Eindhoven University of Technology

MASTER

Buffer management at ASML

Gursoy, P.

Award date:
2011

Eindhoven, July 2011

Buffer Management at ASML

by

Poyraz Gursoy

BSc Industrial Engineering — Istanbul Technical University, 2009
Student identity number 0729078

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
dr. ir. H.P.G. van Ooijen, TU/e, OPAC
dr.ir. R.J.I. Basten , TU/e, OPAC
W. Postma, ASML, Business Engineering
P. Lipsch, ASML, Business Engineering

TUE. School of Industrial Engineering.
Series Master Theses Operations Management and Logistics

Subject headings: time buffer, safety time, lead time planning, cycle time forecasting, process uncertainty, learning, reliable service levels, capital goods manufacturing

Abstract

This master thesis describes a research project conducted on the field of buffer management in a high-tech company. Buffers are needed in this company to set reliable work center lead times. A buffer strategy was developed in order to determine manufacturing lead times for a given service level in unreliable, unbalanced production systems that are subject to learning effects. In addition, the formulas for the cost optimal service levels were given. Finally, the proposed buffer strategy was applied to a part of the manufacturing system of the company.

Preface

This document presents the results of my master thesis project finalizing the MSc. Operations Management and Logistics program at Eindhoven University of Technology. This study was carried out for five months from February 2011 to June 2011 at ASML Company in Veldhoven, the Netherlands. In the following paragraphs, I would like to thank people who helped me throughout this project.

First of all, I would like to thank my first supervisor at TU/e, Henny van Ooijen for his ideas, contribution and guidance. His feedbacks and continuous support were very valuable for me and for this project. Additionally, I would like to thank my second supervisor, Rob Basten for his constructive comments and feedbacks.

Furthermore, I would like to thank my first supervisor at ASML, Walter Postma for encouraging me to think creatively. His extensive knowledge and experience enabled me to improve many aspects of this project. Moreover, I would like to thank my second supervisor at ASML, Peter Lipsch for his support, feedback, comments, tips and all the time he has spent for our regular meetings about the project. Without his support, I would not be able to look back on this project with great satisfaction as I do now. Although he is not my official supervisor, I would like to thank Teun Burgers as well, who also helped me with his knowledge and skills.

Additionally, I would like to thank all the members of the Business Engineering and Business Reporting departments at ASML for supporting me and motivating me throughout this project.

Besides, I am grateful to all my friends in Eindhoven. They motivated me during the difficult moments and shared my joy during my happy moments.

Last but not least, I would like to thank the people that are most valuable to me, my mother and my father. Thank you very much for encouraging me to come to the Netherlands, for supporting me continuously during these two years, and most importantly, for being there to help me whenever I need.

Poyraz Gursoy,
Eindhoven, July 2011

Executive Summary

ASML is currently the market leader for photolithography machines production. These machines are capital goods, which are used by the semiconductor industry. The production processes at ASML faces many uncertainties, due to the nature of the photolithography sector. The main characteristics of ASML production are as follows:

- Unpredictable and stochastic demand
- Short product life cycles
- Uncertainty in cycle times
- Existence of the learning effects
- Low production volume
- Space constraints

These specialties force the company to determine reliable lead times. This can only be achieved by using buffers within production. Therefore, good estimations about the future time buffers are required. These needs of ASML lead us to conduct a research study about buffer management in photolithography producers. The purpose of the study is defined as follows:

"Design a method for calculating time buffers for production environments which have the characteristics of uncertain demand, low production volume, steep learning curve, fluctuating processing times, and space constraints, so that the production units have sufficient delivery performance while costs associated with buffers are kept as low as possible".

We conducted several analyses in order to develop a buffer methodology. They are explained in the following paragraphs.

As the first step, we analyzed the current situation and examined the buffers that are used in this production system. The initial analysis revealed that cycle times of production processes are mainly affected by the following:

- Disturbances, that occur during production
- Capacity bottlenecks
- Working hours of production

In order to cope with these effects, buffers are needed in the production. Buffers can be used either in terms of physical stocks or in terms of time. As also suggested by the literature, safety time is preferable to safety stock for companies with high demand and process variability. Therefore using buffers in time gives the company certain planning advantages. We also examined which factors should be taken into account during the determination of buffers. Our analysis indicated that the distribution of cycle times, learning effects, working hours and outlier determination methods have substantial impacts on buffer calculations.

After the initial analysis, we reviewed the literature of lead time planning and buffers in time. Although we could not reference any study which can be directly applied to our situation, we decided to base our methodology on the works of Radovilsky (1998) and Kuo et al. (2009).

As the next step, a method was developed for calculating and allocating time buffers. The method answers the following questions about buffers:

- Which buffer strategy to be used in order to calculate buffers?

The selection was made from two available options. We decided to calculate the total buffer, considering the process as one production unit, and then divide this buffer within operations if necessary.

- Where to put time buffers?

We decided to place buffers before the bottleneck station and at the end of the whole process as suggested in the literature.

- How to calculate the size of the buffers?

We selected the statistical approach among three possible options. Statistical approach suggests fitting the cycle times into a distribution, determining a service level and then calculating the required buffer time. Besides, we decided to make several additions to this approach. We developed a procedure for outlier elimination. Moreover we determined candidate distributions for cycle times and introduced a way to include learning effects to the buffer calculations.

- How to allocate the buffers to the determined locations?

We determined two possible methods for allocation. The selection was made during the application of the method. We decided to distribute buffers according to the coefficient of variations of the operations.

Furthermore, we investigated the costs of using different service levels. We defined the costs of being early and late. As a result, we concluded that these costs can be represented by the newsvendor equation and thus the optimal service levels can be computed.

As the last step, we applied the developed buffer calculation method to a part of the ASML production system (namely Assembly part). We calculated the lead times for selected processes at this part of the production system. During the application we compared our methodology with the tool in use and also we examined the effects of the learning curves. The main results are as follows:

- The company policy of using 99% service level for all modules would lead to non optimal results.
- Proposed method performs better than the methodology which is currently in use. In terms of buffer times, the proposed method leads to the results that are lower than the company strategy. It is on average 1.55 days lower than the company method if 99% of service level is used and it is on average 3.85 days lower than the company method if the optimal service levels are used.
- The method without taking learning effects into account results higher buffer time values, namely, 2.6 days on average.

The following conclusions were drawn from the results of the application:

- Buffer management decisions in production environments such as photolithography are affected by several factors such as the use of learning curves, selection of cycle time distributions and use of other parameters.
    - The statistical fit between the past processing times and the future predicted lead times is very important in this type of a prediction method.
    - If learning effects exist in a production environment, this should be taken into account while calculating future lead times/buffers. If the calculation is done without any adaptations, the planned times would not be optimal (higher than the necessary value).
    - Learning percentages are situation dependent, therefore they should be re-computed before the calculation/decision on the buffers.
- If buffers are calculated according to the cost optimal service levels, lower costs can be achieved. On the other hand, this would lower the service level. The decision on using the optimal service levels or using a predefined level such as 99% should be made by the user/owner of the method.

Table of Contents

List of Figures

List of Tables

Introduction

For the manufacturing companies it is crucial to deal with the variance in customer demands, processing times and supplier performance. “Nowadays, in order to remain competitive, manufacturing companies have to offer their clients high-quality products at low cost and short lead times (very short, in fact)” (Faria et al., 2006). In order to reach the goal of low cost, companies need to reduce their buffer stocks and buffer times as much as possible, but on the other hand, they cannot reach the other goal of reliable lead times if they completely eliminate the concept of buffer. So, it can be said that the optimum strategy lies somewhere in between. However, mostly it is not known where the optimum lies. This has been examined for decades in the fields of production planning, inventory control, production flow lines, assembly systems etc. It is important to create an efficient strategy for buffer placement because of the fierce competition.

In the second half of the twentieth century, a huge amount of theories and methods have been developed in the fields of inventory and production control. Thanks to these developments, nowadays it is possible to design not only the production system of a company, but also the required buffers in that system. However, throughout the history of operations management the basic manufacturing / assembly system has been the most popular system to be analyzed (Hopp and Spearman, 2008). As a result of this, the developed methods usually consider processes that are similar to mass assembly.

In the light of the previous paragraph it can be said that the capital goods industry is usually neglected in the research on buffers. Capital goods companies manufacture and install large and complex technical systems such as turbine generators, cranes, boilers, computer networks, defence systems, material handling systems and medical machines (Hicks and Pongchareon, 2006; Oner et al., 2010). Hicks and Pongchareon (2006) and Hicks (2004) define the most important characteristics of the capital goods production as follows:

- Low production volume, small number of orders
- Each order require a huge amount of capacity
- Process routings are complex and take long time
- Low stock turn ratio (turnover/inventory)

As it can be seen above, the nature of the capital goods production is quite different from the other types of production. Therefore, the design rules developed for mass production do not generally apply for the capital goods industries.

Besides, the semiconductor sector has specialties which differ from those of other sectors. First, the demand in this sector is very unpredictable and stochastic (Swaminathan, 2000). Firms are subject to the changes in demand; the fluctuations in demand make impossible to plan the production in a precise way. Secondly, the technology in this sector changes rapidly (Swaminathan, 2000). This means that new products are introduced very often, and the product life cycles are short. So, if too much stock is kept in the inventory, there is a risk that they become obsolete and cannot be sold. This dictates the companies to use a Make to Order strategy. Thirdly, the semiconductor sector has a long term trend of growth. As stated by the Moore’s law; “the number of transistors that can be placed inexpensively on an integrated circuit has doubled approximately every two years". This means that the demand is not stable, and the design of the production system should cope with this future trend, otherwise the market share of the company would be smaller even if the production volume remains the same. These issues make it difficult to design a production system in that area.

The combination of the capital goods and semiconductor sectors, namely photolithography systems production, has the core characteristics of both. That means, on one hand the production volume is low and lead times are long, on the other hand the companies are subject to heavy variation / fluctuation in demand. Moreover, the fierce competition dictates the company to increase research and development activities which eventually reduces the life cycles of products (Gursoy, 2010). The photolithography sector produces capital goods for the semiconductor industry. Therefore it is subject not only to the characteristics of the capital goods production, but also to the demand changes in the semiconductor industry. Since the production processes differ significantly from mass production, a new perspective is needed in order to analyze these types of systems.

Organization of the Report

The rest of this paper is organized as follows: Chapter 1 presents the photolithography industry, ASML and a basic overview of ASML's production process. Chapter 2 explains the research design by presenting the problem, the research questions, the scope, the project deliverables and research method. Chapter 3 focuses on initial analysis of work centers, their specialties and the attention points. Chapter 4 is about the buffer calculation method. Chapter 5 evaluates the performance of the developed method. Last, but not least, Chapter 6 concludes the study and presents recommendations about the case.

1. Industrial Partner of the Research

This chapter aims to provide an overview of the business environment, in which the research study is conducted. The study is done at ASML, which produces photolithography systems. This chapter is organized as follows. In Section 1.1, the general characteristics of photolithography industry are given. It is followed by Section 1.2, where the description of ASML is made. In Sections 1.3 and 1.4 general information about relevant departments and processes in the project are given.

1.1. Photolithography Industry

Photolithography (also referred as optical lithography) is the process of transferring geometric shapes on a mask to the surface of a silicon wafer[1]. Optical lithography has been the dominant patterning process for semiconductor fabrication for over 40 years. The patterning process evolved from methods used in the printing industry, but new methods were developed as integrated circuits[2] (ICs) became more complex. Optical lithography systems represent the highest resolution, most accurate optical imaging systems ever produced (Bruning, 1997).

Photolithography production can be described as a complex capital goods production. Therefore, photolithography industry has the common characteristics of capital goods industry. Besides, photolithography industry is a supplier for semiconductor industry. As a result, the demand pattern for lithography systems has a strong relationship with the demand for semiconductors. That means, on one hand the production volume is low and lead times are long, on the other hand the companies are subject to heavy variation / fluctuation in demand. Moreover, the fierce competition dictates the company to increase research and development activities which eventually reduces the life cycles of products (Gursoy, 2010).

1.2. The Company

ASML (originally named ASM Lithography) was founded in 1984 as a joint venture between the Advanced Semiconductor Materials International (ASMI) and Philips. It is currently the world's biggest producer of photolithography systems for the semiconductor industry. ASML manufactures complex machines which are used in the production processes of integrated circuits or microchips. The company is the leader of its niche sector, and it has the responsibility to design, produce, develop, sell and service these machines (ASML internet site, 2011).

Many of the major global IC manufacturers such as Intel, Toshiba and Samsung are customers of ASML. These companies provide the chips used in a wide array of electronic, communications and information technology products such as computers, mobile phones, MP3 players.

---

[1] A wafer is a thin slice of semiconductor material, such as a silicon crystal, used in the fabrication of integrated circuits and other microdevices.

[2] An integrated circuit or monolithic integrated circuit (also referred to as IC, chip, and microchip) is an electronic circuit manufactured by the patterned diffusion of trace elements into the surface of a thin substrate of semiconductor material. Integrated circuits are used in virtually all electronic equipment today and have revolutionized the world of electronics. Computers, cellular phones, and other digital appliances are now inextricable parts of the structure of modern societies, made possible by the low cost of production of integrated circuits.

Customers of ASML use these machines for projecting IC blueprints on wafers. IC blueprints are projected multiple times on a wafer during this process. Wafers then undergo a number of steps to transform this image into a real network of lines that are able to store electricity - and thus information (Schepens, 2009).

In the competitive business environment, the complexity of producing integrated circuits with more functionality increases over time. Semiconductor manufacturers need suppliers that provide technology and complete process solutions. ASML aims to serve these customers not only by providing innovative and high tech products, but also giving customers after sales support.

Since ASML is in the lithography production sector, it is subject to the demand changes in the semiconductor sector. Therefore it is important for the company to periodically control production, improve production processes and update cycle times.[3]

1.3. Manufacturing Process

The schematic overview of the manufacturing process, that is examined, is given in Figure 1.

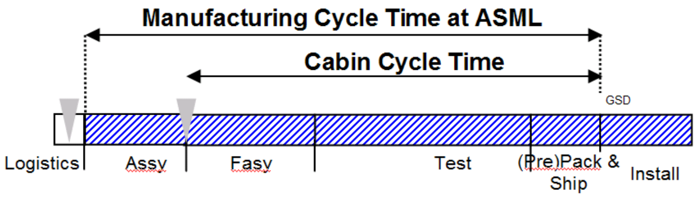


Figure 1 - Production Process of ASML

The manufacturing of ASML starts after the acceptance of materials into the warehouse. These materials are the parts which will be assembled and become a module for the final product. All of these parts are outsourced, so they are produced externally. Acceptance of materials is followed by the modules production (also called Assembly and abbreviated as ASSY) processes. In ASSY, the parts (modules) of the ASML machine are produced. It is consisted of the assembly of parts in different work centers. Usually each work center is responsible of the production of one module and is independent from other work centers.

After all necessary modules are produced; the Final Assembly (FASY) can start. In FASY, the machine is built by assembling the previously finished modules. FASY takes place in the so-called cabins, which are rooms specifically designed for the final assembly. The second process which takes place in cabins is Test. During Test, the machine is tested and also important configurations are done. In general, Test takes quite a long time compared to other types of production.

After Test is finished, the machine is packed and shipped to the customer. Shipping is usually done by planes since the machine is big and most of the customers are in Asia and America. The process is finished when the machine is installed at the customer side by ASML employees.

---

[3] For more information about the department in which the study is conducted, the reader should refer to Appendix I.

The steps of FASY, Test and Pre(Pack) are done in cabins. Therefore the cycle times of these operations are called “Cabin Cycle Time”, as shown in Figure 1. When added ASSY operations, one can reach “Manufacturing Cycle Time at ASML”. Since the Install operation is performed at the customer it is excluded from this time.

Even though this process looks similar to other types of production, photolithography production processes differ from other production processes. Production at ASML is subject to the following:

- Uncertainty in demand
- Uncertainty in cycle times
- Low production volume
- Learning curves
- Constraints associated with production in clean rooms

Furthermore, there is another specialty of production at ASML; the whole manufacturing is performed in the clean room. Clean room is an environment with low level of pollutants such as dust etc. The air in clean room is usually filtered and entrance of staff and material is controlled. Clean rooms are used if a high level of sensitivity needed in production, as typical examples being semiconductor manufacturing or biotechnology sectors. The space in clean rooms is much more expensive than a regular production area (Shi and Gershwin, 2009).

1.4. Assembly Unit

Assembly (ASSY) unit produces modules for the final assembly (FASY), where the whole machine is built. The main aim of ASSY is to supply modules to FASY when required. Therefore FASY is a customer for ASSY, and many suppliers of ASML are also suppliers for ASSY. ASSY also involves assembly operations, but they are parallel rather than sequential. Each of the parallel operations (called as work centers), aims to assemble and produce one specific module. The main specialties of ASSY can be defined as follows:

- High delivery performance: Because of the capacity problems in FASY, the company wants to use all of the available capacity in FASY as much as possible. Therefore FASY should never face starvation. This requirement clearly affects ASSY, since it is the direct supplier of FASY. As a result, ASSY should have a very high delivery performance.

- High variability: The production processes in ASSY are generally not very reliable, many types of incidents might occur and disrupt production. This is mainly because of the complex nature of photolithography production.

- Learning curves: Learning curves for workers in ASSY does exist, especially for newer machines.

In general, ASSY is consists of six work centers. Their names and their features are given below:

1. <u>Assembly Optical, Metroframe:</u> Medium lead time, includes only sequential processes

2. Assembly Optical, Illumination: Short lead time, includes both parallel and sequential processes, high physical buffer
3. Assembly Optical, Laser: Short lead time, includes only sequential processes, high physical buffer
4. Assembly Optical, Lens: Short lead time, includes only sequential processes, high physical buffer
5. Assembly Mechanical, Electrical: Short lead time, includes only sequential processes, low on time performance (uses associated buffer times constantly)
6. Assembly Mechanical, Wafer Stage: Long lead time, includes both parallel and sequential processes, low delivery performance

1.5. Selection of the Work Centers

Although ASSY unit consists of six main processes, it is not possible to include all of them due to the time restrictions. Instead, a choice is made to include the processes which represent the most important characteristics of ASSY. This choice depends on the information about the characteristics of the work centers. The processes are either only sequential or a mix of sequential and parallel. Therefore it is decided to take one process of each type. As a result, we decided to focus on two work centers. Therefore other work centers are considered out of scope. Furthermore, throughout the report, the names of the selected work centers are not used because of the confidentiality rules of the company. Instead, we named selected work centers as Work Center 1 and 2.

2. Research Design

This chapter describes the research method and the solution approach which was followed during the master thesis project. In Section 2.1, the description of the problem which is studied is given. In Section 2.2, the research assignment is presented. In Section 2.3, the research methodology is described. In Section 2.4, the solution approach is explained. In Section 2.5, research questions are given. In Section 2.6, the deliverables of the project are listed. Last but not least, the scope of the project is given in Section 2.7.

2.1. Problem Description

The literature study on buffers in production by Gursoy (2011) reveals that buffers are used extensively in production and have various functions. In general, their goal is to compensate possible losses due to machine breakdown, material failure, problems in process and maintenance. Buffer management techniques in production environments aim to find the cost optimal configuration in terms of costs by changing the location and quantity of buffers (Gursoy, 2011).

Gursoy (2011) pinpoints four drawbacks of the literature on this topic. First, even though different types of techniques can be used in buffer allocation problems, they are not suitable for production types with different kinds of characteristics since these production types have different specialties than the so called “regular” production processes. Second, the literature lacks proper investigations about the specific features of the production processes with low production volume, long lead times, short product life cycles, high variation on demand and cycle times. Therefore design rules developed for mass production does not generally apply for industries with the characteristics described above. Third, the literature focuses on buffers in terms of physical inventory except a few articles. However, buffers in time could be preferred over inventory since it gives the producer certain advantages. Therefore it is important to examine systems which aim to determine how much time to allocate as their buffer. Fourth, buffer management approaches usually applied to flow lines, e.g. workstations in sequential order. There are not many articles which propose a solution for the workstations in parallel. Even though that would make the problem more complex, parallel placement of workstations are used in many types of production systems and the solution methods should also include the possibility of parallel stations. In the light of these points, it can be stated that there are gaps in the literature of buffer management.

As pointed out by Gursoy (2011), there is still room to further research buffer management, in production sectors with different characteristics. Photolithography production is one of those sectors. This thesis aims to serve the need for close cooperation between academia and the industry to address the questions that arise in practice.

The practical study if this master thesis research was done at ASML, that produces photolithography systems for the semiconductor industry. The production process of ASML has the following characteristics: It is subject to fluctuation in demand, uncertainty in cycle times, learning curves, constraints associated with clean rooms. Therefore, the underlying constructs of the production should be carefully examined such that the manufacturing has a sufficient delivery performance while costs associated with buffers are kept as low as possible.

The main purpose of this research is to develop a methodology for calculating buffers in production systems with the characteristics described above. Therefore the underlying constructs of the production should be carefully examined such that the manufacturing has a sufficient delivery performance while costs associated with buffers are kept as low as possible.

2.2. Research Assignment

As mentioned by Gursoy (2011) there are gaps in the buffer management literature. The main points of these gaps are described in Section 2.1.

These points show that design rules developed for other types of production does not generally apply for this sector. The difference in characteristics of production types and the lack of articles about the lithography industry leads us to the research assignment. The aim of the research is to close the gaps in the literature.

Industries like photolithography are not examined in depth in the buffer management. Therefore it is important to conduct a study in a company which produces photolithography machines. The company which has been described in the second section, namely ASML, matches with the requirements of this research. The company has a suitable environment to conduct a study about buffer management. Moreover, the company needed to know how effective their current buffer management methods are. As a result, the study is both beneficial for the literature about buffer management and for the company itself in its buffer planning operations.

The research assignment is defined as follows:

"Design a method for calculating time buffers for production environments which have the characteristics of uncertain demand, low production volume, steep learning curve, fluctuating processing times, and space constraints, so that the production units have sufficient delivery performance while costs associated with buffers are kept as low as possible."

2.3. Research Method

This project is an example of Business Problem Solving (BPS). According van Aken et al. (2005) BPS projects should aim designing sound solutions (design-based) and reaching performance improvements (performance focused). Additionally, rigour of the project will be strengthened by having the previously conducted literature review as a basis (theory based) and showing the reasoning behind the proposed solutions (justified). The process steps of the project, that are based on Aken et al. (2005) are depicted in Figure 2.

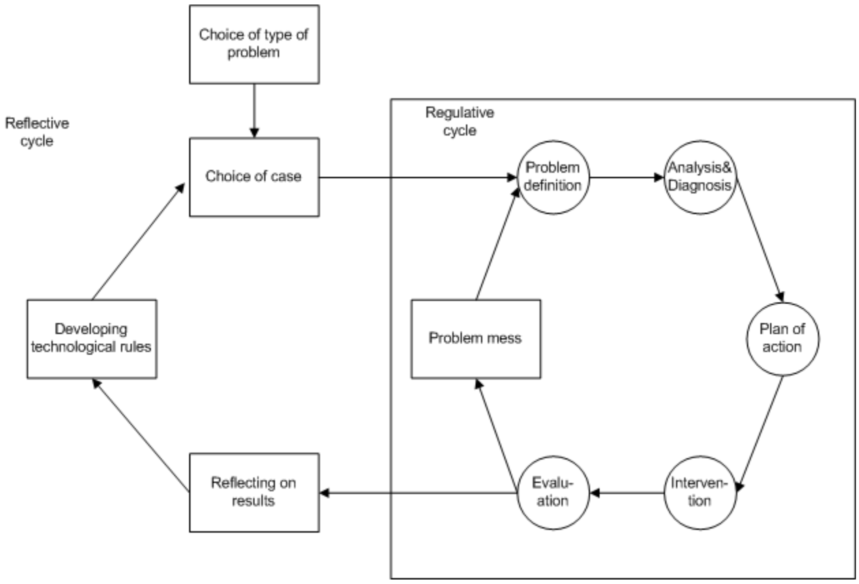


Figure 2 - Research Design

The project focuses on the first three stages of the regulative cycle: problem definition, analysis & diagnosis, and plan of action. These stages are associated with design, whereas the remaining two stages are associated with implementation. The length of the project (5 months) does not allow for both design and implementation of the solution, therefore intervention stage is omitted. This stage is replaced by insights about the implementation of the solution.

It should be noted that several stages of the project are already completed. Firstly, the type of problem and the case was chosen. The reader should refer to literature review of Gursoy (2011) for the choice of the type of the problem. Secondly, the problem to be solved was defined. More information about the problem definition can be found in Section 2.1.

Moreover, another classification could be made in terms of Operations Management (OM) literature. According to Bertrand and Fransoo (2002) OM research methodologies can be described as being axiomatic or empirical; normative or descriptive and quantitative or qualitative. This research project is quantitative, because it is based on the assumption that objective models can be built by researchers that explain (at least a part of) the behaviour of the real-life operational processes, thus the relationships between the variables are described as casual. Moreover it is axiomatic since it aims to obtain solutions within the defined model, get insights about the structure of the problem as defined within the model and produce knowledge about the behaviour of certain variables in the model. Furthermore, the researcher of the project is interested in developing not only policies and strategies about the problem, but also to improve over the results available in the current literature and therefore it is considered normative.

2.4. Solution Approach

Bertrand and Fransoo (2002) define two different approaches for axiomatic quantitative research. These approaches differ in the use of simulation. One of these approaches uses pure mathematical methods, while the other uses simulation in validation and analysis. If simulation is used, additional steps are needed for validation of results. Even though they have different execution steps, Bertrand and Fransoo (2002) take the first approach as the default way of doing research. Therefore this report is based on the axiomatic quantitative research. The approach of Bertrand and Fransoo (2002) is as follows:

- Conceptual model of the process or the problem: Use of selected concepts, accepted standards in relevant literature (buffer management in that case) in order to position the problem in the scientific literature.
- Scientific model of the process or the problem: Presentation of the problem in formal, mathematical terms such that either mathematical or numerical analysis is possible, or computer simulation can be carried out.
- Solution
- Proof of solution
- Insights relating the solution to the conceptual model

2.5. Research Questions

The research questions of this study are as follows:

1) What are the purposes of using buffers and safety times?

2) What are the main parameters (factors) that affect buffers?
    a) What are the variables, and what are the constraints?
    b) What is the relationship between these parameters?
    c) What are the effects of process uncertainties on the buffer calculations?

3) What method (or a tool) can be used to calculate buffers for modules production by using the parameters and relationships identified above?
    a) What is the objective of the method? Which performance measures represent the nature of the system?
    b) What is the best approach to model the relationships between parameters and other effects?
    c) Is it simple enough to be used in practice?

4) What is the difference between the proposed method and the calculation method that is currently in use at ASML?
    a) Is the solution proposed by the method better than the current situation?
    b) What is the improvement amount in terms of previously identified performance measures?

2.6. Deliverables

The deliverables of the project are be listed as follows:

- An overview of the current system, current situation of buffers
- A proposal of a new buffer management strategy for production environments which have the characteristics of low production volume, complex process routings, steep learning curve and fluctuating processing times
- Application of the proposed strategy for the photolithography producer ASML
    - A planning tool to determine buffer times for the production processes
    - Performance analysis of the new method in terms of delivery performance and cost
- Future research directions
- A Master Thesis report

2.7. Research Scope

As mentioned earlier, the project will focus on buffers at ASSY department. Therefore the suppliers and other production units (e.g. FASY and Test) are omitted. The research on buffer management in FASY and Test constitutes another graduation project, conducted by Perez Zavala (2011).

Furthermore, as explained in Section 2.4, the project does not aim to examine all of the ASSY processes. Instead, we have decided to include the processes which represent the most important characteristics of ASSY. Moreover, the actual aim of the project is to design buffers for all types of machines and processes. However, because of time limitations the most important machines and processes are selected to examine. This selection is explained in the following paragraphs.

In general, ASML produces two types of products such as PAS and TWINSCAN. The demand for PAS machines are severely reduced in the last years, so the project will only focus on TWINSCAN. There are different types of TWINSCAN machines and they are named the following: TWINSCAN product names start with some letters showing the type of the platform (bottom part of the machine) (e.g.: XT, NXT etc.), followed by a number that designates the lens (e.g.: 860, 1950 etc.). Following this representation for example, XT860 is a machine type, which is a member of XT8x0 machine family (ASML intranet site, 2011).

During the project, only the demand and production for two product families are considered, namely NXT 19x0 and XT 8x0. Several arguments could be made in order to justify this selection. Firstly, these two product families have the biggest demand; therefore most of the production is included by examining these families. Secondly, these two product families differ in terms of complexity of production processes. NXT 19x0 has a complex production process, high variability and long cycle time. On the other hand XTx80 has less variability and shorter cycle time. Therefore these two represent the most important and different processes at ASML.

Furthermore, the project does not aim to solve (or decrease) the variability issues in cycle times. Therefore the root causes of variability issues (referred as B-times) are out of scope.

3. Initial Analysis

As discussed in Section 2.4, the Assembly process consists of different work centers, which produce modules for Final Assembly. In this section, detailed information about the operations in these two work centers is given.

The remaining of this chapter is organized as follows: In Section 3.1, the key terms about the processes which are studied is given. Section 3.2 focuses on the selected processes. In Section 3.3, factors that have effects on the production and the cycle times are presented. In Section 3.4, the reasons behind the need of buffers are discussed. Section 3.5 describes the current situation of buffers at ASSY. In Section 3.6, the current buffer determination process is examined and the factors related to that process are explained.

3.1. Key Terms

This section gives information about the ASML terminology, key terms and its operations which will be analyzed further.

<u>Process:</u> The whole tasks of work centers, each of which produce one specific module. For example, “Wafer Stage” is considered as process.

<u>Operation:</u> The sub tasks within processes. For example, “Assembly” within Wafer Stage process is considered as one operation. Processes can include one or more operations. The relationship between a process and an operation is shown in Figure 3.

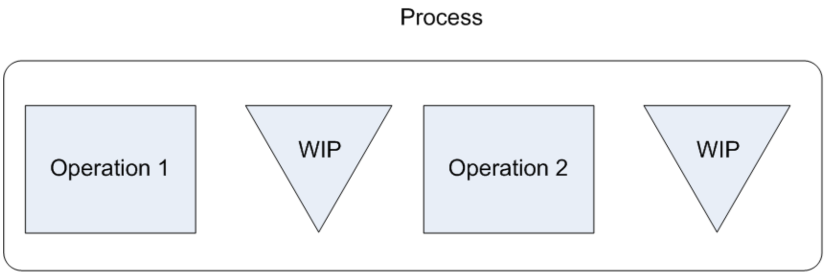


Figure 3 - Distinction between Process and Operation

<u>Work groups:</u> It is the number of groups of operators which work in that specific work center. Every work center consists of several work groups. Usually a work group consists of experienced and less experienced operators. The most experienced operators, which can do more than one job, would be the replacement for other people.

<u>Takt:</u> The work package which consists of 8,5 hours of work (at most), which is also equivalent to a shift. ASML executes 2 takts per working day and 1 takt on Saturday and Sunday. The length of the processing time is defined in terms of takts.

<u>A-times:</u> The normative, minimum elapsed time under optimal conditions (standard working speed, staffing, availability of materials etc.) between start and finish of a process.

B-times: Elapsed time for disturbances during manned machine hours between start and finish of an activity.

Disturbances: Disruptions in the operations because of the material issues, people issues and other issues. More information about the disturbances can be found in Section 3.3.1.

C-times: Idle time during holidays, weekends, and nights.

Cycle time (CT): The actual time a job takes to traverse a routing, from begin to finish. Cycle time of a part is the sum of A-time, B-time and C-time spent for that specific part. Since A-time of a part is always the same, cycle time varies with the changing values of B and C-times.

Processing time (PT): The worked time a job takes to traverse a routing, from begin to finish. Non-worked times (such as nights etc.) are not counted in processing time. Therefore processing time is not equal to cycle time. The relationship between cycle time and processing time can be expressed in the following equation:

CT = PT + C-time

Figure 4 also shows the relationship between the cycle time and processing time.

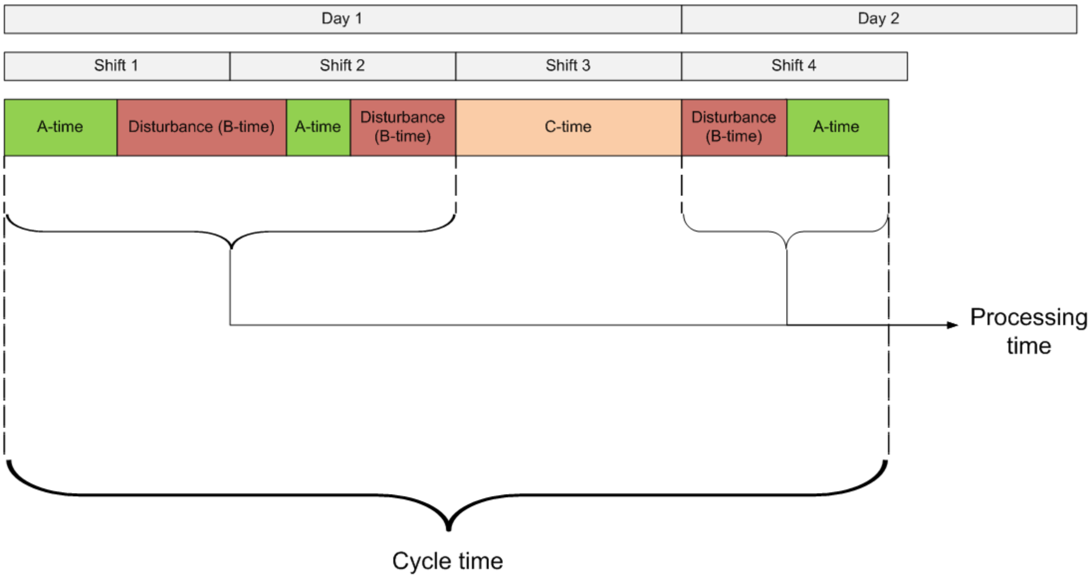


Figure 4 - Actual Cycle Time

Queue time (QT): The time, which is added to each operation after its A-time. According to ASML's way of working, after adding the queue time, the planned cycle time, which should cover 80% of all future realizations of cycle time.

It should be noted that the queue time that is used in that report is different from the queue time that is used in the literature. Queue time in the literature is the time that a part waits before a server until it is started to be processes. Queue time in this report, as well as the queue time according to the definition of ASML, is defined as safety time or buffer time in the literature.

Safety time (ST): The time, which is added to each process after its queue time. According to ASML's way of working after adding the safety time, the planned cycle time should cover 99% of all future realizations of cycle time. The sum of safety time and queue time should cover the B and C-times of the processes.

It should also be noted that the safety time that is used in that report is taken from the ASML definition. However, this terminology is also defined as safety time or buffer time in the literature. The reader should be aware that both of the discussed terms of queue and safety time are defined as the safety time (or buffer time) in the literature. We decided to use the definitions of ASML in order to diminish the confusions which might arise.

Planned cycle time: The planned time for each process from the beginning to the finish. Planned cycle time of a process is the sum of A-time, queue time and safety time of that process. Figure 5 shows the planned cycle time, the queue and safety time.

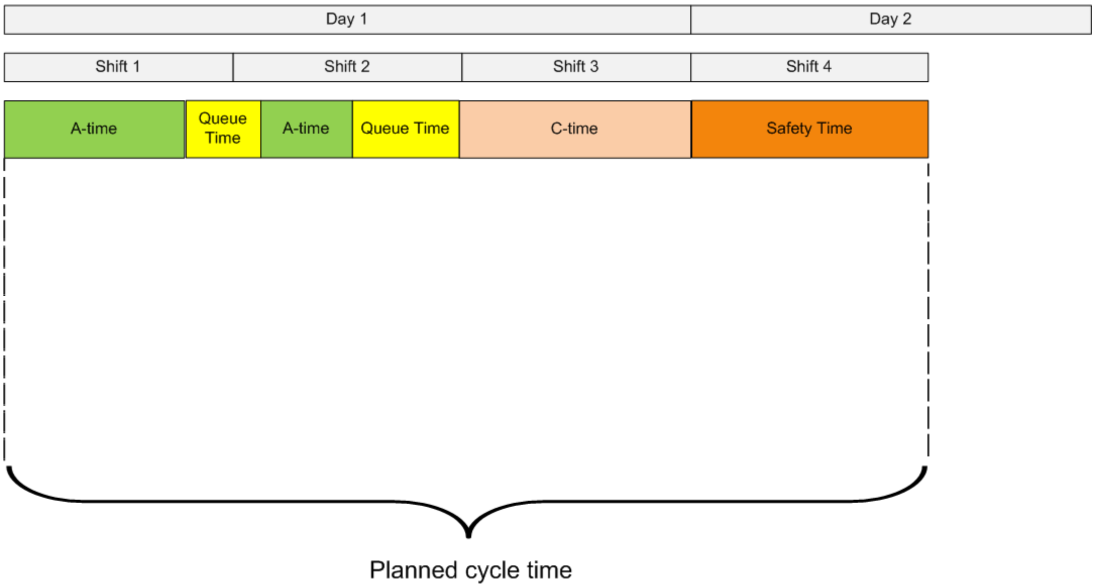


Figure 5 - Planned Cycle Time

3.2. Selected Processes

This section gives more detailed information about the processes which are analyzed further. As discussed earlier, work centers other than the selected two (Work centers 1and 2) are out of scope. These two work centers produce different types of parts, which are used in FASY. For Work Center 1, two most important parts are considered (Part 1 and Part 2). For Work Center 2 there is only one end product.

3.2.1. Work Center 1, Part 1

The team for this part is responsible for all TWINSCAN machines. It has its main focus for the XT and the NXT platforms. The production process of Part 1 is shown in **Error! Reference source not found.**.

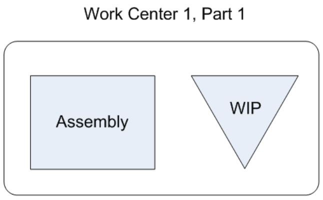


Figure 6 - Production Process of Part 1

The production consists of only Assembly operation, however the production processes are different for XT and NXT machines. For the NXT the average and standard deviation of the cycle time is bigger compared to XT. Furthermore, since NXT is a newly introduced product, its learning percentage is bigger than XT, which is a volume product.

3.2.2. Work Center 1, Part 2

The team for this part is responsible for the production for all machines. It focuses mainly on the XT and the NXT platforms. The production process is shown in **Error! Reference source not found.**:

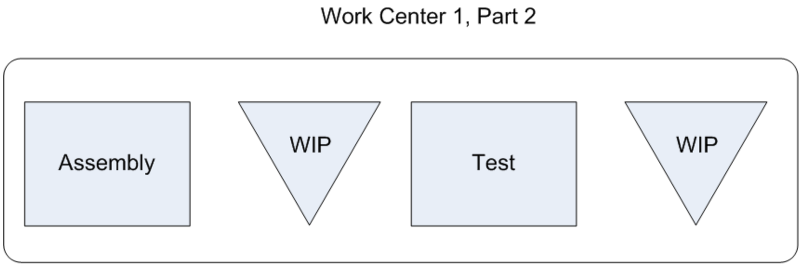


Figure 7 - Production Process of Part 2

The production consists of Assembly and Test operations. Although the production processes for XT and NXT are not the same, the difference between the average and standard deviation of the cycle time is not very high.

3.2.3. Work Center 2

The team of this work center has the job to produce the module which is necessary for the alignment of the wafer. **Error! Reference source not found.** shows the production process.

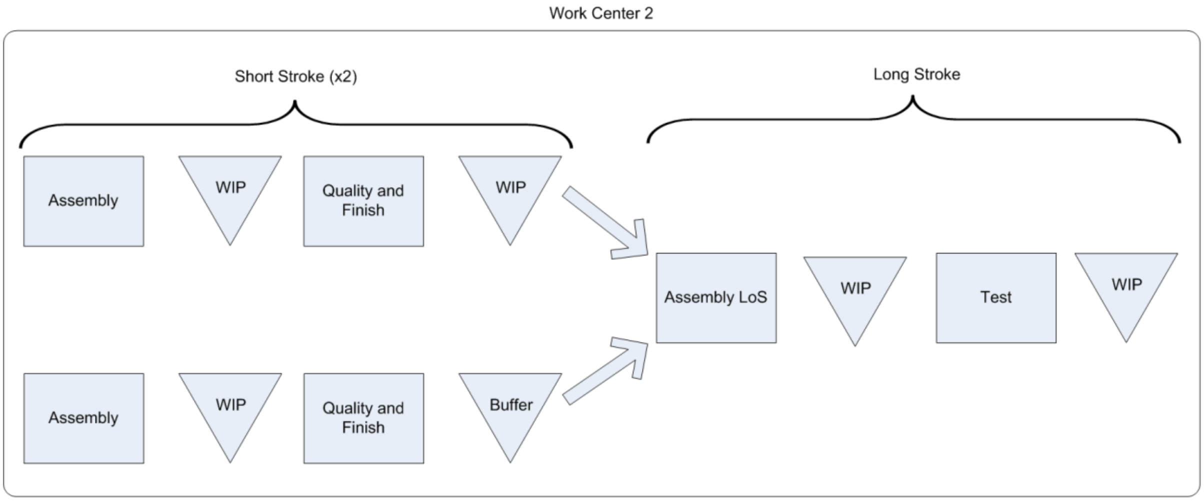


Figure 8 - Production Process of Work Center 2

In this process, first two Short Strokes are built in parallel. Afterwards they are assembled to create one Long Stroke. Therefore the operation of Assembly Long Stroke can only begin if two Short Strokes associated with this part are ready.

3.3. Factors Affecting the Cycle Times of Operations

This section describes the most important factors that affect the cycle times of the previously discussed processes. Because of the nature of the manufacturing operations at ASML, there are many disturbances during the processes. These disturbances are usually sudden and therefore affect the production heavily. Moreover, the capacity bottlenecks in the production and the working hours of the company have also an impact over cycle times. In the following sections these factors are discussed in detail.

3.3.1. Disturbances

The photolithography production is a complex process and it is subject to heavy failures or other problems. Therefore in a considerable amount of the cases it is not possible to finish the part in the previously defined amount of time (A-time). These problems, also called as disturbances, which cause the cycle time to get longer, are defined as B-times. There are several causes for disturbances in the process. The most important ones are as follows:

- Material issues: This includes all of the problems caused by missing material or other problems with the quality of the material. The material might be missing because of the supplier, as well as because of another internal production process. Moreover the material might not be in the right place; this would also cause fluctuations in the cycle time. The causes of the quality problems are diverse but they affect the production significantly. In general, longer disruptions are usually caused by material issues.

- People issues: This includes all of the problems caused by missing operators, wrongly followed procedures, quality issues related to operators etc. Operators might be missing because of illness etc. or because he/she is helping to another process. Other two issues might be caused by several reasons. In general it can be said that shorter disruptions are usually caused by people issues.

- Other issues: Except the material and people issues, other problems might also occur. For example, the tools which are required to do the process might be missing or occupied, there might be other problems in the facility or there might be other technical issues. All of these are grouped under the name of other issues, since their occurrence is not that high compared to the first two issues.

As a result of these disturbances, processes might not be finished within planned processing time (A-time).

3.3.2. Capacity bottlenecks

As any other production system, ASSY processes also suffer from bottlenecks, starvation and blocking effects.

In Work Center 1, blocking and starvation are not present. The main reason for this is that there is not a clear bottleneck in the system. Furthermore the processing times are short and there are not big deviations in the system. Therefore the process runs more smoothly compared to other work centers.

In Work Center 2 however, the Test operation is the bottleneck. In fact, this process is the bottleneck of the entire ASSY Unit. The main reason for this is the need of so called ‘Testrigs’ in the Test operation. Testrigs are unfinished machines and they are used to test the finished modules. However, these Testrigs should be replaced every couple of months and because of space restrictions the company cannot maintain more than 3 Testrigs at a time. Therefore before the Test operation there are physical buffers to put finished Work Center 2 modules. The company aims to make sure that Testrigs are almost always utilized.

3.3.3. Working hours

Cycle times are calculated according to the full days. Therefore cycle times are equal to the sum of A,B and C-times.

C-times are consisted of weekends and nights, during which no production is made. It should be noted that operators work two shifts per day during weekdays, and one shift per day during weekends. The working hours are not equally distributed during the week (and the day); therefore the cycle times might follow a non-continuous pattern. The lead times of work centers within ASSY are relatively short, therefore working hours of the company can have a large impact on cycle time reported in calendar days.

It should also be noted that in order to calculate stock values, inventory turnover rates etc., and cycle times need to be used, because even in the night the materials spend time at the clean room. However in order to calculate buffers, processing times (sum of A and B times) should be used.

3.4. The Need for Buffers at ASSY

This section explains the necessity and the reasons to use of buffers at ASSY work centers.

3.4.1. The challenges of planning lead times at ASSY

As defined in Section 1.4, the ASSY module faces big challenges while planning modules production for FASY. The main problems stem from demand fluctuations, the cycle time variability and the existence of the learning curves.

The company is subject to heavy fluctuations in demand because of it is a supplier for semiconductor industry. Therefore the demand follows a pattern with big variation. The meaning of this is that, for some period of time the demand is very high, and after that for another period it is very low. During demand peak times, which were the case in the last year, it is very important for the company to produce and sell as much as possible. This can only be done with utilizing FASY cabins as much as possible. Therefore in demand peak times, ASSY should be on time for almost every time. During low demands however, it is important for the company to minimize its operating costs. Therefore in those times, the company would like to work with a lead time which ensures the minimum cost.

These requirements force ASSY to set reliable lead times, so that it can deliver on time (and/or with low cost) to FASY.

Besides the demand variability, there are two important aspects which make the job of ASSY difficult. First of them is the variability in cycle times, which was discussed in the Section 3.3. The second is the existence of the learning curves. The ASSY processes are subject to learning curves and therefore cycle times are decreasing in time. More information about this is given in Section 3.6.

3.4.2. The purposes of buffers

The production planning process at ASSY is affected by the challenges explained in the previous section. There are typically two important goals to be reached while planning production. There are as follows:

- Low lead times
- High delivery reliability

It is possible to reach low lead times, however this would clearly decrease the delivery reliability, or vice versa. So, very high delivery reliability can be reached but in that case the lead times would be very high. It is clear that there is a trade-off between these two goals.

In the optimal case, the company wants that the production of each module goes as planned, so that the lead times would be as low as possible and each module is delivered to FASY on time. However, this is not the case, because, as explained earlier, ASML processes are subject to many disturbances and therefore cycle times are fluctuating heavily. This fluctuation might affect FASY, the so called "customer" of ASSY.

In order to prevent the delivery reliability problems to FASY, two types of lead times are used, such as internal and external cycle times. Internal cycle time is the actual planned time for the process. This includes cycle times of operations and the buffers within the operations. However, because of the disturbances, if these lead times are used, it is not possible to reach to high delivery reliability. Therefore the company uses another concept, namely external cycle time, which is the supply lead time for FASY. The relationships between internal and external lead times are shown in Figure 9:

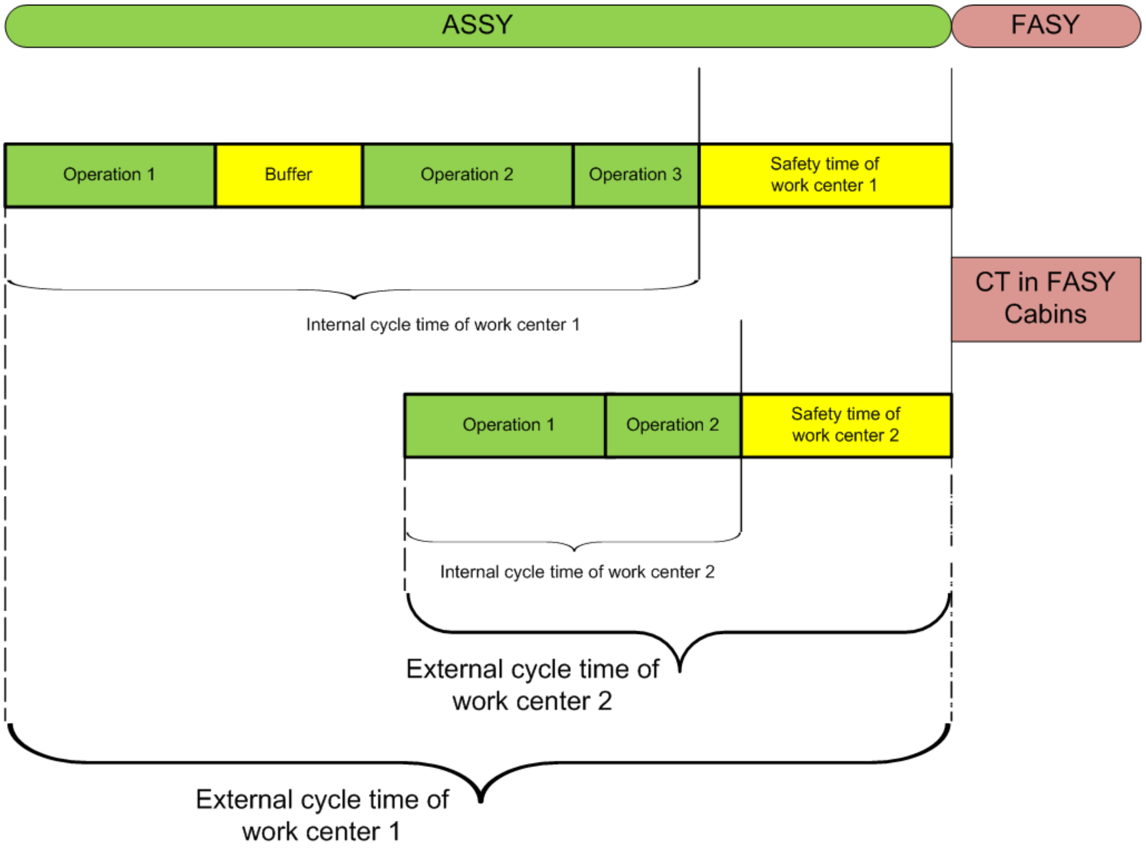


Figure 9 - Current Situation of ASSY

The safety time, which is put at the end of the operation, determines the difference between the internal and the external cycle time. The reasons of using time buffers at ASSY is described in the following section.

3.4.3. The types of buffers

In most of the cases buffers at ASML are defined in terms of time. The buffers are kept in time because it gives the company planning advantages:

- Each type of end product (machine) has its own specific parts (modules). That means, even at the beginning of ASSY, each module is associated with a specific end product type. As a result of this, the company usually does not use common parts in production.

- The semiconductor industry has a very the dynamic nature in terms of demand. Thus, if modules are kept in stock, there is a possibility of being obsolete and the necessity of rework.

As stated by Molinder (1997), safety time is preferable to safety stock for companies with high demand and process variability. He argues that in those cases the lowest cost is attained by using safety times. Moreover, Molinder (1997) also proposes that low stockout/inventory ratio has an effect on the choice of safety time/safety stock. For the low level of stockout/inventory ratio, safety stock is a better alternative compared to the safety time. However, for the higher value of this ratio, the safety time should be preferred. At ASSY system, both costs are high compared to other industries. However, stockout costs are derived from cabin costs at FASY and they are very costly compared to the holding costs of modules. Molinder's conclusions show that it is logical to use safety times (time buffers) instead of physical stocks in production environments similar to ASSY work centers.

Time buffers also result in physical stocks, if the production finishes before predefined time. For example, if a process is planned to be finished in 10 days (including buffers), there will be cases that the process finishes in less than 10 days, e.g. 7 days. This will mean that it has to wait 3 days until it is picked by the next production unit, namely FASY. As a result of the buffer times, some finished parts need to wait for some time before they continue to be processed. This issue might create a space problem and it costs many because capital is tight up in production environment if it is not handled with care. Therefore the company specified the physical storages to put finished modules. For each module there are some predefined places. However, physical buffers are an important problem for ASML, because of the space restrictions. Currently there is a huge demand of machines, so the capacity of storages for modules is may not be enough. Moreover, modules also need to be stored in the clean room, and the holding cost in the clean room is quite high compared to outside.

3.5. Current Situation of Buffers at ASSY

The current design of buffers includes two different aspects, namely queue time (also referred as queue buffer) and safety time (also referred as safety buffer). Queue time is the buffer time between operations. It is needed to make sure that each operation starts on time. According to ASML strategy, queue time should ensure %80 of delivery reliability for every work center. Safety time on the other hand is the buffer time at the end of each process (module production). So, in this case, it is the time between ASSY and FASY. ASML strategy states that after adding safety time to processes %99 of delivery reliability should be reached. It should be noted that, these two time aspects are not substitute of each other, thus in each work center both of these aspects could found. Figure 5 explains the two types of buffers.

Currently the company is deciding on the size of buffers and on the location of buffers. These decisions are related to the buffer allocation problem (BAP), since BAP is about deciding on how much buffer to allocate and where to allocate. The decisions are explained below.

3.5.1. Size of buffers

The company would like to have some buffer time after each process. Therefore they use the concepts of Queue and Safety Times, which are explained in Section 3.1. How long these times should be and how they are calculated is explained below.

As a first step, the cycle times of the last 30 parts are taken from SAP. Older cycle time data are omitted from calculation. Then, the outliers are determined and eliminated. It should be noted that currently the company only analyzes values which are very big compared to the average cycle times. Current process does not aim to adapt the values of nights and weekends to reach the processing times. Therefore cycle times are used in calculations. In the next phase, the cycle times are assumed to be normally distributed and the queue and safety times are calculated accordingly.

3.5.2. Placement of buffers

As explained in Section 2.5, the placement of buffers is done by using a predefined rule/definition, which was set by the company managers. According to this, the queue buffer should absorb variations in the operations; therefore they are put after each operation. On the other hand, the goal of the safety buffer is to reduce tardiness for the whole process, therefore it is put at the end of the processes. That rule implies that if queue buffer is not enough for an operation, safety buffer at the end of the process will be consumed.

3.6. Factors Affecting the Buffer Determination Process

This section describes the most important factors that affect the buffer determination process at the company. As discussed in the previous section, because of the nature of the manufacturing operations at ASML, the cycle times of production processes have high variation. Besides, other issues also affect determining (calculating) buffers such as the working hours of operators, the outlier values and existence of the learning curves in production. In the following sections these factors are discussed in detail.

3.6.1. Distributions of cycle times

One of the main inputs while calculating buffers is the cycle time distributions. Since the company do not use buffer stocks, there are no physical stock issues to be considered.

As the reader already knows, buffer times aim to absorb the effect of disturbances (deviations from the normal process) by creating some slacks between operations. The company would like to reach to a service level by setting buffer times (slacks), such that the costs related buffers (and costs related to tardiness of Assy) would be minimized. The answer to the question of ‘how much slack should be used after a process’ depend heavily on the probability that cycle time is bigger than a certain value. In that context, how cycle times are distributed has a big importance in the buffer determination process.

Because of the complexity of production at ASML, it is not easy to determine distributions of cycle times. Moreover, since the specialties of processes are completely different from each other, it is not possible to use one distribution for all processes. In most of the cases, cycle times of operations can be fitted to a distribution such as exponential or gamma. As an example, the current distributions of the selected modules of Work Center 1 are shown in Figures 10-13:

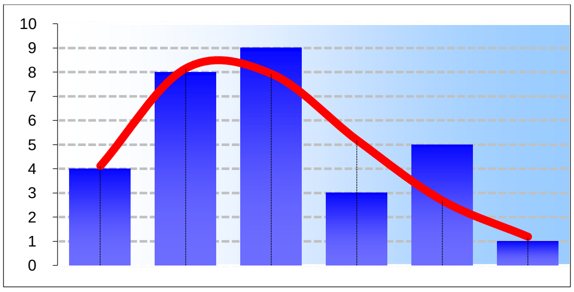


Figure 10 - Cycle Time Distribution of Part 1 for XT

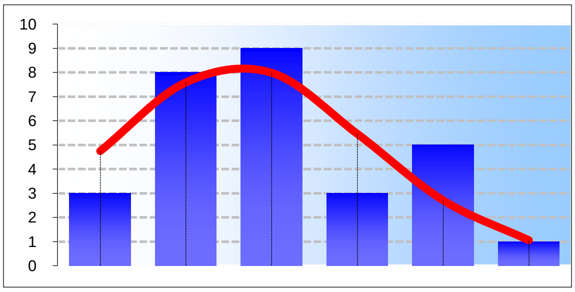


Figure 11 - Cycle Time Distribution of Part 2 for XT

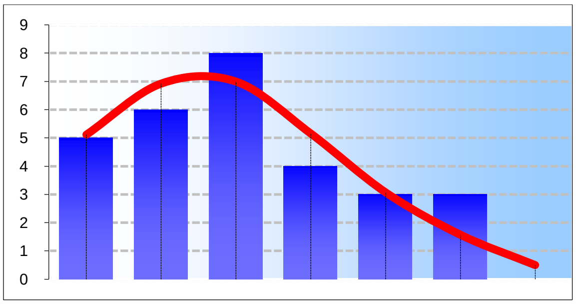


Figure 12 - Cycle Time Distribution of Part 1 for NXT

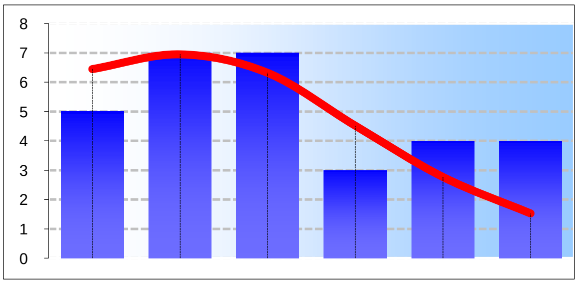


Figure 13 - Cycle Time Distribution of Part 2 for NXT

Currently the company is using the normal distribution as the default way to fit cycle times. This assumption has two drawbacks. Firstly, as discussed above processes have different specialties, therefore they should be examined one by one before deciding on distributions. Secondly, with normal distribution, values can be both negative and positive. Since real life operations can only have positive cycle time values, the use of the normal distribution might lead to suboptimal decisions.

3.6.2. Working hours

As explained in Section 3.3.3, working hours have impacts on the cycle times. This would also affect buffer calculation process; therefore the situation of working hours should be also taken into account while calculating buffers. The issues which need to be taken into account are time spend in the nights (because there are no night shifts in ASSY) and time spend during the weekends (because there is only one shift at the weekends instead of two). If cycle times are used without any adaptation, the used times might follow a non-continuous pattern and this would lead to using wrong distributions.

3.6.3. Outliers

Another issue which has an effect on the buffer calculations is the determination of outliers. The company has the data for cycle times of operations, but this data set cannot be used without doing any preliminary analysis since it contains values which need to be eliminated. In the production environment, both very high and very low values are observed and therefore both types should be analyzed.

It is somewhat common to observe unrealistically long cycle times compared to average. There are two reasons for higher values. Firstly, it could be a recording mistake, e.g. the operator might forget to record that the production of that particular part has been finished. Secondly, there might be extraordinary instances such as accidents, emergencies etc., which usually are the cause for the very long cycle times. Although they are real (which is different than the first case), because they are usually one time instances, they might be neglected while calculating buffers. The decision to eliminate a cycle time is taken either after discussing it with production or applying a statistical methodology.

It is also possible to observe low cycle times values. It is obvious that cycle time values cannot be negative, thus negative values should be excluded. On the other hand, positive but very small values can also be observed. For each process there is a predefined time (A-time) and under normal circumstances a particular process cannot be finished earlier than this A-time. Therefore values which are smaller than A-time should be analyzed and a decision should be taken whether to exclude them. Again, there are two reasons to observe lower values. Firstly, it could be a recording mistake e.g. the operator might forget to record when the production of that particular part has started. Secondly, a deviation from normal circumstances such as an experienced operator, faster working pace etc. might cause the cycle time to be smaller. The values which resulted from the second case should not be excluded. The decision to eliminate a cycle time is taken either after discussing it with production.

Both very high and very low values affect buffer calculations. They cause deviations in the average, variance and also in the histogram of cycle times. Therefore in order to create a good buffer management strategy, detection of outliers is important.

3.6.4. Learning

The production at the company is affected heavily by the learning curves. Since the operations are complex, it takes a very long time to produce a machine for the first time. However, after some time, the processing time is reduced drastically. Therefore most of the cycle times are not static, especially for the machines that have been newly introduced to the market. After some time, the cycle times reach a more or less steady state. The cycle times usually follow a pattern similar to which is shown in Figure 14. The reader should keep in mind that this graph shows the change in the average cycle times over time.

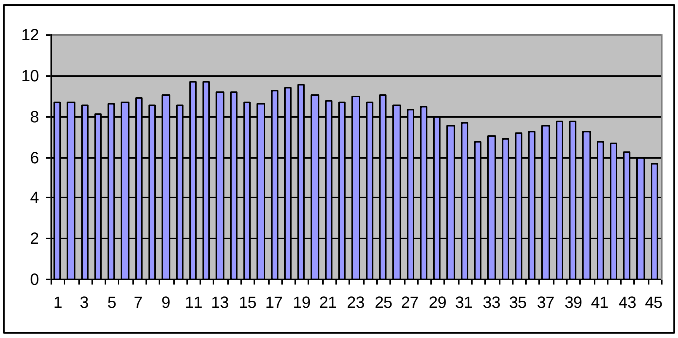


Figure 14 - Average Cycle Times of a Work Center

Learning has important effects on buffer calculations. Because of the changing nature of the cycle times, buffers need to be reviewed in a timely basis. Currently the company is reviewing buffers in every 3 months, but for some cases it is not enough. Furthermore, the effects of learning can only be seen over time, therefore it is not easy to estimate the future effects.

The use of the older cycle times during buffer calculations creates another problem. If the time span which is used as base in buffer calculations is too large, that would lead using very old data (cycle times) and therefore the correct buffer values cannot be found. On the other hand, if the time span is too small, then enough data cannot be collected to estimate the future buffer times. Therefore the company uses the last 30 machines or parts in order to make buffer calculations. This includes the values from approximately 3-4 months.

4. Buffer Calculation Method

The previous chapter revealed the factors that affect the cycle times of operations and factors that have effects on the buffer calculation process. These two groups of factors represent the main parameters that affect buffers, thus the first research question of this study is answered. The second research question aims to design a method for calculating and managing time buffers.

Throughout this chapter we need to define and use a number of parameter values such as learning percentages and service levels of operations. However these parameters heavily depend on the situation. In this chapter we have described the method without defining the values. Chapter 6 shows how to determine the values for a given system.

The remainder of this chapter describes the proposed buffer calculation method in detail. It is organized as follows: In section 4.1, the assumptions, that are needed to be made to develop a buffer calculation method, are presented. Section 4.2 discusses the most relevant studies that are conducted in this area. In Section 4.3, our options for designing a buffer calculation methodology are presented. In Sections 4.4, 4.5, 4.6, 4.7 and 4.8, these options were examined and our selections are explained. Section 4.9 gives information about the extension of the model, namely finding optimal service levels

4.1. Assumptions

In order to create a model for the buffer calculation process, some assumptions need to be made, especially because the study is conducted in a business environment. These assumptions are presented below.

- All operations within the process are independent, thus the cycle time of an operation does not depend on another.
- The data of past cycle times is known.
- Since the high-tech machines production is subject to the developments, the cycle times of modules might vary in time because of technical modifications (to product) or updates. However, throughout the analysis, this possibility is omitted (We consider a steady state in terms of technological developments).
- The costs of modules becoming obsolete or damaged while waiting in stock are assumed to be zero.

4.2. Related Works

As discussed in Section 2.1., the studies conducted in the buffer management literature do not focus on the characteristics of the photolithography manufacturing systems. Although the methods that are developed do not match completely with our case, the most relevant works are described in this section, which constitute the base of our methodology.

The earlier works in the area of buffer times/external lead times focused on the problem of finding the optimal lead times in serial production systems such as Yano (1987) and Gong et al. (1994). More recent studies on buffer times are based on the theory of constraints (TOC), developed by Goldratt (1984). Two of these papers are described in the following paragraphs.

Radovilsky (1998) developed a methodology to calculate the size of time buffers based on TOC. He argues that time buffers should be put into the production system by planning operations with extra time such that the system is protected from the fluctuations of processing time. According to his study, there are two important issues which need to be solved in terms of buffer times. There are as follows:

1. The location of the time buffer: According to the paper the most suitable places to put time buffer are before the bottleneck station and after the final operation.
2. The size of the time buffer: Even though previous works suggest that there is no need to calculate the time buffer (the basic recommendation is to determine the half of the total manufacturing as buffer), Radovilsky (1998) proposes a queuing theory based model.

Kuo et al. (2009) focused on time buffers at a wafer fabrication factory. They assume that the locations of the buffers are already determined and therefore their methodology aims the following:

1. Calculating the size of the total time buffer: They argue that four main approaches can be used for calculating the buffer sizes, such as direct approaches, simulation method, analytical method and statistical analysis. In their analysis, the authors used a combination of simulation and analytical approaches.
2. Calculating the buffers within operations: Kuo et al. (2009) suggest that the lead times for operations should be estimated based on the methodology used in 1.

The works of Radovilsky (1998) and Kuo et al. (2009) have made the most important contributions to the buffer (time) management literature. However, neither of these papers can be applied directly applied in our case, since the characteristics of the manufacturing system considered in this study are different than the settings in those papers. Therefore a new methodology is developed based on these works in literature.

4.3. The Methodology

As it can be seen in the previous section, the (time) buffer management literature aims to find the answers to the following three questions:

1. Where to put time buffers?
2. How to calculate the size of the buffers?
3. How to allocate the buffers to the determined locations?

On top of that, we think that there is another question to be answered. Articles in this area use the following strategy: They calculate the total buffer, considering the process as one production unit, and then they divide this buffer within operations if necessary. However, no researcher in this area investigated if it is beneficial to use a reversed strategy, which is as follows: Calculate buffers separately per process, and then sum them up in order to find the total buffer. We formulated this issue as another question to be answered.

- Which buffer strategy to use in order to calculate buffers?

The methodology which is proposed in this thesis answers all of the four questions. The selection of methods is explained in the following sections.

However, in order to proceed to the next steps, we need to determine also the sequence of these decisions. The articles in this literature show us the following:

- Radovilsky (1998) used the sequence of first determining the location and then the size. The reason behind this is that the location of buffers will depend on parameters (or specialties) which do not change very often. Therefore a decision on the sizes of buffers will stay the same for a long period. However, the decision on buffer sizes can change more often. Therefore the location of buffers can constitute a "default" starting point for buffer size calculations.
- Kuo et al. (2009) put the allocation of buffers to the last sequence, making the size calculation earlier then allocation. The reason behind this is that allocation decision might depend on the total size of the buffers. Therefore first the size of the whole buffer should be known.

Thus, the sequence we used is as follows:

1. Which buffer strategy to use in order to calculate buffers? (Section 4.4)
2. Where to put time buffers? (Section 4.5)
3. How to calculate the size of the buffers? (Section 4.6)
    - Additions to the calculation method (Section 4.7)
4. How to allocate the buffers to the determined locations? (Section 4.8)

Following sections explain the decisions which were taken in order to answer the questions given above.

4.4. The Buffer Calculation Strategy

As we stated in section 4.3, we were investigated which buffer strategy is suitable for the production system in consideration. Our options are as follows:

- Option 1: Calculate the process as a whole: Calculate the total buffer, considering the process as one production unit, then divide this buffer within operations if necessary
- Option 2: Calculate operations independently: Calculate buffers separately per process, and then sum them up in order to find the total buffer.

All articles in the literature use the first option as their buffer calculation strategy without considering another alternative. The reasons behind this choice are explained in the following paragraphs.

The variations (or tardiness) of the earlier operations can be compensated by the later operations if the operations complete their work on time (or with very small variation). Therefore the total process may not be late, even though some of the operations are late. The opposite is also the case; the variations of the later operations can be compensated by the earlier operations. That means, in order to be on time as a whole process, it is not necessary that all of the small parts (operations) are on time.

Using option 1, (calculating the process as a whole) assures that the whole process will finish on time according to the predefined service level. On the other hand using option 2 (calculating buffers of operations independently), will assure that all of the small parts (operations) are satisfying the predefined service level, which results in the overestimating buffers and lead to suboptimal solutions. Therefore we decided to use option 1, calculating the buffers in a process with considering the disturbances of the whole process.

4.5. The Location of the Buffers

In this section, the locations of the buffers are determined. Placing buffers between the processes are equivalent to decoupling the production units (processes) by using the decoupling points. According to the methodology of Bertrand (n.d.) these decoupling points are usually the physical stock points, however the production units could also be decoupled by other means, which is the 'time buffers' in this case.

As also discussed by Radovilsky (1998) the most suitable places to put time buffer are the following:

- Before the bottleneck station
- After the final operation

As stated by Bertrand (n.d.) the bottleneck station should be decoupled and buffers should be placed in front of and after the bottleneck. This is also in line with the definitions of Radovilsky (1998). In the case of the photolithography modules production, many of the operations within a process are interconnected because they use the same resources (material, worker etc.). For example the workers of one operation might also be involved in the next (or previous) operation. Besides, putting time buffers at the end of the process is logical, because this would allow the buffer cover all of the disturbances which occur while producing the module.

As a result of this, we decided to use the suggestion of Radovilsky (1998), putting buffers before the bottleneck and after the process.

4.6. The Method for the Buffer Size Calculation

As stated by Radovilsky (1998), most of the articles in that area do not focus on the buffer size calculation. The work of Radovilsky (1998) considers some easy measures as well as advanced ones. One of the most basic methods suggests adding half of the manufacturing time as a buffer. Kuo et al. (2009) proposed an analytical calculation methodology. Moreover, they introduce "statistical analysis" as another approach, but they do not use this in their paper. As having considered the methodologies used in these papers, we decided to discuss three possible methods to use in order to calculate buffers in photolithography industry. They are as follows:

- Method 1: Direct approach (e.g. adding half of the process time)
- Method 2: Analytical approach (Kuo et al. (2009))
- Method 3: Statistical approach (fitting distribution)

These three methods are explained in detail in the following parts. After the introduction of the methods, we selected the most suitable methodology.

4.6.1. Direct approach

According to Radovilsky (1998) the direct approach determines the buffer size by adding half of the process time as the buffer time. Thus this approach is represented as follows:

$$B_i = 0.5 * Ta_i \qquad (1)$$

where;

B_i : Total buffer time of process i
Ta_i : A-time of process i

This can be extended by allowing choosing the multiplier a number which is different from ½. In that case the formula turns into:

$$B_i = n * Ta_i \qquad (2)$$

There is not a generally acknowledged way to determine the multiplier n. Past data could be used to find a suitable value. According to the article of Kuo et al. (2009) this ratio lies between 2.5 and 10.

4.6.2. Analytical approach

The analytical approach, proposed by Kuo et al. (2009), determines the buffer time by using the average and standard deviation of the process. In their study, they use buffer time as an intermediary parameter to set the due date and they also use the release date of a lot in their formula. However, since our aim is to determine the buffer time only, these additions are excluded. The analytical formula is as follows:

$$B_i = PT_i^{av} + \Theta * PT_i^{\sigma} - Ta_i \qquad (3)$$

where;

B_i : Total buffer time of process i
PT_i^{av} : Average processing time of process i
PT_i^{σ}: Standard deviation of the processing time of process i
Ta_i : A-time of process i
Θ : Safety allowance

Like in the first method, there is not a clear way to determine the value of θ how to determine the θ.

4.6.3. Statistical approach

This approach aims to determine the buffer time by using statistical distributions and setting a service level. In that method, a statistical distribution that can describe the characteristics of cycle times of processes is selected. Furthermore, a service level is determined. The method calculates the required buffer time in order to reach that predefined service level. The buffer time is found by using the below given formula:

$$P(Ta_i + B_i > CT_i) = \alpha \qquad (4)$$

where;

B_i : Total buffer time of process i
CT_i : Cycle time of process i
Ta_i : A-time of process i
$P(x)$: The probability of occurrence of a cycle time level of x

It is uncertain how to determine the service level, as in the case for the first two methods. An attempt was made to determine the cost optimal service level for this method in the Section 5.4.

4.6.4. Selection of the Method

In the previous sections, three methods were discussed which can be used for calculating buffers. Selecting one method mainly depends on the following:

- Does the method fit the characteristics of the photolithography industry? (Research questions 3.a and 3.b)
- Is it simple enough to be used in practice? (Research question 3.c)

One of the main characteristics of the photolithography industry is the existence of the learning effects. These effects are added to the average of the cycle times (in other words the average cycle time is reduced according to the result of the learning effect calculation), therefore all of the methods can be modified to be used in an environment where learning exists.

However, the first two methods do not take the distributions of the cycle times into account. They use only the average and standard deviation (thus only the first and second moments) and calculate the buffer times by using these parameters. Therefore, with different distributions and with the same parameters, they give the same result. On the other hand, the third method can be adapted to different distributions.

In order to see why the use of distributions is important, the reader can refer to the following example. In Table 1 it is shown the quantiles of modules of Work Center 1, which are in the scope of this study. We use the average and standard deviations of the production processes of these modules (in terms of shifts) and calculated the values for 0.99 probability for three different distributions (without any other adaptation).

Table 1 - 99% Quantile for Different distributions

		Part 1 XT	Part 1NXT	Part 2 XT	Part 2 NXT
PARAMETERS	AV	2,5	5,3	6,36	6,12
	STDEV	1,2	1,8	3,21	2,64
RESULT FOR %99 QUANTILE	NORMAL	5,29	9,49	13,83	12,26
	GAMMA	6,10	10,36	16,08	13,86
	EXPON	11,51	24,41	29,29	28,18

The results of the different distributions for the same average and standard deviation parameters differ on average at least 1.38 shifts, which is 0.69 days. The difference for each distribution pair is shown in Table 2.

Table 2 – Average differences

	AVERAGE DIFFERENCE
NORMAL-GAMMA	1,38
GAMMA-EXPON	11,75
NORMAL-EXPON	13,13

This small example shows the different the results that can be achieved with the same parameters, even for the 99% service level. Therefore it is crucial for these types of situations to select and use a suitable distribution. The first and second models do not rely on the distribution, therefore their quality of decisions are limited. It is possible to get some adequate results with these methods for one module or for one situation, but their use cannot be generalized for more modules or for more cases.

All of the three methods use parameters which need to be calculated with decision makers. The multipliers in the first two methods can be interpreted as the “service level”, because the aim of the multipliers is to determine on the weight of the part that is essential in the calculation. The service level, which is used in the statistical approach leads to more efficient decision making processes because of two reasons. Firstly, it is easier to determine a service level (compared to the other two parameters). Secondly, it is also an output of the buffer calculation model, and therefore the decision maker has more control over the general process.

As having considered these arguments, we decided to select and use the statistical approach as the buffer calculation method to be used in production processes with the characteristics that are similar to photolithography industry.

4.7. Additions to the Selected Buffer Size Calculation Method

In the previous section, the statistical method is determined as the most suitable approach to calculate buffers. However, without modifying the method according to the characteristics of photolithography industry, the method cannot give sufficient results. additions need to be made to the selected method, and they are presented in the following sections.

4.7.1. Selection of the Distribution

This part constitutes the selection of a distribution for the data set. Since the cycle time data is used in order to calculate the buffers, it is expected to fit into a distribution from theory. If it would be possible to create a match between the data set and a distribution, a prediction about the buffer times could be made. Although, it is time consuming to try to fit the data to a various number of distributions, a couple of the distributions can be determined as 'candidate distributions' for fitting. Following paragraph explains the choices which were made in terms of the distributions and the reasoning behind these choices.

Kuo et al. (2009) suggest that Beta distribution is widely used in project management types of jobs because of its left skewed shape and suggest using it in buffer time calculations as well. Thus, beta is considered as a candidate distribution.

Moreover, the length of a process that is a sequence of several independent tasks can also be modelled by a variable following the Erlang or Gamma distributions. Since the Erlang distribution is a special case of the Gamma distribution where the shape parameter k is an integer, the Gamma distribution is selected to be used as a candidate distribution. Moreover, gamma distribution is the conjugate prior for the exponential distribution, as well as for many other distributions such as normal, Poisson, inverse gamma etc.

In queuing theory, the service times of agents in a system are often modelled as exponentially distributed variables, because of the simplicity of using this distribution. Gamma distribution might substitute Exponential, because with shape parameter it can be defined as Exponential. However, this distribution is selected as a candidate distribution because it can be used easily decision makers.

As a result, we decided to use these three distributions as candidates for our analysis, namely the beta, gamma and exponential distributions. This part does not determine one 'suitable' distribution; it leaves the decision to the decision maker, which should fit the real data to the one of the three candidate distributions.

Moreover, in Section 3.3 we showed how the cycle times of processes are distributed. This initial analysis show us that our choice of 'candidate distributions' are in line with the company characteristics.

While fitting the data to the distribution a statistical test should be used to measure the 'goodness of fit' between the distribution and the shift data. Kolmogorov-Smirnov test is chosen as the method to measure the goodness of the fit. The Kolmogorov–Smirnov test is a nonparametric test for the equality of the continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution or to compare two samples. It can also be modified to serve as a goodness of fit test. Even if the selection of distribution would made via using a histogram, Kolmogorov-Smirnov test would be used to validate this decision.

4.7.2. Learning effects

It is known that because of the complexity of this kind of production, all of the modules are under the effect of the learning curves. Therefore, before predicting the future buffers, the future average and standard deviation of the shifts should be computed.

The "learning effect" states that the number of direct labour hours to produce a unit decreases over time according to a function (Towill, 1990). There are specific and widely known formulas to calculate learning effect in terms of labour hours; but they can be adapted to cycle times. The literature review of Anzanello and Fogliatto (2011) shows the different types of learning models. In short, the learning models are grouped as follows:

<u>Log-linear models:</u> These types of learning curve models are considered the most basic ways to model learning effects. Wright's model, as being one of the first learning models, is also referred as the only log-linear model by some sources. It has the following mathematical representation:

$$Y_x = K * x^{\log_2 lp} \tag{5}$$

where;

Y_x: The number of labour hours to produce the xth unit
K: The number of labour hours to produce the first unit
x : The number of units
lp : The learning percentage

<u>Exponential models:</u> They rely on a more complete set of parameters compared to log-linear models. The most basic formulation of exponential models is as follows:

$$Y_x = K * x^{\log_2 lp} * e^{(c*x)} \tag{6}$$

where;

c : The second constant

<u>Hyperbolic models:</u> They relate the number of conforming units to the total number of units produced. The formulation is as follows:

$$Y_x = k * \left(\frac{x}{x+lp}\right) \tag{7}$$

where;

k : Maximum performance level

Each type of model works well in certain situations. These situations depend on the industry or other situational aspects. Anzanello and Fogliatto (2011) state that log-linear models are used heavily and perform well in several production sectors, semiconductor and electrical components manufacturing are two of these sectors. Photolithography sector has the specialties of those two sectors and therefore we decided to use log-linear models in our analysis.

In our case, we are aiming to calculate cycle times of operations, so an adoption is needed. The formula for the average number of direct labour hours to produce the first specified number of units is adapted to our situation. The used formula for the average cycle times is given below:

$$avct_{fut} = avct_{pas} * \frac{\frac{1}{1+\log_2 lp} * x^{1+\log_2 lp}}{x} \tag{8}$$

where;

CT^{av}_{fut}: The predicted average cycle time for the production of the following x units
CT^{av}_{pas}: The current average cycle time
x : The number of units
lp : The learning percentage

There are not many studies conducted to determine the learning percentage. Moreover, the learning percentages should be sector specific; therefore for each type of production a different value should be used. Therefore it is not known what the learning percentage for the photolithography production sector is. In Chapter 5, the learning percentages for the given case was calculated.

4.8. Allocation of the Buffer Times

In the literature of buffer management, there are different types methods to allocate buffers, however all of these methods consider buffers in physical stocks. Time buffers are not considered in those methods.

The articles which consider time buffers mainly use basic rules for allocation. In the articles of Radovilsky (1998) and Kuo et al. (2009), the buffers within operations (or the buffer before the bottleneck) are calculated by basic analytical or direct approaches and then it is subtracted from the whole buffer, and the rest is put at the end of the process. We have already decided in Section 4.5 to put buffers only before the bottleneck and after the whole process. Thus, the method of these authors turns into the following:

$$B_i^{end} = B_i^{tot} - B_i^{bbn} \tag{9}$$

where;

B_i^{tot} : Total buffer time of process i
B_i^{bbn}: Buffer time before the bottleneck
B_i^{end}: Buffer time after the last operation

As we discussed in 5.3.3., bottleneck should be protected from the variations of the process and this method assures that, first calculating the buffer of the bottleneck. Thus, this method is formulated as follows:

- Method 1: Use the statistical approach to calculate the buffer time before the bottleneck and subtract it from the whole buffer.

Besides, Powell and Pyke (1998) developed an allocation strategy for the physical buffers; however their reasoning to make the allocation can be used in time buffers as well. The authors suggest that the buffer allocation should depend on the mean and the standard deviation of the operations, or the ratio between them (namely coefficient of variation).

$$B_i^{bbn} = B_i^{tot} * \frac{CV_i^{bbn}}{CV_i^{bbn} + CV_i^{end}} \tag{10}$$

$$B_i^{end} = B_i^{tot} * \frac{CV_i^{end}}{CV_i^{bbn} + CV_i^{end}} \tag{11}$$

where;

B_i^{tot} :	Total buffer time of process i
B_i^{bbn}:	Buffer time before the bottleneck
B_i^{end}:	Buffer time after the last operation
CV_i^{bbn}:	Coefficient of variation before the bottleneck
CV_i^{end}:	Coefficient of variation after the bottleneck

Thus the method is as follows:

- Method 2: Allocate time buffers to the operations according to their coefficient of variation

The choice between these two methods was made in Chapter 6.

4.9. Costs of Service Levels

Previous section described a buffer calculation method, which uses the cycle times as an input and gives the buffer time which is required in order to reach to a desired (predefined) service level. However, it is not known if this service level would lead to a cost-optimal solution. There is no study which investigated the goodness of the predefined service level (which is set as 99% for all of the modules). This part aims to find the cost-optimal service levels and the buffer times associated with these service levels. In the following paragraphs the costs components are analyzed.

4.9.1. Cost components

The buffer times which are placed after the process would lead to the physical stocks in production. Each time a production process finishes before the buffer time, the module have to wait until it is transferred to the next station. The time between the early finish of a module and the shipment date would represent the holding cost for the company. Three parameters need to be known in order to calculate the average holding cost of a module. They are as follows:

- Cost of a module: Each module has a different cost; therefore this component differs from one module to another.
- Weighted average cost of capital (WACC): It is the same for all of the modules; 11% of WACC is used.
- Average earliness in terms of time (Representing average stock): Average earliness shows how much each module would wait on average on stock. It expected to be different if different distributions and buffer times are used.

On the other hand, being late constitutes another cost, which is the cost of being late (or tardy). We require two parameters to calculate the average holding cost of a module. They are as follows:

- Cabin cost: If the module would be late, the cabin which is allocated for production would be empty until the module is ready. There is a constraint on the number of cabins and therefore it is very costly to hold an empty cabin. The cabin cost is different for the different type of end products (machines).
- Average tardiness in terms of time: Average tardiness shows the expected amount of time, during which the cabin would wait for the module. It expected to be different if different distributions and buffer times are used.

Since the cabin cost is very high compared to the stock costs, the optimal service levels are expected to be high, minimizing the risk of being late. The calculations for finding the optimal level can be done in spreadsheet programs by evaluating the total costs and the service levels (and buffers associated to it).

4.9.2. Mathematical representation

This section presents the mathematical representation of the extension of the model in terms of optimal service levels. For each process *i*, the objective is:

$$\min \quad C_i^{wip} * C^c * Ts_i^{av} + C_j^{cab} * Tt_i^{av} \tag{12}$$

$$\text{s.t.} \quad Ts_i^{av} = \int_0^{x=Tai+Bi} \left(Ta_i + B_i - \mathrm{x} \right) * f(x)\, dx \tag{13}$$

$$Tt_i^{av} = \int_{x=Tai+Bi}^{\infty} \left(x - Ta_i + B_i \right) * f(x)\, dx \tag{14}$$

$$SL_i = \int_{x=Tai+Bi}^{\infty} F(x) dx \tag{15}$$

$$SL_i > \alpha_{min} \tag{16}$$

where;

C_i^{hol}: Holding cost; $C_i^{hol} = C_i^{wip} * C^c * Ts_i^{av}$

C_i^{tar}: Tardiness cost; $C_i^{tar} = C_j^{cab} * Tt_i^{av}$

C_i^{tot} : Total buffer cost of the work center i; $C_i^{tot} = C_i^{hol} + C_i^{tar}$

Ts_i^{av} : Average days spent in stock

Ta_i : A-time of process i
B_i : Total buffer time of process i
α_{min}: Predefined minimum service level
x: Random variable of cycle time
SL_i : Service level of the work center i
$f(x)$: The distribution of the cycle time

The aforementioned situation can also be defined in terms of newsvendor problem. Hill (2010) defines this as the one time business decision problem, which can be used in various types of decision making issues such as setting safety stock levels, inventory target levels, buying seasonal goods and overbooking customers. Our problem with newsvendor context is shown as follows:

$$\text{E(Cost(}CT_i^{pl}) = \int_{x=0}^{CT_i^{pl}} \left(CT_i^{pl} - \mathrm{x} \right) * f(x)\, dx + \int_{x=CT_i^{pl}}^{\infty} \left(\mathrm{x} - CT_i^{pl} \right) * f(x)\, dx \qquad (17)$$

where;

CT_i^{pl}: Planned cycle time
E(Cost(CT_i^{pl}) Expected cost associated with the planned cycle time of CT_i^{pl}
C_u: Underage cost; $C_u = C_i^{wip} * C^c$
C_o: Overage cost; $C_o = C_j^{cab}$
C_i^{tot} : Total buffer cost of the work center i; $C_i^{tot} = C_i^{hol} + C_i^{tar}$

This formulation leads to the mathematical determination of the optimal service level SL_i^*. This is shown below:

$$\text{F(}CT_i^{pl}) = \frac{C_u}{C_u + C_o} \qquad (18)$$

$$SL_i^* = F^{-1}\left(\frac{C_u}{C_u + C_o}\right) = F^{-1}\left(\frac{C_i^{wip} * C^c}{C_i^{wip} * C^c + C_j^{cab}}\right) \qquad (19)$$

4.10. Summary

This chapter presented a buffer calculation method for the production systems with the specific characteristics. Throughout the chapter, the details of the methodology are explained both verbally and mathematically.

In general, the method aims to calculate the buffer sizes for the each production process and divide them between the operations in order to maximize the performance of the whole system. This is done by deciding on the following:

- Buffer strategy: Calculate the total buffer, considering the process as one production unit.
- Location of buffers: Place buffers before the bottleneck station and at the end of the whole process.

- Calculation of buffer sizes: Use statistical approach. Make additions to the approach by adding learning effects, and determining candidate solutions.
- Allocation of buffers: Distribute buffers according to the coefficient of variations of the operations.

Furthermore, we investigated the costs of using different service levels and suggested that the cost optimal service levels can be computed.

5. Application at ASML

In the previous chapter, a buffer calculation method is proposed for the production systems which have the characteristics of low production volume, steep learning curve and fluctuating processing times. Even though the method is developed by taking these characteristics into account, it is not known if the method is suitable for an actual system with the aforementioned specialties. The production system of the photolithography producer ASML is used in order to measure and evaluate the performance of the method. The specialties of the photolithography production suits to the characteristics that are considered during the development of the methodology.

The remaining of this chapter is organized as follows: In Section 5.1 the performance measures of the production system and for the method are presented. Section 5.2 aims to explain the first study which has conducted in order to assess the performance of the developed buffer calculation method for buffer size calculation. In Section 5.3 the second study is explained that is done in order to select one of the buffer allocation methods. Last but not least, Section 5.4 gives detailed information about the analysis.

5.1. Preparations for the Analyses

This section describes the performance preparations which are conducted before the analysis of the proposed method.

5.1.1. Performance measures

As explained in Section 4, each work center is responsible for the production of one specific module. They are expected to reach a required service level, while minimizing the costs related to the earliness and lateness of the modules. Moreover, because of the space constraints in the clean room, the level of physical stocks is also important. However, this issue is out of scope and therefore is not discussed in detail. Detailed information about the first two performance indicators of the production system is given below:

1. Cost: The main goal of the system is to reach lowest possible cost. This includes two components such as earliness and lateness cost.
2. Service level: The other goal of the system is to reach to a high service level while keeping costs as low as possible.

5.1.2. Calculation of the parameters

In this section we determine the parameters which are used in the following parts of the thesis.

<u>5.1.2.1. The learning percentages:</u> The learning percentage for each module is calculated by examining the change in the average of the cycle times of the past data. The steps that are taken during the calculation of the learning percentage are as follows:

- Cycle times of past data for each module is taken and the moving average of every 10 modules are calculated
- The changes between the first moving average and the each other moving average is found in terms of percentages. These values are representatives for learning percentages

- The future learning percentage is predicted by using exponential smoothing.

However, since the size of the data set is limited, the predicted learning percentages might be very different from the actual values and this could heavily effect or analysis. Therefore the found values are also discussed with engineers from production. They acknowledged that it is highly possible that two of the found values are lower than it should actually be.

Moreover, a study about learning was already conducted in ASML production. Although, this study did not focus on lead times, but on manpower need, their results can be used by other studies. As a result of this, we made changes on two learning percentage values. Thus, the calculated and agreed learning percentages are shown in Table 3.

Table 3 - Learning Percentages

	Calculated Lp	Agreed Lp
WC 1, Part 1 XT	0,96	1
WC 1, Part 1 NXT	0,86	0,85
WC 1, Part 2 XT	0,85	0,85
WC 1, Part 2 NXT	0,78	0,85
WC 2	0,67	0,80

During the analysis, we decided to use the learning percentages agreed with engineers from production.

5.1.2.2. Optimal service levels: The optimal service levels of the modules are computed by using the formulas of newsvendor problem ((13) and (14)). Thus the optimal service level for a module is found as follows:

$$SL_i^* = \frac{C_u}{C_u + C_o} = \frac{C_i^{wip} * C^c}{C_i^{wip} * C^c + C_j^{cab}} \qquad (20)$$

The results of this calculation are given in Table 4.

Table 4 - Optimal Service Levels

Modules	Optimal SLs
WC 1, Part 1 XT	96,42%
WC 1, Part 1 NXT	95,14%
WC 1, Part 2 XT	92,55%
WC 1, Part 2 NXT	93,34%
WC 2	78,83%

5.2. Simulation Analysis for the Buffer Sizes

In order to assess and validate the performance of the buffer size calculation method an analysis is conducted with the actual cycle time data of the modules. The aim of the analysis is to calculate buffer times for a given period and compare it with the actual data. The specialties of the analysis are given in the following sections.

5.2.1. Input data

As the first step of the analysis, the data is collected and prepared for the next steps. The input data have the following characteristics:

- The cycle time data from the period of June-April are taken as the input data.
- The data in hand constitutes 60 cycle time data in terms of shifts for the selected modules. This data is taken from the so called OBIEE software, which retrieves data from SAP.
- Outliers were identified and eliminated by using the rules that are described in Appendix II and Appendix III.
- The statistical characteristics of the data are given in Appendix IV.

5.2.2. Study design

The steps which constitute the base of the analysis are as follows:

- The input data is divided into 3 parts. The aim was to conduct two simulations (replications) with different input and output data. First the simulation is conducted with the first and second part; and then with second and the third part.
- The first part of the input data (20 observations) is identified as the so called "past data", they are used in the buffer calculations.
- The prediction of the buffer time (planned cycle time) for the coming months is done by using the developed buffer time calculation method. Thus, the steps are as follows:
    - The distribution of the data set of each module was found by using Easyfit software in Excel. It shows that the all of the data set follows either gamma or exponential distribution.
    - The learning effects are added to the data.
    - The buffer times for 99% service level are computed.
    - The optimal service level is found and the buffer times associated with this service level is computed.
- The results of the method are named as "buffer times" and they are compared to the next part of the data (20 observations), which is named as the so called "future data".
- After this comparison the earliness and tardiness of the cases are identified and service levels and costs are computed.
- As explained above, the prediction and calculation is repeated with second and third parts.

The graphical representation of the study is shown in Figure 15.

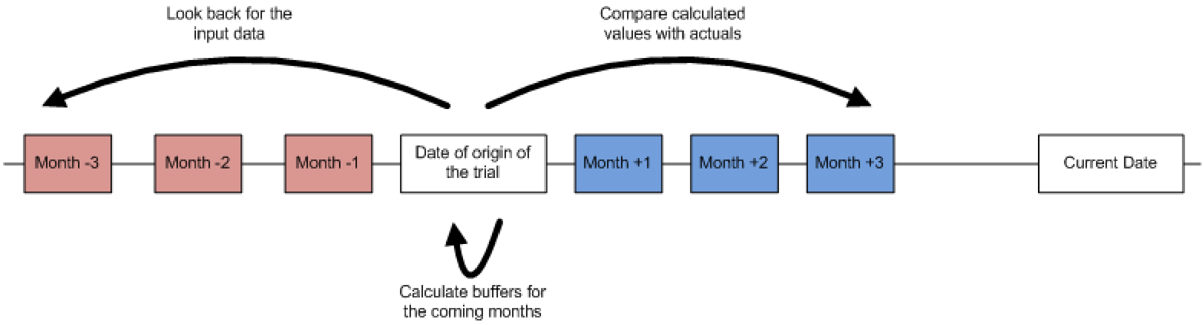


Figure 15 - Explanation of the Study

5.2.3. The results

In this section the results of the method for each module are presented and these results are compared with the case of 99% service level.

5.2.3.1. Work Center 1, Part 1 for XT

The proposed method suggests using the planned lead times of 6,43 and 6,77 days for the setting of 99% service level. In both replications, these lead times result in no lateness. For the optimal service level on the other hand (which is 96,42%), the planned lead times are 5,07 and 5,20 days. If these lead times are used in the simulation, 1 module is late for both replications. However, the costs using lead times of optimal service levels are lower than costs of lead times with 99% service level. The results are also shown in the table below:

Table 5 - Results of WC1 Part 1 XT

Service Level	99%	Optimal (96,42%)
Planned LT	6,60	5,14
Number late	0	1
Cost	2.731,76	1.952,27

5.2.3.2. Work Center 1, Part 1 for NXT

For the 99% service level, the planned lead times are calculated as 6,31 and 5,76 days. These lead times cause no lateness in both of the replications. For the optimal service level on the other hand (which is 95,14%), the planned lead times are 4,99 and 4,26 days. The costs using lead times of optimal service levels are lower than costs of lead times with 99% service level. The results are also shown in the table below:

Table 6 - Results of WC1 Part 1 NXT

Service Level	99%	Optimal (95,14%)
Planned LT	6,04	4,63
Number late	0	1
Cost	4.391,51	3.571,15

5.2.3.3. Work Center 1, Part 2 for XT

The proposed method calculates the planned lead times as 11,55 and 9,90 days for the setting of 99% service level. For the optimal service level on the other hand (which is 92,55%), the planned lead times are 8,30 and 6,78 days. For this module, none of the calculated lead times cause tardiness. The costs using lead times of optimal service levels are lower than costs of lead times with 99% service level. The results are also shown in the table below:

Table 7 - Results of WC1 Part 2 XT

Service Level	99%	Optimal (92,55%)
Planned LT	10,73	7,54
Number late	0	0
Cost	10.233,65	5.756,03

5.2.3.4. Work Center 1, Part 2 for XT

For the 99% service level, the planned lead times are calculated as 10,18 and 11,01 days. These lead times cause no lateness in both of the replications. For the optimal service level on the other hand (which is 93,34%), the planned lead times are 7,07 and 6,74 days. If these times are used, the system is late for 3 occasions in total (2 in the first replication and 1 in the second replication). The costs using lead times of optimal service levels are lower than costs of lead times with 99% service level. The results are also shown in the table below:

Table 8 - Results of WC1 Part 2 NXT

Service Level	99%	Optimal (93,34%)
Planned LT	10,60	6,91
Number late	0	1,5
Cost	10.770,26	6.326,88

5.2.3.5. Work Center 2

The proposed method suggests using the planned lead times of 28,28 and 28,20 days for the setting of 99% service level. In both replications, these lead times result in no lateness. For the optimal service level (78,83%), the planned lead times are 20,51 and 17,56 days. If these lead times are used in the simulation, 2 modules are late for the first replication and 7 modules are late for the second replication. However, the costs using lead times of optimal service levels are lower than costs of lead times with 99% service level. The results are also shown in the table below:

Table 9 - Results of WC 2

Service Level	99%	Optimal (78,83%)
Planned LT	28,24	19,04
Number late	0	4,5
Cost	62.195,38	43.331,73

5.2.4. Comparison with the company methodology

In this section, the results of the tool in use at the company and the results of the proposed method are compared. The costs of both methods are summarized in Table 10:

Table 10 - Cost Comparison of Methods

	Proposed	ASML
WC 1, Part 1 XT	1.952,27	2.106,24
WC 1, Part 1 NXT	3.571,14	3.698,82
WC 1, Part 2 XT	5.756,03	7.177,90
WC 1, Part 2 NXT	6.326,87	6.370,23
WC 2	43.331,73	72.222,45

As it can be seen in the table, the proposed method outperforms the method in use in terms of costs. The proposed method results in fewer costs in each replication for each module. The cost comparison can also be seen in **Error! Reference source not found.**.

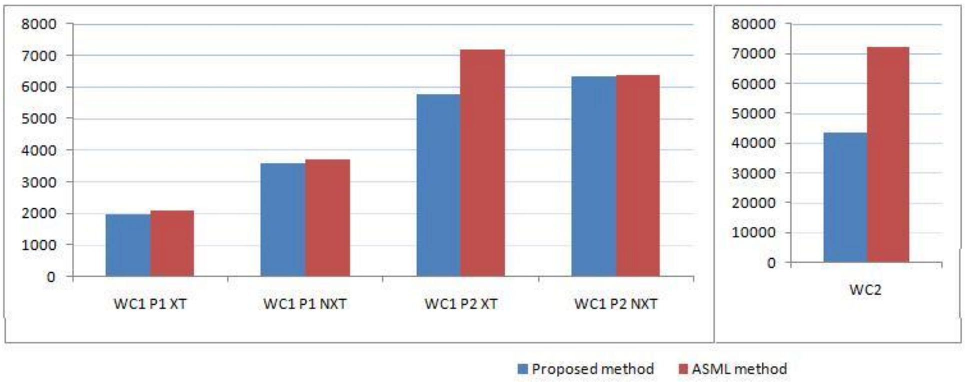


Figure 16 - Cost Comparison of Methods

To conclude this section, we can say that if the proposed method is used, cost reduction would be achieved for the modules shown above.

5.2.5. Effects of the learning curves

This section compares the proposed method with the other usage alternatives of learning curves and shows how effective they are in the buffer calculations. The situations that are considered in this section are shown below:

1. Effects of using/not using learning curves
2. Effects of the learning percentage

In the following two sections, these effects are discussed.

5.2.5.1. Effects of using learning curves

In order to find the how learning curves affect our results and our methodology, the buffer calculations are done with and without adding learning effects. As expected, the method without adding learning effects assigns higher values to buffers and planned lead times. The comparison of lead times between calculations with and without using learning curves is as shown in Table 11.

Table 11 - Effects of Using Learning Curves

	Replication 1		Replication 2	
	Proposed	No Learning	Proposed	No Learning
WC 1, Part 1 NXT	4,99	5,48	4,26	4,53
WC 1, Part 2 XT	8,30	9,87	6,78	7,81
WC 1, Part 2 NXT	7,07	8,08	6,74	7,56
WC 2	20,51	27,23	13,97	17,56

The reader should note that the Part 1 for XT at Work Center 1 is not on the table, since in its default setting the learning percentage is 1, thus the learning effects did not taken into account in the buffer calculations of that module.

Moreover, if calculated by the method without learning, the buffers cause higher costs compared to the buffers calculated by using the method with learning. The cost comparison of these two situations is given in Table 12.

Table 12 - Cost Effects of Using Learning Curves

	Replication 1		Replication 2		Average	
	Proposed	No Learning	Proposed	No Learning	Proposed	No Learning
WC 1, Part 1 NXT	2.440,51	3.051,04	4.701,78	5.090,08	3.571,14	4.070,56
WC 1, Part 2 XT	4.279,30	6.445,85	7.232,76	9.399,30	5.756,03	7.922,58
WC 1, Part 2 NXT	6.406,65	6.870,62	6.247,10	7.383,95	6.326,87	7.127,28
WC 2	32.768,21	59.608,68	53.895,25	80.563,05	43.331,73	70.085,87

Furthermore, in terms of average costs, the buffer calculation method with the learning curve performs 33,8% better than the method without learning curves.

All in all, we can say that using learning curves are very important both in terms of lead times and also in terms of costs.

5.2.5.2. Effects of the learning percentage

We decided to further investigate the effects of learning curves and learning percentages with the following study. In order to find the how learning percentages affect our results and our methodology, we calculated buffers by using different values of learning curves. The company usually expects learning percentages higher than 79%, therefore we used following values between 79% and 100%. The reader should note that using the learning percentage of 100% is equivalent to not using the learning effects (rate of improvement=0%)

The calculated average planned lead times, that are calculated by the given learning percentages are as shown in Table 13.

Table 13 - Effects of the Learning Percentages

LP	79%	82%	85%	88%	91%	94%	97%	100%
WC 1, Part 1 XT	4,66	4,73	4,79	4,85	4,91	4,98	5,06	5,14
WC 1, Part 1 NXT	4,38	4,57	4,63	4,69	4,76	4,84	4,92	5,01
WC 1, Part 2 XT	7,16	7,35	7,54	7,76	8,00	8,25	8,54	8,84
WC 1, Part 2 NXT	6,61	6,76	6,91	7,07	7,24	7,42	7,61	7,82
WC 2	16,77	17,40	18,08	18,82	19,62	20,48	21,41	22,40

It is obvious that the calculated lead times are lower if the learning percentage is lower (the effect of learning curves higher). The costs associated with the lead times above are compared in Figures 17 and 18.

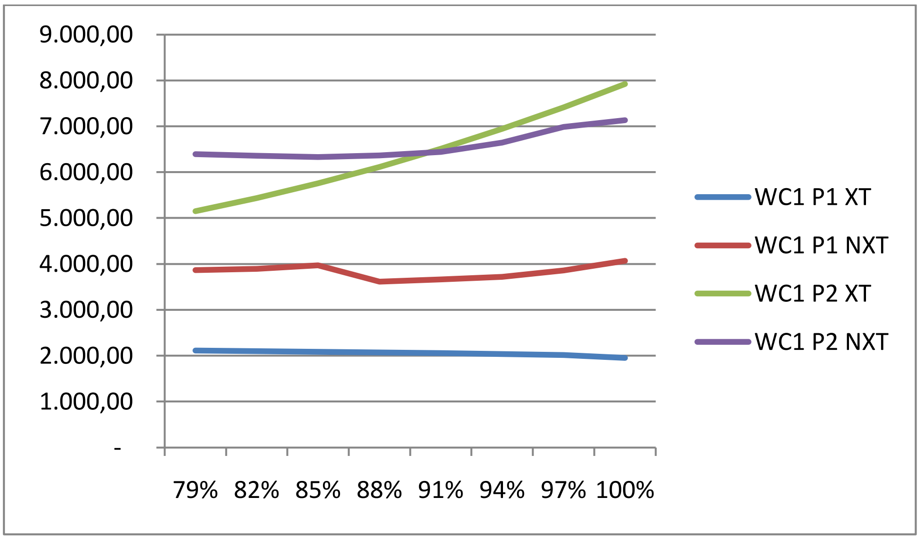


Figure 17 - Comparison of Learning Percentages (WC 1)

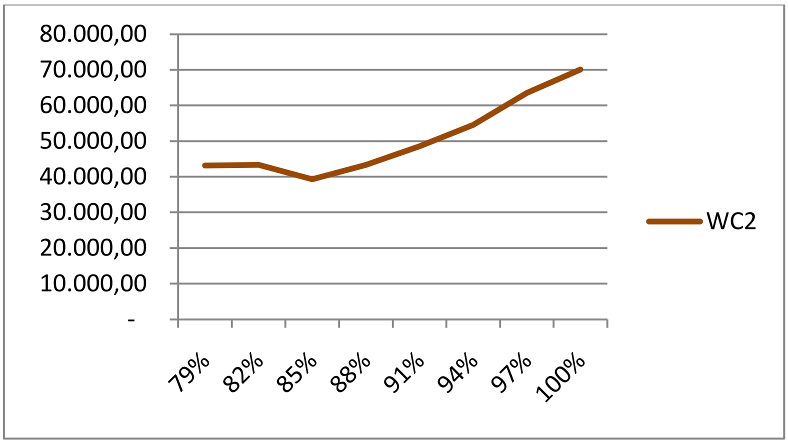


Figure 18 - Comparison of Learning Percentages (WC 2)

Following conclusions can be drawn from this analysis:

- WC 1 Part 1 XT module is not subject to learning effects. Therefore the costs are higher if a value other than 100% is used as learning percentage.
- For modules of WC 1 Part 1 NXT, WC 1 Part 2 NXT and WC 2, there is a minimum value in terms of costs. This minimum shows the optimal level of learning percentage. For WC Part 1 NXT it is 88%, for WC 1 Part 2 NXT it is 94% and for WC 2 it is 85%.
- For WC 1 Part 2 XT module the cost function is increasing. We can say that the optimal value for learning percentage lies below 79%. Further analyses showed that the optimal level of learning percentage for this module is 67%.

5.3. Simulation Analysis for the Buffer Allocation

In order to select one of the two proposed of the buffer allocation methods another simulation analysis is conducted only for Work Center 2 module. The aim of the analysis is to compare the buffer allocation methods. The specialties of the analysis are given in the following sections.

5.3.1. Study design and input data

This study simulates the production environment of the Work Center 2 with defining three production units. Therefore the system in consideration is as in **Error! Reference source not found.**.

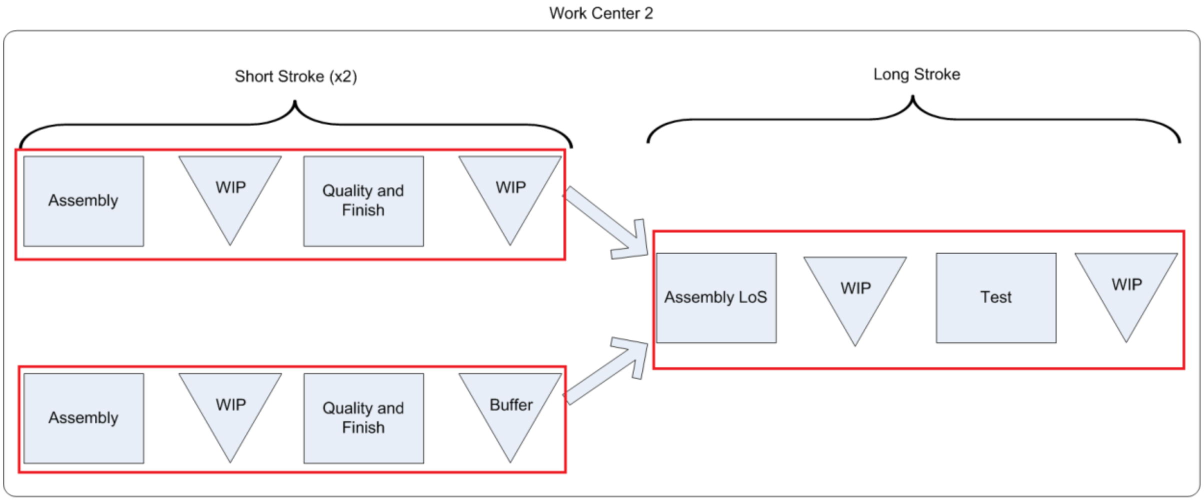


Figure 19 - Situation for the Simulation

As the first step of the analysis, the data is collected and prepared for the next steps. The input data have the same characteristics as the one in Section 5.2. The only difference is that, this time instead of the parameters for the whole process, the parameters for operations (SS1, SS2, and Test) are used.

5.3.2. The results and comparison

First we recall the methods which we are assessing:

- Method 1: Use the statistical approach to calculate the buffer time before the bottleneck and subtract it from the whole buffer.
- Method 2: Allocate time buffers to the operations according to their coefficient of variation

Moreover we compared these methods with the situation of the system before the allocation, e.g. all the buffer put at the end of the process (We named this situation Method 0)

5.3.2.1. Results

This section briefly describes the results of the allocation decision.

In the situation before the allocation, the entire buffer (46 shifts) is put at the end of the process (after the bottleneck). This causes 5 late deliveries to FASY and a 92% utilization of Test operation.

The statistical approach suggests that the buffer before the bottleneck should be 28 shifts and therefore the buffer at the end would be 18 shifts. This causes also 5 late deliveries to FASY and a 96% utilization of Test operation.

The allocation based on coefficient of variation of processes result in putting a buffer of 20,7 shifts before the bottleneck and 25,3 shifts after the bottleneck. In this case, 14 modules are late, and the utilization of the bottleneck is %99.

The results are presented in Table 14.

Table 14 - Results of Simulation on Buffer Allocation

Method	0	1	2
Planned LT before Test	-	28	20,7
Planned LT after Test	45,00	18,00	25,30
Number late	5	14	5
Utilization of Test	92%	99%	96%

5.3.2.4. Selection of the method

The main idea behind the allocation of buffer times is to utilize the bottleneck as much as possible. The reasons behind this were given in Section 4.2 and 4.5.

As it can be seen in the previous section, in method 1 the bottleneck is highly utilized. However, method 1 results in 14 late deliveries. Therefore, using method 1 will clearly decrease the service level of this module.

On the other hand, method 2 also results in a high utilization of the bottleneck (96%). Besides, the number of late deliveries is the same as the situation before allocation. So, we can say that the second method is performing better than the first method.

As a result of this, we decided to choose method 2, allocating time buffers according to their coefficient of variation.

5.4. Qualitative Analysis

This section gives more information about the results of the analysis of the work centers.

5.4.1. Service levels

The results about the service levels are interpreted as follows:

- The optimal service level lies between 92-97% for all Work Center 1 modules. For Work Center 2 module it is 78.83%.
- In Work Center 1, the WIP costs of Part 1 modules are lower than the WIP costs of Part 2 modules. Therefore Part 1 modules have higher calculated optimal service levels than Part 2 modules. The earliness cost is relatively low for Part 1 modules even if the service level (and accordingly the average earliness of the module) is relatively high.
- The company policy of using 99% service level for all modules would lead to non optimal results.
- The relationship between service levels, calculated lead times and the costs components can be seen in Figure 20. The reader should note that, although we present only the graph for one module for visualisation, a similar shaped graph was found also for other modules.

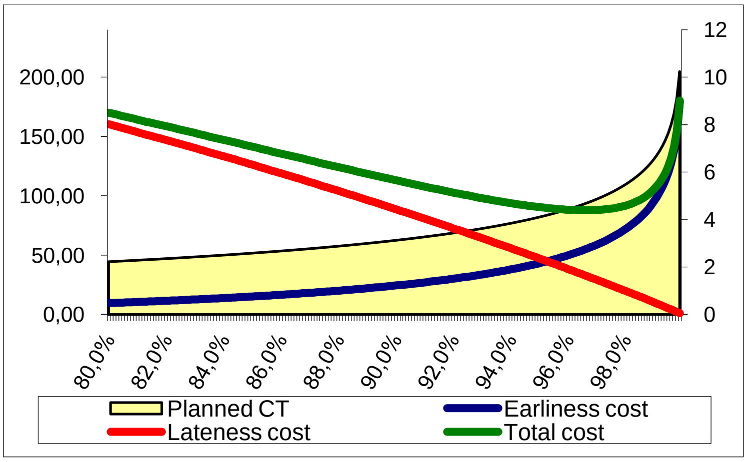


Figure 20- Relationship between Costs and Service Levels (WC 1 Part 1 XT)

5.4.2. Buffer times

The results about the buffer times are interpreted as follows:

- The ways of choosing and using the learning curves have big effects on buffer time calculations.
- We have also experienced that outlier elimination have also an impact on results.
- In order to reach 99% of service level in reality (e.g. the company target), buffer times should be on average %15 higher than optimal values of buffer times.
- Because of the nature of the exponential distribution, the buffer times are very high if this distribution is used during calculations and a high service level is required.

5.4.3. Analysis and comparison of methods

The results about the performance of the methods and alternatives are interpreted as follows:

- Proposed method performs better than the methodology which is currently in use. In terms of buffer times, the proposed method leads to results that are lower than the company strategy. It is on average 1.55 days lower than the company method if 99% of service level is used and it is on average 3.85 days lower than the company method if the optimal service levels are used.
- Learning effects have big impacts on results. The method without taking learning effects into account results higher buffer time values, 2.6 days on average.

5.5. Summary

This chapter presented the validation and the comparison studies of the buffer calculation methods. Throughout the chapter, the details and the results of the study are given, and the outcomes are interpreted. The main results are as follows:

- The company policy of using 99% service level for all modules leads to non optimal results.
- Proposed method performs better than the methodology which is currently in use.
- The method without learning effects results higher buffer time values.

6. Conclusion

This chapter summarizes the whole report. We summarize our findings in Section 6.1., give recommendations in Section 6.2. Last but not least, future research options are discussed in 6.3.

6.1. General Conclusions

- In a production environment with a very high process variation, the use of buffers in terms of time is appropriate.
- Buffer management and buffer decisions in production environments with the characteristics of low production volume, steep learning curve and fluctuating processing times are affected by several factors such as the use of learning curves, the determination of learning percentages, selection of distributions and use of other parameters.
- It is better to calculate the buffers for the whole process instead of calculating it separately for the individual operations.
- Bottlenecks should be protected by placing buffers before them. In the situation of buffers in time, the other place to put a buffer is after the last operation in this process.
- The allocation on time buffers should be made on their ratio of coefficient of variation.
- The statistical fit between the past processing times and the future predicted lead times is very important in this type of a prediction method.
- If learning effects exist in a production environment, this should be taken into account while calculating future lead times/buffers. If the calculation is done without any adaptations, the planned times would not be optimal (higher than the necessary value).
- Learning percentages are situation dependent, therefore they should be re-computed before the calculation/decision on the buffers.
- If buffers are calculated according to the cost optimal service levels, lower costs can be achieved. On the other hand, this would lower the service level. The decision on using the optimal service levels or using a predefined level such as 99% should be made by the user of the method.

6.2. Recommendations

6.2.1. Academic Recommendations

- Buffers in terms of time are not studied in depth in the literature of buffer management. Existing studies either consider physical buffers or the control aspects of the ‘time buffer management’. More studies would be beneficial to improve the applications in practice on buffer management in the capital goods industry.
- The learning curves in industries similar to photolithography are very effective. The learning effects and learning percentages should be taken into account in cycle time and manpower capacity studies that are done in sectors that have similar characteristics as photolithography industry.

6.2.2. Recommendations for ASML

- The buffer time calculation method, that is developed in this study, should be used by the company during its buffer review process.
- The reliability of the data and the database is crucial in order to conduct studies in the industry. Therefore more attention should be given to the data management.
- Since the parts are expensive compared to other industries, using cost optimal service levels can cause considerable cost reductions.

6.3. Future Research Options

- In this study, the developed method is tested for five modules, in two work centers. The method can be applied to other work centers, other modules. Besides, if more data would be available, more and frequent tests can be conducted.
- The effect of the downturn and upturn periods on the learning effects and disturbances should be investigated.
- More case studies can be conducted in buffer (and lead time) management in capital goods industry. For example other designs might be feasible for production environments with lower process variability, lower demand variability and different effects of learning curves.
- Learning curves are used extensively in the literature, however the applications in the area of buffer management are limited. This study used learning curves for a specific type of production environment. More studies on learning curves in buffer management would create more general results for the use of learning curves in this area.

References

Anzanello, M., J., and F., S. Fogliatto. “Learning curve models and applications: Literature review and research directions.” *International Journal of Industrial Ergonomics*, 2011: 1-11.

ASML Intranet website. (accessed April 25, 2011).

Bertrand, J.,W.,M., and J.,C. Fransoo. “Operations management research methodologies using quantitative modeling.” *International Journal of Operations and Production Management 22*, 2002: 241-264.

Bruning, J. H. “Optical lithography - thirty years and three orders of magnitude: the evolution of optical lithography tools.” *The International Society for Optical Engineering*, 1997.

Faria, J., M. Matos, and E. Nunes. “Optimal design of work-in-process buffers.” *International Journal of Production Economics*, 2006: 144-155.

Goldratt, E. M. *The Goal.* 1994.

Gong, L., T. de Kok, and J. Ding. “Optimal leadtimes planning in a serial production system.” *Management Science*, 1994: 629-632.

Gursoy, P. “Buffer Management: Literature Review.” 2011.

Hicks, C. “A genetic algorithm tool for designing manufacturing.” *International Journal of Production Economics*, 2004: 199-211.

Hicks, C., and P. Pongchareon. “Dispatching rules for production scheduling in the capital good industry.” *International Journal of Production Economics 104*, 2006: 154-163.

Hill, A., V. *The newsvendor problem.* Clamshell Beach Press, 2010.

Hopp, W. C., and M. L. Spearman. *Factory Physics.* McGraw Hill, 2008.

Kuo, T., S. Chang, and S. Huang. “Due-date performance improvement using TOC's aggregated time buffer method at a wafer fabrication factory.” *Expert Systems with Applications 36*, 2009: 1783-1792.

Molinder, A. “Joint optimization of lot-sizes, safety stocks and safety lead times in an MRP system.” *International Journal of Production Research 35*, 1997: 983-994.

Oner, K. B., G. P. Kiesmuller, and G. J. Houtum. “Optimization of component reliability in the design phase of capital goods.” *European Journal of Operational Research*, 2010: 615-624.

Powell, S., G., and D., F. Pyke. “Buffering unbalanced assembly systems.” *IIE Transactions 30*, 1998: 55-65.

Radovilsky. “A quantitative approach to estimate the size of the time buffer in the theory of constraints.” *International Journal of Production Economics*, 1998.

Schepens, S. *NXE volume organisation.* Master Thesis, Eindhoven: Eindhoven University of Technology, 2009.

Schragenheim, E., and B. Ronen. “Buffer management: A diagnostic tool for production control.” *Production and Inventory Management Journal*, 1991.

Shi, C., and S. Gershwin. “An efficient buffer design algorithm for production line profit maximization.” *International Journal of Production Economics*, 2009: 725-740.

Swaminathan, J. M. “ Tool capacity planning for semiconductor fabrication facilities under demand uncertainty.” *European Journal of Operational Research*, 2000: 545-558.

Towill, D., R. “Forecasting learning curves.” *International Journal of Forecasting 6*, 1990: 25-38.

Van Aken, J. E., H. Berends, and H. van der Bij. “Problem solving in organizations: a methodological handbook for business students.” Lecture Handout, 2005.

Yano, C., A. “Setting planned leadtimes in serial production systems with tardiness costs.” *Management Science*, 1987.

Appendices

Appendix I. Business Engineering Department

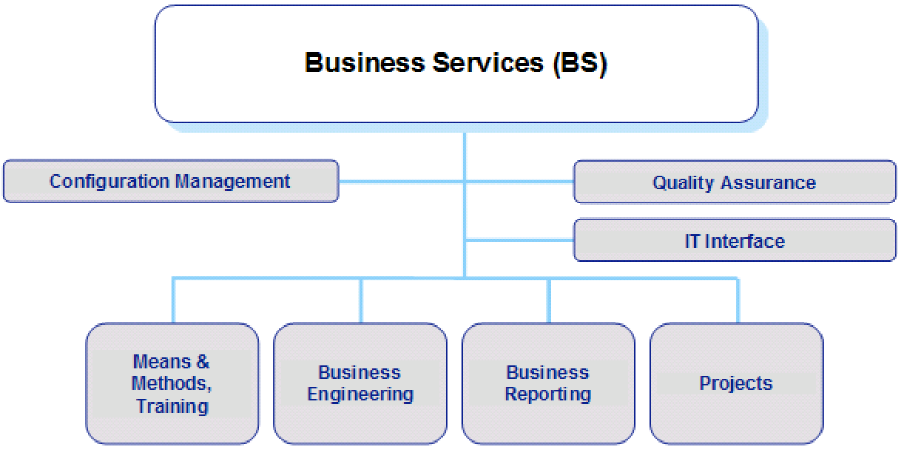
On the very highest level ASML is divided into five different functional units such as Corporate Support, Product, Support, Market and Operations.

Operations unit is mainly responsible for the supplying materials and manufacturing of the machines. Production planning, manufacturing operations, supply chain operations, master planning etc. functions are in this unit. Furthermore, there are different departments in Operations unit, which run projects to improve quality of production, cycle times, delivery performance of production accuracy of production planning and quality of supplier.

The thesis was be conducted in Business Engineering department, which is a part of Operations unit. It is a support department for production and runs improvement projects. Some of the responsibilities and improvement projects of this department can be found below:

- Cycle time updates in work centers
- Capacity and resource planning
- Improve upstream quality improvements
- Layout improvements

Business Engineering department has close relationships with Business Reporting and Projects department. Business Reporting is publishing periodical updates for KPIs and other performance information. The current organizational relationships of Business Engineering are as follows:



Appendix II. Use of Data

Since the method uses the processing times as the only input for calculations, it is very important to have a reliable dataset before starting to calculate buffers. We have made three suggestions in order to reach to a more reliable dataset.

1) Input data: As also discussed in the Sections 4.2.3 and 4.3.2, using cycle times without any adaptation, might lead to observing a non-continuous pattern of data and this would lead to using the wrong distributions. It is decided to use the data which only includes the time which has spend on the machine; the processing times in terms of shifts.

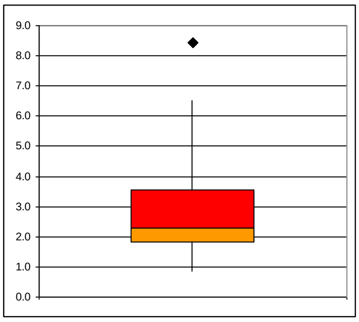
2) Span of data: The span of data represents the decision on how much data values to use in the analysis of buffers. This is extremely important since using too much or too less data might cause to the wrong results and conclusions in terms of buffers.

- It is known that production processes at high-tech companies are subject to the learning effects and therefore the cycle times are not stable throughout the time. If the time span of the data is very big (for example a year), than the data, which are used in the analysis would not be a correct representation of the current situation, since the processing times (or shift times) from one year ago would be larger compared to times from one month ago. Moreover, it is also possible that the standard deviation would change over time. Therefore the time span of the used data should not be bigger than a certain value.
- On the other hand, if the span is very small, then the collected data would not be enough to be analyzed, a clear distribution could not be observed, and the average and the standard deviation of the data set would not represent the actual situation. Therefore the time span of the used data should not be lower than a certain value.
- Because of the reasons given in the above bullets, statistical and quantitative analyses cannot be applied directly to this case. Therefore we decided to use a way of qualitative reasoning to decide on the span of data. The production volume of the high-tech machines production is low and therefore the time span should be at least two months so that adequate number of data is collected. On the other hand, the learning effects do not allow looking back more than three months. Therefore it is decided to use the produced parts of the last three months to predict the future cycle times and buffers.

3) Outliers: Because of the nature of the high-tech machines production, the processing times might be very high because of an extraordinary issue, a problem caused by customer wishes or bad logging of data etc. A string of rules need to be developed to tackle these issues. These rules are summarized below:

- Low values: It is assumed that the actual data (shifts) cannot be smaller than the predefined values of A-times since the A-times show how much time the process should take under the normal circumstances. It should be noted that an experienced worker might work faster and therefore in some cases actual data might be lower than A-times. Therefore it is decided to include the actual data which are slightly lower than A-times. Thus the developed rule is as follows: *Exclude the data which are smaller than (0.5 * A-time)*

- Big values: Very big values, which are caused by extraordinary events or bad loggings, might affect the analysis drastically. Therefore they should also be determined and eliminated. It is decided to use a two step procedure to evaluate big values. In the first step, a small discussion with group leaders of the each module production should be made. This is crucial because sometimes what happened in the production environment cannot be seen or realised by the engineers or planners. Group leaders at production environment might detect some of the obvious outliers and help to eliminate them. In the second step, a statistical analysis is used, by drawing a box plot (also named as box and whisker diagram). In this diagram the ends of the upper whisker represent the highest datum still within 1.5 interquartile range of the upper quartile. The data which are bigger than upper whisker are eliminated.



Appendix III. Use of Parameters

The average and standard deviation of the data set are used as the input values. These two parameters are determined by using the data set and adopted to the future periods by using the learning curves. However, another adaptation could still be made to the average. Using the parameters on the distributions with or without making this adaptation would lead to two different methods of using parameters. Both of these methods are explained below:

- Option 1: Use the average without further computations. Thus, the whole part of the cycle time data could be used in the distributions.
- Option 2: Subtract A-times of processes from the average, since this part is not variable. Thus, only the varied part (e.g. disturbances) is used in the distributions.

Sharagenheim and Ronen (1991) state that the processing times or cycle times should be neglected during buffer time calculations. The fluctuations in processes (disturbances) are more important in terms of buffer time estimations. Therefore we decided to use the Option 2, thus subtract A-times from the data.

Appendix IV. Characteristics of the Dataset

	WC1 P1 XT	WC1 P1 NXT	WC1 P2 XT	WC1 P2 NXT	WC2
mean	4,1	5,8	8,8	7,8	31,4
minimum	1,2	2,9	2,7	2,7	14
maximum	9	9,9	18,6	14	49,1
std. deviation	2	1,7	3,8	3	9,5
distribution	Gamma	Gamma	Gamma	Gamma	Gamma