Integrating Production Planning in a Manufacturing Execution System (MES)

João Pedro da Costa Silva Pereira

Master's Dissertation

Supervisor: Prof. José Fernando Oliveira

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia e Gestão Industrial

2021-06-28

Abstract

As markets become more and more competitive, companies look at decision-support systems as a valuable asset, helping them utilize their resources effectively in meeting demand at minimum cost. In the context of manufacturing, planning is traditionally done using several systems arranged hierarchically. One of those systems placed at the very bottom of the hierarchy is the Manufacturing Execution System (MES). This focuses on operational tasks such as the routing of lots through the shop floor.

This thesis arises from the necessity of expanding the scope of a MES. Firstly, this would better support its current scheduling functionality, by defining the moment when production should start under finite capacity planning. And secondly, it would make it more attractive to buyers due to its integrated character. In this sense, three planning levels were introduced using mathematical programming models, with the first two providing perspective and context to the solution. The first is the Master Planning model and it aims at solving the multi-site planning problem as well as decide on capacity expansion for factories (either through equipment purchase or changes in the workforce levels). The second model, which focuses on one factory, is entitled Production Planning, and it evaluates capacity at main production phases. The third model, named Detailed Lot Sizing, would determine daily production for both end-items and components while assessing capacity in resource groups (work centers). For each level, the decisions and objectives were identified, and considerations were made regarding how these levels could relate with each other and with the production scheduling module.

The last part of this work exemplifies the application of the proposed Detailed Lot-Sizing model to the assembly phase of the semiconductor industry. Its objective was to generate the production plan that establishes the guidelines for the scheduling. Changes in the values of the parameters of this model, such as the capacity of a bottleneck work center, the availability of overtime, the setup, and inventory holding costs constituted different scenarios, showing their effect on the production plan.

Resumo

A crescente competitividade dos mercados leva as empresas a considerarem os sistemas de apoio à decisão como um ativo valioso que as auxilia na eficaz utilização dos recursos para satisfazer a procura ao menor custo. Num contexto de produção, o planeamento é tradicionalmente feito recorrendo a diversos sistemas dispostos de um modo hierárquico. Um desses sistemas situado na base da hierarquia é o Manufacturing Execution System (MES). Este foca-se em tarefas operacionais como o roteamento dos lotes no chão de fábrica.

Esta tese surge da necessidade de expandir o alcance de um MES. Em primeiro lugar, esta expansão melhoraria o suporte à atual funcionalidade de escalonamento, através da definição do momento de início de produção tendo em consideração restrições de capacidade. E em segundo lugar, tornaria a solução mais atrativa para os compradores devido ao seu carácter integrado. Neste sentido, foram apresentados três níveis de planeamento recorrendo a formulações matemáticas, sendo que os dois primeiros fornecem contexto e perspetiva à solução. O primeiro modelo designa-se Master Planning e pretende resolver o problema de planeamento multi-site, bem como decidir sobre expansão de capacidade das fábricas (ou através de compras de equipamentos ou através de modificações na força de trabalho). O segundo modelo, já vocacionado para uma única fábrica, designa-se Production Planning e avalia a capacidade nas principais fases de produção. O terceiro modelo designa-se Detailed Lot Sizing e determina a produção diária quer de produtos finais quer de componentes através da avaliação da capacidade em grupos de recursos. Para cada nível foram identificadas as decisões e objetivos, e foram feitas considerações relativamente a como os níveis se poderiam relacionar entre si e com o módulo de escalonamento da produção.

A última parte do trabalho exemplifica a aplicação do modelo de Detailed Lot Sizing desenvolvido à fase de assemblagem da indústria de semicondutores. O objetivo foi gerar um plano de produção que constituiria as diretrizes para o escalonamento. Alterações nos valores dos parâmetros deste modelo, tais como a capacidade do grupo de máquinas, a disponibilidade de horas extra, os custos de setup e custos de inventário, constituiram cenários analisados que permitiram demonstrar o seu efeito no plano de produção.

Acknowledgements

This thesis marks the last step in my five-year journey studying Industrial Engineering and Management at FEUP, for which I could not have been more grateful and satisfied.

Firstly, I would like to thank my supervisor José Fernando Oliveira for his availability, patience, and diligence in answering my doubts and concerns. Without his support, this thesis would not have been possible.

I also thank Critical Manufacturing for welcoming me and presenting such a challenging task. To my company supervisor Nuno Vitorino, I thank his availability and cooperation.

This whole journey would not have been the same without my family. To my mother, father, and brothers, I deeply express my thanks for always having my back and being role models of perseverance and hard work. Thank you for supporting me and for always encouraging me to dream big.

I also want to express my gratitude to my close friends for all the experiences we shared, learned, and laughed at together during these times. Even those moments that seemed meaningless at the time revealed to have their own special place.

Lastly, to all people that made these years so special and that I will forever remember with affection, a big thank you!

"Every journey that is successful has culs-de-sac and speed bumps. I carry a wisdom gene through my life through the good, the bad, and the ugly."

Peter Guber

Contents

Acronyms and Symbols

APS	Advanced Planning System
BOM	Bill-of-Materials
CMF	Critical Manufacturing
ERP	Enterprise Resource Planning
MES	Manufacturing Execution System
MRP	Materials Requirements Planning
SCPM	Supply Chain Planning Matrix
WIP	Work-in-Process

List of Figures

List of Tables

Chapter 1

Introduction

With globalization and continuous advances in technology, the competitiveness in the manufacturing sector has increased, and with it the complexity of the operations involved. Nowadays, just being efficient in performing manufacturing operations on the shop floor is not enough, since companies have to compete in many other dimensions such as quality, delivery, cost efficiency, and flexibility (Olhager, 2013). This means that the focus cannot be only on the factory but rather on the supply chain. This change in perspective is depicted in the evolution of planning systems.

The first systems for production planning and control (until the 1960s) were manual and were based on replenishing, together with the suppliers, the quantities for the most used items. Then, Materials Requirements Planning systems (MRP) appeared to face the increasing diversity of production that made the previous method unpractical. These enabled planning the required production of components in a timely manner to obtain the final product when desired. Its successor, the Manufacturing Resource Planning (MRP-II), is an extension that integrates other functional manufacturing areas by, for instance, incorporating an additional level - a long-term Aggregate Production Plan. In the 90s, Enterprise Resource Planning systems (ERP), further widened the scope by extending into other business areas of the enterprise besides production, such as accounting, distribution, or sales. More recently, Advanced Planning Systems (APS) appeared, constituting the state of the art. A market research evaluated the APS software market in 2020 at US$ 1,491.22 million and estimated that its value could reach US$ 2,941.27 million by 2028 (Partners, 2021).

A recurring issue in production planning by the different systems is the consideration of capacity. Many authors state that only APS are capacity focus, while the other systems do not consider capacities adequately (Tempelmeier, 2001). This is a critical consideration that, when not taken into account results in low service levels and long customer waiting times. Therefore, being able to ensure both the right availability of resources to meet demand on time or regard its capacity limitations are requirements that increase the resilience of companies.

This thesis arises in the context of an ongoing project in a major Portuguese Software company in the Manufacturing Execution Systems (MES) field. This project consists of specifying a planning solution that comprises several levels while considering capacities to be incorporated in MES. It intends to enhance the scope of the current solution that is dedicated to performing scheduling and that therefore only acts at an operational level.

The remainder of this chapter continues to introduce the topics addressed by this dissertation. Section 1.1 provides the project context and its underlying motivation. Section 1.2 presents the goals for this thesis and Section 1.3 addresses the approach used to achieve them. This chapter finalizes with the structure outline for the rest of the dissertation, in Section 1.4.

1.1 Project context and Motivation

This project was developed in Critical Manufacturing (CMF), a company that provides a manufacturing software solution designed to manage production, denominated Critical Manufacturing MES. This is a computerized system that helps to document the transformation of raw materials into finished products (product tracking) through collecting and managing data. It works in real-time, thereby supporting decision-making and quickly identifying opportunities for improvement on the job floor. It has a modular design capable of performing several functionalities that can specifically be chosen by clients to meet their specific needs.

By being an execution system, it operates under a very short-term level, meaning that the range of decisions that can be made are limited in scope and are affected by decisions made by other systems located at a higher aggregate level. Moreover, since production planning is not currently incorporated in MES, the system relies on an external system, typically, an ERP to perform it. The issue is that capacity might not be considered when production planning at the MRP level is performed, contributing to several problems, stated in Chapter 3, after running the scheduling functionality of MES.

The need to include a planning module to complement CMF's offer has also been identified by the product implementation team through contact with clients already using MES or through contacts with potential clients in fairs and conferences. Therefore, it constitutes a necessity not just to support the current solution of the company but also to increase the competitiveness of its product in the future.

1.2 Goals

The main goal of this thesis is to specify the requirements for the planning solution to be included in Critical Manufacturing MES, to then be either developed internally by CMF or integrated from a commercial provider. It is important to understand the different levels that should be considered, the inputs needed, the outputs, and how they will be shown to the user.

Since the scheduling module is already developed, the focus is not on how the scheduling module will function but rather on how it will relate with other levels considered.

As an additional objective, it is referred the application of the developed planning solution to an example.

1.3 Approach

In order to define the planning features with relevance to be included in the solution, a literature review was performed on the topic, together with discussions made with CMF members who presented requirements asked by clients. Afterward, each level was individually defined, according to the scope of decisions to be included. In the end, the back-end-semiconductor industry served as a use case to assess the results given by the level which would connect with the scheduling module, since this is the most relevant to CMF.

1.4 Thesis outline

This thesis is organized into six chapters. Its outline is as follows:

Chapter 1 introduces the topic of this dissertation by providing context on the problem to be addressed and defining the desired outcomes.

Chapter 2 provides a theoretical background on relevant concepts for this work. First, planning is positioned in the context of the supply chain, with issues such as the division of the planning task in different levels highlighted. Then the focus is pointed towards hierarchical planning, finishing with a review of the characteristics of the state-of-the-art planning systems.

Chapter 3 presents the company, describes the functionality of its scheduling module, and then presents the problem and its consequences. Afterward, some desired planning functionalities to be included in a planning system are overviewed.

Chapter 4 identifies relevant planning levels for the functionalities overviewed in chapter 3 and proposes a mathematical programming model for each of them. Considerations on how these levels are connected, especially between the scheduling level and the level directly above it are made. It finishes by identifying graphs and statistics relevant to a user.

Chapter 5 presents a case study in the back-end semiconductor industry for which the planning level above the scheduling was implemented to simulate the outputs that would be introduced to it. Also, several computer experiments were performed to assess the validity of this model and verify the changes in the production plan.

Chapter 6 reflects on the work done and suggests ideas for the future.

Chapter 2

Theoretical Background

This chapter presents, from a theoretical point of view, the main topics discussed in this dissertation, aiming to give the overall picture without being fully extensive. Section 2.1 provides an overview of fundamental planning concepts. It starts by highlighting the relevance of planning in a supply chain context, moving on to a description of the three planning levels. The rolling horizon is presented as a method to deal with uncertainty in planning, and the supply chain planning matrix is exhibited to contrast supply chain processes with planning tasks. Among the many planning decisions, emphasis is given to lot-sizing due to its difficulty and importance in planning production. In Section 2.2, the main principles of hierarchical planning are analyzed, focusing on the communication between different levels. Lastly, in Section 2.3, Advanced Planning Systems (APS) are introduced by describing its characteristics, structure, advantages relative to predecessor planning systems, and how necessary information is exchanged between systems.

2.1 Planning in a Supply Chain context

Stadtler (2015) presented the task of Advanced Planning as one of the building blocks that sustain coordination of material, information, and financial flows in a supply chain (Figure 2.1). The idea is that Supply Chain Management (SCM), defined as "the effort involved in producing and delivering a final product from the supplier's supplier to the customer's customer" (Larson and Rogers, 1998), essentially aims at increasing competitiveness of the products or services offered by the whole supply chain at minimum cost. This can be achieved by satisfying adequate customer service levels, either through better integration of the linked organizations or by better coordinating the transferred flows (Stadtler, 2005). Planning, since it enables better preparation, helps with the coordination part (de Kok and Fransoo, 2003).

Yet, a generated plan is only valid during the planning horizon. Based on this concept and influenced by the organizational structure of enterprises, Anthony (1965) defined three planning levels that differ on the objectives and type of decisions made:

- Long-term planning: Deals with strategic decisions that significantly affect the course and character of the organization (Anthony, 1965) from a design and structure point of view

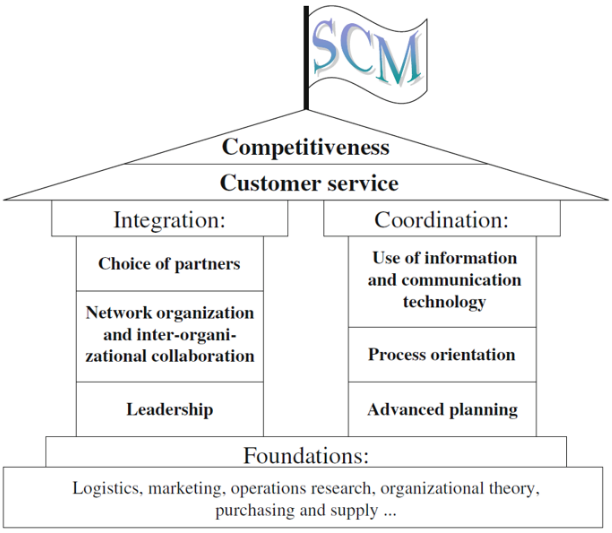


Figure 2.1: House of SCM (Stadtler, 2015)

(Fleischmann et al., 2015). For instance, it involves defining the objectives, the resources requirements, the location of new production facilities, the incorporation of new suppliers, the introduction of new products, etc. Normally defined on a year granularity.

- Medium-term planning: Also known as tactical decisions, concern the usage of resources previously defined at the long-term level to fulfill the objectives (Anthony, 1965). Examples of these decisions are related to inventory policy, workforce levels, subcontracting, and overtime (Chen et al., 2009). Normally defined on a monthly granularity, from 6 months to 24 months (Fleischmann et al., 2015).

- Short-term planning: These decisions are operational in nature and mostly relate to performing detailed tasks effectively and efficiently (Anthony, 1965). This could be the sequencing of the manufacturing tasks (while considering upper levels defined constraints) (Gupta and Maranas, 2003). Normally defined on a daily or weekly granularity, from days to 3 months (Fleischmann et al., 2015).

Note that the three planning levels not only differ on the planning horizon and on the purpose but also on characteristics such as the level of aggregation (larger in the long-term and smaller in short-term); frequency of replanning (more frequently in the long-term and less often on short-term); precision required (higher on long-term and lower on short-term); and uncertainty and risk (higher on long-term and lower on short-term) (Fleischmann and Meyr, 2003).

Uncertainties in a production process can be categorized into two groups: (i) environmental uncertainty and (ii) system uncertainty (Mula et al., 2006). The first group comprises those uncertainties that go beyond the production process, such as uncertainty in supply and demand. The second group refers to those that occur within the production process, examples being: "operation yield uncertainty, production lead time uncertainty, quality uncertainty, failure of production system and changes to product structure" (Mula et al., 2006).

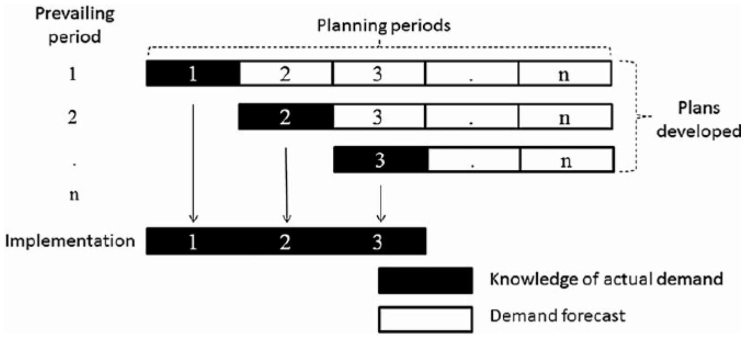


Figure 2.2: Rolling horizon planning concept (Gautam et al., 2015)

A common way to deal with these uncertainties in planning is using the rolling horizon approach (Figure 2.2). This concept assumes that decisions to be made on the most immediate period are based on forecasts of relevant information for a certain number of periods in the future (planning horizon) (Sethi and Sorger, 1991). Every time the most immediate period is concluded, forecasts are updated based on the new information available for the new relevant period (that partially overlaps with the previous one but includes one extra period). This implies that part of the plan concerning the most immediate period, which will be implemented, is considered *frozen* (meaning it will not change), but the rest of the plan (that comprises all other periods) can change to reflect updates.

Vogel et al. (2017) identified a trade-off between frequently updating the plan (resulting in more flexibility and cost savings) and generating planning nervousness (which should be limited). Nervousness is reduced by freezing more periods than just the most immediate one.

An alternative way of updating plans is the event-driven approach. Instead of updating the plan on a regular interval basis (as in the rolling horizon), the plan is reformulated whenever an important event occurs (e.g. unexpected sales, major changes in customer orders, breakdown of a machine) (Stadtler, 2005).

The planning tasks can be arranged according to the corresponding process in the supply chain (procurement, production, distribution, and sales) in what is known as the "Supply Chain Planning Matrix" (SCPM) (Steven, 2004). This is, however, a general representation because depending on the type of supply chain considered, the importance of the planning task or its positioning in the SCPM can be different. For instance, lot-sizing decisions can be considered on a mid-term level if setup costs are significant, as it happens in process industries (Albrecht, 2010).

In fact, lot-sizing is an important and difficult planning decision (Bian, 2018). It determines when and in what quantities to produce so that setups and inventory holding costs can be minimized (Brahimi, 2004). Karimi et al. (2003) identified the following characteristics that affect the complexity of the lot-sizing problem (and are in the basis of its classification in literature):

- **Planning horizon**: It can be *finite* (usually assumes dynamic demand) or *infinite* (usually assumes static demand). Finite problems are divided into discrete buckets, that depending

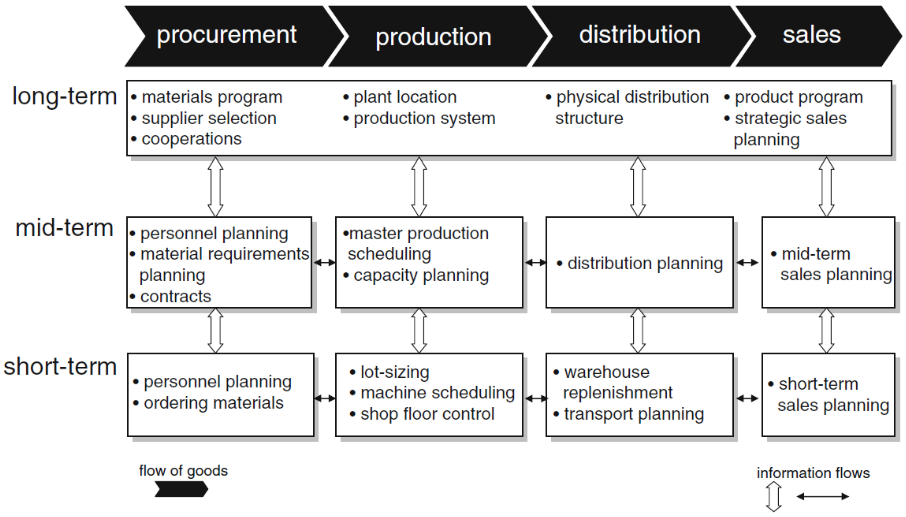


Figure 2.3: Supply Chain Planning Matrix (Fleischmann et al., 2015)

on their length are classified into *Big-bucket* (when the time period is long enough to produce multiple items per period) or *Small-bucket* problems (if short production periods are considered and only one item can be produced per period).

- **Number of levels** in the production process: When final products are directly obtained from raw materials without intermediate sub-assemblies being involved (case of independent demand), it is called a *Single-level* problem, otherwise it is *Multi-level* problem (case of dependent demand).
- **Number of end-products** considered in production process: When only one final item is considered, the problem is referred to as *Single-item*, otherwise it is a *Multi-item* problem
- **Capacity constraints:** *Capacitated* problems consider resource constraints, whereas *Uncapacitated* do not.
- **Deterioration of items**: It restricts inventory holding time when considered
- **Demand**: It can be *Static* (if it is constant during the planning horizon) or *Dynamic* otherwise. The prior knowledge of future demand implies the designation of *Deterministic* (*Probabilistic* otherwise).
- **Setup structure**: Setups have a *Simple-structure* if they are not influenced by the sequence of production, or *Complex-structure* otherwise (in this case it is common to use the term *sequence-dependent* setups). *Complex-structures* can also refer to *setup carry-over* (when a production run can continue from a previous run without incurring an extra setup) or *minor/major* setups (when setups are insignificant when occurring between changes of items belonging to the same family group but significant between items of different families).
- **Inventory shortage**: It either results in *Backorder* (when orders can be delivered late) or *Lost-sales*.

For a detailed description of the other planning tasks presented in the SCPM see Fleischmann et al. (2015). Note that in the SCPM, the different planning tasks are interlinked with each other, through horizontal and vertical information flows.

Horizontal Informational Flows: These mainly concern "customer orders, sales forecasts, internal orders for warehouse replenishment and for production in the various departments, as well as of purchasing orders to the suppliers, ..., actual stocks, available capacity lead-times and point of sales data" (Fleischmann et al., 2015). This sharing of information along the supply chain members contributes to minimizing the bullwhip effect (that states that fluctuation of orders increases from sales to the procurement) (Ščukanec et al., 2007).

Vertical Informational Flows: Two types of vertical informational flows can be distinguished: downward and upward flows. Downward flows of information restrict the range of decisions made at lower planning levels and it can be "aggregate quantities, allocated to production sites, departments or processes" (Fleischmann et al., 2015). Examples of upward informational flows are "actual costs, production rates, utilization of equipment and lead-times" (Fleischmann et al., 2015), which provide detailed feedback to upper planning levels. Vertical information flows are related to the idea of hierarchical planning (Bian, 2018), a fundamental concept in the planning theory and widely used in a supply chain context.

2.2 Hierarchical Planning

Hierarchical Planning is based on five key principles: (1) decomposition and hierarchical structure, (2) aggregation, (3) hierarchical coordination, (4) model building, and (5) model solving (Stadtler, 2015).

Decomposition and hierarchical structure refer to dividing the planning task into different levels so that the top level reflects the most influential decision and the lower levels are composed by several decision units (Stadtler, 2015).

Aggregation enables both simplifying a complex task and reducing uncertainty, and it can be done along 4 dimensions: time (e.g. hourly, daily, weekly, monthly, or yearly time buckets), place (e.g. individual customer, zip code, country, sales region), products (e.g. final product, product family) and resources (e.g. individual machines, groups of alternative machines, plant as a whole) (Fleischmann and Meyr, 2003).

Hierarchical *coordination* refers to the communication between levels. It is crucial since just solving models sequentially and independently without considering interactions between them can lead to sub-optimality (a common critic made to hierarchical planning) as well as to infeasibilities in the disaggregation process (resulting in a non-executable plan at the operational level) (Vicens et al., 2001). Schneeweiß (1995) proposed a framework for communication established on 3 steps: anticipation, instruction and reaction (Figure 2.4).

- Anticipation occurs when the upper level considers in its decision characteristics of the lower level but in a level of aggregation compatible with the upper level (feedforward

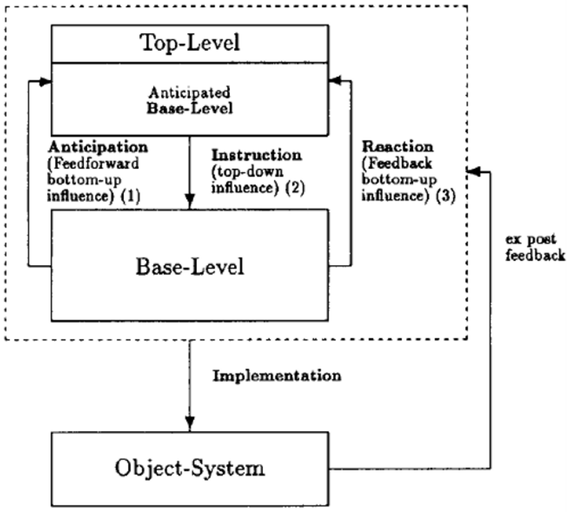


Figure 2.4: Coordination between different hierarchical levels (Schneeweiß, 1995)

bottom-up influence). An example of anticipation is increasing the processing time of a product group, in upper levels, to account for setups that would become visible after disaggregating them into product families in the lower level (Vogel et al., 2017).

- Instruction constitutes of directives, resulting from decisions made at an upper level, sent to a lower level (influence top-down). An example is setting an inventory target level for an end product at the planning horizon of the lower level (Stadtler, 2015).

- Reaction consists of feedback information provided by the lower level that guides the future plans and instructions of the upper level (influence bottom-up).

Note that some consequences resulting from decisions made can only be detected after the plan has been implemented (by an object system) (Fleischmann and Meyr, 2003). This way ex-post feedback helps to influence the next decision of the top level to consider similar situations. This reinforces the idea that an iterative process is required to reach a feasible disaggregated plan, that satisfies both levels (Schneeweiß, 1995).

Finally, the decentralization of decision making implies that each decision unit has an associated *model* (can be a mathematical model) that represents the complexities associated with it (and possible anticipations to lower-level reactions), which is then *solved* (Stadtler, 2015). This greatly contrasts with the monolithic approach, which emphasizes the centralization of information and therefore it is only based on one model. The monolithic approach is seen in literature as “unrealistic” (to include all supply chain planning tasks) and requiring high computational effort (Bian, 2018). However, if applied to only a part of the supply chain like the production side as it was done in Vogel et al. (2017) , the integrated model can yield better results than the hierarchical model within acceptable time intervals (being also less error-prone during implementation since no disaggregation is required).

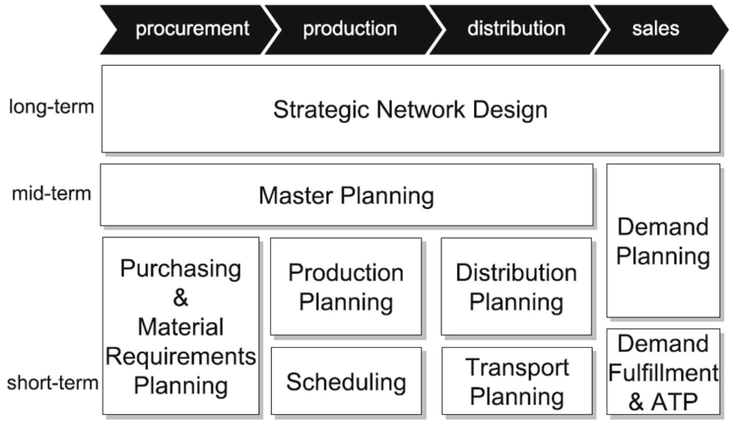


Figure 2.5: Typical structure of an APS (Meyr et al., 2015)

2.3 Advanced Planning Systems

Advanced Planning (or Advanced Planning and Scheduling) Systems (APS) can be defined in three different ways: (1) by identifying its typical characteristics, (2) by describing its structure through the functionalities associated with its software modules, or (3) in terms of the well-known deficiencies of its predecessor systems (Kjellsdotter Ivert, 2012).

2.3.1 Characteristics of APS

The three main characteristics of an APS are, according to Fleischmann et al. (2015), (1) integral planning (either of an enterprise with several facilities or ideally of the whole supply chain), (2) true optimization on solving the production and scheduling activities (based on the mathematical programming, constraint programming, and heuristics) and (3) hierarchical planning.

It supports decision-making at the three levels of planning, by quickly creating new plans (what-if simulation). The term "advanced" stems from the simultaneous consideration of constraints in order to improve the production plans and schedule (Bermudez, 1996) or because well-known planning concepts were for the first time were implemented in a software (Fleischmann and Meyr, 2003).

2.3.2 Structure of APS

Meyr et al. (2015) presented the general structure of an APS (Figure 2.5) consisting of several software modules aimed at tackling the planning tasks stated in Figure 2.3. Typically not all planning tasks are covered by an APS solution.

Strategic Network Design: It is dedicated to long-term planning tasks, essentially, determining the location of plants and the design of the physical distribution network (procurement and distribution channels) (de Santa-Eulalia et al., 2011). It also supports strategic marketing decisions such as the positioning of products in certain markets (Meyr et al., 2015).

Demand Planning: It forecasts future demand on an aggregate mid-term and detailed short-term basis, and it acts as an input for decisions made on other modules. It also determines the optimal safety stock level (to hedge against forecast error) and performs what-if analysis to see the impact of promotions, price levels, and discounts (Fleischmann and Meyr, 2003).

Demand Fulfillment & Available to Promise (ATP): It tracks customer orders, and comprises tasks such as order promising, due dates settings and shortage planning (Fleischmann and Meyr, 2003).

Master Planning: Coordinates procurement, production, and distribution (between sites) by balancing demand and capacity over a mid-term planning horizon (de Santa-Eulalia et al., 2011).

Production Planning and Scheduling: The Production Planning module is responsible for lot-sizing, while the Scheduling module is in charge of machine scheduling and shop floor control (Meyr et al., 2015). The planning activities in the Production Planning module are done within a single site (as opposed to the Master Planning) (de Santa-Eulalia et al., 2011).

Distribution Planning and Transport Planning: Distribution Planning deals with material flows in a more detailed manner than what is performed by the Master Planning. It includes mid-term constraints within the distribution system such as regular transport links, the delivery areas of warehouses, the allocation of customers to sources, and the use of service providers (Kjellsdotter Ivert, 2012). The Transport Planning module plans short-term dispatches of shipments both in the distribution and the procurement sides (Fleischmann and Meyr, 2003).

Purchasing & Material Requirements Planning: It supports purchasing planning for materials and components, enabling tasks such as choosing alternative suppliers, quantity discounts, and lower (mid-term supply contracts) or upper (material constraints) bounds on supply quantities. These tasks are not supported by traditional ERP systems, which only provide BOM explosion and ordering of materials (Meyr et al., 2015).

In literature, there are other representations for the structure of an APS, mainly motivated by the unclear boundaries of the modules. In Albrecht (2010) version, the Master Planning module extends into sales (partially overlapping with the Demand Planning module) because "Master Planning frequently involves sales-related decisions like backorders and lost sales".

Lastly, Tempelmeier (2001) highlights drawbacks in the common structure that hinder the identification of specific planning requirements by a prospective APS user. These are, for example, the not visibility of the multi-site-based structure in the planning system and the fact that "the problems treated by the different modules are not clearly differentiated with respect to the type of the production and logistics systems considered" (Tempelmeier, 2001). The author also criticizes the fit of the Demand Planning and Demand Fulfillment & ATP modules in the planning matrix since they are not based on a decision model (they are only supporting modules).

2.3.3 Deficiencies of predecessor systems

The predecessor systems of an APS were the Material Requirements Planning (MRP), which evolved into Manufacturing Resource Planning (MRP-II) and then into Enterprise Resource Planning (ERP) systems.

The logic associated with MRP systems is vastly criticized in literature mainly because capacities are assumed infinite and production lead times are considered fixed (Arica and Powell, 2014). Both these parameters are seen, in these systems, as a characteristic of the product when, in fact, they result from planning (Sequeira, 2009). This results in problems at the shop floor level such as excess inventories, poor customer service, and insufficient capacity utilization (de Sousa et al., 2014). APS, on the other hand, considers material and capacity constraints in its optimization process (de Sousa et al., 2014), and so production lead times can be reduced (ideally resulting in an order-based pull production) (Kjellsdotter Ivert, 2012).

The MRP-II can detect material and capacity violations but cannot resolve them automatically due to the sequential process in which capacity is only assessed after planning. This, in practice, results in no finite capacity loading and no integration of these planning functions (Fleischmann and Meyr, 2003). Moreover, it gives equal importance to all customers, and material is allocated based on a first-come-first-served basis (Van Eck, 2003). In APS, these issues are solved. For further comparison between APS and MRP-II systems see Table A.1 of Appendix A.

In ERP systems, the problems mentioned for the other systems (fixed lead times and capacities not considered in bill of materials processing) also apply, which implies that the different planning tasks are modeled inadequately (Stadtler, 2015). Moreover, ERP focus on a single firm, not the whole supply chain. Hence, ERP systems are typically only used as transaction systems (Mauergauz, 2016). However, APS cannot be used independently since they require information provided by the ERP, and so APS are seen as “add-ons” to these systems, filling its gaps. For further comparison between APS and ERP systems see Table A.2 of Appendix A.

2.3.4 Interaction with other planning systems

Besides APS and ERP systems, Manufacturing Executing Systems (MES) is another category of information systems used in planning. MES aids production execution, monitoring, and control of shop floor activities by performing functionalities such as resource allocation, scheduling, data collection on process and resource status, product tracking, and performance analysis (Arica and Powell, 2014). An APS can support MES in evaluating unanticipated events and suggesting corrective actions (Arica and Powell, 2014).

There are currently two different views on the position of APS in relation to MES (Mauergauz, 2016). Frolov, EB and Zagidullin (2008) state that APS should be applied at the long-term and mid-term planning levels, with short-term planning being the responsibility of MES. This view portraits the two systems as separated entities, and this is close to what happens in practice (Mauergauz, 2016). While Meyer et al. (2009) suggests that APS should be directly included in MES (which is the view shared by Critical Manufacturing).

The communication between ERP systems and APS is based on the information transferred between its databases. By holding an independent database, an APS can simulate different plans without affecting the main database (in the ERP system), but it can lead to problems of redundancies and inconsistencies (Mauergauz, 2016). This way, it is recommended to define an integration and an exchange model (Reuter and Rohde, 2015). An integration model defines the planning

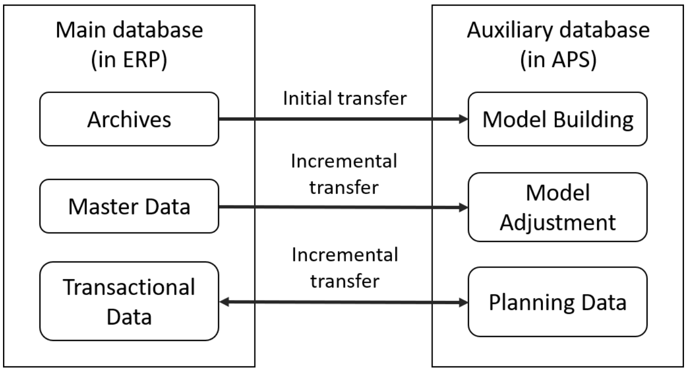


Figure 2.6: Transferring data between ERP and APS (Mauergauz, 2016)

tasks to be performed in each system, which objects will be exchanged, and where they come from. Only planning tasks related to critical products, materials and resources should be done in an APS (Reuter and Rohde, 2015). The exchange model specifies how the flow of information will be performed. The data required by an APS can be exchanged from an ERP system in two steps (Figure 2.6):

1. Initial data transfer: In this step, data needed for building the Master Planning, Production Planning and Scheduling, Distribution and Transport Planning models is transferred (Reuter and Rohde, 2015). This could be information on the properties of potential bottleneck resources, regular capacities, bill of materials (BOM), etc.

2. Incremental data transfer: In this step, only net changes to the initial data should be transferred. Changes to the master data require adjusting the models in the APS. For instance, this occurs when purchasing a new production resource or when introducing a new production shift for the long-term (Reuter and Rohde, 2015). Transactional data corresponds to information that is transferred to and from the APS as a result of planning tasks. Examples of exchanged transactional data are current inventories, current orders, availability of resources, planned production quantities, and stock levels (Reuter and Rohde, 2015).

Chapter 3

Problem Context & Description

This chapter presents the company and describes the problem that motivated the topic of this dissertation.

In section 3.1 it is introduced the company, by exploring the sectors it operates with and the solution it currently offers to clients. In section 3.2, the problem is outlined, and section 3.3 overviews some features for a planning solution. Finally, section 3.4 highlights the main conclusions from this chapter.

3.1 Company description and current solution

Critical Manufacturing (CMF) is a company, founded in 2009 with headquarters in Maia, that provides software solutions for advanced manufacturing industry segments. Currently it is in five countries (Portugal, Germany, USA, China and Taiwan), and it aims to "make Industry 4.0 a reality for all manufactures" (Manufacturing, 2021).

It operates in several segments, namely the semiconductor, electronics, medical devices, automotive, and other discrete manufacturing segments. Among these, the semiconductor is one of its core segments. This is an industry that typically operates under a hybrid make-to-order and make-to-stock strategy due to the challenge of constantly needing to develop diversified products while keeping some level of inventory to face high demand uncertainty and long manufacturing lead times (Kim and Kim, 2012).

The product from CMF is a Manufacturing Execution System (MES). It focuses on ensuring traceability and quality assurance and it is composed of several modules. One of its modules is the "Advanced Planning & Scheduling". This module focuses on giving a low-level (operational), high-detail plan of the operations to be executed on the shop floor. It thrives in environments in which operational factors, such as the sequence of setups or maintenances, are significant and are difficult to be estimated at a mid-term planning level. This means that only after having done the sequencing of lots on the shop floor, one can really understand the restrictions that affect production and how to get around them.

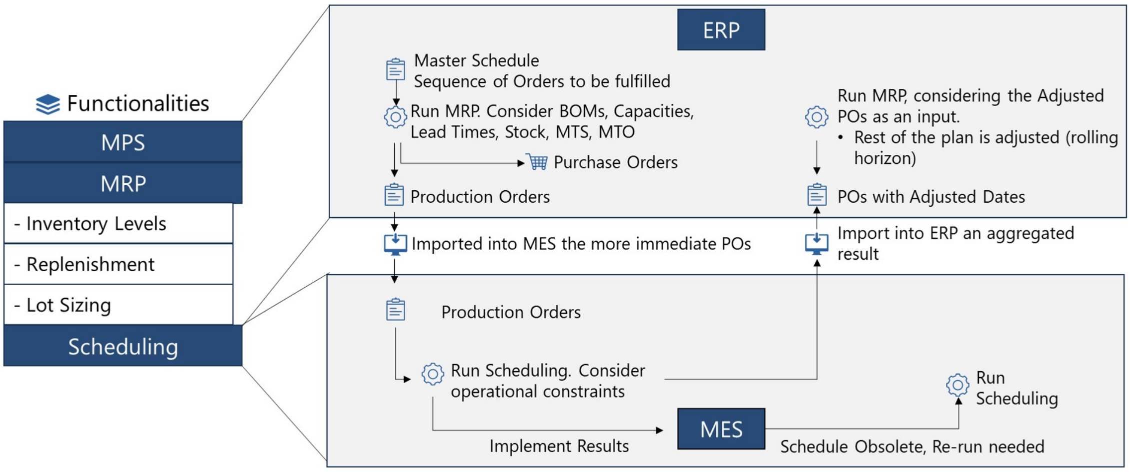


Figure 3.1: Assumed interaction logic between ERP and CMF MES Scheduling

Sequencing and scheduling are two operations performed by this module. Sequencing refers to determining which component lot will be processed first, while scheduling refers to determining the actual machine (among several machines that can perform the task) that will process the component lot. It is then related with the routing of lots through the shop floor so as to optimize the production flow, considering several objectives such as minimizing cycle times, lateness or waiting times. Hence, the level of aggregation present in this module is low, as it deals with individual lots and individual resources (machines), and the re-planning frequency is high due to its operational character. The planning horizon considered in the module is short.

Currently, planning functionalities determining production and inventory levels in discrete time-buckets, such as days, weeks, or months, are out of the scope of Critical Manufacturing MES. This means that it relies on planning being done at an ERP level. Then, from now on, the "Advanced Planning and Scheduling" module from CMF MES will just be referred to as "Scheduling" module to avoid ambiguity. The assumed flow of information can be seen in Figure 3.1. The ERP will supply MES with production orders (PO) affected by capacity planning. Each production order has a start and due date, as well as the product with a quantity to be produced. Then the Scheduling module from MES takes these orders to specify the sequence and job allocation while considering operational constraints. An integration message can be sent back to ERP to adjust the MRP based on the outputs of scheduling. This output will be aggregated at either a production order level or at a production order and operation level.

3.2 Problem description

In reality, it is empirically observed that capacity is not considered while running the MRP in the ERP, impacting the start date of production orders (Figure 3.2). When these production orders are imported into MES with an erroneous start date, a forward schedule is generated assuming they are correct. If the aggregated results from MES, influenced by the start date, are then sent back to the ERP for adjusting the plan in the next planning run, the ERP generates a new plan significantly

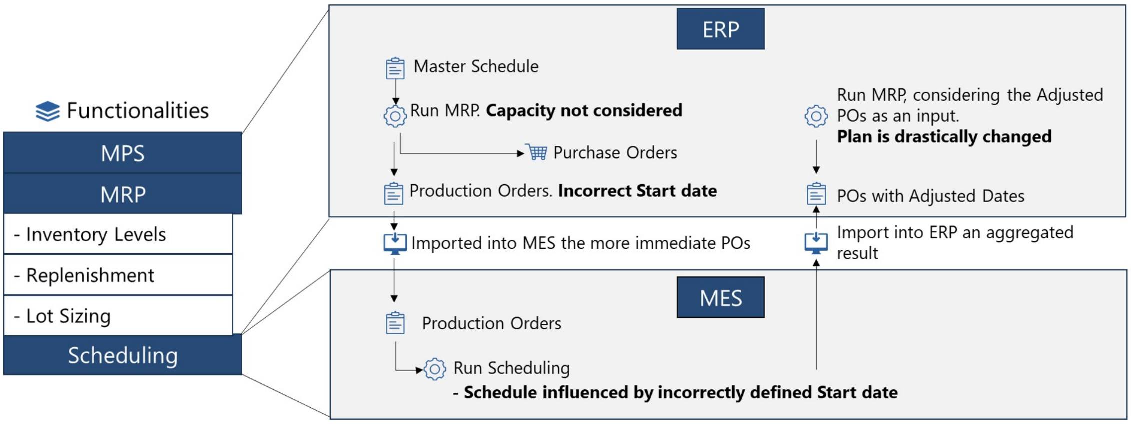


Figure 3.2: Problems empirically identified in the assumed interaction logic between ERP and CMF MES Scheduling

different from the previous one. This instability impacts decisions at the tactical level, such as purchasing or order promising.

Moreover, it is impossible for Scheduling to achieve a cost-effective plan by working with the wrong input. The consequences could be having too much work-in-process (WIP) on the line or having items produced late.

Planning has been more and more a functionality appreciated by customers. Although scheduling is essential in any manufacturing context, customers desire a system that ideally has all planning decisions related to manufacturing as concentrated as possible. This means not just provide information on how to best execute a given plan while considering priorities, constraints, and production conflicts - aspects dealt with by the scheduling functionality – but also understand the real capacity of the production plant by being able to answer questions such as *How much equipment is needed to achieve a certain volume?* or *How to use inventory to cope with lack of capacity?*. This issue underlines the different purposes for using planning and scheduling. The planning is much more focused on the preparation and coordination of decisions that will later affect the scope of scheduling.

Furthermore, one of the most significant advantages of acting on a tactical planning level, instead of just in an operational one, is the limited effort and inputs required to be provided to the system to make quick decisions and test different scenarios. This is highly valued by a manufacturing planner.

Therefore, the previously described factors point to the requirement to consider, as part of Critical Manufacturing MES, higher levels of supply chain planning, which can perform the necessary capacity validations and can support the current scheduling functionality. In the next section, some features for this planning solution will be analyzed.

3.3 General features for a Planning Solution

Firstly, it is important to mention that, ultimately, planning features depend on the complexities of the industry for which the planning solution is designed. However, some common issues exist, which will be addressed in this section.

Secondly, one must take into account that in the context of CMF's MES, whose scope focuses on the production itself, only planning aspects that directly affect production decisions are to be regarded. This means that the intricacies related to procurement and distribution are not to be considered.

After looking at planning solutions provided by competitors from CMF in the APS field such as Preactor, Adexa, or JDA (Lebreton et al., 2015), the following issues were identified:

Multi-site planning

This refers to the ability to simultaneously planning production in several factories located in different places. It is a relevant feature when one plant has lower production output than the others, possibly due to having older technology or just because it is smaller in size, but the extra available capacity from other factories can be used to bridge the gap between supply and demand.

What-if simulation

This refers to the possibility of running the plan with different planning parameters. This way, it is possible to see the impact of small modifications. It supports decision-making by providing insights on, for example, if a company is still able to meet demand after changes occur whether in terms of quantity or delivery dates.

Handling backorder and backlog

Backorder is an important consideration whenever there are capacity limitations that prevent all demand to be satisfied on time. Backlog is different from backorder, as it concerns the products ordered that have not yet been shipped to the customers but are within their defined due date. Hence, only backorders negatively affect a business and should be considered.

Finite and Infinite Capacity Planning

Finite capacity planning considers capacity restrictions while infinite does not. It is also possible to run the planning system in unconstrained mode at first, and then perform capacity leveling to adjust quantities produced to the available capacities of the resources.

Rough and Detailed Planning

This is related with the character of planning under different planning horizons.

3.4 Synthesis

Given that the planning solution from the ERP assumes infinite capacities when establishing the production plan and since this production plan is a pre-requirement for using the Scheduling module, CMF has added to its long-term roadmap having an integrated capacitated planning solution in MES.

Some general features for this planning solution were addressed such as multi-site planning, what-if simulation, handling backorders, finite and infinite capacity planning, and the ability to generate a rough or a detailed plan.

In the next chapter, the design of this planning solution will be specified, highlighting the interactions between the different levels considered as well as the decisions made in each level according to an adequate level of aggregation.

Chapter 4

Proposed Solution

In this chapter three planning levels are proposed. The top two provide context and perspective, with more focus being placed on the lowest level. It starts in Section 4.1, by presenting the general architecture, together with the level of aggregation. Then, the following three Sections, 4.2, 4.3, 4.4 will address each level individually. This chapter finishes in Section 4.5, by providing insights on how the planned information could be shown to the user.

4.1 General architecture

The proposed solution is composed of three different planning levels, as can be seen in Figure 4.1. The level of aggregation in time and product for each level, as well capacity decisions involved are presented in Table 4.1. In this table, the Scheduling level is shown just for comparison purposes as it will not be the object of analysis.

The Master Planning level is the most aggregate level. It has a planning horizon of one year divided into monthly time-buckets. The product is grouped into product families and whole factories are considered. Product families can be made of products with similar setup costs and identical

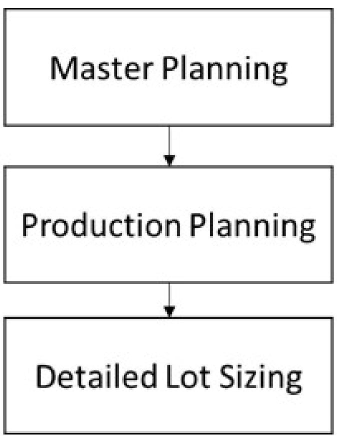


Figure 4.1: Architecture of the levels considered in the planning solution

Table 4.1: Level of aggregation in each planning level

Characteristic	**Master Planning**	**Production Planning**	**Detailed Lot-sizing**	**Scheduling**
Planning Horizon	1 year	3 months or 4 months	4 weeks to 8 weeks	few days
Time-bucket	Month	Week	Day	Continuous
Demand	Mid-term Forecasts	Short term forecasts and customer orders	Customer orders	Production orders
Product	Product family	End-Product	BOM	BOM
Resource	Factory	Production phase	Work center (Resource Group)	Machine
Capacity decisions	Subcontracting, capacity adjustment by acquiring machines (machine driven industry) or by workforce levels (labor driven industry)	Subcontracting, overtime, backorder	Overtime, backorder	Backorder

BOMs (Albrecht et al., 2015). Since it is such an aggregated level, it deals with forecasts of customer orders, and it is assumed that at this level all demand can be met without delay. It aims to solve the Multi-site planning problem and plan on capacity extensions, which usually cannot be changed in the short term.

The next level, entitled Production Planning considers a planning horizon of three or four months, divided into weeks. Here, product families are decomposed into end products and the shorter horizon translates into more certainty in terms of the demand assigned to a factory, then dealing with both short-term forecasts and customer orders. The inclusion of this level in the solution relates to the need to better specify the characteristics of the production process by considering the main production phases.

The last level considered, right above the Scheduling solution from CMF, was termed Detailed Lot Sizing. This terminology was inspired by the APS structure found in Fleischmann and Meyr (2003), in which an additional step is required to bridge the gap between the traditional APS Production Planning level and its Scheduling level. At this level, machines that perform exactly the same service, and are then related to the same operation in the production process, are grouped, constituting a work center. This level has a shorter horizon, typically of four to eight weeks, with daily time buckets. The product structure is considered by using the Bill-of-Materials (BOM).

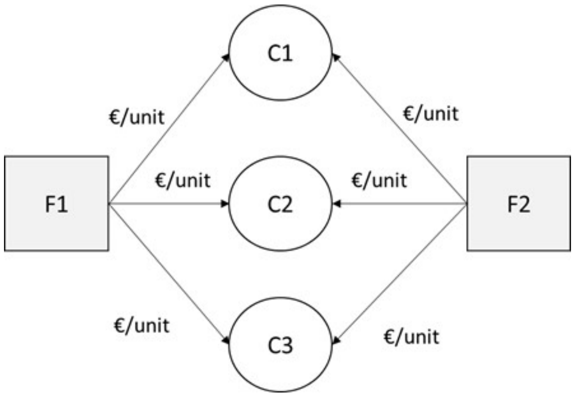


Figure 4.2: Multi-site planning problem

4.2 Master Planning

The Master Planning level involves at least two types of entities in its planning process: factories and customers. The process of determining which factories will produce to which customers (selling locations) is an allocation problem commonly designated as the Multi-site planning problem. This is a network design problem that can be visualized in Figure 4.2.

In this figure, it is portrayed the case in which plants are parallel to each other and every factory can produce and distribute its products to customers. Each node has a transportation cost associated with it, normally linearly related with the distance from the factory to the customer (Albrecht et al., 2015). There is another version for this problem in which factories are represented in series, being that the first factory produces intermediate products that are then sent into the next factories to complete the production process into finished items (Badhotiya et al., 2019).

The allocation of a customer to a factory depends on several factors such as the production capacity of each factory, the transportation costs, and on characteristics of the demand of the customer, such as the type of products requested or the size of the order. If the demand of a customer could be allocated to more than one factory at different moments in time, then it is referred to as a dynamic allocation problem. This version is regarded as more difficult than its opposed fixed version (Badhotiya et al., 2019) and it enables a reduction on total costs (e.g. transportation and inventory holding costs) alongside a better utilization of resources (Kanyalkar and Adil, 2007).

Beyond multi-site planning, other issues should fall in the domain of a Master Planning level. These are related to those decisions that need to start being planned at a sufficient aggregated level (such as monthly) so that they are available in shorter-term periods. Subcontracting and other capacity enhancement decisions such as purchasing new equipment are some examples.

Following, it is presented a general mathematical formulation for the parallel and dynamic allocation version of the multi-site problem, with the inclusion of subcontracting and capacity enhancement decisions:

Indices

i: Product Family, $i = 1,..., I$

p: Factory, $p = 1,..., P$

t_m: Month, t_m = 1,..., T^{MP}
l: Customer, l = 1,..., L
$\phi(p)$: Set of Product Families that can be produced in Factory p

Decision Variables

X_{ipt_m}: Quantity produced of Product Family i in Factory p in Month t_m
T_{iplt_m}: Quantity transported of Product Family i from Factory p to Customer l in Month t_m
I_{ipt_m}: Inventory quantity of Product Family i in Factory p at the end of Month t_m
S_{ipt_m}: Quantity subcontracted of Product Family i by Factory p in Month t_m
$E^+_{pt_m}$: Capacity expansion of Factory p at the beginning of Month t_m (in time units)
$E^-_{pt_m}$: Capacity reduction of Factory p at the beginning of Month t_m (in time units)
K_{pt_m}: Capacity of Factory p in Month t_m (in time units)

Parameters

C^{Prod}_{ip}: Production cost per unit of Product Family i in Factory p
C^T_{pl}: Transport cost per unit from Factory p to Customer l
C^H_i: Holding cost per unit of Product Family i (per month)
C^S_i: Subcontracting cost per unit of Product Family i
C^{E+}: Cost for expanding capacity per time unit
C^{E-}: Cost for reducing capacity per time unit
I_{ipt_0}: Initial inventory quantity of Product Family i in Factory p
K_{pt_0}: Initial capacity of Factory p (in time units)
F_{ilt_m}: Forecasted demand of a Product Family i related to a Customer l in Month t_m
a_{ip}: Time needed to produce one unit of Product Family i in Factory p
I^{MIN}_{ip}: Minimum stock level of Product Family i in Factory p
I^{MAX}_{ip}: Maximum stock level of Product Family i in Factory p
E^{MAX+}_p: Maximum expansion of capacity allowed in Factory p
E^{MAX-}_p: Maximum reduction of capacity allowed in Factory p
$S^{MAX}_{ipt_m}$: Maximum subcontracting quantity for Product Family i in Month t_m requested by Factory p
Δ_{pt_m}: Not used capacity in Factory p in Month t_m
C_{excess}: Cost for having excess capacity

Objective Function

$$Min \sum_{t_m=1}^{T^{MP}} \sum_{p=1}^{P} \sum_{i=1}^{I} \left(X_{ipt_m} C^{Prod}_{ip} + (I_{ipt_m} - I^{MIN}_{ip}) C^H_i + S_{ipt_m} C^S_i + \sum_{l=1}^{L} (T_{iplt_m} C^T_{pl}) \right) + \sum_{t_m=1}^{T^{MP}} \sum_{p=1}^{P} \left(E^+_{pt_m} C^{E^+} + E^-_{pt_m} C^{E-} + \Delta_{pt_m} C_{excess} \right) \quad (4.1)$$

Subject to

$$I_{ipt_m} = I_{ip,t_{m-1}} + X_{ipt_m} + S_{ipt_m} - \sum_{l=1}^{L} T_{iplt_m}, \qquad \forall i, p, t_m \tag{4.2}$$

$$F_{ilt_m} = \sum_{p=1}^{P} (T_{iplt_m}), \qquad \forall i, l, t_m \tag{4.3}$$

$$K_{pt_m} = K_{p,t_{m-1}} + E^{+}_{pt_m} - E^{-}_{pt_m}, \qquad \forall p, t_m \tag{4.4}$$

$$\sum_{t_m=1}^{T^{MP}} E^{+}_{pt_m} \leq E^{MAX+}_{p}, \qquad \forall p \tag{4.5}$$

$$\sum_{t_m=1}^{T^{MP}} E^{-}_{pt_m} \leq E^{MAX-}_{p}, \qquad \forall p \tag{4.6}$$

$$S_{ipt_m} \leq S^{MAX}_{ipt_m}, \qquad \forall i, p, t_m \tag{4.7}$$

$$\sum_{i=1}^{I} (a_{ip} X_{ipt_m}) + \Delta_{pt_m} = K_{pt_m}, \qquad \forall p, t_m \tag{4.8}$$

$$I^{MIN}_{ip} \leq I_{ipt_m} \leq I^{MAX}_{ip}, \qquad \forall i, p, t_m \tag{4.9}$$

$$X_{ipt_m}, S_{ipt_m} = 0 \qquad \forall i \notin \phi(p),\ p, t_m \tag{4.10}$$

$$X_{ipt_m}, I_{ipt_m}, S_{ipt_m}, T_{iplt_m}, E^{+}_{pt_m}, E^{-}_{pt_m}, K_{pt_m} \geq 0, \qquad \forall i, p, l, t_m \tag{4.11}$$

The objective function 4.1 aims at minimizing costs of production, inventory holding, subcontracting, transportation, and costs incurred due to expanding and diminishing of capacity. It is considered a penalty for excess capacity in a factory that is not utilized. The fact that different factories have different efficiencies in producing product families is expressed by the different production cost C^{Prod}_{ip}.

The inventory balance for product families in each factory is presented in Constraints 4.2. Note that a factory can produce and subcontract a quantity of a product family, that can then be transported to meet the forecasted demand of several customers. Constraints 4.3 imposes that the forecasted quantity of a product family requested by a customer in a given month can be satisfied by the quantities transported from different factories. At this level of aggregation, it is assumed that the forecasted demand will be satisfied without any backorder.

Constraints 4.4 states that the capacity of a factory in a month can change depending on it being expanded or reduced. In constraints 4.5 and 4.6, a limit on capacity variation (expansion and diminishing) is imposed over the considered planning horizon. In constraints 4.7 the amount that a factory can subcontract of a product family in a month is limited. This limit $S_{ipt_m}^{MAX}$ can either be related to the capacity of the supplier in providing that product family or with budgetary or strategical decisions in which the company wants to limit the influence of third-party companies - possibly to avoid losing control over performance quality or the ownership of delegated activities (Guers et al., 2014).

In a factory, the quantity produced of all product families is limited on constraints 4.8. The capacity of the factory which is not utilized will provide an incentive to be reduced. Constraints 4.9 places the inventory quantities between their minimum and maximum quantities. Constraints 4.10 force the production amount to zero if the factory is not capable of producing that product family. It was assumed that if a factory is not capable of producing a product family, then it should not make subcontracted decisions regarding that product family. The last constraints 4.11 define the domain of the variables.

Normally the Master Planning level works with deterministic inputs and uncertainty can be considered based on rolling schedules (Albrecht et al., 2015). Several methods can be used to solve the problem proposed, both exact and heuristics methods. Building on the mathematical programming model, optimization methods as branch and bound, and heuristics such as Lagrangian relaxation can be used. But also metaheuristics as the genetic algorithm, particle swarm optimization, or hybrid mathematical simulation may be used. (Badhotiya et al., 2019).

This model would provide several targets and decisions for the next planning level, enabling the coordination of decisions that would be made decentrally in a hierarchical planning system. The targets, which would give visibility to the lower planning level, are the inventory quantities of product families stored in a factory at the end of a month I_{ipt_m}. In addition, decisions, which constraint the next level, include the available capacity in the factory during the month K_{pt_m} as well as the part of the customer demand that would be assigned to each factory T_{iplt_m}/F_{ilt_m}. Subcontracting, similarly to production quantities, would be sent indirectly through the inventory target.

4.3 Production Planning

After planning at a multi factory level in the Master Planning model, more emphasis on a specific factory is needed. This is achieved with the Production Planning model that is responsible for determining the quantities that a factory should produce weekly to meet the demand of customers. At this level, the demand of a customer is related to the part of its whole demand assigned to this factory during the Master Planning level.

Since, at this level, the product aggregation is of end-items, as opposed to product families in the Master Planning level, it would be assumed that the percentage of the demand of a product family attributed to a factory would apply equally to all end-items belonging to that family.

To simplify the processes that take place inside the factory, the main phases of production are considered. These could be modelled as potential bottleneck operations that would limit the quantities of end items produced.

Following it is proposed a model formulation that represents the scope and decisions made at this level. In terms of capacity, it is highlighted the possibility of not fully meeting the demand, by incorporating backorder. The demand considered is for end items in a week, which implies that before running this model, the demand placed by all customers in the same week must be aggregated for each end item. The formulation also considers overtime and further details if and how subcontracting should be done (per week and end-item).

Indices and Sets

j: End-item , j = 1,..., J

t_w: Week, t_w = 1,..., T^{PP}

s: Production phase, s = 1,..., S

$\lambda(i)$: Set of End-items j belonging to Product Family i

Decision Variables

X_{jt_w}: Quantity produced of End-item j in Week t_w

I_{jt_w}: Inventory quantity of End-item j at the end of Week t_w

O_{st_w}: Overtime used in Production phase s in Week t_w (in time units)

B_{jt_w}: Backorder quantity of End-item j in Week t_w

S_{jt_w}: Subcontracted quantity for End-item j in Week t_w

Parameters

K_{st_w}: Capacity in Production phase s in Week t_w

D_{jt_w}: Demand of End-item j in Week t_w

C_j^B: Backorder cost per week and unit for End-item j

C_j^H: Inventory holding cost per unit of End-item j (per week)

C_s^O: Overtime cost per time unit in Production phase s

C_j^{Prod}: Production cost per unit of End-item j

a_{sj}: Time needed in Production phase s for producing one unit of End-item j

C_j^S: Subcontracting cost per unit of End-item j

O_s^{MAX}: Maximum overtime available per week in Production phase s (in time units)

I_j^{MAX}: Maximum stock level of End-item j

I_j^{MIN}: Minimum stock level of End-item j

$S_{jt_w}^{MAX}$: Maximum subcontracted quantity of End-item j available in Week t_w

I_{jt_0}: Initial inventory quantity of End-item j

B_{jt_0}: Initial backorder quantity of End-item j

Objective Function

$$Min \sum_{t_w=1}^{T^{PP}} \sum_{j=1}^{J} \left((I_{jt_w} - I_j^{MIN}) C_j^H + X_{jt_w} C_j^{Prod} + S_{jt_w} C_j^S + B_{jt_w} C_j^B \right) \quad + \quad \sum_{s=1}^{S} \sum_{t_w}^{T^{PP}} \left(O_{st_w} C_s^O \right) \tag{4.12}$$

Subject to

$$I_{jt_w} = I_{j,t_w-1} + X_{jt_w} + S_{jt_w} + B_{jt_w} - B_{j,t_w-1} - D_{jt_w}, \qquad \forall j, t_w \tag{4.13}$$

$$\sum_{j}^{J} (X_{jt_w} a_{sj}) \leq K_{st_w} + O_{st_w}, \qquad \forall s, t_w \tag{4.14}$$

$$O_{st_w} \leq O_s^{MAX}, \qquad \forall s, t_w \tag{4.15}$$

$$S_{jt_w} \leq S_{jt_w}^{MAX}, \qquad \forall j, t_w \tag{4.16}$$

$$I_j^{MIN} \leq I_{jt_w} \leq I_j^{MAX}, \qquad \forall j, t_w \tag{4.17}$$

$$X_{jt_w}, I_{jt_w}, O_{st_w}, B_{jt_w}, S_{jt_w} \geq 0, \qquad \forall j, s, t_w \tag{4.18}$$

The objective function 4.12 aims to reduce the costs of holding inventory, production, subcontracting, backorder, and overtime. Constraints 4.13 refer to the inventory balance for end-items. Note that the backorder quantity from a period enters as an additional demand in the next period (-B_{j,t_w-1}). Constraints 4.14 impose a limit on the number of end-items produced. The overtime used is limited in Constraints 4.15, and in constraints 4.16 subcontracting quantities are restricted. Inventory levels must be between their minimum and maximum values, in constraints 4.17. Lastly, the domain of the variables is specified in constraints 4.18.

Having characterized the scope of the Production Planning level, it is important to understand how it will relate to the targets provided by the Master Planning. The inventory targets of product families received must be disaggregated into inventory targets for end items. In Vogel et al. (2017), one can find an approach to do so by using a disaggregating factor estimated based on demand (equations 4.19 and 4.20). This factor is then multiplied by the inventory quantity of the product family i in the factory p under analysis (equation 4.21) to define the new target with the desired aggregation. This new target $I_{jt_w}^{MP}$, at the Production Planning level, has to be satisfied at least by a certain amount f (equation 4.22), on the last week of each month (set of target weeks TW).

$$V_j = \frac{\sum_{t_w=1}^{T^{PP}} D_{jt_w}}{\sum_{j \in \lambda(i)} \sum_{t_w}^{T^{PP}} D_{jt_w}}, \qquad \forall j \tag{4.19}$$

$$\sum_{j \in \lambda(i)} V_j = 1, \qquad \forall i \tag{4.20}$$

$$I_{jt_w}^{MP} = V_j I_{iptm}, \qquad \forall t_w \in TW \tag{4.21}$$

$$I_{jt_w} \geq f I_{jt_w}^{MP}, \qquad \forall t_w \in TW \tag{4.22}$$

The Production Planning also sends targets to a lower level. These would be the inventory levels I_{jt_w} and backorder quantities B_{jt_w}. The amount subcontracted S_{jt_w} would constitute a decision.

4.4 Detailed Lot Sizing

This level is the most relevant to the proposed solution since it is the one that would directly interface with the Scheduling module from CMF. The main goal of this level is in determining the quantities of each item that should be produced each day while considering the capacity of the resource groups. By doing so, it is expected that a better date for start production is defined when compared with the date currently provided by the ERP. In addition, the initial setup activities proposed at this level will also be important in providing these accurate start dates.

Following it is presented a formulation that could be used to describe this level, based on the Multi-level Capacitated Lot Sizing Problem (Prakaiwichien and Rungreunganun, 2018):

Indices and Sets

u: Item
t: Day
m: Work center
μ: Set of items, $\mu = \{1, ..., U\}$
$\mathscr{M}$: Set of work centers, $\mathscr{M} = \{1, ..., M\}$
τ: Set of Days, $\tau = \{1, ..., T\}$
μ_m: Set of items u produced in work center m
S_u: Set of direct successors of item u
End_Items: Set of End-items
$Components$: Set of Components

Variables

Y_{ut}: Binary setup variable of item u in day t
X_{ut}: Production quantity of item u in day t
I_{ut}: Inventory quantity of item u at the end of day t
O_{mt}: Amount of overtime used in work center m at day t
B_{ut}: Backorder quantity of end-item u at the end of day t

Parameters

r_{uj}: Quantity of item u directly required to produce one unit of item j (Gozinto factor)
K_{mt}: Available capacity of work center m in day t
d_{ut}: External demand of end-item u in day t
C_u^H: Holding cost of item u per unit and day
C_u^S: Setup cost of item u
t_u^P: Production time per unit of item u (in time units)
t_u^S: Setup time for the production of item u (in time units)
b_u: Sufficiently big number

C_m^0: Overtime cost in work center m per time unit
C_u^B: Backorder cost per unit of end-item u
S_{ut}: Subcontracted quantity of end-item u in day t
$MaxI_u$: Maximum stock of item u (per day)
$MinI_u$: Minimum stock of item u (per day)
I_{ut_0}: Initial inventory of item u
B_{ut_0}: Initial backorder of end-item u
O_m^{MAX}: Maximum overtime available for work center m (in time units)

Objective Function

$$Min \quad Z1 = Min \sum_{u \in Components} \sum_{t \in \tau} \left(Y_{ut} C_u^S + (I_{ut} - MinI_u) C_u^H\right) + \sum_{u \in EndItems} \sum_{t \in \tau} \left(Y_{ut} C_u^S + (I_{ut} - MinI_u) C_u^H + B_{ut} C_u^B\right) + \sum_{m \in \mathscr{M}} \sum_{t \in \tau} \left(O_{mt} C_m^O\right) \tag{4.23}$$

Subject to

$$I_{ut} = I_{u,t-1} + X_{ut} - \sum_{j \in S_u} \left(r_{uj} X_{jt}\right), \qquad \forall u \in Components, t \tag{4.24}$$

$$I_{ut} = I_{u,t-1} + S_{ut} + X_{ut} + B_{ut} - B_{u,t-1} - d_{ut}, \qquad \forall u \in Enditems, t \tag{4.25}$$

$$\sum_{u \in \mu_m} \left(t_u^P X_{ut} + t_u^S Y_{ut}\right) \leq K_{mt} + O_{mt}, \qquad \forall m, t \tag{4.26}$$

$$X_{ut} \leq b_u Y_{ut} \qquad \forall u, t \tag{4.27}$$

$$MinI_u \leq I_{ut} \leq MaxI_u, \qquad \forall u, t \tag{4.28}$$

$$O_{mt} \leq O_m^{MAX}, \qquad \forall m, t \tag{4.29}$$

$$X_{ut}, I_{ut} \geq 0, \qquad \forall u, t \tag{4.30}$$

$$Y_{ut} \in \{0, 1\}, \qquad \forall u, t \tag{4.31}$$

The objective function 4.23 aims to reduce total costs, composed by the inventory holding and setups costs (both for components and end-items), backorder costs for end-items, and overtime costs used in resource groups. An interesting consideration is that although a given end-item may be worth more than other end-item (C_u^B), its quantity to be backorder will also depend on the

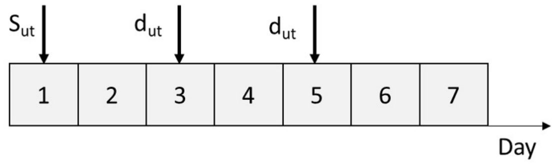


Figure 4.3: Representation of Subcontracting decision sent from the Production Planning level

capacity that it consumes. This means that an item that requires a lower capacity utilization could be the one selected to be produced, despite being less valuable – the objective function tries to minimize the backorder cost for all units delayed.

Constraints 4.24 refer to the inventory balance for components, with the term $\sum_{j \in S_u} (r_{uj} X_{jt})$ referring to the dependent demand induced by the parent-component relation between items. Constraints 4.25 also show inventory balance but for end-items. The choice to use two different sets of constraints is due to additional variables that would only exist for end-items, namely the subcontracting S_{ut} and backorder B_{ut}, and because it is assumed a case in which external demand solely applies to end-items. Subcontracting quantity is a decision that would be sent from the Production Planning level. This quantity is assumed to arrive on the first day of the week so that an inferior quantity of that end-item would be produced throughout the week (Figure 4.3). Then, it helps to satisfy the demand for that product without having to use capacity on resource groups. It is considered that backorder only exists for end-items since having backorder for components could be problematic and induce a wrong production plan in which the parent product would start being produced without physically having its required components.

Constraints 4.26 impose the limit for using the available capacity of the work center, which could be extended by using overtime. In this formulation, each item, produced in only one work center, would consume capacity in direct relation to the quantity produced and induce a setup on the resource group every time it would be produced. Note that the notion of setup carryover - which states that no setup would be performed in producing an item in a machine if it continues being produced since a previous period - requires the introduction of state variables and can only be considered at a level of aggregation of individual machines, not at the level of resource groups (Buschkühl et al., 2010). By considering the capacity of a work center in a day K_{mt} it is possible to account for the effect of not having all machines available due to maintenance or downtime.

Constraints 4.27 state that an item is produced only if there is a setup. The variable bu imposes an upper bound for the quantity produced. For end-items, the value of this variable can be estimated by adding the whole demand for that end-item during the planning horizon $\sum_{t \in \tau} d_{ut}$. For components, the estimation of bu would be done by multiplying the demand of end-items (that uses that component) by the successive incorporation factors connecting the component in the BOM level to that end-item. Constraint 4.28 imposes minimum and maximum quantities on the inventory of an item, and constraint 4.29 limits the overtime that could be used in the work center in a day. Lastly, constraints 4.30 and 4.31 state the domain of the variables.

The value of the parameters for this model related to business decisions essentially costs such as C_u^H, C_m^0, could be obtained from the company ERP, while other more operational such as t_u^P could be found in a MES. By aggregating information at the scheduling level, other parameters can be determined. Examples are the capacity of the work center, K_{mt}, and the setup time of an item in the work center, t_u^S. K_{mt} is obtained by adding the capacity of each machine in the same time-bucket t, Cap_{maq_t} as shown in equation 4.32. For estimating t_u^S, it can be used the setup time of the item in one machine, $t_{u_{maq}}^s$, together with the number of machines in the work center, $Number_{maq}$, as represented in equation 4.33. Hence, it is assumed that when an item is produced it induces a setup on all machines of the work center used for processing it.

$$K_{mt} = \sum_{maq \in m} (Cap_{maq_t}), \qquad \forall t \tag{4.32}$$

$$t_u^S = t_{u_{maq}}^S Number_{maq}, \qquad \forall u \tag{4.33}$$

Regarding the inventory and backorder targets provided by the Production Planning level, the Detailed Lot Sizing model could handle them by penalizing in the objective function 4.23 their deviation from the value of the same variables in the homologous periods. These homologous periods correspond to the last day of each week, defining the set of target days TD. Since inventory targets sent from the Production Planning level are defined in terms of end-items, no disaggregation would be needed. So, for inventory, a penalization by a big number M would occur only if the inventory quantity I_{ut} falls shorts from its target I_{ut}^{target}, as expressed in equation 4.34. While for backorder, this penalization would only apply if the value of the variable B_{ut} would be greater than its target B_{ut}^{target} (equation 4.35).

$$Min\, Z2 = Z1 + \sum_{u \in EndItems} \sum_{t \in TD} \left(M(I_{ut}^{target} - I_{ut}) \right) \tag{4.34}$$

$$Min\, Z3 = Z1 + \sum_{u \in EndItems} \sum_{t \in TD} \left(M(B_{ut} - B_{ut}^{target}) \right) \tag{4.35}$$

The interaction between the Detailed Lot Sizing and the Scheduling levels could be summarized in Figure 4.4. The lot size of an item with a date for production, X_{ut}, would be passed on for execution, constituting a production order. This production order would be characterized by a product, which determines the processing steps and the types of machines required for processing, a quantity, as well as the date of production. Then the Scheduling would determine the sequence of the products produced in the day and the actual machines used for the processing.

After running the Scheduling, which accounts for operational constraints, feedback would be sent to the Detailed Lot Sizing on whether the capacity of work centers K_{mt} should be adjusted. Quadt and Kuhn (2009) present a methodology to adjust the capacity of a bottleneck work center when this variable is overestimated in a lot-sizing model, meaning that not all quantities defined to be produced can be scheduled. The key takeaway to perform this adjustment is based on a backtracking mechanism with the aggregate capacity in lot sizing being reduced in an amount corresponding to the need for producing the not scheduled units. In the paper presented by these

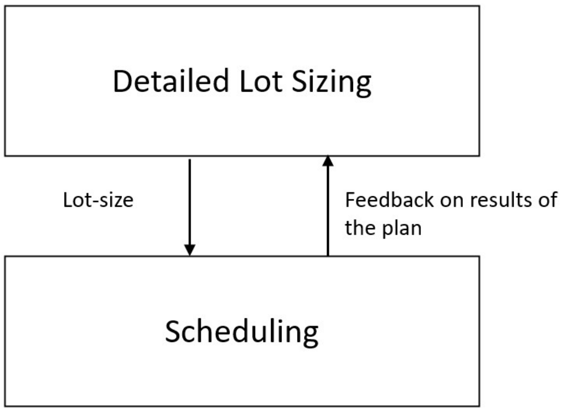


Figure 4.4: Interaction between Detailed Lot Sizing and Scheduling

authors, the lot-sizing model would iterate in a forward period-by-period manner and covering the planning horizon, with the scheduling procedure only being applied to a single period (coinciding with the first period of the lot-sizing). Hence, when no feasible solution could be found (in the scheduling procedure) for a period t, the capacity in the lot sizing for that period t would be reduced, and a period backtracking is performed so that in the next iteration, the model could run with capacity adjusted and quantities previously attributed to being produced in period t could be shifted to period t-1.

4.5 Display of information to the user

From the levels presented, the production planner wants to have the big picture of the results. Two of the most common graphs presented in commercial planning solutions, such as *Preactor APS*, are capacity usage graphs and graphs showing the evolution of the inventory level for a given product over time, designated by stock profile.

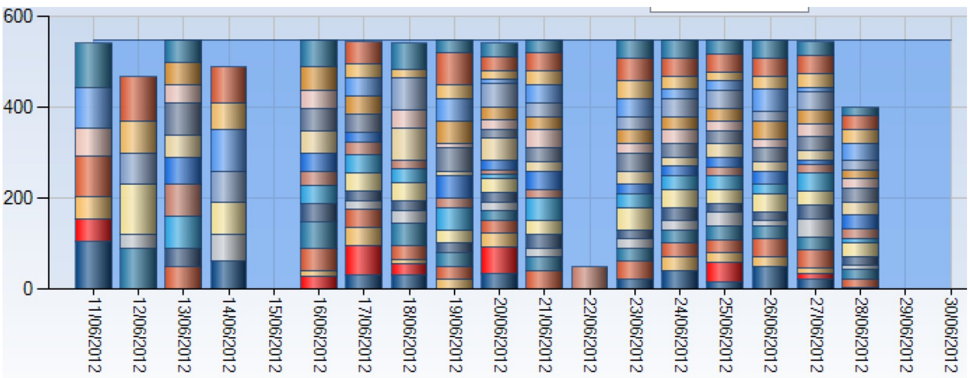


Figure 4.5: Capacity usage graph from Preactor APS system (Siemens, 2016)

In the first, shown in Figure 4.5, the capacity used is compared against the maximum capacity available (represented by the blue background area). The stacked columns represent production volumes. Note that this graph can be shown for the three levels proposed since depending on

Figure 4.6: Stock profile graph from Preactor APS system (Connect, 2021)

the level of aggregation, capacity used in factories, production phases, or work centers could be evaluated in the corresponding time buckets of its level. This graph is a visual representation of equations 4.2, 4.13, 4.24, 4.25.

The stock profile graph, in Figure 4.6, can be displayed for any product, depicting how the inventory will change following the planned quantities produced (represented by the blue bars) and the expected demand for the product (represented by the red bars). The inventory line (represented in black) will be placed between the maximum and minimum stock levels, here illustrated by the upper blue and lower yellow areas, respectively. The yellow line represents the shelf life of the product.

Costs are also crucial information to any planner since that is perhaps the most relevant criteria when choosing between different planning alternatives. Therefore, reports could be made stating how total costs are distributed along its components (setups, holding, backorder and overtime). These costs could then, for example, be shown in a bar chart.

Finally, insightful statistics related to business decisions could be explored such as the percentage of demand that could not be fulfilled on time due to capacity limitations, the percentage of subcontracting versus quantity produced, or the amount of overtime required by period.

Chapter 5

Case Study

In this chapter, the proposed Detailed Lot Sizing model will be exemplified in a part of the back-end semiconductor industry. Starting in Section 5.1, an overview of the main phases of the production process will be presented, followed by the computational experiments accomplished in Section 5.2.

5.1 Overview of the production process

In the semiconductor industry, the production process can be divided into four main phases (Figure 5.1): WaferFab (referring to the fabrication of the Wafers), Probe (in which the dies integrating the wafers are electrically tested), Assembly (in which integrated circuits, IC, are assembled and packaged) and Test, that makes sure ICs meet customer requirements (Mönch et al., 2013). Wafers are thin disks made of silicon in which many identical chips, also known as dies, are fabricated. The front-end part of the process comprises the first two phases, while the back-end takes care of the last two.

The global character of this industry makes multi-site planning a relevant feature. In the past, the four production phases used to take place inside the same factory. Then, it is not unreasonable to imagine a "virtual" factory that would comprise the four phases. In this sense, the proposed Master Planning level would plan production for several "virtual" factories located at different places. The Production Planning level would focus on one "virtual" factory, evaluating capacity at the four production phases. And the Detailed Lot Sizing level would consider the various groups of resources related to a production phase.

This case study will only address the assembly phase of the back-end process, which corresponds to a concrete request from a client from CMF. In this example, two different types of final products are produced: a Dynamic Random-Access Memory (DRAM) and an embedded Multi-Chip Package (eMCP). To produce these final products, two types of components, both having an in-house preparation process, must be considered. Firstly, the wafers, having already undergone the frontend part of the semiconductor process, must be prepared to have their individual dies assembled into either type of final product. Secondly, a substrate, responsible for providing the

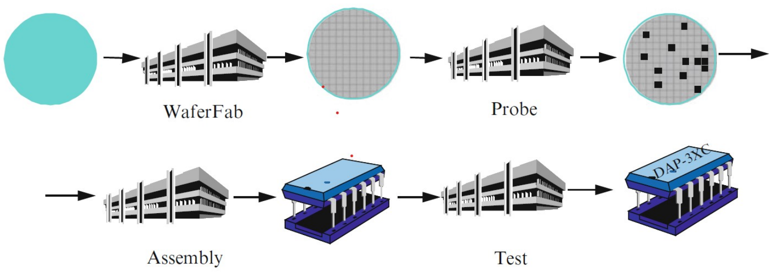


Figure 5.1: Stages of semiconductor manufacturing (Mönch et al., 2013)

physical support to the assembled must be prepared as well. Once the various components are ready for assembly, an automated equipment will handle incoming wafers and substrates and will match each die of the wafer into a specific position within the substrate (Instruments, 1999). In the case of the example studied, a DRAM product only requires one die, whereas, in an eMCP product, more than one die is attached to each substrate position.

In figure 5.2 the main steps for preparing a wafer to obtain the separated dies are shown. The process starts with the tape being applied to the front of the wafer. This tape enables the dies in the wafer to hold together during the cutting process, by mounting them into a frame. It also helps to make the wafer less vulnerable to break during the next phase. Then, grinding is performed on the backside of the wafer to reduce its thickness. Following, there is the grooving step, in which grooves are made in the wafer with a laser to ease the wafer sawing phase. Here, the wafer is cut to get the dies separated. The process terminates with a UV cure. Usually, each of these operations is performed in a group of machines. For some wafers, however, some operations are combined and performed on independent machine groups under a different technology. For example, “Stealth Dicing” uses a laser to perform the sawing process, with the advantage of not generating debris, thereby replacing the grooving and wafer sawing operations. Throughout the Wafer Preparation process, the wafers are transported within a Front Opening Universal Pod (FOUP) carrier, with the capacity for 25 wafers (Hu et al., 2009).

A substrate is generally prepared according to the steps in Figure 5.3. Firstly, the substrate is prepared, meaning that debris is removed from its surface so that it can receive electrical components in the SMT (Surface Mount Technology) phase. Since the SMT is executed in an integrated and uninterrupted production line, it will be considered as a single operation for the purposes of the current case study. Then the substrate undergoes a cure and a cleaning phase (plasma), that will improve the adhesion of the die attach epoxy used when assembling the die to the substrate. For DRAM types of final products, the substrates used to package the dies typically do not go through the SMT and Pre-Bake phases.

Figure 5.2: Steps in preparing a wafer to obtain separated dies

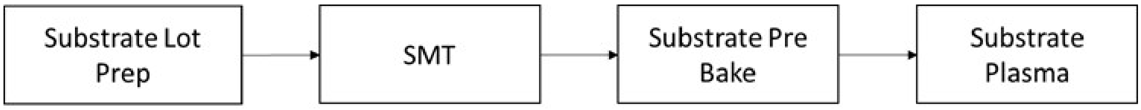


Figure 5.3: Steps in preparing a substrate

The assembly process of the die and the substrate is different depending on the type of product, due to the technology used to interconnect the two. In eMCP products, electrical wires are used for the connection (wire bonding), whereas in DRAM products, conductive solder bumps are used (flip-chip technology) (Datta, 2020). In Figures 5.4 and 5.5 , it is presented the process flow for obtaining DRAMs and eMCPs, respectively. The Flip Chip Bonder step for obtaining a DRAM, similarly to the SMT line, is composed of operations occurring in an integrated and uninterrupted production line, and therefore will be regarded as a single operation performed in one work center.

For an eMCP, the process starts with the Die attach phase, in which the attachment between the die and the substrate is performed. During this phase, the substrate is also coated with a material, such as epoxy, in preparation for the die, to ensure adhesion between it and the substrate (Instruments, 1999). Then the dies undergo a cure process, which is followed by plasma treatment to remove any contaminants within the die's bond pads, and that could impede the formation of the wire bond in the last phase - wire bonding (Nowful et al., 2001). The stages from Die Attach to Wire Bonding are visited repeatedly in a number proportional to the number of dies to be stacked. This process occurs in a complex job shop manufacturing environment, with multiple parallel machines able to perform each individual production stage, considering among other constraints the reentry of the same final product lot into the same work centers (Gupta and Sivakumar, 2006).

5.2 Computational experiments

5.2.1 Data and application of the algorithm

The data used was obtained from Critical Manufacturing MES and then aggregated to be compatible with the proposed model. Six end-items were considered, two belonging to the DRAM type and four to the eMCP type (Table B.1 of Appendix B). For each end-item, the BOM, represented in Table B.2 of Appendix B, has only one level made of a substrate and wafer dies.

Re-entrant flows in the same work center for processing the same product were regarded as one big operation at the Detailed Lot Sizing level. Hence, the processing times of these individual

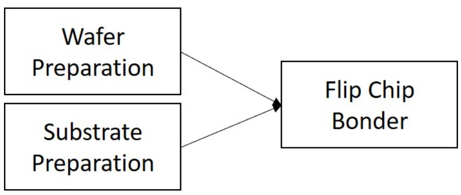


Figure 5.4: Steps for obtaining a DRAM product

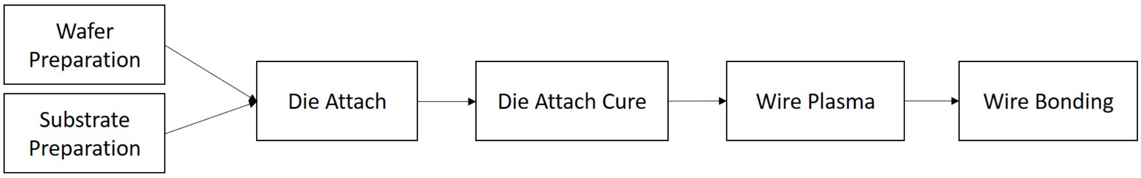


Figure 5.5: Steps for obtaining an eMCP product

steps, provided by MES, were added together and one setup for the work center was considered when producing that product.

Another issue relates to the same item passing through several work centers. So, for each item one must have both its production and its setup time on the work centers used to produce it – variables t_u^P and t_u^S in equations 4.26, are changed to t_{um}^P and t_{um}^S , respectively. The consequence is that no WIP is considered in the formulation since for an item to be produced, it has to pass through all the intermediary stages of production in the day to be completed. In Appendix C, it is shown the items that are processed in each work center along with its variable processing and setup times. Note that for work centers related to operations performed on whole wafers, such as the Tape Laminator, the variable processing time corresponds to an individual die of the wafer. Also, other operations such as the Substrate Pre-Bake and the Die Attach Cure require a fixed cycle time, independent of the number of pieces processed. As the proposed formulation, only deals with a variable processing time for each item, those fixed times were approximated to variable ones by dividing them by the maximum number of units that could simultaneously be processed in one of those machines.

The period of demand analyzed was from 11/01/2021 to 07/02/2021, which corresponds to four weeks (Table B.3 of Appendix B). Its pattern can be characterized as lumpy with more concentration of orders in the middle of the planning horizon as seen in Figure B.1 of Appendix B.

The factory analyzed operated at two shifts per day, each with eight hours. These two figures together with the number of resources in each work center (presented in Appendix C) were used to compute the daily capacity of each work center. This capacity value was reduced to 70% in order to anticipate potential infeasibilities at the scheduling level. This reduced value is referred to as the standard capacity of a work center K_{mt}. The maximum overtime available of a work center O_m^{MAX} was set to be 20% of its standard capacity. These capacity values are represented for each work center in Table B.5 of Appendix B.

No information on costs was available. Since these are inputs to the model, it was considered that an end-item would have lower holding costs (0.5 monetary units per unit per day) than substrates (1 monetary unit per day) and dies from wafers (5 monetary units per unit and day). The reason for considering higher costs for components was to try to depict the fact that components are perishable, with wafer dies being more critical components than substrates since they are not produced in-house and have to come from a front-end operation. The cost per minute of overtime used was set to 10 for all work centers. Regarding setup costs of an item C_u^S, they were estimated based on the opportunity cost related to the number of units that could be produced during the setup time of the work center, according to equation 5.1 :

$$C_u^S = \sum_{m \in Mu} \frac{t_{um}^S}{t_{um}^P} \qquad \forall u \tag{5.1}$$

With Mu representing the set of work centers used to process the item u.

Finally, the backorder cost of each end item was defined according to its demand. Products with a lower demand were assumed to be less relevant for the company and thus would have a lower backorder cost. A summary of the costs obtained for each item can be found in Table B.4 of Appendix B.

5.2.2 Scenarios explored

The model was solved using Gurobi Optimizer 9.1, modeled in Python using Spyder IDE, and importing the *gurobipy* package. The experiments were executed on a computer from Hewlett Packard with a processor Intel(R) Core(TM) i7-4702MQ CPU @ 2.20GHz 2.20 GHz and 16 GB RAM. The time limit for running the experiments was set to 300 seconds.

With the aim of assessing how the results obtained by the Detailed Lot Sizing algorithm varied under different circumstances, the following situations were tested:

- Reduction of the daily available capacity of one of the work centers known as a bottleneck from 70% (assumed in the base scenario A) to 50% (scenario B). This could be motivated by feedback received from a previous scheduling run and it is explored in the sub-sub section 5.2.2.1;
- The impact of both reducing (scenario C) and increasing (scenario D) setup costs by 50% in relation to base scenario A. Results are presented in sub-sub section 5.2.2.2;
- The inventory holding cost for wafer dies was reduced by 25% (scenario E), by 50% (scenario F), and 75% (scenario G) relatively to base scenario A, still remaining higher in comparison with the inventory costs of other items. This sensitivity analysis is relevant since the initial input value could have been set too high. The reader is directed to the sub-sub section 5.2.2.3;
- No overtime would be available for use (scenario H). More on sub-sub section 5.2.2.4;
- Reducing the standard capacity of work centers by the planned daily maintenance time of machines (scenario J). Results are shown in sub-sub section 5.2.2.5.

5.2.2.1 Impact of reducing capacity in the Die Attach work center

The Die Attach process is usually referred as the bottleneck in the back-end process (Park and Hur, 2020). This is because, in the shop floor, a large number of wafers (usually 25 wafers), is suddenly carried over to the die attaching process, resulting in an overload of the equipment (Park and Hur, 2020). In addition, many times, the arriving wafers are not properly separated during the

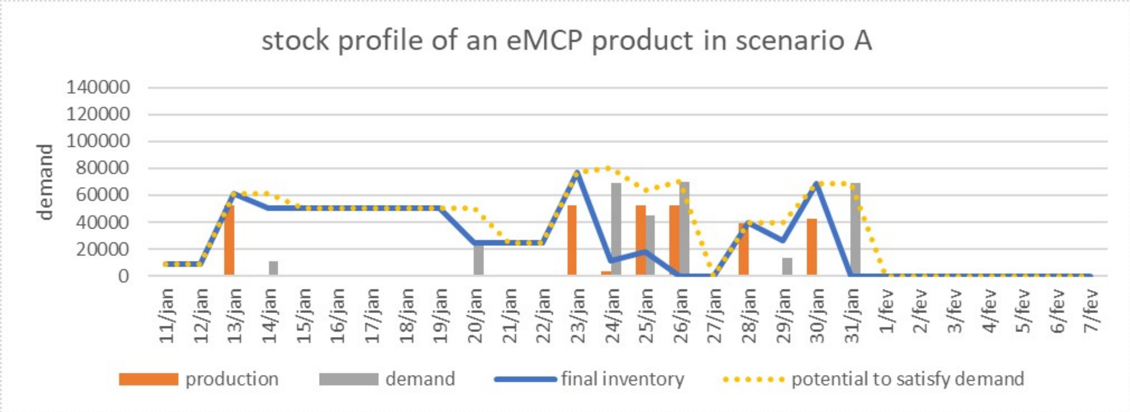


Figure 5.6: Stock profile graph of an eMCP product in scenario A

wafer sawing process, resulting in an unexpected equipment fault in the Die Attach, with loss of productivity (Park and Hur, 2020). To anticipate the severance of this lost time, the capacity of the Die Attach work center (8 hours X 2 shifts X 60 min X 17 machines), was reduced to 50% in scenario B. This would contrast with base scenario A which considers a limit of 70% on the capacity of all work centers.

Note that by changing the standard capacity, the availability for using overtime is different in the two scenarios. With the assumed value of 20% for the limit of maximum overtime, maximum capacity in the Die Attach work center (with overtime included) is, for scenario A, 84% of capacity (calculated as 70% + 70% X 20%), and for scenario B it is 60% (calculated as 50% + 50% X 20%).

Figures 5.6 and 5.7 represent the evolution of inventory quantities for an eMCP product (SPMCP224533) under scenarios A and B respectively. The line representing the potential to satisfy demand was calculated by adding the inventory quantity at the end of the previous day with the quantity produced that day. If this line is above the requested demand, then backorder is not generated. If it is below, backorder will be represented on the following day. The impact of reducing the capacity of the Die Attach work center translated into the generation of higher

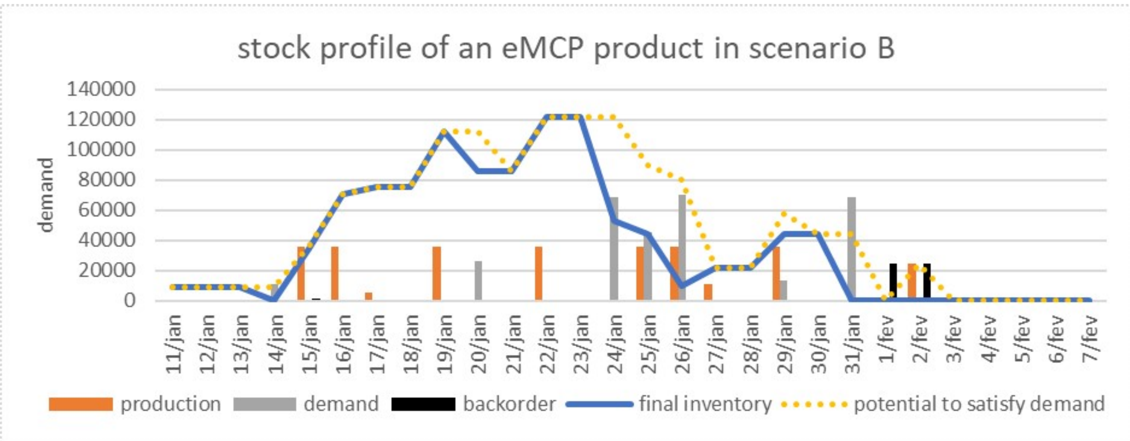
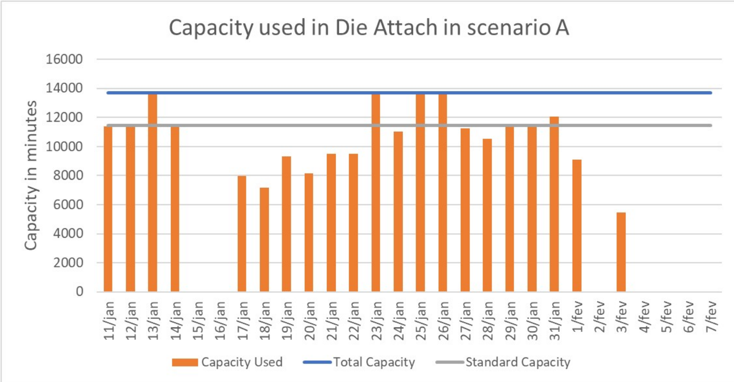
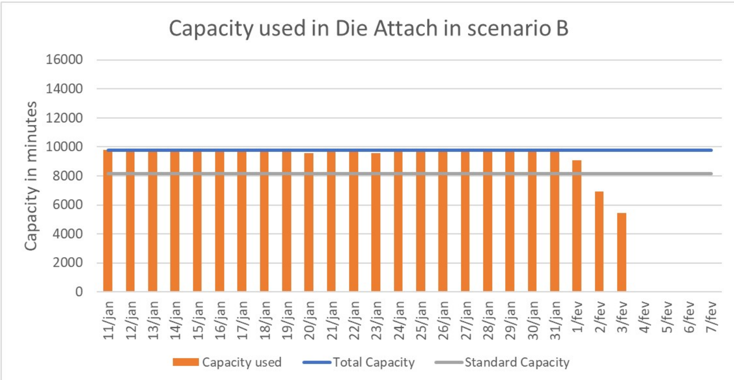


Figure 5.7: Stock profile graph of an eMCP product in scenario B



(a) Capacity used in Die Attach at 70% of available capacity (scenario A)



(b) Capacity used in Die Attach at 50% of available capacity (scenario B)

Figure 5.8: Impact of reducing the available capacity of the Die Attach work center

inventory quantities for the product throughout the first half of the planning horizon, during which demand levels were lower. Although on 15th of January, it was registered a small quantity of backorder (1811 units), the backorder quantity was considerably more expressive on the 31st of January and 1st of February (represented on the 1st and 2nd of February), being of 24418 units and corresponding to 36% of unsatisfied demand. In scenario B, the capacity of the Die Attach work center would be fully utilized during most days of the planning horizon (Figure 5.8), contrasting with the available capacity in scenario A.

The impact also manifested in the quality of the solution obtained. In scenario A, the gap was 1.1464% whereas the gap in scenario B was 5.8781%. Globally, total costs, in scenario B, increased by 236%, in comparison to scenario A. Even without considering backorder in total costs, whose evaluation can be rather subjective, the increase would be by 217%. In terms of inventory, its costs increased by 36%, with overtime costs increasing by 167% and setup costs increasing by 14%.

5.2.2.2 Impact of the setup cost when producing an item

One of the considerations arising from the Detailed Lot Sizing model formulation, was that to produce an item, all machines of a work center are used to produce it and thus would have to incur in a setup. This has the consequence of high setup times involved in the production and therefore its associated setup cost should also be high. By using the estimation presented in equation 5.1, large setup costs were obtained (Table B.4 of Appendix B), especially for producing wafer dies. This makes sense since the setup times regarded in its calculation are for a group of wafers, composed of many dies. This way there is a strong incentive to only produce large quantities of dies. For substrates, the setup costs are zero and so the metric "total number of setups", used to evaluate the impact of the setup cost, only refer to wafer dies and end items.

The results obtained for the base scenario A show a production plan (represented in Tables D.1 and D.2 of Appendix D), in which the quantities produced were concentrated in days close to demand. The overtime used in the different work centers mainly occurred in the second half of January due to the higher demand registered in that period, and it is shown in Tables F.1 and F.2 of Appendix F. Regarding inventory quantities (represented in Tables E.1 and E.2 of Appendix E), they mostly concern end-items produced in anticipation so that high demand in future periods could be satisfied. Inventory for dies was practically nonexistent most of the time. The reason for this is that their inventory cost considered is higher that makes generating inventory, not an attractive alternative. This is verified by the trade-off between producing extra units in one day to be stored and having a setup in the following day. For instance, the number of dies of WAFFLCH44789 produced on 29/jan was 109844. Having capacity in the work centers to produce these units alongside the 109844 units produced in the previous day, the setup cost of 121390 would still be considerably lower than the inventory cost of 549220 (5 monetary units times 109844 units). Moreover, due to the excess of capacity in work centers that process wafer dies, together with the relatively high holding cost, dies produced in a day are consumed in satisfying the dependent demand in that day, not generating inventory.

When the setup cost of wafer dies and end products are reduced by 50% (scenario C), total holding costs decreased by 17%, as seen in Table 5.1. This can be justified by a smaller quantity of end-items stored - having a total of 1831714 cumulative units of end-items stored in comparison to 2285436 cumulative units in scenario A - and having no wafer dies in inventory. This implies that wafer dies are produced and incorporated into the final product closer to the due date, requiring more overtime to complete production. The fact that the reduction by 50% did not increase the total number of setups considerably, highlights that the initial setup costs (in scenario A) were low compared to inventory costs.

On the other hand, when the setup cost of wafer dies and end products are increased by 50% (scenario D) more production is generated per production run, given that the total number of setups was greatly reduced and inventory costs increased (Table 5.1). Also, more overtime was used since the setup cost would compensate the extra overtime incurred.

5.2.2.3 Impact of reducing the inventory holding cost of a wafer die

To assess the influence that the inventory holding cost of wafer dies would have on the solution, its inventory holding cost was reduced by 25% (scenario E), by 50% (scenario F), and by 75% (scenario G). As expected, it was verified that when reducing the unitary inventory holding cost of

Table 5.1: Impact of increasing and decreasing setup costs relative to base scenario A

Setup Cost	Total Holding Cost	Total Setup Cost	Total Overtime Cost	Total Number Setups	Computational time (limit 300 seconds)	gap
Scenario C (50% reduction)	- 217 539 (-17%)	- 2 670 258 (-47%)	+14 600 (+10%)	98	77.46 seconds	0.000%
Base Scenario A	-	-	-	97	Time limit reached	1.1464%
Scenario D (50% increase)	+ 343 486 (+27%)	+ 1 803 737 (+32%)	+85 260 (+57%)	89	146.87 seconds	0.0003%

a wafer die, the relative importance of its setup cost would increase, producing more dies in one day to be assembled in the next day. As seen in Table 5.2, the savings in setup cost, due to the reduction in the number of setups for wafer dies, outweighed the increase in holding and overtime costs. The increase in the number of wafer dies, that led to the increase in inventory holding costs present in the same table, can be assessed in Figure 5.9, which represents the breakdown of the total cumulative units in inventory by product, during the planning horizon.

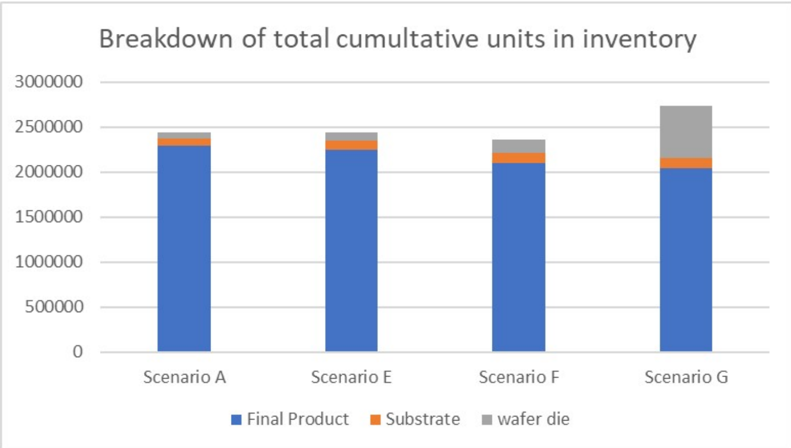


Figure 5.9: Breakdown of total cumulative units in inventory

The extra units of wafer dies produced in one day are not immediately incorporated into a final product (whose holding cost would be lower), because when wafer dies are produced in great amounts, the capacity of the work centers used to process final products were used to a maximum (with overtime included). Therefore, the final product is also produced in the day following the production of wafer dies, consuming these components from inventory.

Besides, it was verified that better gaps were obtained when the holding cost of a wafer die was reduced (Table 5.2).

Table 5.2: Impact of decreasing inventory holding cost of a wafer die relative to base scenario A

Wafer die holding cost	**Total Holding cost**	**Total Setup cost**	**Total Overtime cost**	**Number Setups for wafer dies**	**Computational time (limit 300 seconds)**	**gap**
Base Scenario A	-	-	-	65	Time limit reached	1.1464%
Scenario E (25% reduction)	+ 320 234 (+25%)	- 364 170 (-6%)	+ 30 850 (+20%)	62	Time limit reached	0.8425%
Scenario F (50% reduction)	+ 260 004 (+21%)	- 482 692 (-9%)	+ 44 220 (+29%)	61	Time limit reached	0.4517%
Scenario G (75% reduction)	+ 586 941 (+46%)	- 1 774 384 (-31%)	+ 68 660 (+46%)	49	169.33 seconds	0.0001%

5.2.2.4 Impact of not using overtime in work centers

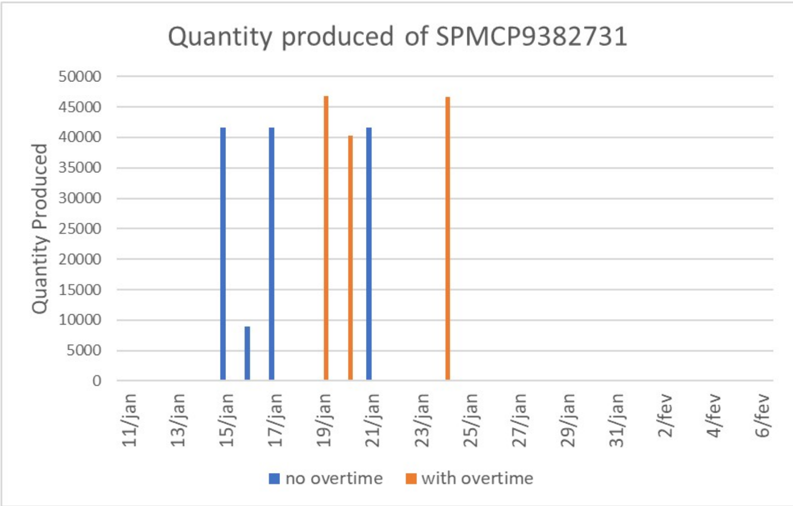
Overtime was a variable introduced in the model to allow capacity extension under a given limit. However, it might not always be available for use, due to regulations. Therefore, it was tested a scenario in which no overtime could be used and then compared with base scenario A (in which overtime was allowed at 20% of standard capacity). In Figure 5.10 this effect is exemplified in the distribution of the quantity produced for the product SPMCP9382731 and on the percentage of capacity utilized on the Die Attach work center.

When overtime is not used, more inventories will be generated since production occurs earlier, and more often, thereby resulting in a higher number of total setups (total of 106 in comparison with the 97 from scenario A). The number of setups increased for seven of the items analyzed, increasing the setup costs by 11%. Inventory costs increased by 35%, with total costs increasing by 13%. However, with no overtime and at the limit of 70% of available capacity in work centers, no backorder was generated.

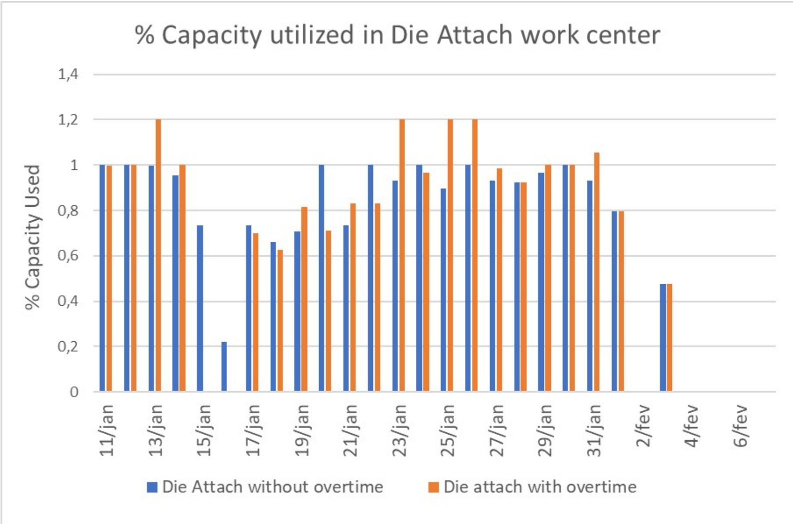
5.2.2.5 Impact of considering the preventive maintenance plan of machines

When an equipment is not available, either due to scheduled preventive maintenance or corrective maintenance, the normal production plan will be affected. In the semiconductor industry, equipment is expensive and therefore maintenance has to be regularly performed, affecting capacity (Li and Ma, 2017).

When incorporating the planned maintenance of machines in the Detailed Lot Sizing model, the value of capacity in work centers would vary each day. And since maintenance might just be performed on some machines of the work center, it was necessary to estimate how many machines would be available each day. This approach would provide a better value for calculating the daily



(a) Impact on the distribution of quantities produced



(b) Impact on the percentage of capacity used in Die Attach work center

Figure 5.10: Impact of overtime consideration

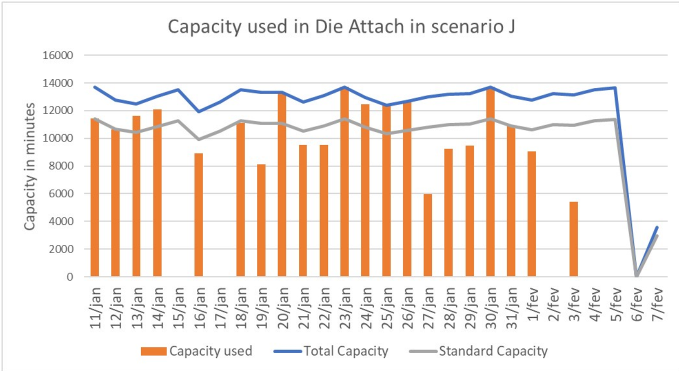


Figure 5.11: Capacity utilized in Die Attach work center when the preventive maintenance plan of its machines is considered

setup time that an item u should incur when being processed in a work center m, t^S_{umt}. The daily number of available machines in a work center was estimated using the following expression (and considering its value continuous since rounding to the closest integer would over or under-estimate capacity):

$$Machines\_available_t \times total\_time_t = Total\_machines_t \times total\_time_t - Maintenance_t$$

In this expression, the $total\_time_t$ corresponds to the capacity of the day (8 hours X 2 shifts X 60 min) reduced to 70% to try to anticipate operational constraints at the scheduling level, and for comparison purposes with the base scenario A, in which the same approach was followed. The maintenance time ($Maintenance_t$) associated with each work center in each day of the planning horizon, was taken from Critical Manufacturing MES. Its effect can be visualized in the irregular capacity lines from Figure 5.11 for the Die Attach work center, contrasting with the constant capacity lines shown in Figure 5.8, for the base scenario A.

It can be concluded that with lower available capacity, more frequent productions are induced, with total setup costs increasing by 4% in comparison to base scenario A (Table 5.3). Overtime costs increased by 40%, since using extra time more frequently would be a requirement to compensate for the downtime. This was verified by the increase in the number of times overtime was used in work centers (from 35 to 39). Regarding inventory, its cost increased by 16%. No backorder was generated, stressing that limiting capacity to 70% might not be enough to capture what occurs in the shop-floor, given that usually tight capacities exist in this industry. However, running the scheduling would be required to confirm this affirmation since the demand for products in the period might have been simply low due to the pandemic. Globally, this plan was 7% more costly than the plan generated in scenario A.

Table 5.3: Impact of considering preventive maintenance plan relative to base Scenario A

Maintenance Plan	**Total Holding cost**	**Total Setup cost**	**Total Overtime cost**	**Total Cost**
Scenario J	+204 713 (+16%)	+241 312 (+4%)	+60 060 (+40%)	+506 084 (+7%)

Chapter 6

Conclusion and Future Work

This study enabled a deeper knowledge of the structure of a planning system that performs capacity validation at different levels. Its aim was to understand the scope of each level, together with the decisions that should be made, and the inputs and outputs.

Firstly, and in order to identify best practices in the topic of production planning, it was conducted a literature review identifying the characteristics of modern planning systems, namely Advanced Planning Systems, which constitute the state of the art. Its hierarchical modular structure highlighted by authors such as Lebreton et al. (2015) and Fleischmann and Meyr (2003) served as inspiration for the proposed solution.

This solution would be composed of three planning levels, tackling issues under different aggregation: Master Planning, Production Planning, and Detailed Lot Sizing. The level of Master Planning essentially aims at solving the multi-site planning problem, in which the demand of customers would be allocated to production facilities. The Production Planning level would focus on a factory and simplify the production process to consider suspected bottlenecks in that factory in the form of production phases. Based on them, a production plan would be defined for each end product per week. Then, the Detailed Lot Sizing level would determine the daily lot size for each item in BOM while considering the capacity of groups of resources (work centers) in which items are processed. This solution constitutes a first attempt to incorporate planning to the current very operational MES solution provided by CMF. By being a general framework, its purpose was not to fall into the intricacies of any specific industry since that would most likely make it unusable for other industries.

A crucial understanding that should forward from the present work is on the communication between upper and lower levels, especially between the Detailed Lot Sizing level and the scheduling level. The feedback loop would enable capacity in work centers to be adjusted accordingly to the ability to schedule all production quantities defined.

As a second phase, the Detailed Lot Sizing model was applied to a case in the back-end semiconductor industry because it constitutes a major client segment of CMF, and thus data was easily obtained. It was verified that the initial inputs provided to the model have a great impact on the production plan obtained, whether products will be produced earlier to inventory or more frequently closer to due dates. Another key insight was that when capacity was reduced on a bottleneck work center, the ability to satisfy demand was compromised, with the generation of backorder, despite the increase in overtime. The impact of using overtime was also contrasted against the case of no overtime available, highlighting a greater number of setups with more inventory generated. Finally, the inclusion of the preventive maintenance plan for machines in a work center, which traditionally affects available capacity, was assessed on cost distribution, with setup, holding, and overtime costs increasing.

Limitations and Future Work

This work is not without its limitations and opportunities for improvement.

In the Master Planning model formulation, the value of capacity is reduced to a single value, representing the capacity of the factory. This would be valid if all products would consume capacity from the same pool of resources, which is a rather simplification of what happens in a factory. This could, however, be easily solved by introducing further indices related to the product family. Moreover, at this level, reducing capacity not utilized might not be a desired decision for capital-intensive industries. This is usually dealt with by management by increasing marketing expenses to stimulate demand, and so in the model, a high value could be given to the cost of this parameter to limit its influence.

In the Detailed Lot Sizing, the consideration of setups performed on every machine of the resource group every time a product is produced (in constraints 4.33) can be restrictive and lead to sub-optimal solutions, especially when the demand for a product is low. A possible alternative to consider the setup effect more straightforwardly would be to perform the lot-sizing with individual machines. This would, however, increase the complexity of the level because it would be required to determine which machine would produce the product.

Regarding the application of the Detailed Lot Sizing model to the semiconductor example, no backorder was generated in most experiments done, even when capacities in work centers were limited to 70%, and no overtime was allowed, or when the preventive maintenance plan for machines was included. This is quite remarkable since usually tight capacities are present in this industry. The main reason, allied to the low demand due to the pandemic, could lie in the high backorder costs defined. Although the value for the backorder cost of a product was differentiated in proportion to its demand, the base cost was subjectively defined, which might have been too high for the ideal trade-off between the company's costs. Therefore, in future work, the base value of the backorder cost could be further analyzed according to the company's backorder policies.

A further improvement to the applicability of the Detailed Lot sizing model to the semiconductor industry, could be the inclusion of setup carryover. Haase (1998) states that when setup carryover is not considered, a feasible solution might not be obtained, which explains multiple formulations in the semiconductor field that account for setup carryover either in a simultaneous lot-sizing and scheduling model or at the lot sizing level as presented in Quadt and Kuhn (2009). This would, however, require the consideration of individual machines to regard their setup state.

For the future, it would be interesting to further detail the proposed logic for the communication between the suggested levels along with a rolling horizon approach. In fact, interaction should not just occur between levels but also between iterations of the plan.

Another feature that could be explored and introduced to the planning functionalities is the Order Promising, which involves establishing a date when customers will receive their orders. It could be done using the Available to Promise (ATP) or the Capable to Promise (CTP) concept. ATP uses uncommitted available stock (affected by planned supply) to define this date, whereas CTP goes further and also considers production capacity and material availability when defining this date (Van Eck, 2003).

Finally, and since the ultimate goal for CMF is the commercialization of a production planning system, it is not desired that models are solved using software based on mathematical programming techniques, such as *Gurobi*. These require expensive solvers which would have to be charged to clients. Therefore, the next steps can go through the application of meta-heuristics such as genetic algorithms to the proposed model formulations. But first, maybe even more relevant would be thinking on how the production plan could be stored in a MES. Defining a data model would contribute to achieving this goal.

Bibliography

Albrecht, M. (2010). *Supply Chain Coordination Mechanisms*, Volume 628 of *Lecture Notes in Economics and Mathematical Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Albrecht, M., J. Rohde, and M. Wagner (2015). Master Planning. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies*, Chapter 8, pp. 155–175. Springer.

Anthony, R. N. (1965). *Planning and control systems: a framework for analysis*. Division of Research, Graduate School of Business Administration, Harvard.

Arica, E. and D. J. Powell (2014). A framework for ICT-enabled real-time production planning and control. *Advances in Manufacturing 2*(2), 158–164.

Badhotiya, G. K., G. Soni, and M. L. Mittal (2019). Multi-site planning and scheduling: state-of-the-art review and future research directions. *Journal of Global Operations and Strategic Sourcing 13*(1), 17–37.

Bermudez, J. (1996). Advanced Planning and Scheduling Systems: Just a Fad or a Breakthrough in Manufacturing and Supply Chain Management? Technical report, Advanced Manufacturing Research.

Bian, Y. (2018, dec). Tactical production planning for physical and financial flows of supply chain in a multi-site context. *4OR 16*(4), 445–446.

Brahimi, N. (2004). *Production Planning : New Lot-Sizing Models and Algorithms*. Ph. D. thesis, Université de Nantes.

Buschkühl, L., F. Sahling, S. Helber, and H. Tempelmeier (2010, apr). Dynamic capacitated lot-sizing problems: a classification and review of solution approaches. *OR Spectrum 32*(2), 231–261.

Chen, C. S., S. Mestry, P. Damodaran, and C. Wang (2009). The capacity planning problem in make-to-order enterprises. *Mathematical and Computer Modelling 50*(9-10), 1461–1473.

Connect, S. (2021). Preactor APS.

Datta, M. (2020). Manufacturing processes for fabrication of flip-chip micro-bumps used in microelectronic packaging: An overview. *Journal of Micromanufacturing 3*(1), 69–83.

de Kok, T. G. and J. C. Fransoo (2003). Planning Supply Chain Operations: Definition and Comparison of Planning Concepts. *Handbooks in Operations Research and Management Science 11*(C), 597–675.

de Santa-Eulalia, L. A., S. DAmours, J.-M. Frayret, C. Cesar, and R. Cambiaghi (2011, aug). Advanced Supply Chain Planning Systems (APS) Today and Tomorrow. In *Supply Chain Management - Pathways for Research and Practice*, pp. 137–144. InTech.

de Sousa, T. B., C. E. S. Camparotti, F. M. Guerrini, A. L. da Silva, and W. Azzolini Júnior (2014). An Overview of the Advanced Planning and Scheduling Systems. *Independent Journal of Management & Production 5*(4), 1032–1049.

Fleischmann, B. and H. Meyr (2003). Planning Hierarchy, Modeling and Advanced Planning Systems. In *Handbooks in Operations Research and Management Science*, Chapter 9, pp. 455–523. Elsevier Science.

Fleischmann, B., H. Meyr, and M. Wagner (2015). Advanced Planning. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies*, Chapter 4, pp. 71–95. Springer.

Frolov, EB and Zagidullin, R. (2008). MES-systems as they are or evolution of production planning systems. *Machine park 10*(55), 31–7.

Gautam, S., L. LeBel, and D. Beaudoin (2015). Value-adding through silvicultural flexibility: An operational level simulation study. *Forestry 88*(2), 213–223.

Guers, C., C. Martin, and J. L. Wybo (2014, sep). The impact of the use of subcontracting on organizational reliability and safety. In *Safety and Reliability: Methodology and Applications*, pp. 1043–1046. CRC Press.

Gupta, A. and C. D. Maranas (2003). Managing demand uncertainty in supply chain planning. *Computers and Chemical Engineering 27*(8-9), 1219–1227.

Gupta, A. K. and A. I. Sivakumar (2006). Job shop scheduling techniques in semiconductor manufacturing. *International Journal of Advanced Manufacturing Technology 27*(11-12), 1163–1169.

Haase, K. (1998). Capacitated Lot-Sizing with Linked Production Quantities of Adjacent Periods. In *Beyond Manufacturing Resource Planning (MRP II)*, pp. 127–146. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hu, S. C., C. W. Ku, Y. C. Shih, and K. Hsu (2009). Dynamic analysis on particle concentration induced by opening the door of a front opening unified pod (FOUP) that loaded with 25 pieces of 300 mm wafer manufacturing processes. *Aerosol and Air Quality Research 9*(1), 139–148.

Instruments, T. (1999). Semiconductor Packaging Assembly Technology. Technical report, Texas Instruments.

Kanyalkar, A. P. and G. K. Adil (2007). Aggregate and detailed production planning integrating procurement and distribution plans in a multi-site environment. *International Journal of Production Research 45*(22), 5329–5353.

Karimi, B., S. M. Fatemi Ghomi, and J. M. Wilson (2003). The capacitated lot sizing problem: A review of models and algorithms. *Omega 31*(5), 365–378.

Kim, J. and S. Kim (2012). Positioning a Decoupling Point in a Semiconductor Supply Chain under Demand and Lead Time Uncertainty. *International Journal of Advanced Logistics 1*(2), 31–45.

Kjellsdotter Ivert, L. (2012). *Use of Advanced Planning and Scheduling (APS) systems to support manufacturing planning and control processes*. Ph. D. thesis, Chalmers University of Technology.

Larson, P. D. and D. S. Rogers (1998). Supply Chain Management: Definition, Growth and Approaches. *Journal of Marketing Theory and Practice 6*(4), 1–5.

Lebreton, B., H. Meyr, H. Rosič, C. Seipl, and U. Wetterauer (2015). Architecture of Selected APS. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies*, pp. 341–361. Springer.

Li, R. and H. Ma (2017). Integrating Preventive Maintenance Planning and Production Scheduling under Reentrant Job Shop. *Mathematical Problems in Engineering 2017*, 1–9.

Manufacturing, C. (2021). Strategy.

Mauergauz, Y. (2016). *Advanced planning and scheduling in manufacturing and supply chains*. Springer.

Meyer, H., F. Fuchs, and K. Thiel (2009). *Manufacturing execution systems: optimal design, planning, and deployment*. McGraw-Hill Education.

Meyr, H., M. Wagner, and J. Rohde (2015). Structure of Advanced Planning Systems. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies*, Chapter 5, pp. 99–106. Springer.

Mönch, L., J. W. Fowler, and S. J. Mason (2013). Semiconductor Manufacturing Process Description. In *Production Planning and Control for Semiconductor Wafer Fabrication Facilities*, pp. 11–28. Springer.

Mula, J., R. Poler, G. S. García-Sabater, and F. C. Lario (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics 103*(1), 271–285.

Nowful, J. M., S. C. Lok, and S. W. Ricky Lee (2001). Effects of plasma cleaning on the reliability of wire bonding. *Advances in Electronic Materials and Packaging 2001 00*(C), 39–43.

Olhager, J. (2013). Evolution of operations planning and control : from production to supply chains. *International Journal of Production Research 51*(23-24), 6836–6843.

Park, Y.-J. and S. Hur (2020, mar). Improvement of Productivity through the Reduction of Unexpected Equipment Faults in Die Attach Equipment. *Processes 8*(4), 394.

Partners, T. I. (2021). Advanced Planning and Scheduling (APS) Software Market Forecast to 2028 - COVID-19 Impact and Global Analysis By Deployment (On-Premise, Cloud), Organization Size (Large Enterprise, Small & Medium Enterprise), and Industry (Manufacturing, Pharmaceutical &. Technical report, The Insight Partners.

Prakaiwichien, S. and V. Rungreunganun (2018). Solving Dynamic Multi-Product Multi-Level Capacitated Lot-Sizing Problems with Modified Part Period Balancing Heuristics Method. *International Journal of Applied Engineering Research 13*(6), 3350–3360.

Quadt, D. and H. Kuhn (2009). Capacitated lot-sizing and scheduling with parallel machines, back-orders, and setup carry-over. *Naval Research Logistics 56*(4), 366–384.

Reuter, B. and J. Rohde (2015). Coordination and Integration. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies*, Chapter 13, pp. 241–256. Springer.

Schneeweiß, C. (1995). Hierarchical structures in organisations: A conceptual framework. *European Journal of Operational Research 86*(1), 4–31.

Ščukanec, A., K. Rogić, and D. Babić (2007). Bullwhip effect in supply chains. *Promet - Traffic - Traffico 19*(5), 289–293.

Sequeira, F. P. (2009). *O Papel Estratégico dos Sistemas APS na Gestão da Produção*. Ph. D. thesis, Universidade de Aveiro.

Sethi, S. and G. Sorger (1991, dec). A theory of rolling horizon decision making. *Annals of Operations Research 29*(1), 387–415.

Siemens (2016). SIMATIC IT Preactor APS 2016. Technical report, Siemens.

Stadtler, H. (2005). Supply chain management and advanced planning - Basics, overview and challenges. *European Journal of Operational Research 163*(3), 575–588.

Stadtler, H. (2015). Supply Chain Management: An Overview. In *Supply Chain Management and Advanced Planning Concepts, Models, Software, and Case Studies* (5 ed.)., Chapter 1, pp. 3–28. Springer.

Steven, M. (2004). Hierarchical Planning Structures in Supply Chain Management. In *Modern Concepts of the Theory of the Firm*, pp. 301–312. Berlin, Heidelberg: Springer Berlin Heidelberg.

Tempelmeier, H. (2001). Supply chain planning with advanced planning systems. Technical report, University of Cologne, Cologne.

Van Eck, M. (2003). *Is logistics everything, a research on the use (fullness) of advanced planning and scheduling systems*. Ph. D. thesis, Universiteit Amsterdam.

Vicens, E., M. E. Alemany, C. Andrés, and J. J. Guarch (2001). A design and application methodology for hierarchical production planning decision support systems in an enterprise integration context. *International Journal of Production Economics 74*(1-3), 5–20.

Vogel, T., B. Almada-Lobo, and C. Almeder (2017). Integrated versus hierarchical approach to aggregate production planning and master production scheduling. *OR Spectrum 39*(1), 193–229.

Appendix A

Comparison between planning systems

Table A.1: Comparison between APS and MRP-II systems (Kjellsdotter Ivert, 2012)

APS system	**MRP II system**
Customer preference may be varied depending on the business importance of the customer	All customers are given equal preference in the system
Lead times can be dynamically entered by contacting the customers	Lead times are fixed and known a priori
APS applications dynamically calculate a plan and schedule within minutes of any change being made to them	MRP runs are usually batch time and have longer duration times
Support superior decision making by what-if analysis and simulations	Does not support any decision making aids
Smart and easy to drill down reporting based on the identification of exceptional conditions	Detailed reports, which are hard to read and decipher
Material allocation according to availability and according to the criterion specified	Material allocation done on a first come first service basis

Table A.2: Comparison between APS and ERP systems (Kjellsdotter Ivert, 2012)

Areas	**APS system**	**ERP system**
Planning philosophy	Planning provides feasible and reasonable plans based on the limited availability of key resources	Planning without considering the limited availability of key resources required for executing the plans
	Goal: Optimal plans	Goal: Feasible plans
	Pull	Push
	Integrated and simultaneous	Sequential and top-down
Business driver	Satisfaction of customer demand	Manufacturing coordination
Industry scope	All industries	Primarily discrete manufacturing
Major business area supported	Planning: Demand, Manufacturing, Logistics, Supply chain	Transaction: Finance, Controlling, Manufacturing
Information flow	Bi-directional	Top down
Simulation capabilities	High	Low
Ability to optimize cost, price, profit	Available	Not available
Manufacturing lead times	Flexible	Fixed
Incremental planning	Available	Not available
Speed of replanning	High	Low
Data storage and calculations	Memory-resident	Database

Appendix B

Data used in the case study

Table B.1: Products considered and their types

Type	Product
DRAM (end-item)	SDDR4MD.P1
	SDDR4MD.P1TD9
eMCP (end-item)	SPMCP224533
	SPMCP224542
	SPMCP9382731
	SPMCP9382997
Substrate (component)	SUBDDR44553
	SUBDDR49775
	SUBDDR49821
	SUBDDR49981
Wafer die (component)	WAFCTRL21567
	WAFFLCH44789
	WAFLP861134
	WAFLP866678
	WAFLP883117
	WAFLP884792
	WAFLP884891
	WAFSPCR2244

Table B.2: BOM for the products considered

End-item	Component	Quantity
SDDR4MD.P1	SUBDDR44553	1
	WAFFLCH44789	1
SDDR4MD.P1TD9	SUBDDR44553	1
	WAFFLCH44789	1
SPMCP224533	SUBDDR49981	1
	WAFSPCR2244	1
	WAFLP883117	1
	WAFCTRL21567	1
SPMCP224542	SUBDDR49775	1
	SUBDDR49821	1
	WAFLP884792	2
	WAFLP884891	2
SPMCP9382731	SUBDDR49981	1
	WAFLP883117	1
	WAFLP866678	1
	WAFLP861134	1
SPMCP9382997	SUBDDR49981	1
	WAFLP883117	1
	WAFLP866678	1
	WAFLP861134	1

Table B.3: Demand data from 2021-01-11 to 2021-02-07 (4 weeks)

Due Date	Product	Quantity
2021-01-12	SPMCP9382997	81377
2021-01-14	SDDR4MD.P1	31023
2021-01-14	SPMCP224533	10868
2021-01-14	SPMCP224542	12303
2021-01-14	SPMCP9382997	74710
2021-01-20	SPMCP224533	25910
2021-01-20	SPMCP224542	36622
2021-01-20	SPMCP9382731	52140
2021-01-24	SDDR4MD.P1TD9	112973
2021-01-24	SPMCP224533	68859
2021-01-24	SPMCP224542	73931
2021-01-24	SPMCP9382731	81567
2021-01-25	SDDR4MD.P1	20800
2021-01-25	SDDR4MD.P1TD9	11832
2021-01-25	SPMCP224533	45259
2021-01-26	SDDR4MD.P1TD9	93536
2021-01-26	SPMCP224533	70203
2021-01-26	SPMCP224542	32548
2021-01-29	SDDR4MD.P1TD9	91674
2021-01-29	SPMCP224533	13325
2021-01-29	SPMCP224542	60283
2021-01-31	SDDR4MD.P1TD9	81184
2021-01-31	SPMCP224533	68743
2021-01-31	SPMCP224542	90271
2021-02-01	SPMCP9382997	45650
2021-02-03	SPMCP224542	21302

Table B.4: Costs for each item in scenario A

Item	**Holding Cost/unit**	**Setup Costs**	**Backorder cost/unit**
SDDR4MD.P1	0.5	2868	100
SDDR4MD.P1TD9	0.5	2868	400
SPMCP224533	0.5	7204.45	300
SPMCP224542	0.5	7794.08	400
SPMCP9382731	0.5	9684.55	200
SPMCP9382997	0.5	9684.55	300
SUBDDR44553	1	0	-
SUBDDR49775	1	0	-
SUBDDR49821	1	0	-
SUBDDR49981	1	0	-
WAFCTRL21567	5	121390.00	-
WAFFLCH44789	5	121390.00	-
WAFLP861134	5	121390.00	-
WAFLP866678	5	23332.50	-
WAFLP883117	5	121390.00	-
WAFLP884792	5	23332.50	-
WAFLP884891	5	23332.50	-
WAFSPCR2244	5	121390.00	-

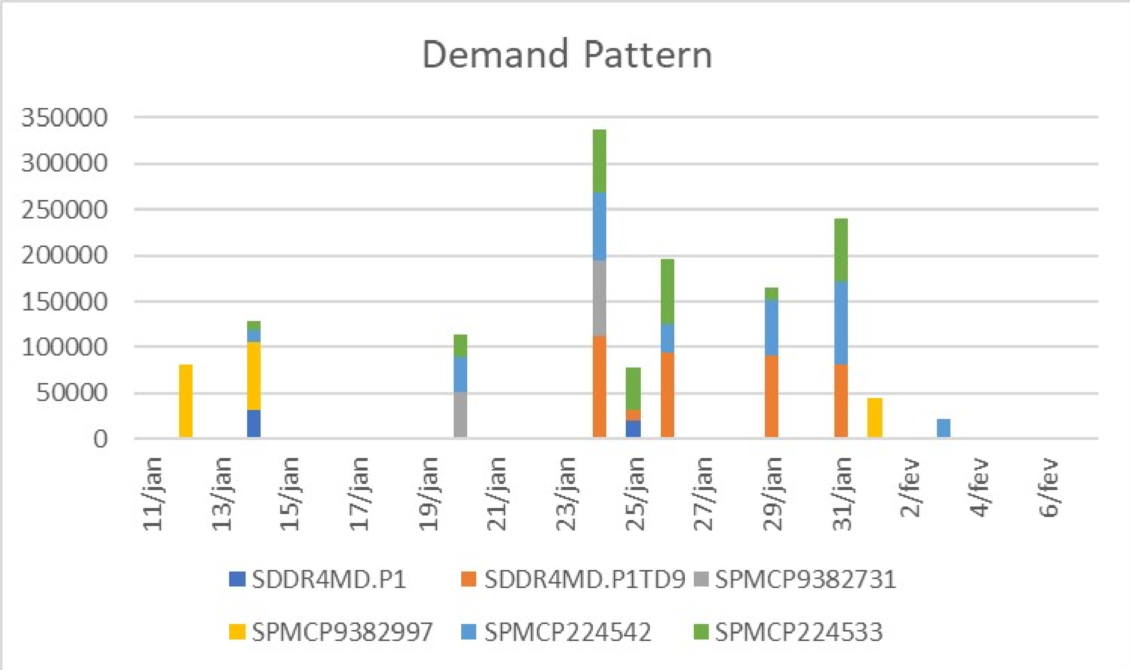


Figure B.1: Demand Pattern of end-items requested from 2021-01-11 to 2021-02-07 (4 weeks)

Table B.5: Standard capacity and Maximum overtime available for each work center (Scenario A)

Work center	**Standard Capacity (min)**	**Max Overtime (min)**
ClusterFlipChip	1344	268.8
DieAttach	11424	2284.8
Pressure Oven	2016	403.2
Plasma	1344	268.8
WireBonder	54432	10886.4
TapeLaminator	672	134.4
Grinding	2016	403.2
LaserGrooving	1344	268.8
WaferSaw	4704	940.8
UV Cure	672	134.4
Die Stretch	672	134.4
SubstrateLotPrep	2688	537.6
SMT	672	134.4
BakeOven	2016	403.2
Stealth Dicing	672	134.4

Appendix C

Processing and setup time of each product in the work center

In parenthesis, it is represented the operations of the production process performed by each work center.

Work center	# Machines	Products Processed	Variable processing time (min)	Setup time per machine (min)
ClusterFlipChip (Flip Chip Bonder)	2	SDDR4MD.P1	0.0279	40
		SDDR4MD.P1TD9	0.0279	40
DieAttach (Die Attach)	17	SPMCP9382731	0.1787	55
		SPMCP9382997	0.1787	55
		SPMCP224542	0.2121	55
		SPMCP224533	0.2449	55
Pressure Oven * (Die Attach Cure)	3	SPMCP9382731	0.0225	0
		SPMCP9382997	0.0225	0
		SPMCP224542	0.0300	0
		SPMCP224533	0.0300	0
Plasma (DieAttach Plasma + Substrate Plasma)	2	SPMCP9382731	0.0039	0
		SPMCP9382997	0.0039	0
		SPMCP224542	0.0052	0
		SPMCP224533	0.0052	0
		SUBDDR44553	0.0007	0
		SUBDDR49981	0.0010	0
		SUBDDR49775	0.0010	0
		SUBDDR49821	0.0010	0
WireBonder (Wire Bonding)	81	SPMCP9382731	0.5459	30

		SPMCP9382997	0.5459	30
		SPMCP224542	0.7175	30
		SPMCP224533	0.7175	30
TapeLaminator (Tape Laminator)	1	WAFLP883117	0.0007	0
		WAFLP861134	0.0007	0
		WAFCTRL21567	0.0007	0
		WAFFLCH44789	0.0007	0
		WAFLP884792	0.0007	0
		WAFLP884891	0.0007	0
		WAFLP866678	0.0007	0
		WAFSPCR2244	0.0007	0
Grinding (Grinding)	3	WAFLP883117	0.0016	25
		WAFLP861134	0.0016	25
		WAFCTRL21567	0.0016	25
		WAFFLCH44789	0.0016	25
		WAFLP884792	0.0032	25
		WAFLP884891	0.0032	25
		WAFLP866678	0.0032	25
		WAFSPCR2244	0.0016	25
LaserGrooving (Laser Grooving)	2	WAFLP883117	0.0043	0
		WAFCTRL21567	0.0043	0
		WAFFLCH44789	0.0043	0
WaferSaw (Wafer Saw)	7	WAFLP883117	0.0066	70
		WAFLP861134	0.0066	70
		WAFCTRL21567	0.0066	70
		WAFFLCH44789	0.0066	70
		WAFSPCR2244	0.0066	70
UV Cure (UV Cure)	1	WAFLP883117	0.0007	0
		WAFFLCH44789	0.0007	0
		WAFSPCR2244	0.0007	0
Die Stretch (Die Stretch)	1	WAFLP861134	0.0061	0
		WAFLP861134	0.0031	0
		WAFCTRL21567	0.0031	0
		WAFLP884792	0.0003	0

		WAFLP884891	0.0003	0
		WAFLP866678	0.0003	0
SubstrateLotPrep (Substrate Lot Prep)	4	SUBDDR44553	0.0031	0
		SUBDDR49981	0.0023	0
		SUBDDR49775	0.023	0
		SUBDDR49821	0.0023	0
SMT (SMT)	1	SUBDDR49981	0.0100	0
		SUBDDR49775	0.0100	0
		SUBDDR49821	0.0100	0
BakeOven * (Substrate Pre Bake)	3	SUBDDR49981	0.0281	0
		SUBDDR49775	0.0281	0
		SUBDDR49821	0.0281	0

* Fixed cycle times in this work center were converted to variable processing times

Appendix D

Production Plan provided by Detailed Lot Sizing model for the base scenario A

Table D.1: Quantities produced during first two weeks of the planning horizon in Scenario A

Item	11/jan	12/jan	13/jan	14/jan	15/jan	16/jan	17/jan	18/jan	19/jan	20/jan	21/jan	22/jan	23/jan	24/jan
SDDR4MD.P1	0	0	0	49238	0	0	0	0	0	0	0	0	2585	0
SDDR4MD.P1TD9	0	0	0	0	0	0	0	0	0	0	0	0	49469	54922
SPMCP9382731	0	0	0	0	0	0	0	0	46822	40300	0	0	0	46585
SPMCP9382997	38671	58708	0	58708	0	0	0	0	0	0	0	0	0	0
SPMCP224542	12303	0	0	0	0	0	33183	29318	0	0	40300	40300	0	0
SPMCP224533	0	0	52153	0	0	0	0	0	0	0	0	0	52153	3430
WAFLP883117	38671	58708	52153	58708	0	0	0	0	93644	80600	0	0	52153	96600
WAFLP861134	38671	58708	0	58708	0	0	0	0	46822	40300	0	0	0	46585
SUBDDR44553	0	0	0	49238	0	0	0	0	0	0	0	0	101523	109844
SUBDDR49981	38671	58708	52153	58708	0	0	0	13044	80600	80600	0	0	68153	80600
SUBDDR49775	12303	0	0	0	0	0	33183	29318	0	0	40300	40300	0	0
SUBDDR49821	6526	0	0	0	0	0	33183	29318	0	0	40300	40300	0	0
WAFCTRL21567	0	0	52153	0	0	0	0	0	0	0	0	0	55583	0
WAFFLCH44789	0	0	0	49238	0	0	0	0	0	0	0	0	101523	109844
WAFLP884792	24606	0	0	0	0	0	66366	58636	0	0	80600	80600	0	0
WAFLP884891	24606	0	0	0	0	0	66366	58636	0	0	80600	80600	0	0
WAFLP866678	38671	58708	0	58708	0	0	0	0	46822	40300	0	0	0	46585
WAFSPCR2244	0	0	52153	0	0	0	0	0	0	0	0	0	55583	0

Table D.2: Quantities produced during last two weeks of the planning horizon in Scenario A

Item	25/jan	26/jan	27/jan	28/jan	29/jan	30/jan	31/jan	1/fev	2/fev	3/fev	4/fev	5/fev	6/fev	7/fev
SDDR4MD.P1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SDDR4MD.P1TD9	54922	54922	0	54922	54922	0	54922	0	0	0	0	0	0	0
SPMCP9382731	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPMCP9382997	0	0	0	0	0	0	0	45650	0	0	0	0	0	0
SPMCP224542	0	0	48625	0	49442	0	52487	0	0	21302	0	0	0	0
SPMCP224533	52153	52153	0	39242	0	42826	0	0	0	0	0	0	0	0
WAFLP883117	52153	52153	0	39242	0	42826	0	45650	0	0	0	0	0	0
WAFLP861134	0	0	0	0	0	0	0	45650	0	0	0	0	0	0
SUBDDR44553	109844	109844	0	109844	109844	0	109844	0	0	0	0	0	0	0
SUBDDR49981	52153	52153	0	39242	0	42826	0	45650	0	0	0	0	0	0
SUBDDR49775	3	15997	32625	16000	33442	8374	44113	0	0	21302	0	0	0	0
SUBDDR49821	0	650	47975	2284	47158	16000	36487	0	0	21302	0	0	0	0
WAFCTRL21567	52153	52153	0	39242	0	42826	0	0	0	0	0	0	0	0
WAFFLCH44789	109844	109844	0	109844	109844	0	109844	0	0	0	0	0	0	0
WAFLP884792	0	0	97250	0	98884	0	104974	0	0	42604	0	0	0	0
WAFLP884891	0	0	97250	0	98884	0	104974	0	0	42604	0	0	0	0
WAFLP866678	0	0	0	0	0	0	0	45650	0	0	0	0	0	0
WAFSPCR2244	52153	52153	0	39242	0	42826	0	0	0	0	0	0	0	0

Appendix E

Inventory levels provided by Detailed Lot Sizing model for the base scenario A

Table E.1: Inventory quantities during the first two weeks of the planning horizon in Scenario A

Item	Initial Inv	11/jan	12/jan	13/jan	14/jan	15/jan	16/jan	17/jan	18/jan	19/jan	20/jan	21/jan	22/jan	23/jan	24/jan
SDDR4MD.P1	0	0	0	0	18215	18215	18215	18215	18215	18215	18215	18215	18215	20800	20800
SDDR4MD.P1TD9	12198	12198	12198	12198	12198	12198	12198	12198	12198	12198	12198	12198	12198	61667	3616
SPMCP9382731	0	0	0	0	0	0	0	0	0	46822	34982	34982	34982	34982	0
SPMCP9382997	0	38671	16002	16002	0	0	0	0	0	0	0	0	0	0	0
SPMCP224542	0	12303	12303	12303	0	0	0	33183	62501	62501	25879	66179	106479	106479	32548
SPMCP224533	9057	9057	9057	61210	50342	50342	50342	50342	50342	50342	24432	24432	24432	76585	11156
WAFLP883117	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP861134	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR44553	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR49981	0	0	0	0	0	0	0	0	13044	0	0	0	0	16000	0
SUBDDR49775	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR49821	5777	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFCTRL21567	0	0	0	0	0	0	0	0	0	0	0	0	0	3430	0
WAFFLCH44789	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP884792	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP884891	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP866678	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFSPCR2244	0	0	0	0	0	0	0	0	0	0	0	0	0	3430	0

Table E.2: Inventory quantities during the last two weeks of the planning horizon in Scenario A

Item	25/jan	26/jan	27/jan	28/jan	29/jan	30/jan	31/jan	1/fev	2/fev	3/fev	4/fev	5/fev	6/fev	7/fev
SDDR4MD.P1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SDDR4MD.P1TD9	46706	8092	8092	63014	26262	26262	0	0	0	0	0	0	0	0
SPMCP9382731	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPMCP9382997	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPMCP224542	32548	0	48625	48625	37784	37784	0	0	0	0	0	0	0	0
SPMCP224533	18050	0	0	39242	25917	68743	0	0	0	0	0	0	0	0
WAFLP883117	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP861134	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR44553	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR49981	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUBDDR49775	3	16000	0	16000	0	8374	0	0	0	0	0	0	0	0
SUBDDR49821	0	650	0	2284	0	16000	0	0	0	0	0	0	0	0
WAFCTRL21567	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFFLCH44789	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP884792	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP884891	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFLP866678	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WAFSPCR2244	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix F

Overtime provided by Detailed Lot Sizing model for the base scenario A

Table F.1: Overtime used (in minutes) at each work center during the first two weeks of the planning horizon in Scenario A

Workcenter	11/jan	12/jan	13/jan	14/jan	15/jan	16/jan	17/jan	18/jan	19/jan	20/jan	21/jan	22/jan	23/jan	24/jan
ClusterFlipChip	0	0	0	110	0	0	0	0	0	0	0	0	268	268
DieAttach	0	0	2284	0	0	0	0	0	0	0	0	0	2284	0
Pressure Oven	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plasma	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WireBonder	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TapeLaminator	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Grinding	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LaserGrooving	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaferSaw	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UV Cure	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Die Stretch	0	67	0	67	0	0	0	0	0	0	0	0	0	0
SubstrateLotPrep	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SMT	0	0	0	0	0	0	0	45	134	134	134	134	10	134
BakeOven	0	0	0	0	0	0	0	0	251	251	251	251	0	251
Stealth Dicing	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table F.2: Overtime used (in minutes) at each work center during the last two weeks of the planning horizon in Scenario A

Workcenter	25/jan	26/jan	27/jan	28/jan	29/jan	30/jan	31/jan	1/fev	2/fev	3/fev	4/fev	5/fev	6/fev	7/fev
ClusterFlipChip	268	268	0	268	268	0	268	0	0	0	0	0	0	0
DieAttach	2284	2284	0	0	0	0	646	0	0	0	0	0	0	0
Pressure Oven	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plasma	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WireBonder	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TapeLaminator	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Grinding	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LaserGrooving	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaferSaw	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UV Cure	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Die Stretch	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SubstrateLotPrep	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SMT	0	16	134	0	134	0	134	0	0	0	0	0	0	0
BakeOven	0	0	251	0	251	0	251	0	0	0	0	0	0	0
Stealth Dicing	0	0	0	0	0	0	0	0	0	0	0	0	0	0