Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

A STEP TOWARD CAPACITY PLANNING AT FINITE CAPACITY IN SEMICONDUCTOR MANUFACTURING

Emna Mhiri
Mireille Jacomino
Fabien Mangione

University of Grenoble Alpes,
G-SCOP, F-38000
Grenoble, FRANCE

Philippe Vialletelle
Guillaume Lepelletier

STMicroelectronics
850, Rue Jean Monnet
F-38926 Crolles Cedex, FRANCE

ABSTRACT

Production planning in the Semiconductor Industry (SI) has emerged as the most complex process due to its process complexity, technological constraints and high-mix low-volume production characteristics. In this paper, we present two different production planning approaches, developed by STMicroelectronics and G-SCOP research laboratory, to better control the production in 300mm production line at Crolles. At first, a mixed integer program (MIP) is proposed that projects the production lot trajectories (start and end dates) for the remaining subsequent steps, taking into account finite production capacities. A heuristic is then proposed to simplify the problem of finite capacity by neglecting equipment capacity. This approach results in the development of an infinite capacity WIP projection engine that complies with lots due dates and takes into account cycle time variability.

1 INTRODUCTION

Semiconductor Industry (SI) is one of the most complex manufacturing domain where constant evolution in the production technologies and products result in decreasing prices. Therefore, it is critical to rapidly install capacity to deliver cutting edge products with shorter cycle times (Leachman 2012). This requires significant attention to scheduling and dispatching tools. At present, production planning and control, in the SI, mainly relies on simple tools such as spreadsheets and discrete event simulation methods. These approaches lack precision as they always overestimate capacities and underestimate cycle times (Ignizio 2012). Hence, both classical (MIP, SP,…) and new methods and tools, leveraging Operational Research techniques and modern computation power, could lead to significant benefits.

The main limitation of current capacity and production planning tools is the evaluation of variability. In the "Front-End" production lines, a.k.a. wafer fabs, several hundreds of individual production steps are required for the manufacturing of electronic components on the silicon wafers. These steps are performed on production equipment that have various design and operating modes e.g. single wafer, lot or batch, serial or parallel, and sequence dependent set-up (Mönch et al. 2011). The number of equipment (a.k.a. tools) for each type impacts the production capacities and downstream inter-arrival rate, a first dimension of variability. The production equipment breakdowns or failures are the second dimension of variability that do not only reduce production capacities but result in unstable production capacities. The reentrant flow, specificity of semiconductor production, is the third variability dimension that impacts dispatching policies. It refers to the process where same tool may be used several times during the manufacturing of a product, or that several products at different stages may be waiting in front of the same tool.

The semiconductor manufacturing production line is characterized by low or high product mix and production volume because it serves various market segments with different business models. The

production of memory components is an example of the low-mix high-volume production line where equipment redundancy enables significant reduction in the variability. However, in high-mix production, the product variety involves multiple options (number of lithography steps, ion implantation and metallization layers). This results in high level of variability induced by dispatching policies and tools that may not be qualified for the desired recipe and given production step. The production lines, operating at low volume, require tools with complex configurations (heterogeneous clusters). The investment of toolsets is very limited. As a result, we find a heterogeneous equipment park which results in high variability.

Most of European wafer fabs operate in the high-mix low-volume mode. Moreover, their capacities are generally built through incremental investment. Several technologies are developed in the same line using several types of equipment. This further complicates both execution (problem of mix and match for process control, heterogeneity of throughputs and recipe qualifications for dispatching) and capacity planning (big amount of data, unique machines and significant impact of variability). New challenges linked to the generalization of Time Constraints Tunnels, and management of small to medium volume products with different customer priorities/expected cycle time, highlight the need for a breakthrough in terms of production and capacity planning.

The ENIAC European Project "INTEGRATE" aims at the development of new tools and techniques to enable semiconductor production lines to efficiently manage a high product and technology mix. For capacity and production planning activities, main challenges to be addressed are the:

- choice of technical options when developing/industrializing new technology
- early identification of blocking points and definition of adjustment/investment strategy
- validation of start plan/commitment on due date for customer deliveries
- cycle time challenge with lots/products flowing at different speeds
- execution/scheduling and dispatching policy for line balancing

To answer these questions, three different notions and tools are generally involved:

1. **Capacity Planning**: What is the capacity needed to answer the demand?
2. **Production Planning**: What is the best production plan knowing the demand and installed capacity?
3. **WIP projection**: What is the activity needed to ensure the WIP delivery?

While (1) is generally answered using infinite capacity computation, (2) and (3) need to consider the relationship between capacity limitation and cycle time where bottleneck increase product cycle time, and delayed products create "WIP bubble" a.k.a. conflicts. In this paper, we mainly focus on WIP projection whereas traditional approaches involve two main techniques: projection at infinite capacity and discrete event simulation.

The paper is organized as follows. The standard tools and techniques used today by semiconductor manufacturers are presented in Section 2. Then Section 3 introduces a brief review of the various approaches proposed in the literature. Sections 4 and 5 present proposed approaches (i) MIP and (ii) heuristic. Finally, in Section 6, we summarize the conclusions and future perspectives.

2 CURRENT TOOLS AND METHODS USED IN SEMICONDUCTOR FACTORIES

The entrance of semiconductor manufacturing in industrial era, 30 years before, highlighted the need for capacity studies a.k.a. capacity analyses. These evaluate the ability of a production line to serve its client needs with respect to product quantities, mix and volume. These are based on the detailed computer models where production tools and steps are described with different characteristics including typical tools availability, tools qualification matrix and process throughput.

The first set of tools used for capacity computations are the spreadsheets. These evaluate an average loading of tools over time by combining computer model with a production plan, expressed as quantities

to be delivered on period and product basis. The key advantages of spreadsheets are *(i)* calculation speed, *(ii)* 'what-if' analysis may be done with instant answer, *(iii)* easiness of bug-identification *(iv)* possibility to compute optimal production plan by adding a solver to the spreadsheet and *(v)* simple answers that are easy to understand for accurate decisions. These tools do not consider cycle-time (up to three months) of semiconductor manufacturing that is significant to avoid bottleneck conflicts. The capacity spreadsheets are of interest for the production lines having few different products and stable product mix (Ignizio and Garrido 2012).

An evolution of capacity spreadsheets has given the second set of capacity computation tools. These tools, like CAPACE (acronym for Capacity Explorer) in STMicroelectronics, add cycle-time dimension to the capacity spreadsheets. It enables production plan analysis for different products, with significantly different cycle-times and product mix; however, they work well only if a correct cycle-time is included in the computer model. As a result, spreadsheet advantages are affected: *(i)* calculation speed is reduced, *(ii)* unexpected errors are difficult to analyze due to cycle-times variations that blend data between products, *(iii)* computation of optimal production plan is almost impossible due to the number of 'variables' and *(iv)* results are difficult to explain. The capacity saturation estimates by these tools require cycle time stability over one year period. Despite all of these drawbacks, these are mostly used and have over passed capacity spreadsheets.

The above presented two sets of tools belong to 'infinite capacity projection', as they proceed in two steps: *(1)* projection of products based on the cycle-time, *(2)* estimation of toolsets saturation based on the projection results. Other commonly used tools belong to the 'finite capacity projection' category, as they compute cycle-time and toolset saturation, simultaneously. The discrete-event simulation methods for the capacity analysis fall in this category. These are used with computer models as the capacity spreadsheets where production plan is input as the 'start plan'. The main advantage of these tools is that modeler do not have to model the cycle time in detail because it is an output from the computation whereas their main drawback is to push the WIP and number of starts without considering any objective function. These tools give feasible projection plans but they cannot answer to the most important question: is the *target* production plan feasible?

3 LITERATURE REVIEW

In addition to tools used in the semiconductor industry, we may find, in the literature, numerous methods and techniques used for capacity planning in the semiconductor manufacturing. Uzsoy, Lee, and Martin-Vega (1992, 1994), Gupta et al. (2006), and Mönch, Fowler, and Mason (2013) have mentioned in their reviews different capacity planning techniques used in the semiconductor environment which can be divided into infinite and finite capacity planning models.

3.1 Infinite Capacity Planning Tools

Among the methods used for infinite capacity planning, we find classical techniques which are considered most famous techniques, successfully used in many industries, such as Material Requirement Planning (MRP) (Vollmann 2005); Manufacturing Resource Planning (MRPII), push system (Rondeau and Litteral 2001); pull technique Just In Time (JIT) (Golhar and Stamm 1991) and Theory Of Constraints (TOC) (Goldratt 1990).

The application of these traditional tools for capacity planning in semiconductor industry presents some shortcomings. Indeed, it is proven that MRPII method can be inefficient and may produce unrealistic Master Production Schedules (MPS) when applied in the semiconductor manufacturing. It does not take into account capacity constraints and cannot handle uncertainty caused by machine breakdowns or other unexpected events (Rupp and Ristic 2000). The JIT technique proves its strengths (Levitt and Abraham 1990); however, it presents some limitations in the high-mix low-volume production systems. It seems to be more suitable for a repetitive production environment with stable demand and low mix product (Carlson and Yao 1992). The TOC seems efficient capacity planning technique in

semiconductor industry (Rippenhagen and Krishnaswamy 1998) but it considers only bottleneck resources.

In addition to these classical methods, new techniques are also proposed such as the method of Guan et al. (2008) that is based on a mathematical programming for a dynamic formation of virtual flow paths and TOC/DBR (Theory Of Constraints/Drum-Buffer-Rope) mechanism with group scheduling used for the control over each flow path.

There are also other capacity planning systems developed to determine lot release time, start fab, and capability of the equipment for multiple semiconductor production lines on the basis of pull philosophy and the assumption of infinite equipment capacity (Chen et al. 2005; Chen, Fan, and Chen 2009). Recently, authors have shifted their focus to linear programming models for capacity planning problems. The goal is to find production quantities taking into account demand, capacity restrictions, and current WIP status with an objective to maximize revenue for forecasted orders by minimizing production costs, inventory holding costs, and costs for unfulfilled orders (Habla and Mönch 2008).

These techniques present several advantages such as the determination of bottleneck resources and compliance with due dates. However, they consider a constant average cycle time for processing steps which is independent of the workstation utilization.

3.2 Finite Capacity Planning Tools

Under finite capacity planning, the tool workload must be lower than its capacity whereas orders defined in the initially forecasted production plan may not respect predefined sequences or even meet due dates. The most relevant techniques to establish finite capacity schedules for the semiconductor production lines are linear programming models and heuristics.

3.2.1 Linear Programming Models

The linear programming approach is widely applied to the specific issues encountered in capacity planning for the semiconductor industry (Leachman and Carmon 1992, Hung and Leachman 1996).

The objective functions corresponding to the linear programs consist of optimizing the allocation of strategic resources to maximize long-term profit (Bermon and Hood 1999); minimizing production costs (Catay et al. 2003); minimizing total weighted lots tardiness (Habla, Mönch, and Drießel 2007) or maximizing profit (Ponsignon and Mönch 2012), taking into account capacity constraints.

LPs may require a very long time to generate the input data and huge amounts of memory and disk space to store the data (Sullivan and Fordyce, 1990). Thus, they are generally combined with heuristics (Leachman et al. 1996; Ponsignon and Mönch 2012) or Lagrangian relaxation (Catay, Erenguc, and Vakharia 2003, Habla, Mönch, and Drießel 2007) to reduce the execution time.

3.2.2 Heuristics

Approximation methods have been generally used to develop finite capacity planning systems for the semiconductor industry. These systems consider different inputs e.g. delivery due dates and resource capacity (Rupp and Ristic 2000; Chen, Chen, and Lin 2008) and the WIP status (Liu et al. 2011); and outputs e.g. orders release date at bottleneck positions (Horiguchi et al. 2001); order release time, start date, and equipment capability for each order (Chen, Chen, and Lin 2008); and mean and variance of the total cost (excess inventory, unused capacity, penalties…) associated with a release plan (Liu et al. 2011). Moreover, we find more recourse to meta-heuristics especially genetic algorithms (Ponsignon, Habla, and Mönch 2008; Ponsignon and Mönch 2009) to solve production planning problems in the semiconductor manufacturing.

The finite capacity planning techniques can be divided into two categories as *(i)* long-term strategic capacity planning methods (Bermon and Hood 1999; Catay, Erenguc, and Vakharia 2003) and *(ii)* mid-

term tactical master planning tools (Ponsignon and Mönch 2012). These techniques are developed for specific purposes such as minimizing delays, reducing cycle time, defining the completion date of customer orders, etc.

In this study, the mid-term tactical planning at finite capacity is considered and we base on the current techniques as mentioned in the literature to resolve the problem.

4 THE FIRST APPROACH

Besides the limitations of Mixed Integer Program (MIP) based models, cited in the literature, we consider this method as a first approach and propose a finite capacity production planning model. This technique is used because it is an exact method and allows us to reformulate the problem mathematically. The basic approach is simple and has been implemented with some variations and extensions, in several commercial production planning systems. It is also inspired from the work of Habla, Mönch, and Drießel (2007); however, our model is not limited to bottlenecks positions. The formulation of the MIP model is given by:

Sets

l	$\in$	$\{1..L\}$	Lot
s_l	$\in$	$\{1..S_l\}$	Steps of lot l
i	$\in$	$\{1..I\}$	Station family
t	$\in$	$\{1..T\}$	Period

Parameters

$p_{sl,i}$	$\forall$	$s_l \in \{1..S_l\}$, $i \in \{1..I\}$	Process time of step s_l on station family i
q_{sl}		$s_l \in \{1..S_l\}$	Waiting time on the equipment before the start of the process step s_l
o_{sl}	$\forall$	$s_l \in \{1..S_l\}$	Time to process a step s_l to add to the first wafer
$Tr(s_l,s_l')$	$\forall$	$s_l, s_l' \in \{1..S_l\}$	Transfer time between two successive steps s_l and s_l' of the same lot l where step s_l' is after step s_l
do_t	$\forall$	$t \in \{1..T\}$	Working time during the period t
d_l	$\forall$	$l \in \{1..L\}$	Due date of lot l
W_l	$\forall$	$l \in \{1..L\}$	Weight of lot l (priority)
M_t	$\forall$	$t \in \{1..T\}$	Number of moves in period t
$K_{i,t}$	$\forall$	$i \in \{1..I\}$, $t \in \{1..T\}$	Capacity of a station family i in period t

Decision variables

$x_{l,sl,t}$	$\forall$	$l \in \{1..L\}$, $s_l \in \{1..S_l\}$, $t \in \{1..T\}$	=1 if lot step s_l is executed in period t
t_{sl}	$\forall$	$s_l \in \{1..S_l\}$	Release date of step s_l
C_l	$\forall$	$l \in \{1..L\}$	Completion time of lot l
L_l	$\forall$	$l \in \{1..L\}$	Lateness of lot l

MIP

$$\text{Minimize} \quad \sum_{l=1}^{L} W_l L_l \tag{1}$$

$$\sum_{t=1}^{T} x_{l,sl,t} = 1 \qquad \forall \quad l \in \{1..L\},\ s_l \in \{1..S_l\} \tag{2}$$

$$\sum_{l=1}^{L}\sum_{sl=1}^{Sl} p_{sl,i}\, x_{l,sl,t} \le K_{i,t} \qquad \forall \quad t\in\{1..T\},\ i\in\{1..I\} \tag{3}$$

$$t_{sl} \ge t_{sl-1} \qquad \forall \quad s_l\in\{2..S_l\} \tag{4}$$

$$t_{sl} = \sum_{t=1}^{T} x_{l,sl,t} * t \qquad \forall \quad l\in\{1..L\},\ s_l\in\{1..S_l\} \tag{5}$$

$$C_l = t_{Sl} \qquad \forall \quad l\in\{1..L\} \tag{6}$$

$$\begin{cases} L_l \ge 0 \\ L_l \ge C_l - d_l \end{cases} \qquad \forall \quad l\in\{1..L\} \tag{7}$$

$$x_{l,sl,t} \in \{0,1\} \qquad \forall \quad l\in\{1..L\},\ s_l\in\{1..S_l\},\ t\in\{1..T\} \tag{8}$$

The objective function (1), in the model, minimizes the total weighted tardiness. The term (2) assures that each step is executed only once whereas term (3) defines the capacity constraint. The term (4) assures succession of processing steps and term (5) defines the release date for each step. The process completion time of each lot is presented by term (6) whereas term (7) defines the lot lateness $L_l=max\ (0,\ C_l-d_l)$. The term (8) is the binary constraint for the decision variable.

The mathematical model presented above was solved by ILOG CPLEX solver ILOG OPL STUDIO. We ran our experiments on an Intel ® Core ™ 2 Duo PC with 3 GHz processor and 4 GB of RAM. Good results are obtained while testing MIP on instances of the reduced size: the production schedule obtained respected the capacity of resources and the priority and target delivery dates of the lots. It gave the exact optimal solutions for lots achieved before the end of planning horizon, taking into account the reentrant flow. However, MIP is not implemented for lots with due dates beyond this limit.

Further, increasing the size of the tested instances (up to about 1000 steps plan), the solution of MIP was halted as it requires a large amount of time and computer memory (Figure 1).

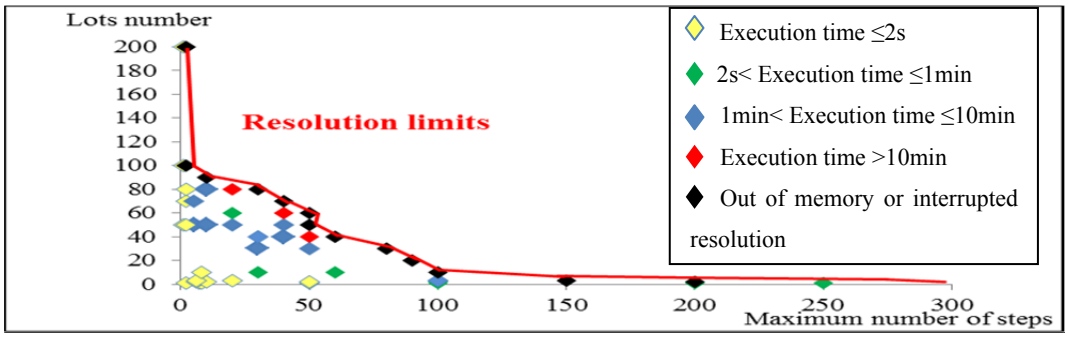

Figure 1: Limits of the MIP solution.

Thus, above proposed MIP do not have a feasible solution to the operational perspectives for our case study at STMicroelectronics due to its large size. Even decomposition approach couldn't allow reducing the computational burden for this large scale production planning problem. It is an approach that divides the problem into small sub-problems and solve them successively (Ponsignon and Mönch 2012).

5 SECOND APPROACH

The MIP based approach, presented in section 4, concluded that exact method could not afford an optimal solution, in terms of execution time, due to its huge number of variables and problem complexity. Thus, we propose another approach to simplify the problem and in first place neglect capacity constraints. An infinite capacity planning engine is then developed that considers cycle time variability, equipment saturation and target delivery dates. It forecasts trajectories (start and end dates) for each lot and computes daily activities. In these projections, each lot has its own cycle-time model that computes process times for individual steps based on necessary and sufficient speed for the remaining steps to achieve lots due dates and common fine-tuned reference cycle time curves. The proposed WIP projection engine is different from the current computation tool CAPACE being used in STMicroelectronics. It takes current WIP at lot level, lots due dates and a new model to compute the objective cycle time per step ($CTobj_{step}$) based on a semi- empirical formula named 'Zachka formula' as inputs. This formula takes into account theoretical cycle time of the step ($CTTH_{step}$) that corresponds to the processing time, theoretical cycle time of route ($CTTH_{route}$) that corresponds to the sum of remaining step times and objective cycle time of each route ($CTobj_{route}$) that takes into account queuing times, as presented below:

$$CTobj_{step} = CTTH_{step} \times Xfactor_{step}^{Zachka}$$

With

$$Xfactor_{Step}^{Zachka} = \frac{CTTH_{route} \times \left(\frac{CTobj_{route}}{CTTH_{route}} - 1\right)}{\sum \sqrt{CTTH_{route} \times CTTH_{step}}} + 1$$

A penalty is added to Zachka Xfactor to take into account the saturation of bottleneck tools that is not considered in the Zachka formula. The Figure 2 illustrates the distribution of final Xfactor considered in STMicroelectronics for cycle time model versus theoretical cycle time in the production process.

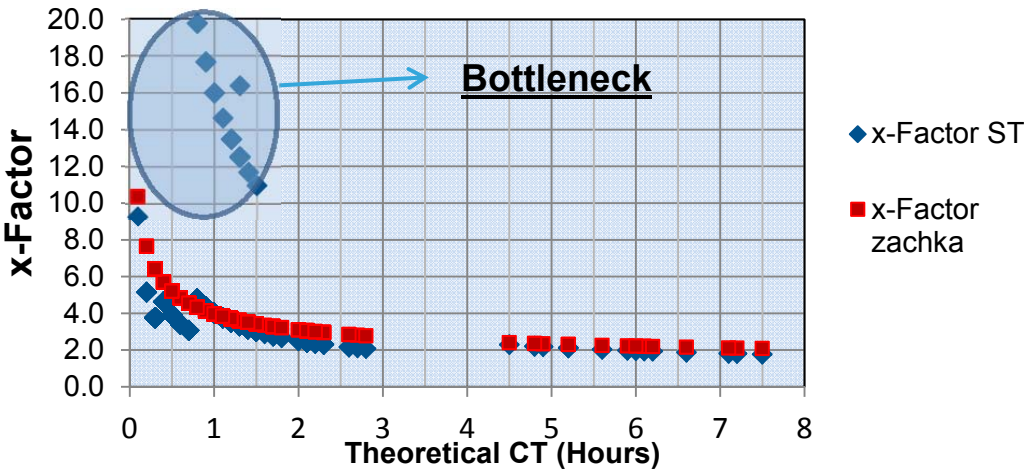

Figure 2: Distribution of Xfactor vs theoretical cycle time.

The proposed WIP projection engine provides number of moves per day (*MovesIn* for wafers entering an operation, *MovesOut* for those completing it), quantity of WIP at the beginning and at the end of each day and cycle time coefficient per lot considering the due date for each lot, as outputs. This is based on an iterative algorithm, as presented below:

1. **For each route**, compute the objective cycle time for each of its steps using the considered cycle time model.

2. **For each lot**, from its position in its route:
   2.1. compute parameters
   - Compute the remaining objective cycle time (*RemCTobj*) from the current step to the end of the route : $RemCTobj = CTobj_{route} - \sum_{s=1}^{s=id_{waiting\ step}-1} CTobj_s$
   - Compute the remaining cycle time (*RemCT*) from the date of extracting WIP (*WIP_date*) to the lot due date (*Due_date)*: *RemCT = Due_date – WIP_date*
   - Compute the cycle time coefficient (*C*) : *C= RemCT/ RemCTobj*

   2.2. Project steps for each lot
   - Determine the start and the end date of each step of the route. The cycle time for each step is equal to $C\ X\ CTobj_{step}$.
3. Compute the number of *MovesIn*, *MovesOut* and the quantity of WIP in the beginning and the end of a day.

This new projection engine is developed in JAVA. After WIP projection, this engine uses current "recipe balancing" and capacity computation modules of CAPACE to define equipment saturation. To explain the difference between WIP projection principles used by MRP, CAPACE and new engine, two simple examples are developed and tested. We consider one lot with three remaining steps A, B and C. The Table 1 contains cycle time for each step using different tools. The cycle time using MRP approach is equal to the sum of process time and the average queuing time derived from historical data. For CAPACE, the cycle time is constant for each step; it is equal to the route's "reference" cycle time divided by the number of its steps. With our new approach, the route's reference cycle time is allocated according to the Zachka + Saturation Factor, which leads to significantly longer steps for bottlenecks.

Table 1: Comparison of cycle times using various tools.

Remaining step	**Cycle Time** (MRP)	**Cycle Time** (CAPACE)	**Cycle Time** (Zachka)	**Cycle Time 1st instance** (New engine)	**Cycle Time 2nd instance** (New engine)
A	• 1,588	2,969	8,909	5,625	1,534
B	0,181	2,969	0,775	0,489	0,133
C	1,383	2,969	7,738	4,885	1,332

For the first instance, we consider a comfortable margin of 11 days between the date of WIP extraction and the lot's due date. So the cycle time coefficient used by our engine is equal to 11/ (8,909+0,775+7,738) =0,631. Figure 3 illustrates the different projection results. It demonstrates that the new projection engine, for this instance, allows the extension of steps queuing times, when we are far from the due date.

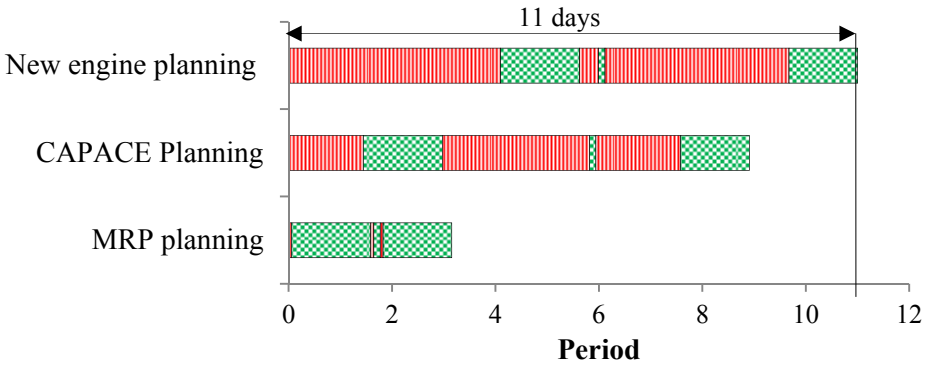

Figure 3: Difference between types of planning in case of large margin between WIP date and due date (processing time in green, waiting time in red).

In the second instance, we reduce the margin between the extraction date of WIP and the lot's due date from 11 to just 3 days. So, the lot's cycle time coefficient becomes equal to 3/(8,909+0,775+7,738)= 0,172. In this case, as illustrated in Figure 4, projections using both MRP and CAPACE lead to delays (significant in the case of CAPACE) while the new engine shrinks steps cycle times in order to satisfy the lot's due date.

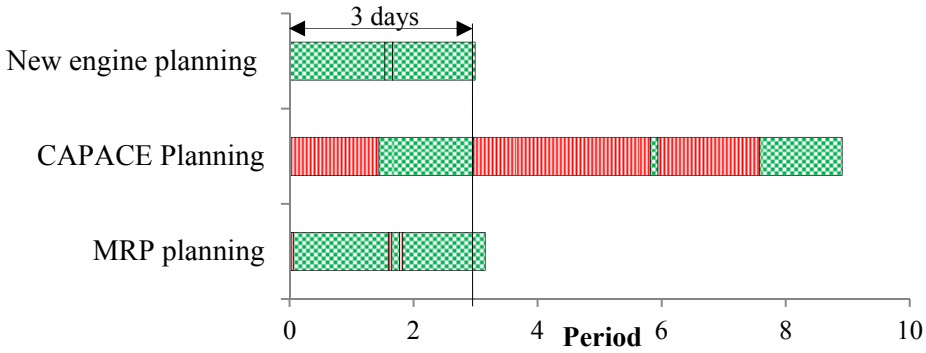

Figure 4: Difference between types of planning in case of strait margin between WIP date and due date (processing time in green, waiting time in red).

The difference between the principles of projection of different tools is explained by the use of different cycle time models for each one. Indeed, MRP uses a cycle time model based on the theoretical one. The cycle time model used by CAPACE is based on a constant homothetic distribution. However, the cycle time model used by the new engine takes into account the variability of the process, equipment saturation and lots due dates.

Hence, new engine leads to a projection which is closer to the targeted behavior than current software tools because of the notion to comply with agreed due dates. In a sense, it delivers the plan that needs to be done to fulfill customer delivery commitments. However, solution still may not be feasible as it runs at infinite capacity and after WIP projection, we may find equipment with workload exceeding capacity.

Nevertheless, this new approach of WIP projection is much faster in terms of execution. It takes approximately five seconds to project more than 2000 lots with 200 to 300 remaining steps for each one, instead of 5 minutes with the current solution. Knowing that cycle time variability is also taken into account as an input, the new tool looks very promising.

6 CONCLUSION AND PERSPECTIVES

In this paper, we proposed two production planning approaches for semiconductor production facilities. The first approach consists of a mixed integer program (MIP) that takes finite capacity considerations into account; however, its resolution requires long execution time. Even decomposition heuristics could not resolve this problem. The results of preliminary computational experiments show that we are far from the execution time required for the realistic cases.

In the second approach, the problem is further simplified and an infinite capacity WIP projection engine is developed. It respects the lots due dates and its design enables to take into account the variability of cycle time. The execution time obtained with this approach isinteresting for its application in real time situations.

Despite the strengths of new projection engine in terms of projection specificities and execution time, it runs at infinite capacity. So, after WIP projection, equipment may be over saturated. Therefore, in future research, we will add a load balancing module after WIP projection to address the oversaturated equipment. This module will sort bottlenecks, identify the oversaturation causes and shift lots to minimize the workload for each period. Then, leveraging on an iterative method, it will compute projection for the next period according to new cycle time estimates.

The new WIP projection engine uses a cycle time model based on a semi empirical formula to take into account the product mix and equipment saturation. Although obtained cycle times are approximately close to the realistic case, it is interesting to define other efficient techniques to predict cycle times for all steps. The iterative simulation-optimization scheme, introduced by Hung and Leachman (1996), where simulation model estimates cycle times that are further used as input parameters in an optimization model, can be an interesting approach. However, we intend to use Entitlement Cycle Time (ECT) estimation technique, defined by Leachman (2012), as it is based on queuing theory, considering process mix and volume.

ACKNOWLEDGMENTS

This work is supported by the ENIAC European Project "INTEGRATE". The authors also acknowledge the industrial engineering team at STMicroelectronics, Crolles for their support on the knowledge of semiconductor industry.

REFERENCES

Bermon, S., and S. Hood. 1999. ”Capacity optimization planning system (CAPS).” *Interfaces* 29(5):31–50.

Carlson, J. G., and A. C.Yao. 1992. *“*Mixed model assembly simulation.” *International Journal of Production Economics* 26(1-3):161–167.

Catay, B., S. S. Erenguc, and A.J. Vakharia. 2003. “Tool capacity planning in semiconductor manufacturing.” *Computers & Operations Research* 30(9):1349–1366.

Chen, J. C., C. W. Chen, C. J. Lin, and H. Rau. 2005. “Capacity planning with capability for multiple semiconductor manufacturing fabs.” *Computers and Industrial Engineering* 48(4):709-732.

Chen, C.W., J. C. Chen, and C. J. Lin. 2008. “Finite capacity requirements planning with equipment capability and dedication for semiconductor manufacturing.” In *Proceedings of the 9th Asia Pacific Industrial Engineering & Management Systems Conference*, 1310-1319.

Chen, J. C., Y.-C. Fan, and C.-W. Chen. 2009. “Capacity requirements planning for twin Fabs of wafer fabrication.” *International Journal of Production Research* 47(16):4473-4496.

Goldratt, E. M. 1990. *Theory of Constraints: What is this thing called Theory of Constraints and how should it be implemented*. Croton-on-Hudson, North River, New York.

Golhar, D. Y., and C. L. Stamm. 1991. "The Just-in-time philosophy: a literature review." *International Journal of Production Research* 29(4):657–676.

Guan, Z., Y. Peng, L. Ma, C. Zhang, and P. Li. 2008. "Operation and control of flow manufacturing based on constraints management for high-mix/low-volume production." *Frontiers of Mechanical Engineering in China* 3(4):454-461.

Gupta, J. N. D., R. Ruiz, J. W. Fowler, and S. J. Mason. 2006. "Operational planning and control of semiconductor wafer fabrication." *Production Planning and Control* 17(7):639–47.

Habla, C., L. Mönch, and R. Drießel. 2007. "A Finite Capacity Production Planning Approach for Semiconductor Manufacturing." In *Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering*, 82-87.

Habla,C., and L. Mönch. 2008. "Solving volume and capacity planning problems in semiconductor manufaturing: a computational study." 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, 2260-2266. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Horiguchi, K., N. Raghavan, R. Uzsoy, and S. Venkateswaran. 2001. "Finite-capacity production planning algorithms for a semiconductor wafer fabrication facility." *International Journal of Production Research* 39(5):825-842.

Hung, Y.-F., and R. C. Leachman. 1996. "A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations." *IEEE Transactions on Semiconductor Manufacturing* 9(2):257–269.

Ignizio, J. P., and H. Garrido. 2012. "Fab Simulation and Variability." *Future Fab International* 41:41-45.

Leachman, R. C., and T. F. Carmon. 1992. "On capacity modeling for production planning with alternative machine types." *IIE Transactions* 24(4):62-72.

Leachman, R. C., R. F. Benson, C. Liu, and D. J. Raar. 1996. "IMPRESS: an automated production planning and delivery quotation system at Harris Corporation-semiconductor sector." *Interfaces* 26(1):6–37.

Leachman, R. C. 2012. "The Engineering Management of Speed." In *Proceedings of the 2012 Industry Studies Association Annual Conference.* Pittsburgh, PA: University of Pittsburgh.

Levitt, M. E., and J. A. Abraham. 1990. "Just-In-Time Methods for Semiconductor Manufacturing." *Proceedings of the 1990 Advanced Semiconductor Manufacturing Conference*, 3-9.

Liu, J., C. Li, F. Yang, H. Wan, and R. Uzsoy. 2011. "Production planning for semiconductor manufacturing via simulation optimization." In *Proceedings of the 2011 Winter Simulation Conference,* edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 3612 – 3622. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mönch, L., S. Dauzère-Pérès, S. J. Mason, and O. Rose. 2011. "A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations." *Journal of Scheduling* 14 (6): 583-599.

Mönch, L., J. W. Fowler, and S.J. Mason. 2013. *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems.* Springer.

Ponsignon, T., and L. Mönch. 2012. "Heuristic approaches for master planning in semiconductor manufacturing." *Computers and Operations Research* 39(3):479–491.

Ponsignon, T., C. Habla, and L. Mönch. 2008. "A Model for Master Planning in Semiconductor Manufacturing". In *Proceedings of the 2008 Industrial Engineering Research Conference*, edited by J. Fowler and S. Mason, 1592-1597.

Rippenhagen, C., and S. Krishnaswamy.1998."Implementing the theory of constraints philosophy in highly reentrant systems." In *Proceedings of the 1998 Winter Simulation Conference,* edited by D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, 993-996. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rondeau, P. J., and L. A. Litteral. 2001. "Evolution of manufacturing planning and control systems: from reorder point to enterprise resource planning." *Production and Inventory Management Journal* 42(2):1–7.

Rupp, T. M., and M. Ristic. 2000. "Fine planning for supply chains in semiconductor manufacture." *Journal of Materials Processing Technology* 107(1):390-397.

Sullivan. G., and K. Fordyce. 1990. "IBM Burlington's Logistics Management System." *Interfaces* 20(1):3-64.

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* 24(4):47–60.

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* 26(5):44–55.

Vollmann, T. 2005. *Manufacturing Planning and Control Systems for Supply Chain Management*. New York: McGraw-Hill.

AUTHOR BIOGRAPHIES

EMNA MHIRI is PhD student in G-SCOP Laboratory (www.g-scop.grenoble-inp.fr). She holds industrial engineering degree from Tunisia, in 2012 and completed her master's degree in industrial engineering from Grenoble University, France in 2013. Her research interests include capacity planning in semiconductor industry. Her email address is Emna.Mhiri@grenoble-inp.fr.

MIREILLE JACOMINO is professor in G-SCOP Laboratory. She is carrying out her research in combinatorial optimization of systems. Her application fields are manufacturing and energy systems, used particularly in execution context of systems. Her works aim at computing control decisions that guaranty performance during execution. Robust control and robust decision are the key research interests of professor Jacomino to address the uncertainties. Her email address is Mireille.Jacomino@grenoble-inp.fr.

FABIEN MANGIONE is assistant professor in G-SCOP Laboratory. He has a PhD in industrial engineering and works on production planning, particularly on industrial case studies. His research also deals with lot sizing problems on supply chain modeling. His email address is fabien.mangione@grenoble-inp.fr.

PHILIPPE VIALLETELLE is principal staff engineer at the Industrial Engineering department of STMicroelectronics Crolles300. He is in charge of the definition and follow-up of collaborative projects in Manufacturing Sciences. His fields of interest cover production planning and management techniques, process control and Big data. His email address is philippe.vialletelle@st.com.

GUILLAUME LEPELLETIER is senior project leader at STMicroelectronics. He received Engineering degree in Operations and Production Management from INSA de Lyon, France and Master of Science in "Advanced Modeling Systems" from Brunel University, Uxbridge, UK in 1997. He has 15 years of professional experience in Industrial Engineering in the semiconductor industry. He is working on capacity planning, cycle time management, discrete event simulation, industrial reporting and equipment performance tracking. His email address is guillaume.lepelletier@st.com.