Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

CAPACITY PLANNING CHALLENGES IN A GLOBAL PRODUCTION NETWORK WITH AN EXAMPLE FROM THE SEMICONDUCTOR INDUSTRY

Sarah Biwer
Michael Filipek

Infineon Technologies AG
Am Campeon 1-15
Neubiberg, 85579, GERMANY

Emel Arikan
Werner Jammernegg

WU Vienna University of Economics and Business
Welthandelsplatz 1
Vienna, 1020, AUSTRIA

ABSTRACT

In the global, volatile semiconductor market classical capacity planning approaches need to be improved to stay competitive. Resource bottlenecks constrain production, limit the availability of finished product which can be promised to the customer and cause different utilization levels. Bottlenecks may vary periodically due to changes in supply and demand. One bottleneck can have an impact on a whole global production network. How can bottlenecks be identified and allocated in the capacity plan of global enterprises with alternative production routes and daily updated, integrated planning across sites? This paper reviews how capacity planning is covered in the literature, focusing on the semiconductor industry. Local vs. global capacity planning in global networks is identified as an area where there is limited research available so far. A case study of a semiconductor company is used to show the relevance of that area and what solutions have been found by practitioners so far.

1 INTRODUCTION

Supply chains in the semiconductor industry are faced with long production lead times of up to 26 weeks. High investment costs require economies of scale and an optimal resource utilization. Costs for a state-of-art wafer fab can reach up to $15-20 billion (Ellis et al. 2017) and even a single machine can cost up to $40 million with lead times of up to 18 months (Barahona et al. 2005; Mönch et al. 2013). A high rate of innovation (Chen et al. 2017) as well as short product life cycles make ongoing investments into new equipment necessary. New machines may be capable to produce older products, but not vice versa (Karabuk and Wu 2003). These high costs and lead times have a specifically high impact in an environment where there is also uncertainty in both supply and demand: With respect to demand, the semiconductor industry needs to cope with variability and the bullwhip effect, whose amplitude is magnified from the end-customer towards the upstream part of the supply chain where semiconductor production is typically located (Chen et al. 2013). In terms of supply, varying yield factors, lead times, potential machine breakdowns and quality issues have to be taken into account for planning. The complexity of production processes and the global supply chain pose more challenges from a planning perspective. The four main production processes in Front-End (wafer fabrication and probing) and Back-End (assembly and testing) need to be well aligned. Apart from in-house manufacturing, silicon foundries and subcontractors are to be integrated in all four phases.

Capacity planning has been recognized as an important factor for addressing these challenges. In short, the supply chain has to be set up at its best and has to run as smoothly as possible throughout the global production network.

Planning in general can be defined as a systematic and rational process based on incomplete information for the solution of decision problems (Domschke and Scholl 2005, p. 25). Capacity Planning is typically used in the mid-term for estimating how many resources are needed to meet a given demand or, respectively, how much a given set of resources can produce with acceptable performance (Uzsoy et al. 2018). This paper also takes a look at shorter term capacity planning, where investment decisions have already been made and the task is to find an optimal way for capturing the capacity of available resources in order to support an efficient allocation of competing demands.

Resources are defined as the entities of a process that a flow unit has to visit as part of its transformation from input to output. Capacity measures the maximum number of flow units that can be supported by a resource at a given point in time. A bottleneck is the resource with the lowest capacity (Cachon and Terwiesch 2011). Resources can be different in each company and each process, but for the semiconductor industry those resources mainly represent tool sets and machines, due to their long lead times; staff availability is typically not a major constraint and therefore not the main focus of this paper.

Figure 1 illustrates the scope of this paper. Many sources can be found for capacity planning in the semiconductor industry, as represented by intersection “C”. While most literature takes on a local perspective for capacity planning though, this analysis includes the aspect of the global network.

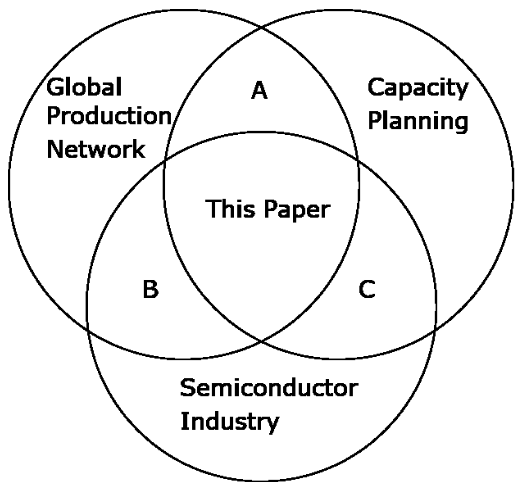

Figure 1: Dimensions and focus of this paper.

2 LITERATURE REVIEW

Cachon and Terwiesch (2011) visualize in Figure 2 how a production process can be constrained by both supply and demand. In the case of a supply constraint as visualized in Figure 2, this bottleneck determines the maximum flow rate.

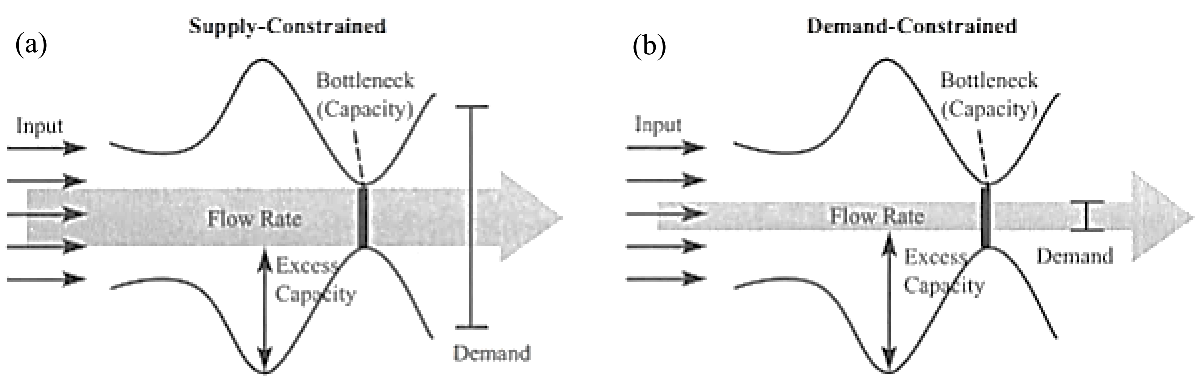

Figure 2: Bottleneck Visualization (a) Supply-Constrained (b) Demand-Constrained (Cachon and Terwiesch 2011).

Planning helps decision makers select the best future activity among a number of alternatives. Any planning activity typically faces three major challenges: Conflicting objectives and preferences between the alternatives, the complexity caused by the number of possible alternative actions and uncertainty. Fleischmann et al. (2008) point out three main characteristics of Advanced Planning Systems: They take an integrative approach to planning, considering a whole network or supply chain; they apply true optimization for a clearly defined problem, objective and constraint; finally a hierarchical planning system is deployed, which is necessary as a trade-off between practicability and the representation of interdependencies in the planning tasks.

Researchers recommend to take an integrated approach to capacity planning, which links to other planning processes such as network design, demand planning, production planning and master planning as described by Fleischmann et al. (2008). In integrated planning, all plans are well harmonized to prevent major mismatches from happening in the first place. A tutorial for central planning, showing current state, including cycle times and capacities, with possible solutions is well illustrated in Fordyce et al. (2012). Fordyce and Milne (2012) provide a decision grid for demand-supply networks which differentiates relevant decision first by planning hierarchy from strategic down to daily operational adjustments and second by the responsible unit, i.e. central planning vs. factory planning.

The concept of an integrated approach advanced the Materials Requirements Planning (MRP) logic, which calculates back production requests based on a product's Bill of Materials without adjusting to production constraints. Such weakness led to the development of the MRP2 logic, which allows for plan adjustments through bottom-up information feedback loops (Mönch et al. 2013).

Broadly speaking, capacity planning approaches can be categorized by the planning horizon and the corresponding method applied.

2.1 Capacity Planning for Different Planning Horizons

Regarding the planning horizon in scope, strategic approaches can be found for the long-term perspective and more tactical/operational approaches have been investigated for the mid- and short-term. On the strategic level, capacity expansion and investment decisions are covered to support network planning. Geng and Jiang (2009) provide a good overview of literature related to strategic capacity planning in the semiconductor industry. Tactical and operational approaches assume the capacity is given and try to optimize the loading, aiming for objectives like high utilization, low cost, high service levels, etc.

Various approaches may appear in use on the strategic vs. operational level, although some are more common for one planning horizon or the other. For the short- to mid-term horizon, capacity planning is often integrated in the form of constraints for master planning. Queueing models and simulations have proven helpful on an operational level while deterministic, mathematical programming is used more commonly for strategic capacity investment decisions. It is generally agreed that an "ideal" capacity planning tool should combine multiple methods in a single tool, which is still easy to implement, understand and use (Fowler and Robinson 1995).

2.2 Capacity Planning Methods

Methods applied range from static approaches, over mathematical programming methods, heuristics such as neighborhood search methods, queueing models and simulations to a variety of mixed approaches. Few sources deal specifically with the determination of bottlenecks in isolation: Cachon and Terwiesch (2011) determine bottlenecks by calculating the implied utilization; Chiang et al. (2000) focus on the sensitivity of the system production rate to the machine's reliability parameter. Static methods are often rather simple and based on a spreadsheet application, like the capacity plan tool by Occhino (2000). Mathematical optimization may include linear programming, as applied for instance in Mönch and Habla (2008) and Romauch and Hartl (2017) or mixed-integer programming as shown in Bermon and Hood (1999), Christoph Habla et al. (2007), Ponsignon et al. (2008) and Tavaghof-Gigloo et al. (2016). Mixed integers are especially common on the strategic level where yes/no investment decisions have to be made. In addition, stochastic programming methods have been introduced to account for non-deterministic variations on both the supply and demand sides, for example in Christie and Wu (2002) and Leung and Wu (2004). Neighborhood search methods have been developed to achieve better computation times. In addition, these models make it easier to reflect the dynamic nature of the real world network. Specifically, a method to capture rolling demand forecasts for solving the capacity planning problems was presented in a dynamic optimization model by Chien et al. (2012). Queuing models deal with the variation of cycle times depending on the utilization, for example in Connors et al. (1996) or Kumar and Kumar (2001). At IBM's semiconductor fab in New York a model called Enterprise Production Planning and Optimization System (EPOS), based on queueing models, was developed and implemented (Brown et al. 2010). Examples for heuristics in Capacity Planning are provided by Kallrath and Maindl (2006), Mhiri et al. (2015), Ponsignon and Mönch (2012), and Ponsignon and Mönch (2014). Heuristics offer the additional advantage that their way of functioning may be easier to be explained and communicated, while mathematical programming is often seen as a black box. In recent years, mixed approaches have become very popular since they aim at combining the strengths of different approaches while mitigating the weaknesses. For example, Kim and Lee (2016) introduce a 5-step integration model; Barahona et al. (2005) mix a stochastic program with a branch-and-bound heuristic; Denton et al. (2006) combine a mixed-integer program with heuristics. But the idea of a mixed approach is not new: Already in 1994 IBM started using Operations Research techniques to develop intelligent models for their supply and demand match, interweaving linear programming with a traditional MRP run and a heuristic matching process (Lyon et al. 2001).

Capacity also depends on the product mix that is being produced on the resources, because different products may consume different capacities of the same resource. One product might be produced faster than another one on the same resource. Product mix optimizations are addressed by linear programming in Romauch and Klemmt (2015) for a single facility. Rowshannahad et al. (2015) present indicators called Capacitated Flexibility Measures and Capacity Deviation Ratios to evaluate the workload balance of a toolset to manage recipe-to-machine qualifications.

2.3 Capacity Planning for Global Production Networks

While it is generally agreed that simplifications have to be made for the problems being manageable and computationally efficient, models have to be evaluated under realistic conditions, where some of the underlying assumptions may not hold true and may require careful refinement. To reduce the problem down to a manageable level, most approaches only look at one facility, not a global network with multiple plants. There is limited research done for capacity planning across the entire supply chain (Uzsoy et al. 2018). This is a challenge, as in reality practitioners face global networks in most cases.

Among the researches who take on a global network perspective are Fowler and Robinson (1995) in their Measurement and Improvement of Manufacturing Capacity (MIMAC) Final Report and Kallrath and Maindl (2006). Wu et al. (2001) present a methodology for arbitrating capacities between multiple plants. Tavaghof-Gigloo et al. (2016) provide a multi-item, multi-facility, multi-stage capacity mixed integer linear programming approach. Fordyce and Milne (2012) explain how missing coupling points between local and central planning can cause losses in responsiveness of factories to central planning and missed efficiency gains when the trade-off between capacity utilization and cycle time according to the operation curve is not considered. The authors demonstrate how central planning can become smarter through improved modelling of factory capacities which captures alternative deployments of tools to manufacturing processes with corresponding consumption rates and the trade-off between capacity utilization and cycle times.

Coming back to Figure 1 in the Introduction, it can be concluded that there is there is a lot of research available for the topic of capacity planning, including works specifically for the semiconductor industry. However, only few sources cover the intersection with the global production network.

Most of the tools suggested in the literature require some sort of capacity constraint for generating a feasible plan. It is agreed that for a solvable model for a global network only the most critical bottleneck resources should be included in the decision process. However, the literature currently provides few recommendations for how to select the correct resources amongst a large number of potential bottleneck resources in a global network.

3 CASE STUDY: CAPACITY PLANNING IN A GLOBAL SEMICONDUCTOR PRODUCTION NETWORK

In practice, many of the principles from literature are being applied, nevertheless new challenges have come up in global networks. In the following sections, the capacity planning challenges and solutions are described on the example of a large, multinational semiconductor company. Today, the company with its four divisions runs 18 manufacturing sites for Front-End and Back-End production in 10 countries with an annual turnover of over 7 billion EUR in 2017.

First, an introduction to the company's Supply Chain Planning framework, Master Planning and Advanced Planning System is provided. Next, some of the company's supply chain planning flexibility measures are described. Finally, this paper discusses the new challenges related to the identification and usage of bottlenecks in a global production network.

3.1 Introduction to Plan

The necessity for integrated planning across different planning horizons has been widely recognized for a volatile environment with long cycle times, such as the semiconductor industry. In the planning scheme of our company example shown in Figure 3, a strategic plan is set up on a yearly basis with a horizon of 5 years. Aggregated Capacity Groups (ACGs) are considered on a global level for a tactical demand and supply match on a quarterly basis. However, ACGs only capture capacities at selected points of the manufacturing process, while in reality bottlenecks may occur at every production stage (Ehm et al. 2010). Capacity corridors in the tactical plan support cross-plant allocation and capacity investment decisions for a time horizon of up to 18 months.

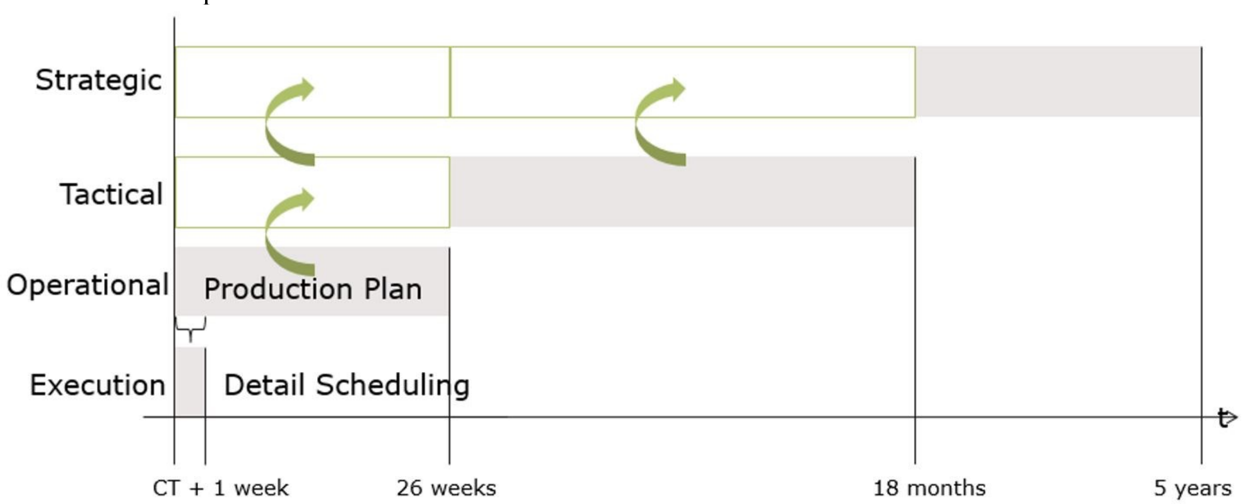

Figure 3: Planning Horizons in the semiconductor industry (Schiller 2017).

Figure 4 illustrates how capacity planning is embedded in the integrated planning landscape for the tactical and operational level. Capacity planning tasks are highlighted in red and include the ACG planning for the tactical level as well as a subset of selected bottleneck resources and their capacities on an operational level.

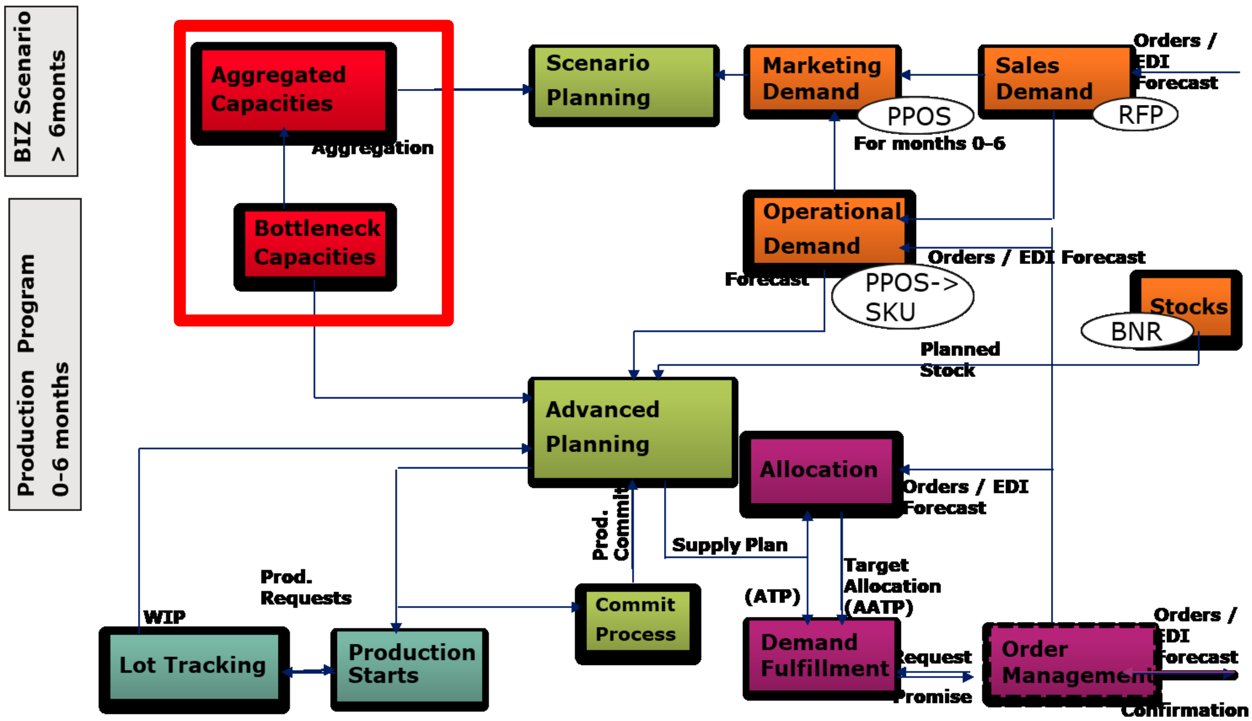

Figure 4: Planning overview in a semiconductor network (Yachi and Schiller 2017).

For managing bottleneck capacities, there is a central database where bottleneck resources on plant level and their capacities are entered, i.e. selected resources which are known to limit the production output. In addition, the database captures information about which products can be produced by the respective resource. For planning, cycle times are of high importance and vary depending on the utilization of a resource. Nonetheless, in our practical example, cycle times are assumed to be fixed in the Master Planning run and may be reviewed and adjusted later on.

In the industry example, Master Planning is done on a global level. The Advanced Planning System runs on a daily basis with a planning horizon of 26 weeks. Bottlenecks, demand forecasts, orders and planned stock levels are considered as an input and used to come up with global production targets and decisions about which products are to be produced where and in which quantities. The main objectives are to maximize equipment utilization and minimize unfulfilled demand through a heuristic procedure. The outputs are weekly production requests per plant and supply volumes which are available-to-promise (ATP) for order confirmation respectively. A production commit process ensures production targets from the APS are transformed into feasible, detailed production plans.

For every level of the planning horizon, different levels of aggregation need to be considered in order to keep the process manageable. A trade-off has to be found between the desired level of detail and the feasibility and speed of the solution. This is especially challenging for Master Panning.

Practitioners decompose the planning problems such as Master Planning into multiple, more manageable sub-problems and try to come up with a smart way of linking the solutions. Hence, planning nowadays is often addressed with a best of breed application with e.g. JDA for the demand supply match and SAP for Order Management in its key role to be aligned with the general ledger.

3.2 Planning Flexibility

In the industry example, plans are integrated and created globally on a daily level to be reactive to the market and enable daily improvements towards the customer wish date. This is possible as a final product may have many production route alternatives which can be considered for improvements. The central planning system allows for demand and bottleneck capacities to be entered and updated on a daily basis. This means the product mix and the base for determining bottlenecks may also change on a daily basis.

A daily batch re-scheduling cycle (BR) redirects supply and demand usage consequently improving the overall outcome in terms of better delivery performance of the available-to-promise (ATP) allocation to the customer. By using a batch mode approach orders are accumulated over a pre-determined time interval and then released for ATP calculation. Such grouping can help reduce infeasible resource allocations, since the order picture might be updated in real-time. The supply information also can be updated regularly to take benefit of the flexibility in rerouting to adapt better to the changing demand. This requires a continuous updating of bottlenecks as well.

3.3 New Challenges in Global Production Networks: How to Identify Bottlenecks

Due to the complexity of the production processes, bottlenecks are determined on a local plant level in the studied case. The local approaches for detecting bottlenecks depend on the specifics of the relevant production process, location and availability of information. In the following, a few practical examples are introduced.

Bottlenecks can appear in different forms in both Front-End and Back-End production. In the Front-End process, semiconductor chips are manufactured by applying layers of conductive and non-conductive materials on raw silicon wafers. This part of the production process is the most complex and time-consuming, including re-entrant process steps, time-bound constraints and varying yield rates. Bottlenecks also depend on the number of layers to be applied for a certain technology or individual machine capacities – e.g. from a stepper. Front-End bottlenecks are mostly commodity-driven for the production of the wafers. In the Back-End process individual semiconductor chips are sawn apart and built into packages which protect them and allow connecting them to electronic devices later on. Typical bottlenecks are assembly or test tools which are either modelled as single machines or as a pool of machines, in which case, multiple machines can produce the same product.

Approaches for identifying bottlenecks can be built on the experience of employees or on sophisticated tools. Employees often know quite well which resources limits the overall output based on experienced historic demand, delivery fulfilment ability patterns and technical limitations. These bottlenecks determined by experts may stay the same for a longer time period, so that changes in their structure are rather rare, whilst bottlenecks determined by tools tend to be more flexible. There is a requirement to assign a timeline to a capacity, due to the changes in available resources when equipment is added or retired.

In the past, different approaches from Operations Research have been assessed for benefits of their application in practice in the company. It was found that modelling the reality of manufacturing frameworks, including production logic and variabilities, needs to be the basis. In a best-of-breed approach the best tools have been developed for different planning tasks. Heuristics can be used for connecting these tools with each other.

On a local level, a standardized, dynamic planning tool is used to account for the fast changing equipment landscape over time. Technically, this tool aims to combine the quality of a solver solution with the speed of spreadsheet applications and the coverage of different possible scenarios through simulations. The tool supports the bottleneck determination by calculating the approximate capacity of each work station along the manufacturing process, considering the Overall Equipment Efficiency (OEE), i.e. the net time that a resource is actually available for production. The tool can reflect complexities such as parallel production lines, resource pools, re-entrant processes or product-specific toolsets which are required in addition to the machines themselves. Moreover, loss factors such as lot sizes, yield, changeover times and breakdowns are modelled. For machines which can produce multiple different products, load factors are derived from the production recipes and define per product how much capacity of a resource it consumes. For instance, if the capacity is defined as available processing time, one piece of a product may consume 8 seconds on one machine, while one piece of another product may only consume 6 seconds on the same machine. The load factor would hence be 8 and 6 respectively.

Alternatively, capacity could be defined as the throughput rate of a reference product, i.e. the number of pieces produced in a defined time period, e.g. 1,000 pieces per hour. The load factor is determined by dividing the capacity of the reference product by the capacity of the product in question. Hence, the load factor for the reference product is always 1. If another product can be processed at 1,000 pieces per hour, then the load factor would be 1 as well. If a separate product can be processed at 2,000 pieces per hour, the load factor for that product would be 2,000/1,000 = 2. This logic is represented in Figure 5.

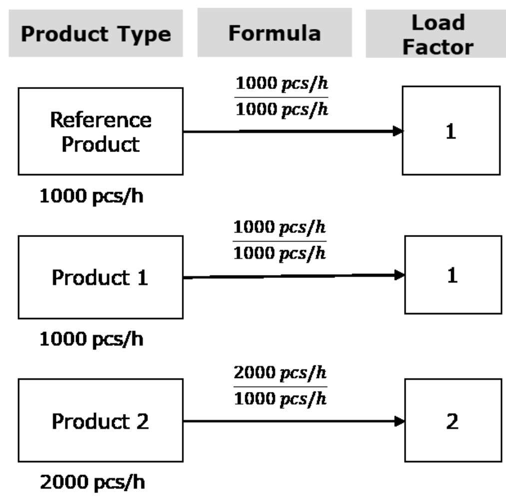

Figure 5: Load Factor example.

Once equipment specifics and rules are modelled in the planning tool, the optimal utilization of each resource is calculated based on the product mix derived from the weekly supply and demand match; if it is higher than a certain percentage a resource is considered a bottleneck. All other resources are considered non-critical and unconstrained (Ehm et al. 2010).

Sometimes, placing additional, artificial capacity constraints can be beneficial – for example, the overall variability can be reduced when production alternatives with exceptionally high load factors are banned. In addition to the operational bottlenecks described so far, so-called business bottlenecks are used by central planning for strategic steering. In the past, one business division in a large corporation might have "owned" one or more dedicated factories. Today, factories and resources are shared between multiple business divisions. For instance, business bottlenecks can be used to reserve certain capacities for a specific business unit or sub unit which the parent company wants to give priority to - even if the orders have not been received yet. Another example is a minimum run rate established to keep production up for a certain product. Hence, resource bottlenecks should not be regarded too negatively – they are simply an integral part of planning and can help meet objectives like customer satisfaction or production stability.

The synchronization of the bottleneck identification and the correct usage of such bottlenecks is critical for the success of planning and optimization.

3.4 New Challenges in Global Production Networks: How to Use Bottlenecks

In spite of all the efforts for modelling resources, it was found that implementing one large planning tool for all needs was not realistic – it is not stable enough, hard to maintain and requires enormous computation times. Accordingly, modelling all the global resources in one tool would quickly exceed today's available computing power. In practical terms, such a system can handle a few thousand bottlenecks but it is not possible to calculate millions of resources in a realistic time. Therefore, today only a subset of relevant, local bottleneck resources are entered into the global APS for Master Planning. They are determined through abstraction of the detailed local capacity planning described in Section 3.3. All other resources are assumed to be unconstrained for the Master Plan or they can be added when they become constrained based on a short control cycle. Thereby, the complexity of the capacity information is reduced, resulting in the local plant as a black box. Figure 6 illustrates how bottlenecks are an output of the individual, local capacity plans as well as an input into the global Master Plan. In the global Master Plan, the global demand is allocated and balanced between different plants. The output are production requests sent back to the plants for execution. The adjustment loop is indicated with a dotted line, since the product mix of the Master Plan is used again on local level as an input for the bottleneck determination.

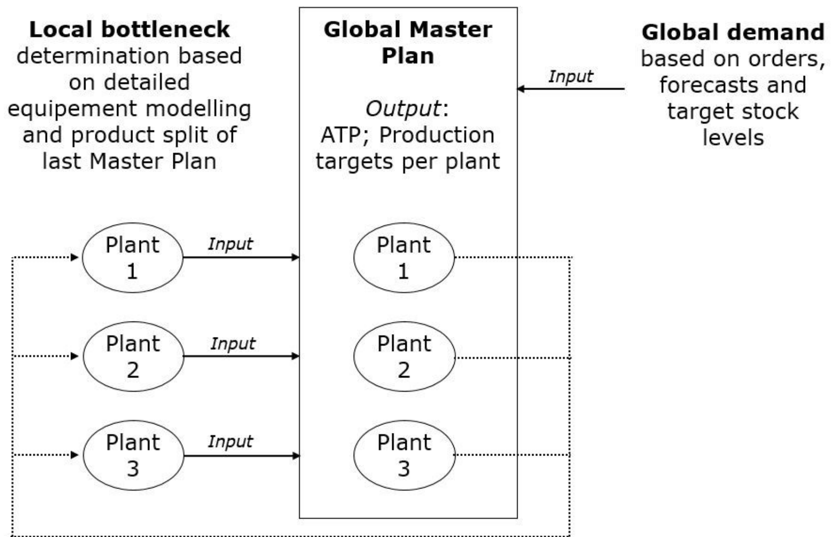

Figure 6: Local Determination of Bottlenecks as an Input for Global Master Planning.

The main advantage of this approach is that capacity planning on the plant level is much easier to manage and compute than on the global level. However, some flexibility is lost by using the Master Plan output as an indicator of the product mix used for bottleneck determination in the plants. If bottlenecks were determined on a truly global level based on the global demand mix, the optimal plant allocation may lead to completely different product mixes, which would then in turn lead to different local bottlenecks. So there is a risk that a local bottleneck determination method can get stuck in a local optimum and miss out on some of these swap opportunities. Besides, the target usage of bottlenecks grants more freedom for tackling other areas where optimization might be required.

4 CONCLUSION

The challenges of capacity planning are not new and many scientific papers have dealt with them. A literature review has been provided. Only a few papers tackle the new challenges in global enterprises with alternative production routes and integrated planning across multiple production sites which is updated on a daily basis. Also, relatively little sources look into how to handle capacities in short- and midterm planning. If they do, it is typically through Master Planning approaches, many of them using sophisticated approaches, but with selected capacity constraints already given and no guidance on how to determine and select these capacities.

Based on a company case study it was shown that reality today requires planners in practice to solve these problems, so a number of practical solutions have emerged which are often based on experience. A local and global planning theory has not been covered in much detail in current literature and might be valuable for further research, especially since this challenge in the semiconductor industry today might be relevant for other industries in the future as well.

ACKKNOWLEDGEMENT

We thank Hans Ehm from Infineon Technologies AG for initiating the idea for this paper and providing valuable input and feedback. We are grateful to Christian Fritzlar of Infineon Technologies AG for sharing his experience with capacity planning and reviewing the paper. Also, we thank Ben Rodriguez from HIPPEROS S.A. for sharing his experience on dynamic and enterprise Capacity Planning.

REFERENCES

Barahona, F., S. Bermon, O. Günlük, and S. Hood. 2005. "Robust Capacity Planning in Semiconductor Manufacturing." *Naval Research Logistics* 52(5):459–468.

Bermon, S., and S. J. Hood. 1999. "Capacity Optimization Planning System (CAPS)." *Interfaces* 29:31–50.

Brown, S. M., T. Hanschke, I. Meents, B. R. Wheeler, and H. Zisgen. 2010. "Queueing Model Improves IBM's Semiconductor Capacity and Lead-Time Management." *Interfaces* 40(5):397–407.

Cachon, G., and C. Terwiesch. 2011, *Matching supply with demand. An introduction to operations management,* 2nd edn., Boston, MA: McGraw-Hill Irwin.

Chen, W., Z. Wang, and F. T.S. Chan. 2017. "Robust Production Capacity Planning under Uncertain Wafer Lots Transfer Probabilities for Semiconductor Automated Material Handling Systems." *European Journal of Operational Research* 261(3):929–940.

Chen, Y.-Y., T.-L. Chen, and C.-D. Liou. 2013. "Medium-term Multi-plant Capacity Planning Problems considering Auxiliary Tools for the Semiconductor Foundry." *The International Journal of Advanced Manufacturing Technology* 64(9-12):1213–1230.

Chien, C.-F., C.-H. Wu, and Y.-S. Chiang. 2012. "Coordinated capacity migration and expansion planning for semiconductor manufacturing under demand uncertainties." *International Journal of Production Economics* 135(2):860–869.

Christie, R. M.E., and D. Wu. 2002. "Semiconductor capacity planning: stochastic modeling and computational studies." *IIE Transaction* 34(2):131–143.

Christoph Habla, Lars Moench, and Rene Driessel. 2007, *A Finite Capacity Production Planning Approach for Semiconductor Manufacturing. CASE 2007 ; 22 - 25 Sept. 2007, Scottsdale, Arizona*, IEEE Service Center: Piscataway, NJ.

Connors, D. P., G. E. Feigin, and D. D. Yao. 1996. "A Queueing Network Model for Semiconductor Manufacturing - Semiconductor Manufacturing, IEEE Transactions on." *IEEE Transaction On Semiconductor Manufacturing* 9.

Denton, B. T., J. Forrest, and R. J. Milne. 2006. "IBM Solves a Mixed-Integer Program to Optimize Its Semiconductor Supply Chain." *Interfaces* 36(5):386–399.

Ehm, H., C. Schiller, and T. Ponsignon. 2010. "Achieving Better Utilization of Semiconductor Supply Chain Resources by Using an Appropriate Capacity Modeling Approach on the Example of Infineon Technologies AG." In *Operations Research Proceedings,* edited by B. Hu, et al., 417–422, Berlin Heidelberg: Springer.

Ellis, S., Y. Gao, and C. Wang. 2017. "TSMC Ready to Spend $20 Billion on its Most Advanced Chip Plant." Last modified 06.10.18 Accessed 25.06.18. https://www.bloomberg.com/news/articles/2017-10-06/tsmc-ready-to-spend-20-billion-on-its-most-advanced-chip-plant.

Fleischmann, B., H. Meyr, and M. & Wagner. 2008. "Advanced Planning." In *Supply Chain Management and Advanced Planning,* edited by H. Stadtler and C. Kilger, 81–109. 4th ed., Berlin Heidelberg: Springer.

Fordyce, K., J. Fournier, R. J. Milne, and H. Singh. 2012. "Tutorial: Illusion of capacity — Challenge of incorporating the complexity of FAB capacity (tool deployment & operating curve) into central planning for firms with substantial non-fab complexity." In *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC),* 1–16: IEEE.

Fordyce, K., and R. J. Milne. 2012. "The Ongoing Challenge for a Responsive Demand Supply Network: The Final Frontier—Controlling the Factory." In *Decision Policies for Production Networks,* edited by D. Armbruster and K. G. Kempf, 31–69, London: Springer London.

Fowler, J., and J. Robinson. 1995. "Measurement and Improvement of Manufacturing Capacity (MIMAC): Final Report." SEMATECH.

Geng, N., and Z. Jiang. 2009. "A review on strategic capacity planning for the semiconductor manufacturing industry." *International Journal of Production Research* 47(13):3639–3655.

Kallrath, J., and T. I. Maindl. 2006, *Real optimization with SAP APO*, Berlin, New York: Springer.

Karabuk, S., and D. Wu. 2003. "Coordinating Strategic Capacity Planning in the Semiconductor Industry." Operations Research 51(6): 839-849.

Kim, S. H., and Y. H. Lee. 2016. "Synchronized production planning and scheduling in semiconductor fabrication." *Computers & Industrial Engineering* 96:72–85.

Kumar, S., and P. R. Kumar. 2001. "Queueing Network Models in Design and Analysis of Semiconductor Wafer Fabs." *IEEE Transactions on Robotics and Automation* 17(5), 548-561

Leung, S. C. H., and Y. Wu. 2004. "A robust optimization model for stochastic aggregate production planning." *Production Planning & Control* 15(5):502–514.

Lyon, P., R. J. Milne, R. Orzell, and R. Rice. 2001. "Matching Assets with Demand in Supply-Chain Management at IBM Microelectronics." *Interfaces* 31(1):108–124.

Mhiri, E., M. Jacomino, F. Mangione, P. Vialletelle, and G. Lepelletier. 2015. "Finite capacity planning algorithm for semiconductor industry considering lots priority." *IFAC PapersOnLine* 48(3):1598–1603.

Mönch, L., J. W. Fowler, and S. J. Mason. 2013. "Production Planning and Control for Semiconductor Wafer Fabrication Facilities." *Operations Research/Computer Science Interfaces Series* 52. Springer Science & Business Media.

Mönch, L., and C. Habla. 2008. "Solving Volume and Capacity Planning Problems In Semiconductor Manufacturing: 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, J. W. Fowler, 2260–2266.

Occhino, T. J. 2000. "Capacity planning model: the important inputs, formulas, and benefits - Advanced Semiconductor Manufacturing Conference and Workshop, 2000 IEEE/SEMI.":455–458.

Ponsignon, T., C. Habla, and L. Mönch. 2008. "A Model for Master Planning in Semiconductor Manufacturing.": In *IIE Annual Conference. Proceedings*. Institute of Industrial and Systems Engineers (IISE), 1592.

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

Ponsignon, T., and L. Mönch. 2014. "Simulation-based performance assessment of master planning approaches in semiconductor manufacturing." *Omega* 46:21–35.

Romauch, M., and R. F. Hartl. 2017. "Capacity planning for cluster tools in the semiconductor industry." *International Journal of Production Economics* 194:167–180.

Romauch, M., and A. Klemmt. 2015. "Product mix optimization for a semiconductor fab: Modeling approaches and decomposition techniques." *Computers & Operations Research* 53:338–352.

Rowshannahad, M., S. Dauzère-Pérès, and B. Cassini. 2015. "Capacitated qualification management in semiconductor manufacturing." *Omega* 54:50–59.

Schiller, C. October 2017, *General Introduction to Plan. Internal document*.

Tavaghof-Gigloo, D., S. Minner, and L. Silbermayr. 2016. "Mixed integer linear programming formulation for flexibility instruments in capacity planning problems." *Computers & Industrial Engineering* 97:101–110.

Uzsoy, R., J. W. Fowler, and L. Mönch. 2018. "A survey of semiconductor supply chain models Part II: demand planning, inventory management, and capacity planning." *International Journal of Production Research* 43:1–19.

Wu, C.-S., Chou, Y-C, and J.-Z. Lin. 2001. "Optimization and economic analysis for tool portfolio planning in semiconductor manufacturing - Semiconductor Manufacturing Symposium, 2001 IEEE International." *IEEE*.

Yachi, G., and C. Schiller. September 2017, *Introduction to Demand Planning. Internal document*.

AUTHOR BIOGRAPHIES

SARAH BIWER is a Master of Science student in Supply Chain Management at WU Vienna University of Economics and Business. She is currently writing her Master Thesis in collaboration with Infineon Technologies AG in Neubiberg, Germany. Before starting the Master of Science program, she worked for four years in international logistics in different roles including process optimization, tender management and business development. She holds a Bachelor of Arts degree in International Cultural and Business Studies from Passau University. Her e-mail address is sarah.biwer@infineon.com.

MICHAEL FILIPEK is a Master of Science student in Supply Chain Management at WU Vienna University of Economics and Business and is currently writing his Master Thesis in collaboration with Infineon Technologies AG in Neubiberg, Germany. He studies in the Bachelor of Science program at Technical University of Vienna and holds a Bachelor of Science degree in Business Economics from Vienna University of Economics and Business. His e-mail address is michael.filipek@s.wu.ac.at / michael.filipek@infineon.com / michael_filipek@hotmail.com.

EMEL ARIKAN FICHTINGER is an Assistant Professor at WU Vienna University of Economics and Business. She holds a doctoral degree in Business Administration from WU Vienna, and a master's degree in Industrial Engineering from Middle East Technical University, Turkey. Before her role as an assistant professor she worked as a researcher in the WWTF-project "Integrated Demand and Supply Chain Management". Her main research interests are in the field of inventory management, with extensions to integrated pricing decisions, environmental aspects of inventory management, and decision making under risk-aversion. Emel teaches several courses in operations and supply chain management. Her e-mail address is emel.arikan@wu.ac.at.

WERNER JAMMERNEGG is Professor of Operations Management in the Institute for Production Management at WU Vienna University of Economics and Business. He studied Business Administration at University of Graz and Technical Mathematics (Information Processing) at Graz University of Technology, Austria. Alongside two master degrees, he holds a doctoral degree in Business Administration and a habilitation degree in Operations Research and Statistics. His main research interests include stochastic inventory models, economic and environmental sustainability in supply chains, dependency of supply chain risks and risk preferences and behavioural inventory decision. He has been Professor of Operations Management at WU (Vienna University of Economics and Business), Austria, since 1990. From 2010 to 2015 he was director of the Master´s of Science program in Supply Chain Management at WU. His e-mail address is werner.jammernegg@wu.ac.at.