in K$,
α:	large number.	

We maximize the following objective function:

$$\sum_{p \in P} \sum_{t \in T} \left\{ rev_{pt} s_{pt}^{(fc)} - \sum_{m \in M} mc_{pmt} x_{pmt} - hc_{p} I_{pt} - \sum_{m \in M} lc_{pm} u_{pmt} - udc_{p} B_{pt} \right\}. \tag{2}$$

The objective function (2) is the difference between revenue for fulfilling additional forecasted demand and manufacturing, location, inventory holding, and backlog costs.

3.2 Constraints

The following constraints are taken into account in our master planning model. First, we have to model the inventory balance. We obtain:

$$I_{pt} = I_{pt-1} - s_{pt}^{(o)} - s_{pt}^{(fc)} + \sum_{m \in M} x_{pmt} + \sum_{m \in M} x_{pmt}^{i}, \quad \forall p \in P, t \in T. \tag{3}$$

We have to relate the sales quantities to the demand quantities and backlog. We get the following constraints:

$$s_{pt}^{(fc)} \le d_{pt}^{(fc)}, \quad \forall p \in P, t \in T, \tag{4}$$

$$s_{pt}^{(o)} + B_{pt} = d_{pt}^{(o)} + B_{pt-1}, \qquad \forall p \in P, t \in T. \tag{5}$$

The next constraint is used to model capacity restrictions:

$$C_{bt}^{min} \le \sum_{p \in P} \sum_{m \in M} \left(\sum_{j=t}^{min(t+k_{max}, t_{max})} cc_{pmbj-t} x_{pmt} + \sum_{j=t}^{min(t+k_{max}, t_{max})} cc_{pmbj-t} x_{pmt}^{i} \right) \le C_{bt}^{max}, \quad \forall b \in B, t \in T. \tag{6}$$

The following constraint is used to model location costs. Whenever a unit of product *p* is completed in location *m* then location costs appear. We obtain:

$$x_{pmt} \le \alpha u_{pmt}, \qquad \forall p \in P, m \in M, t \in T. \tag{7}$$

Constraint (7) makes sure that an additional location is only used in situations when it is necessary. Non-negativity and binary conditions have to be taken into account for the decision variables. We obtain:

$$x_{pmt} \ge 0, s_{pt}^{(o)} \ge 0, s_{pt}^{(fc)} \ge 0, I_{pt} \ge 0, B_{pt} \ge 0, \quad \forall p \in P, m \in M, t \in T, \tag{8}$$

$$u_{pmt} \in \{0,1\}, \quad \forall p \in P, m \in M, t \in T. \tag{9}$$

4. Computational Experiments

We have to assess the performance of our MIP formulation by stochastically generated test instances. Therefore, we start with describing the design of experiments. The results of computational experiments are shown in Section 4.2.

4.1 Design of Experiments

In this section, we describe computational experiments. We are basically interested in the time needed for computation and in the solution quality (for a given amount of time). Therefore, we vary the number of locations, and the number of products. The used design of experiments is shown in Table 1.

Table 1 Experimental Design for Master Planning

Factor	Settings	Count
Number of periods t_{max}	24	1
Number of products P	50,100	2
Number of locations M	6 (in-house), 2 (sub-contractor) 6 (in-house), 4 (sub-contractor) 8 (in-house), 2 (sub-contractor) 8 (in-house), 4 (sub-contractor)	4
Confirmed orders $d_{pt}^{(o)}$	$U[200,300] * m_{max}$	1
Forecast $d_{pt}^{(fc)}$	$U[200,300] * m_{max}$	1
Initial inventory I_{p0}	$500 * m_{max} / p_{max}$	1
Initial backlog B_{p0}	$250 * m_{max} / p_{max}$	1
Work-in process (WIP) x_{pmt}^{i}	$400/p_{max}$ (in-house) $200/p_{max}$ (sub-contractor)	1
C_{bt}^{max}	6720 hours per period (in-house) 2000 wafers per period (sub-contractor)	1
C_{bt}^{min}	$C_{bt}^{min} = 0.9 C_{bt}^{max}$	1
Average cycle time q_{pm}	6	1
mc_{pm}	$U[10,20]$ (for 50% of the products, in-house) $U[20,40]$ (for 50% of the products, in-house) $U[30,40]$ (for 50% of the products, sub-contractor)	1

	$U[40,60]$ (for 50% of the products, sub-contractor)	
Location costs lc_{pm}	$U[300,500]$ (for 50% of the products) $U[500,1000]$ (for 50% of the products)	1
Inventory holding costs hc_p	$U[5,10]$	1
Unmet confirmed orders (backlog) costs udc_p	$U[200,240]$ (for 50% of the products) $U[300,400]$ (for 50% of the products)	
Revenue rev_p	$U[80,120]$ (for 50% of the products) $U[150,200]$ (for 50% of the products)	1
Capacity consumption cc_{pmbk}	two hours per unit (in-house) 1 wafer (sub-contractor)	1
Number of independent test instances per parameter combination		5
Total number of test instances		**40**

4.2 Results of Computational Experiments

The 40 test instances were solved with ILOG CPLEX 10.1 on a Pentium 4 CPU (3.4 GHz, 2.0 GB RAM). The calculation time was set to 10 or 30 minutes, respectively. Figure 1 depicts the solution quality for the two different calculation times by showing the MIP Gap for each test instance. The MIP Gap is the relative difference between the highest objective value and the lowest upper bound that is found during the maximization process, i.e. the guarantied solution quality equals one minus the MIP Gap for each test instance.

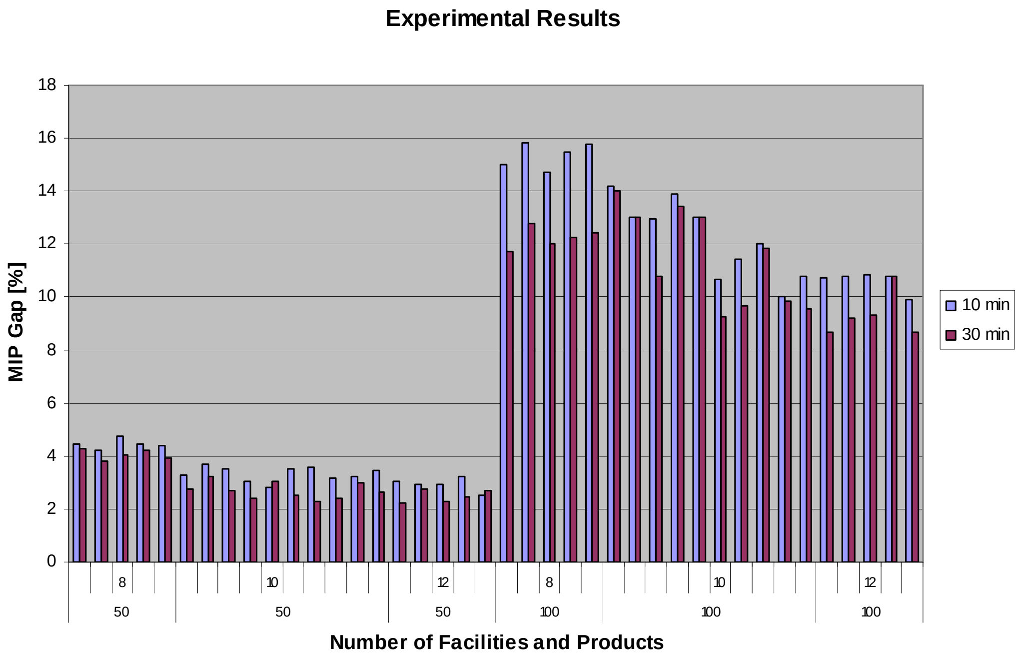

Figure 1: Experimental Results for Master Planning

As it can be seen in Figure 1, the experimental results show a good overall solution quality. The MIP Gap is less than 16 % for all experiments. The difficulty to find near-to-optimal solutions or to prove optimality, respectively,

seems to be mainly affected by the number of products – all experiments with 50 products have MIP Gaps of less than 5% while all experiments with 100 products have at least a MIP Gap of 8 %. The increased computation time of 30 minutes compared to 10 minutes did not have a great effect on the solution quality.
Real world scenarios can include several hundreds of products. To investigate the model behavior for such large scale test instances, some additional experiments with a product number of 300 were carried out. The MIP Gaps of all these test instances were more than 40% after two hours of computation time. These computational results lead to the insight that we have to look for efficient heuristics to solve the master planning problem.

5. Conclusions and Outlook to Future Research

We presented a model for master planning in semiconductor wafer fabrication supply networks. The model is formulated as a MIP. It takes demand in form of confirmed orders and forecast into account. The objective function is the revenue obtained for fulfilling forecasted demand minus the costs for production, inventory holding, and unmet confirmed orders. Furthermore, we consider location costs that are basically fixed production costs for the different facilities where a product is processed. We take the finite capacity of the semiconductor manufacturing system into account during the calculation of master plans. Computational experiments with ILOG were performed that demonstrate that the time needed for computation depends on the number of products. The computational results of the large scale test instances show that the suggested model cannot be directly used for real world scenarios when the number of products is too high, i.e. several hundreds of products.
There are some directions for future research. First of all, it seems to be necessary to include the back-end stage into the model formulation in order to perform an integrated planning of the entire semiconductor production process. Because of the large computational burden of the suggested model, we have to look for efficient heuristics. It seems to be possible to provide heuristics by a decomposition scheme. In the first step of the decomposition, appropriate wafer quantities are assigned to products and periods, while in a second step these quantities are distributed over the different facilities by using techniques from parallel machine scheduling. Finally, it seems to be interesting to consider uncertain forecasts within the model. The interaction of the forecast scheme suggested in [5] with the master planning approach should be investigated within a rolling horizon approach.

Acknowledgements

The authors would like to thank Hans Ehm and Christian Schiller, Infineon Technologies AG, Munich, for fruitful discussions on forecast and production planning issues at Infineon.

References

1. Barahona, F.; Bermon, S.; Günlük, O.; Hood, S. 2005. "Robust Capacity Planning in Semiconductor Manufacturing". Naval Research Logistics, 52, 459-468.
2. Bermon, S.; Hood, S. J. 1999. "Capacity Optimization Planning System (CAPS)". Interfaces, 29(5), 31-50.
3. Chern, C.-C., Hsieh, J.-S. 2007. "A Heuristic Algorithm for Master Planning that Satisfies Multiple Objectives". Computers & Operations Research, 34, 3491-3513.
4. Günther, H.-O. 2005. Supply Chain Management and Advanced Planning Systems: A Tutorial. Supply Chain Management und Logistik: Optimierung, Simulation und Decision Support (Günther, H.-O., Mattfeld, D., Suhl, L. (eds.), 3-40.
5. Habla, C., Drießel, R., Mönch, L., Ponsignon, T., Ehm, H. 2007. "A Short-Term Forecast Method for Sales Quantities in Semiconductor Manufacturing". Proceedings IEEE Conference on Automation Science and Engineering, 94 – 99.
6. Hung, Y.-F., Leachman, R.C. 1996. "A Production Planning Methodology for Semiconductor Manufacturing based on Iterative Calculation and Linear Programming Calculations". IEEE Transactions on Semiconductor Manufacturing, 9(2), 257-269.
7. Stray, J., Fowler, J. W., Carlyle, M., Rastogi, A. P. 2006. "Enterprise-wide Semiconductor Resource Planning". IEEE Transactions on Semiconductor Manufacturing, 19(2), 259-268.
8. Vieira, G. E. 2006. "Understanding Master Production Scheduling from a Practical Perspective: Fundamentals, Heuristics, and Implementations". Handbook of Production Scheduling, J. Hermann (ed.), Springer, 149-176.
9. Xie, J.; Lee, T. S.; Zhao, X. 2004. "Impact of Forecasting Errors on the Performance of Capacitated Multi-Item Production Systems". Computers & Industrial Engineering 46 (2004), 205-219.
10. Zobolas, G. I., Tarantilis, C. D.; Ioannou, G. 2008. "Extending Capacity Planning by Positive Lead Time and Optional Overtime, Earliness and Tardiness for Effective Master Production Scheduling". International Journal of Production Research, accepted for publication.